MARMARA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

FİNANSAL SİSTEMLER İÇİN BÜTÜNLEŞİK BİR AÇIKLANABİLİR YAPAY ZEKÂ ÇERÇEVESİ GELİŞTİRİLMESİ

ABDULLAH EMİR ÇİL

DOKTORA TEZİ
Bilgisayar Mühendisliği
Anabilim Dalı
Bilgisayar Mühendisliği Programı

DANIŞMAN
Prof. Dr. Kazım YILDIZ

İSTANBUL, 2026

MARMARA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

FİNANSAL SİSTEMLER İÇİN BÜTÜNLEŞİK BİR AÇIKLANABİLİR YAPAY ZEKÂ ÇERÇEVESİ GELİŞTİRİLMESİ

ABDULLAH EMİR ÇİL

(723622001)

DOKTORA TEZİ
Bilgisayar Mühendisliği
Anabilim Dalı
Bilgisayar Mühendisliği Programı

DANIŞMAN
Prof. Dr. Kazım YILDIZ

İSTANBUL, 2026

ÖNSÖZ

Bu tez çalışması, finans sektöründe kullanılan yapay zekâ sistemlerinin yalnızca yüksek performans göstermesinin yeterli olmadığını; açıklanabilir, güvenilir ve dayanıklı olmalarının da büyük önem taşıdığını ortaya koyma amacıyla hazırlanmıştır. Finans alanında şeffaf ve hesap verebilir yapay zekâ yaklaşımlarına duyulan gereksinimin giderek arttığı günümüzde, bu çalışmanın alana katkı sağlaması temenni edilmektedir.

Bu tez çalışmasının hazırlanması sürecinde bilgi, deneyim ve yönlendirmeleriyle bana katkı sağlayan değerli hocam Prof. Dr. Kazım YILDIZ'a içten teşekkürlerimi sunarım.

Hayatım boyunca desteklerini her zaman yanımda hissettiğim anneme ve babama; sabrı, anlayışı ve desteğiyle bu süreçte en büyük güç kaynaklarımdan biri olan eşime; sevgileriyle bana umut ve motivasyon veren oğluma ve kızıma gönülden teşekkür ederim.

Abdullah Emir ÇİL

Nisan 2026

İÇİNDEKİLER

ÖNSÖZ i
İÇİNDEKİLER ii
ÖZET vii
ABSTRACT x
YENİLİK BEYANI xiii
SEMBOLLER xv
KISALTMALAR xvi
ALGORİTMALAR xvii
ŞEKİLLER xviii
TABLO LİSTESİ xix
1 GİRİŞ 1
1.1 Tez Çalışmasının Amacı ve Önemi 2
1.2 Tez Organizasyonu 4
2 LİTERATÜR ÇALIŞMASI 6
2.1 Metodoloji 6
2.1.1 Araştırma sorularının belirlenmesi 6
2.1.2 Seçim kriterleri 6
2.1.3 Kuramsal çerçeve 9
2.1.3.1 Modelden bağımsız yaklaşımlar 9
2.1.3.2 Modele özgü yaklaşımlar 18
2.1.3.3 Güvenlik ve gizlilik yöntemi 23
2.2 Literatür Taraması ve Tematik Analiz 24
2.2.1 Literatürün tematik incelemesi 24
2.2.1.1 Finans sektöründe XAI yöntemlerinin kullanımı 25
2.2.1.2 IoT için uç bilişimde açıklanabilir yapay zekâ: finansal uygulamalar için çıkarımlar 27
2.2.1.3 XAI yöntemlerinin performans karşılaştırması 31
2.2.2 XAI için örnek kullanım senaryoları 37
2.2.2.1 Temsili vaka çalışması örnekleri 38
2.2.3 Yöntemsel karşılaştırma 40
2.2.4 Çalışmaların ayrıntılı analizi 51
2.2.4.1 Kredi riski değerlendirmesi ve skorlama 51
2.2.4.2 Dolandırıcılık tespiti uygulamaları 52

2.2.4.3 Portföy yönetimi ve yatırım stratejisi 53
2.2.4.4 Müşteri davranışı ve segmentasyon 54
2.2.4.5 Piyasa tahmini ve ekonomik analiz 54
2.2.4.6 Düzenleyici uyum ve risk yönetimi 55
2.2.4.7 Temel bulgular ve katkılar 55
2.2.5 Boşluklar ve açıklanmamış alanlar 56
2.2.5.1 Kullanım alanlarında algoritma çeşitliliği 56
2.2.5.2 Veriseti çeşitliliği 56
2.2.5.3 Başarı ölçütleri ve performans analizleri 57
2.2.5.4 Etik, gizlilik ve tarafsızlıkla ilgili eksiklikler 57
2.2.5.5 Açıklanabilirlik ve güvenilirlik standartlarının eksikliği 57
2.2.5.6 Gerçek zamanlı açıklanabilirlik eksikliği 57
2.2.5.7 Özelleştirilmiş açıklama modellerinin azlığı 58
2.2.5.8 Veri gizliliği ve güvenliği ile ilgili eksiklikler 58
2.2.5.9 Kapsamlı verisetlerinin eksikliği 58
2.3 Tartışma 58
2.3.1 Araştırma sorularının kapsanmasına genel bakış 58
2.3.1.1 RQ1: Hangi XAI algoritmaları hangi finansal hizmetler için kullanılmaktadır? 59
2.3.1.2 RQ2: XAI algoritmaları finansal hizmetler için uygun mudur? 59
2.3.1.3 RQ3: XAI kullanımının sunulan finansal hizmetlerin performansı üzerindeki etkisi nasıldır? 59
2.3.1.4 RQ4: Finansal hizmetlerde XAI algoritmalarının uygulamalarında ne tür verisetleri tercih edilmektedir? 60
2.4 Sınırlılıklar 60
2.5 Sonuç 61
3 ÖNERİLEN ÖLÇEKLENEBİLİR, FİNANSAL ODAKLI VE YORUMLANABİLİR AÇIKLAMA (SFIX) MODELİ 64
3.1 Metodoloji 65
3.1.1 Dinamik özellik önem skoru (DIS) 65
3.1.1.1 Temel etki skoru (BI) 67
3.1.1.2 Etkileşim etki skoru (II) 67
3.1.1.3 Zamansal ağırlıklandırma (TW) 70
3.1.2 Güven skoru (CS) 72

3.1.2.1 Veri kalitesi skoru (DQ) ... 73
3.1.2.2 Model tutarlılığı skoru (MC) ... 74
3.1.3 Örüntü skoru (PS) ... 76
3.1.3.1 Trend bileşeni ... 77
3.1.3.2 Periyodiklik bileşeni ... 78
3.1.3.3 Oynaklık (Volatilite) bileşeni ... 79
3.1.4 Anomali skoru (AS) ... 80
3.1.5 Risk skoru (RS) ... 82
3.1.6 Toplam XAI skoru ... 85
3.1.7 Parametreler ... 87
3.1.8 Sadeleştirilmiş SFIX ... 90
3.1.9 Önerilen modelin zaman karmaşıklığı ... 92
3.1.9.1 Zaman karmaşıklığı hesaplaması ... 92
3.1.9.2 Model karmaşıklığının değerlendirilmesi ... 94
3.1.10 Modelin özgünlüğünün değerlendirilmesi ... 96
3.1.11 Finansal uygulamalar ... 98
3.2 Önerilen modelin uygulanması ... 99
3.2.1 SFIX performans değerlendirmesi ... 99
3.2.1.1 Finans Verisetinin Tanıtımı ... 99
3.2.1.2 Deney ortamı ... 100
3.2.1.3 Bulgular ... 101
3.2.1.4 Tartışma ... 106
3.2.2 Açıklanabilirlik yaklaşımlarının karşılaştırılması ... 109
3.2.2.1 Sadelik analizi ... 109
3.2.2.2 Dayanıklılık analizi ... 110
3.2.2.3 Karşılaştırmalı analiz ... 116
3.3 Sonuç ... 117
4 ÖNERİLEN SFIX MODELİN FİNANS UYGULAMASI ... 119
4.1 Metodoloji ... 120
4.1.1 SFIX tabanlı derin öğrenme uygulaması ... 121
4.1.2 Veri ön işleme ... 121
4.1.3 Deneysel tasarım ... 122
4.1.4 SHAP ve LIME XAI yöntemleri ... 124
4.1.5 SFIX, SHAP ve LIME'ın karşılaştırılması ... 125

4.2 Bulgular ve Tartışma 125
4.2.1 Orijinal verisetinin SFIX analizi 126
4.2.2 Tamamlanmış verisetinin SFIX analizi 128
4.2.3 Özellik sıralamalarının karşılaştırmalı analizi 129
4.2.4 Makine öğrenmesi algoritmaları ile model performans karşılaştırması 135
4.2.4.1 Orijinal veriseti üzerinde karşılaştırmalı model performansı 136
4.2.4.2 Tamamlanmış veriseti üzerinde karşılaştırmalı model performansı. 137
4.2.4.3 Özellik sayısının model performansına etkisi 137
4.2.5 Modellerin genel performans değerlendirmesi 138
4.2.6 SFIX bileşen analizi ve yorumlanması 139
4.2.7 Özellik kategorilerinin performans analizi 139
4.2.8 XAI yöntemlerinin karşılaştırmalı analizi 140
4.2.8.1 20 özellik seviyesinde model performansı 140
4.2.8.2 15 özellik seviyesinde performans değişimleri 141
4.2.8.3 10 özellik seviyesinde performans kayması 141
4.2.8.4 5 özellik seviyesinde SFIX performansı 141
4.2.9 Sonuçlar ve model etkinliğinin değerlendirilmesi 141
4.2.10 Sınırlılıklar ve başarısızlık durumları 143
4.2.11 Gerçek dünya dağıtımı: finansal denetimde pratik avantajlar ve zorluklar 144
4.2.12 SFIX bileşenlerinin finansal zaman serisi verileri ve denetim görüşü sınıflandırması için kuramsal uygunluğu 146
5 ÖNERİLEN SFIX MODELİN YAPAY ZEKÂ DAYANIKLILIĞI UYGULAMASI 149
5.1 Veri Seti ve Ön İşleme 150
5.1.1 CICIoT2023 veriseti 150
5.1.2 Alt küme oluşturma ve veri dengeleme 151
5.1.2.1 Alt küme oluşturma 151
5.1.2.2 Dengesizlik giderme stratejisi 152
5.1.3 Eğitim, doğrulama ve test ayrımı 153
5.1.4 Ön işleme 154
5.2 Deneysel Ortam ve Deney Tasarımı 155
5.2.1 Önerilen FFNN modeli 159
5.2.1.1 Ağ mimarisi 159
5.2.1.2 Kayıp fonksiyonu ve optimizasyon 160

5.2.1.3 FFNN hiperparametreleri ve seçim gerekçeleri 160
5.2.2 SFIX ve Simplified SFIX modelinin deneysel ortama entegrasyonu .. 162
5.2.3 S-SFIX Eğitimi ve Özellik Sıralaması 162
5.2.4 Gaussian gürültüsü altında S-SFIX dayanıklılık deney tasarımı 165
5.3 Bulgular 167
5.3.1 Performansın değerlendirilmesi 169
5.3.2 Gaussian gürültüsü altında S-SFIX dayanıklılık sonuçları 170
5.4 Tartışma 179
5.4.1 S-SFIX'in özellik seçme yeteneği 179
5.4.2 Alternatif yöntemlerle karşılaştırmalı değerlendirme 180
5.4.3 Açıklanabilirlik ve performans arasındaki ilişki 180
5.4.4 Yapay Zekâ Dayanıklılığı ve gürültü altında kararlılık 181
5.4.5 Alanlar arası uygulanabilirlik 181
5.4.6 Literatüre katkı 182
6 SONUÇ 183
KAYNAKLAR 186
ÖZGEÇMİŞ 209

ÖZET

FİNANSAL SİSTEMLER İÇİN BÜTÜNLEŞİK BİR AÇIKLANABİLİR YAPAY ZEKÂ ÇERÇEVESİ GELİŞTİRİLMESİ

Bu tez, finans sektöründe yapay zekâ kullanımının giderek yaygınlaşmasıyla birlikte daha görünür hâle gelen açıklanabilirlik, güvenilirlik, düzenleyici uyum ve operasyonel dayanıklılık gereksinimlerine odaklanmaktadır. Çalışmanın temel amacı, finansal karar problemleri için yalnızca yüksek doğruluk üreten değil, aynı zamanda karar mekanizmasını çok boyutlu biçimde açıklayabilen ve farklı uygulama alanlarına uyarlanabilen yeni bir açıklanabilir yapay zekâ çerçevesi geliştirmektir. Bu amaç doğrultusunda tez üç ana eksen üzerinde yapılandırılmıştır. İlk eksende, finans alanında açıklanabilir yapay zekâ uygulamalarını sistematik biçimde inceleyen ACM Digital Library, SpringerLink, IEEE Xplore ve ScienceDirect veritabanlarında kapsamlı bir literatür taraması gerçekleştirilmiştir. Bu kapsamda yapılan tarama sonucunda finans alanında SHAP ve LIME gibi yöntemlerin baskın biçimde kullanıldığını, ancak finansal zaman bağımlılığı, veri güvenilirliği, anomali yapısı ve risk odaklı değerlendirmeleri aynı anda ele alan bütünleşik açıklanabilirlik yaklaşımlarının sınırlı kaldığını göstermiştir.

Tezin ikinci ekseninde, bu boşluğu gidermek amacıyla SFIX (Scalable Financial-oriented Interpretable eXplanation) adlı özgün bir açıklanabilir yapay zekâ modeli önerilmiştir. SFIX; Dinamik Önem Skoru, Güven Skoru, Örüntü Skoru, Anomali Skoru ve Risk Skoru bileşenlerini birlikte kullanarak özelliklerin statik önemini, zamansal davranışını, veri kalitesiyle ilişkisini, örüntü yapısını ve risk duyarlılığını değerlendirmektedir. Böylece model, SHAP ve LIME gibi ağırlıklı olarak post-hoc açıklama üreten yöntemlerden farklı olarak, finansal verinin doğasına daha uygun, içsel ve çok boyutlu bir açıklama mekanizması sunmaktadır. Tezde ayrıca, zamansal yapı gerektirmeyen senaryolar için SFIX'in sadeleştirilmiş biçimi olan S-SFIX geliştirilmiş ve alanlar arası uygulanabilirliği test edilmiştir. SFIX'in özgün katkısı, açıklanabilirliği tek bir yerel katkı hesabına indirgememesi; veri güvenilirliği, anomali davranışı ve risk yapısını açıklama sürecine doğrudan dâhil etmesidir. Bu yaklaşım aynı zamanda özellik seçimi için de güçlü bir çerçeve sağlamaktadır. Çalışmada,

SFIX'in SHAP ve LIME'a kıyasla daha sadelik odaklı ve daha yoğun açıklamalar ürettiği; önceki deneylerde bootstrap ve Lorenz/Gini tabanlı değerlendirmelerle daha sade ve daha kararlı açıklama yapısı sunduğu gösterilmiştir.

Tezin üçüncü ekseni, önerilen yaklaşımın iki farklı uygulama alanında ampirik olarak doğrulanmasıdır. İlk uygulama, finansal denetim görüşü sınıflandırması problemidir. Bu bölümde SFIX, ileri beslemeli sinir ağı (FFNN) mimarisi ile bütünleştirilmiş ve hem orijinal veriseti (N=10572) hem de eksik verileri tamamlanmış veriseti (N=23830) üzerinde değerlendirilmiştir. Bulgular, SFIX'in denetim görüşü tahmininde ayırt edici özellikleri etkili biçimde seçebildiğini ve özellikle düşük boyutlu özellik alt kümelerinde SHAP ve LIME tabanlı seçimlere göre daha tutarlı sonuçlar verdiğini göstermektedir. Deney sonuçlarında belirtildiği üzere, SFIX tabanlı yaklaşım yalnızca %0.45'lik marjinal bir performans kaybı ile boyutsallığı %59 oranında azaltmıştır. Bu sonuç, yüksek boyutlu finansal veride önemli bir hesaplama avantajı sağlarken tahmin gücünün büyük ölçüde korunduğunu göstermektedir. Ayrıca özellikle az sayıda özellikle çalışılan senaryolarda SFIX+FFNN yapısının daha belirgin bir üstünlük sergilediği görülmüştür. Örneğin beş özellikli en dar kurulumda SFIX+FFNN modeli %90,14 doğruluk, %88,49 precision ve %90,14 recall değerlerine ulaşmış; bu durum sınırlı özellik uzayında bile bilgi yoğun ve ayırt edici bir alt küme üretilebildiğini göstermiştir. Denetim tarafındaki genel bulgular, SFIX'in finansal denetim bağlamında yalnızca açıklanabilirlik sağlayan bir yöntem değil, aynı zamanda güçlü bir özellik seçimi ve boyut indirgeme mekanizması olduğunu ortaya koymaktadır.

İkinci uygulama, yapay zekâ dayanıklılığı ve alanlar arası uygulanabilirlik çerçevesinde IoT saldırı tespiti problemidir. Bu bölümde CICIoT2023 veriseti kullanılmış; S-SFIX ile seçilen özellik alt kümeleri aynı FFNN mimarisi altında SHAP, LIME ve Random Forest temelli seçimlerle karşılaştırılmıştır. Sonuçlar, tam özellik kümesi ile eğitilen modelin %98.45 doğruluk sağladığını, ancak özellik sayısı önemli ölçüde azaltıldığında performansın keskin biçimde bozulmadığını göstermektedir. S-SFIX ile seçilen 20 özellikli model %98.05, 15 özellikli model %97.93, 10 özellikli model %96.96 ve 5 özellikli model %96.89 doğruluk üretmiştir. Bu, özellik boyutunun 46'dan 20'ye düşürülmesiyle yaklaşık %56.5, 46'dan 15'e düşürülmesiyle yaklaşık %67.4 oranında bir azaltım sağlanmasına rağmen

sınıflandırma başarısının yüksek düzeyde korunduğunu ortaya koymaktadır. Tezde bu sonuçlar, S-SFIX'in düşük boyutlu temsil uzaylarında da en yüksek bilgi taşıyan çekirdek özellikleri önceliklendirebildiği ve özellikle daha agresif boyut indirgeme senaryolarında pratik bir özellik seçme yaklaşımı sunduğu şeklinde yorumlanmıştır. Böylece çalışma, geliştirilen açıklanabilirlik çerçevesinin yalnızca finansal denetim gibi alan özgü bir problemde değil, IoT siber güvenlik gibi farklı bir bağlamda da etkili biçimde kullanılabildiğini göstermektedir.

ABSTRACT

DEVELOPING AN INTEGRATED EXPLAINABLE ARTIFICIAL INTELLIGENCE FRAMEWORK FOR FINANCIAL SYSTEMS

This thesis focuses on the explainability, reliability, regulatory compliance, and operational resilience requirements that have become increasingly visible with the growing use of artificial intelligence in the financial sector. The main objective of the study is to develop a new explainable AI framework for financial decision problems that not only produces high accuracy but also explains the decision-making mechanism in a multidimensional way and can be adapted to different application areas. To this end, the thesis is structured around three main axes. The first axis involves a comprehensive literature review of the ACM Digital Library, SpringerLink, IEEE Xplore, and ScienceDirect databases, systematically examining explainable AI applications in the financial field. This review revealed that methods such as SHAP and LIME are predominantly used in finance, but integrated explainability approaches that simultaneously address financial time dependence, data reliability, anomaly structure, and risk-focused assessments are limited.

The second axis of the thesis proposes an original explainable AI model called SFIX (Scalable Financial-oriented Interpretable eXplanation) to address this gap. SFIX; The Dynamic Importance Score, Confidence Score, Pattern Score, Anomaly Score, and Risk Score components are used together to evaluate the static importance of features, their temporal behavior, their relationship with data quality, their pattern structure, and their risk sensitivity. Thus, unlike methods such as SHAP and LIME, which primarily produce post-hoc explanations, the model offers an endogenous and multidimensional explanatory mechanism more suitable to the nature of financial data. In this thesis, S-SFIX, a simplified form of SFIX for scenarios that do not require temporal structure, has been developed and its cross-domain applicability tested. SFIX's unique contribution is that it does not reduce explainability to a single local contribution account; it directly incorporates data reliability, anomaly behavior, and risk structure into the explanation process. This approach also provides a strong framework for feature selection. The study shows that SFIX produces more parsimony-focused and denser explanations compared

to SHAP and LIME; previous experiments showed that bootstrap and Lorenz/Gini-based evaluations provided a simpler and more stable explanatory structure.

The third axis of this thesis is the empirical validation of the proposed approach in two different application areas. The first application is the financial audit opinion classification problem. In this section, SFIX is integrated with a feedforward neural network (FFNN) architecture and evaluated on both the raw dataset (N=10572) and the dataset with missing data completed (N=23830). The findings show that SFIX can effectively select distinguishing features in audit opinion prediction and provides more consistent results compared to SHAP and LIME-based selections, especially in low-dimensional feature subsets. As stated in the experimental results, the SFIX-based approach reduced dimensionality by 59% with only a marginal performance loss of 0.45%. This result shows that it provides a significant computational advantage in high-dimensional financial data while largely preserving predictive power. Furthermore, it was observed that the SFIX+FFNN structure exhibited a more significant superiority, especially in scenarios where a small number of features were studied. For example, in the narrowest setup with five features, the SFIX+FFNN model achieved 90.14% accuracy, 88.49% precision, and 90.14% recall; this demonstrates that an information-intensive and distinctive subset can be generated even with a limited feature space. Overall findings from the audit side reveal that SFIX is not only a method that provides explainability in the context of financial auditing, but also a powerful feature selection and dimensionality reduction mechanism.

The second application is the IoT intrusion detection problem within the framework of AI resilience and cross-domain applicability. In this section, the CICIoT2023 dataset was used; feature subsets selected with S-SFIX were compared with SHAP, LIME, and Random Forest-based selections under the same FFNN architecture. The results show that the model trained with the full feature set achieved 98.45% accuracy, but performance did not deteriorate sharply when the number of features was significantly reduced. The 20-feature model selected with S-SFIX produced an accuracy of 98.05%, the 15-feature model 97.93%, the 10-feature model 96.96%, and the 5-feature model 96.89%. This shows that despite a reduction of approximately 56.5% when the feature dimensionality was reduced from 46 to 20 and approximately 67.4% when it was reduced from 46 to 15, the classification success was maintained at a high level. In this thesis, these results are

interpreted as S-SFIX being able to prioritize the most informative core features even in low-dimensional representation spaces and offering a practical feature selection approach, especially in more aggressive dimensionality reduction scenarios. The study demonstrates that the developed explainability framework can be effectively used not only in a domain-specific problem such as financial auditing, but also in a different context such as IoT cybersecurity.

YENİLİK BEYANI

Finans sektöründe yapay zekâ tabanlı sistemlerin kullanımının artması; kredi değerlendirme, denetim, risk yönetimi, dolandırıcılık tespiti ve karar destek mekanizmalarında önemli ilerlemeler sağlamıştır. Bununla birlikte, bu sistemlerin önemli bir bölümü yüksek tahmin başarısına sahip olmasına rağmen kararlarını yeterince açıklayamayan kara kutu yapılar olarak çalışmaktadır. Bu durum, özellikle şeffaflık, güven, hesap verebilirlik ve düzenleyici uyum gereksinimlerinin kritik olduğu finans alanında önemli bir sorun teşkil etmektedir. Mevcut açıklanabilir yapay zekâ yöntemleri çoğunlukla genel amaçlı olup, finansal verilerin zamansal yapı, risk duyarlılığı, veri kalitesi, örüntü davranışı ve anomali karakteristiği gibi alan özgü niteliklerini bütüncül biçimde ele almakta yetersiz kalmaktadır. Ayrıca literatürde, açıklanabilirliğin yalnızca yorum üretme düzeyinde ele alındığı; açıklamaların tutarlılığı, kararlılığı ve farklı veri koşulları altındaki dayanıklılığının sınırlı biçimde incelendiği görülmektedir.

Bu tezde, söz konusu boşlukları gidermek amacıyla finans alanına özgü, bütünleşik ve özgün bir açıklanabilir yapay zekâ yaklaşımı geliştirilmiştir. Tezin özgün katkıları aşağıda sunulmaktadır:

1. Finansal verilerin doğasına uygun olarak tasarlanan ve alan bağımlı açıklanabilirlik gereksinimlerini dikkate alan özgün bir model önerilmiştir.
2. Önerilen modelde özellik önemi, güven, örüntü, anomali ve risk boyutları tek bir çatı altında birleştirilerek mevcut yöntemlere göre daha kapsamlı bir açıklama yapısı oluşturulmuştur.
3. Açıklanabilirlik, yalnızca model çıktılarının yorumlanmasıyla sınırlandırılmamış; özellik seçimi, sınıflandırma başarımı, açıklamaların tutarlılığı ve model kararlılığı ile birlikte değerlendirilmiştir.
4. Modelin sadeleştirilmiş bir sürümü geliştirilerek farklı veri yapılarında uygulanabilirliği incelenmiş, ayrıca gürültü altındaki dayanıklılığı ve özellik sıralama kararlılığı deneysel olarak ortaya konulmuştur.
5. Önerilen yaklaşımın finansal denetim verileri üzerindeki uygulamasıyla, daha az sayıda özellikle etkili ve açıklanabilir karar desteği üretilebildiği gösterilmiştir.

6. Tez kapsamında yürütülen sistematik literatür incelemesi ile finans alanındaki açıklanabilir yapay zekâ çalışmalarının mevcut durumu, yöntemsel yönelimleri, uygulama alanları ve araştırma boşlukları bütüncül biçimde değerlendirilmiştir.

Sonuç olarak bu tez, finans sektöründe daha şeffaf, güvenilir, dayanıklı ve uygulanabilir yapay zekâ sistemlerinin geliştirilmesine yönelik özgün bir model, kapsamlı bir değerlendirme yaklaşımı ve alan odaklı bir bilimsel çerçeve sunarak literatüre ve uygulamaya yenilikçi katkılar sağlamaktadır.

SEMBOLLER

x	:	Girdi vektörü
x_j	:	j. özellik değişkeni
X	:	Girdi değişkenleri kümesi / veri matrisi
X_j	:	j. özellik
y	:	Gerçek çıktı / sınıf etiketi
$\hat{y}$	:	Tahmin edilen çıktı
$f(x)$	:	Modelin x girdisi için ürettiği çıktı
$f_j(X_j)$	:	j. özelliğin modele katkısını temsil eden fonksiyon
a	:	Sabit terim
S	:	Özelliklerin bir alt kümesi
$\|S\|$	:	Alt kümedeki eleman sayısı
I_j	:	j. özellik için önem skoru
n	:	Toplam örnek sayısı
$L(.)$	:	Kayıp fonksiyonu
$P(.)$	:	Olasılık fonksiyonu
$N(x)$	:	x etrafında oluşturulan yerel komşuluk
$1[.]$	:	Gösterge fonksiyonu
μ	:	Ortalama
σ	:	Standart sapma
σ_y	:	Tahmin çıktılarının standart sapması
ρ	:	Korelasyon katsayısı
tau	:	Kendall Tau katsayısı / zaman indeksi
z	:	Standartlaştırılmış skor
w_{ij}	:	i. ve j. nöronlar arasındaki ağırlık
ε	:	Sayısal kararlılık için kullanılan küçük sabit
$O(.)$	:	Zaman karmaşıklığı gösterimi
$\int$	:	İntegral işleci
$\sum$	:	Toplam işleci
∂	:	Kısmi türev işleci

KISALTMALAR

AI	**:**	Artificial Intelligence
AI Act	**:**	Artificial Intelligence Act
ALE	**:**	Accumulated Local Effects
AS	**:**	Anomali Skoru (Anomaly Score)
CNN	**:**	Convolutional Neural Network
CS	**:**	Güven Skoru (Confidence Score)
DIS	**:**	Dinamik Özellik Önem Skoru (Dynamic Importance of Feature Score)
EBM	**:**	Explainable Boosting Machine
EdgeAI	**:**	Edge Artificial Intelligence
FFNN	**:**	İleri Beslemeli Sinir Ağı (Feed-Forward Neural Network)
GAM	**:**	Generalized Additive Model
GDPR	**:**	General Data Protection Regulation
GLM	**:**	Generalized Linear Model
IG	**:**	Integrated Gradients
IoT	**:**	Internet of Things
LIME	**:**	Local Interpretable Model-agnostic Explanations
LORE	**:**	Local Rule-based Explanations
LRP	**:**	Layer-wise Relevance Propagation
LSTM	**:**	Long Short-Term Memory
MIA	**:**	Membership Inference Attack
PRISMA	**:**	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PS	**:**	Örüntü Skoru (Pattern Score)
RNN	**:**	Recurrent Neural Network
RS	**:**	Risk Skoru (Risk Score)
SFIX	**:**	Ölçeklenebilir, Finans Odaklı ve Yorumlanabilir Açıklama (Scalable Financial-oriented Interpretable eXplanation)
SHAP	**:**	Shapley Additive Explanations
S-SFIX	**:**	Sadeleştirilmiş SFIX (Simplified SFIX)
TREPAN	**:**	Tree-based Rule Extraction Algorithm

ALGORİTMALAR

Algoritma 3.1. Dinamik Önem Skorunun hesaplanması 67
Algoritma 3.2. Güven Skorunun hesaplanması 73
Algoritma 3.3. Örüntü Skorunun hesaplanması 77
Algoritma 3.4. Risk Skorunun hesaplanması 85

ŞEKİLLER

Şekil 2.1. Çalışma seçim süreci için PRISMA akış diyagramı 7
Şekil 3.1. Önerilen SFIX modelinin bileşen tabanlı mimarisi 87
Şekil 3.2. Kredi başvuru veriseti özellik skorları 101
Şekil 3.3. SHAP modelinin kredi başvuru verisetine uygulanmasının sonucu 103
Şekil 3.4.LIME modelinin kredi başvuru verisetine uygulanmasının sonucu 103
Şekil 3.5.LIME, SHAP, S-SFIX kararlılık sonucu 105
Şekil 3.6. SHAP, LIME ve SFIX için özellik önem skorlarının kümülatif dağılımını gösteren Lorenz eğrileri..... 110
Şekil 3.7. 10 bootstrap örneklemi boyunca SHAP, LIME ve SFIX için kümülatif önem dağılım eğrileri 112
Şekil 3.8. 10 bootstrap örneklemi boyunca SHAP, LIME ve SFIX için özellik bazlı önem dağılım grafikleri 114
Şekil 4.1. SFIX modelinin akış şeması 123
Şekil 5.1. S-SFIX tarafından en yüksek öneme sahip olarak belirlenen ilk 10 özellik 164
Şekil 5.2. Gaussian gürültüsü seviyelerine göre ilk k özellik kümesi kararlılığı (Jaccard benzerliği)..... 175
Şekil 5.3. Gaussian gürültüsü seviyelerine göre ortalama mutlak sıra kayması 176
Şekil 5.4. Gaussian gürültüsü seviyelerine göre S-SFIX sıralama yapısının Kendall Tau korelasyonu..... 177
Şekil 5.5. Gaussian gürültüsü seviyelerine göre S-SFIX sıralama yapısının Spearman korelasyonu..... 178

TABLO LİSTESİ

Tablo 1.1. AI ile ilişkili riskler 2
Tablo 2.1. Çeşitli açıklanabilirlik algoritmalarının sayıları 24
Tablo 2.2. Uygulama alanları ve kullanılan algoritmalar 25
Tablo 2.3. Finansta IoT içerisinde XAI yöntemlerine genel bakış 31
Tablo 2.4. XAI yöntemlerinin katkıları ve performansı 32
Tablo 2.5. XAI yöntemlerinin karşılaştırmalı analizi 33
Tablo 2.6. Veri özellikleri ve zorluklara ilişkin tartışma 35
Tablo 2.7. XAI yöntemlerinin güçlü yönleri ve sınırlılıkları 36
Tablo 2.8. XAI yöntemlerine genel bakış 40
Tablo 2.9. Araştırma soruları ve destekleyici tablolar 59
Tablo 3.1. DIS bileşeninin parametreleri 88
Tablo 3.2. CS bileşeninin parametreleri 88
Tablo 3.3. PS bileşeninin parametreleri 89
Tablo 3.4. AS bileşeninin parametreleri 89
Tablo 3.5. RS bileşeninin parametreleri 89
Tablo 3.6. Zamansız formüller ve değişiklikler 91
Tablo 3.7. Model bileşenlerinin zaman karmaşıklığı özeti 94
Tablo 3.8. Model bileşenlerinin zaman karmaşıklığı değerlendirmesi 96
Tablo 3.9. Kredi başvuru veriseti açıklaması 99
Tablo 3.10. S-SFIX modelinin kredi başvuru verisetine uygulanmasının sonucu 102
Tablo 3.11. SFIX modeli için kredi başvuru verisetindeki en önemli 5 özellik 102
Tablo 3.12. LIME, SHAP ve S-SFIX'ten elde edilen özellik önem sıralamaları 104
Tablo 3.13. Model kararlılık metriklerinin karşılaştırılması 105
Tablo 3.14. SFIX, SHAP ve LIME'ın karşılaştırmalı genel görünümü 107
Tablo 3.15. Bootstrap örneklemeleri boyunca SHAP, LIME ve SFIX yöntemlerinin dayanıklılık karşılaştırması 115
Tablo 3.16. SHAP, LIME ve SFIX'in yapısal ve işlevsel özellikler açısından karşılaştırılması 117
Tablo 4.1. SFIX bileşenlerinin SHAP ve LIME'daki kavramsal olarak ilişkili açıklanabilirlik boyutları ile karşılaştırılması 125
Tablo 4.2. SFIX analiz sonuçları – orijinal veriseti (N=10572) 126
Tablo 4.3. SFIX analiz sonuçları – tamamlanmış veriseti (N=23830) 128
Tablo 4.4. SFIX model performansının karşılaştırmalı analizi 129

Tablo 4.5. SFIX bileşenlerinin orijinal ve tamamlanmış verisetleri üzerindeki karşılaştırmalı istatistiksel analizi 130
Tablo 4.6. SFIX kullanılarak ilk 20 özelliğin karşılaştırılması 131
Tablo 4.7. Finansal özellik kategorilerinin orijinal ve tamamlanmış verisetleri açısından karşılaştırmalı performansı 132
Tablo 4.8. Eksik veri tamamlamanın model performansı üzerindeki etkisi 133
Tablo 4.9. En önemli 5 özelliğin SFIX bileşenleri için ayrıntılı bileşen karşılaştırması 134
Tablo 4.10. Orijinal ve tamamlanmış verisetlerinde ilk 20 özelliğin sıra karşılaştırması 135
Tablo 4.11. Model ve veriseti yapılandırmalarının performans karşılaştırması 136
Tablo 4.12. Özellik sayısının FFNN model performansına etkisi 137
Tablo 4.13. Genel model performans özeti ve SFIX etkililik analizi 138
Tablo 4.14. Finansal verisetinde FFNN tabanlı modellerin performans karşılaştırması 140
Tablo 5.1. CICIoT2023 verisetindeki etiketlerin dağılımı 150
Tablo 5.2. Saldırı üst sınıfı altında yer alan saldırı alt türlerinin dağılımı 152
Tablo 5.3. Eğitim, doğrulama ve test kümelerinde sınıf dağılımı 154
Tablo 5.4. Deneysel veri hazırlama sürecinin aşamaları 155
Tablo 5.5. Karşılaştırmalı özellik seçimi deney düzeni 157
Tablo 5.6. Gürültüsüz verisetinde S-SFIX bileşen skorları ve nihai özellik skorları .. 162
Tablo 5.7. Gaussian gürültüsü altında S-SFIX stabilite deney düzeni 167
Tablo 5.8. Farklı XAI yöntemleri ve özellik sayıları altında FFNN performans sonuçları 168
Tablo 5.9. Gaussian gürültüsüne göre 46 özelliğin S-SFIX sıralama değişimi 170
Tablo 5.10. Gaussian gürültüsüne göre S-SFIX özellik sıralama değişimi 173
Tablo 5.11. Noise = 0.20 seviyesinde en fazla sıra değişimi gösteren özellikler 174
Tablo 5.12. İlk 10 özelliğin Gaussian gürültüsü altındaki konumları 174
Tablo 5.13. Gaussian gürültüsü seviyesine göre S-SFIX dayanıklılık metrikleri 179

1 GİRİŞ

Yapay Zekâ (Artificial Intelligence (AI)), finans sektöründe dönüştürücü bir güç haline gelmiş; müşteri deneyimi, dolandırıcılık tespiti, yatırım yönetimi, kredi skorlama ve risk analizi alanlarındaki uygulamaları köklü biçimde değiştirmiştir (Bahoo et al., 2024; Aldasoro et al., 2025). Küresel benimsenme oranı artmaya devam etmekte olup, tahminler AI'ın finans sektöründeki pazar payının 2023 yılında %18'e ulaştığını ve 2032 yılına kadar 3 trilyon $'ı aşmasının beklendiğini göstermektedir (Fortune Business Insights, 2023). NVIDIA'nın 2024 raporu finansal kuruluşların %90'ından fazlasının AI teknolojilerini ya kullandığını ya da kullanmayı planladığını göstermektedir (NVIDIA, 2024).

AI önemli faydalar sunsa da artan kullanımı aynı zamanda kritik bir sınırlılığı olan açıklanabilirlik eksikliğini ortaya koymaktadır. Birçok AI sistemi, yorumlanması güç sonuçlar üreten "kara kutular" olarak çalışmaktadır. Bu kapalı yapı, kullanıcı güveninin kazanılmasında ve GDPR, AI Act gibi düzenlemelere uyum sağlanmasında zorluklar doğurmaktadır (European Parliament, 2023; Wachter ve ark., 2017). Düzenleyici kurumlar artık AI tarafından alınan finansal kararların yorumlanabilir, hesap verebilir ve adil olması gerektiğini vurgulamaktadır.

Bu ihtiyaçları karşılamak amacıyla, Açıklanabilir Yapay Zekâ (XAI) merkezi bir kavram olarak ortaya çıkmıştır. XAI yöntemleri, performanstan ödün vermeden AI destekli kararların insanlar tarafından anlaşılabilir olmasını amaçlamaktadır. McKinsey'ye göre açıklanabilirlik, kurumların AI sistemlerine duyulan güveni artırmasına, riskleri azaltmasına ve düzenleyici uyumu güçlendirmesine yardımcı olmaktadır (McKinsey ve Company, 2022). Bu bağlamda, XAI uyum ve etik zorunluluk olmaktadır.

Finansta açıklanabilir olmayan AI'ın önemli riskleri mevcuttur. Bu riskler arasında, özellikle kredi skorlama, portföy optimizasyonu veya sigorta onayları gibi alanlarda modellerin "kara kutu" niteliği yer almakta; bu durum şeffaflığı azaltmakta ve hesap verebilirliği zorlaştırmaktadır. Bir diğer endişe algoritmik önyargıdır. Hatalı verisetleriyle eğitilen modeller belirli grupları sistematik olarak dezavantajlı duruma düşürebilmekte ve bu da hukuki ve etik sorunlara yol açabilmektedir. Ayrıca, karmaşık ve kara kutu AI sistemleri tespit edilmesi ve savunulması güç güvenlik açıkları içerebilmekte; bu durum finansal kurumları dolandırıcılık ve siber saldırılara maruz

bırakabilmektedir.

Tablo 1.1'de, The Economist Intelligence Unit tarafından gerçekleştirilen anket, finansal hizmetlerde AI kullanımına ilişkin temel endişeleri ortaya koymaktadır.

Tablo 1.1. AI ile ilişkili riskler

Risk Kategorisi	**Yüzde (%)**
Güvenlik ve Gizlilik İhlalleri	46.6
AI Sistemlerinin Başarısızlığı	40.5
AI Kararlarından Kaynaklanan Hukuki Sorumluluk	32.2
İş Gücü / İş Gücü Yer Değişimi	19.5
Müşteri Güveninin Kaybı	17.6
Etik Riskler	16.6
Düzenleyici Uyumsuzluk	16.1

Tablo 1.1'de görüldüğü üzere, AI teknolojilerine duyulan ihtiyaç arttıkça veri koruma ve yetkisiz erişime ilişkin endişeler ön plana çıkmıştır. Benzer şekilde, AI sistemlerinin teknik olarak ya da beklenen çıktıları sunma açısından başarısız olma potansiyeli de ciddi bir risk olarak görülmektedir (The Economist Intelligence Unit, 2022).

1.1 Tez Çalışmasının Amacı ve Önemi

Finansal karar süreçlerinde açıklanabilirlik gereksiniminin artmasının temel nedenlerinden biri, bu alandaki kararların yüksek risk içermesi ve doğrudan ekonomik sonuçlar doğurmasıdır. Kredi onayı, risk sınıflandırması, denetim görüşü belirleme ya da şüpheli işlem tespiti gibi uygulamalarda model tarafından üretilen sonuçların gerekçelendirilebilir olması büyük önem taşımaktadır. Aksi durumda, model çıktılarının kullanıcılar, uzmanlar ve düzenleyici otoriteler tarafından değerlendirilmesi güçleşmekte; bu durum güven, adalet, uyum ve operasyonel uygulanabilirlik açısından önemli sorunlara yol açmaktadır. Dolayısıyla finans sektöründe yalnızca doğru çalışan değil, aynı zamanda karar mantığını anlaşılır biçimde ortaya koyabilen yapay zekâ çözümlerine ihtiyaç duyulmaktadır.

Bu tez çalışmasının başlangıç motivasyonu, mevcut açıklanabilir yapay zekâ yaklaşımlarının finans alanındaki ihtiyaçları her zaman yeterli düzeyde karşılayamamasından ortaya çıkmıştır. Literatürde yaygın olarak kullanılan yöntemler

önemli katkılar sunmakla birlikte, çoğu zaman genel amaçlı bir yapıda kalmakta; finansal verilerin risk duyarlılığı, veri kalitesi, örüntü davranışı, anomali yapısı ve bazı durumlarda zamansal karakteristiği gibi alan özgü özelliklerini bütüncül biçimde ele alamamaktadır. Ayrıca mevcut çalışmaların önemli bir bölümünde açıklanabilirlik, çoğunlukla model çıktılarının sonradan yorumlanması şeklinde ele alınmakta; açıklamaların tutarlılığı, kararlılığı ve farklı veri koşulları altındaki dayanıklılığı sınırlı biçimde incelenmektedir. Bu durum, finans gibi güvene dayalı ve düzenlemelere tabi alanlarda daha kapsamlı ve alan odaklı açıklanabilirlik yaklaşımlarının geliştirilmesini gerekli kılmaktadır.

Bu kapsamda tez çalışmasının temel amacı, finans sektöründe kullanılan yapay zekâ tabanlı sistemler için daha açıklanabilir, güvenilir ve dayanıklı bir yaklaşım geliştirmektir. Bu amaç doğrultusunda çalışmada öncelikle finans alanında açıklanabilir yapay zekâ uygulamalarının mevcut durumu incelenmiş, kullanılan yöntemler, uygulama alanları, verisetleri ve araştırma boşlukları sistematik biçimde değerlendirilmiştir. Söz konusu literatür incelemesi tez kapsamında ayrı bir bölümde ayrıntılı olarak sunulmakta ve alanın mevcut durumuna ilişkin kapsamlı bir çerçeve ortaya koymaktadır. Ardından, finansal verilerin yapısal özelliklerini dikkate alan özgün bir açıklanabilirlik modeli önerilmiş; bu model özellik önemi, güven, örüntü, anomali ve risk boyutlarını birlikte ele alan bütünleşik bir yapı içerisinde geliştirilmiştir. Sonraki aşamada önerilen yaklaşımın finansal uygulamalardaki etkinliği, açıklama gücü, performans katkısı ve farklı koşullar altındaki kararlılığı deneysel olarak değerlendirilmiştir.

Tez çalışmasının önemi hem kuramsal hem de uygulamalı düzeyde ortaya çıkmaktadır. Kuramsal açıdan çalışma, finans alanında açıklanabilir yapay zekâ literatüründe öne çıkan boşluklara yanıt vermekte ve alan özgü bir model önererek mevcut bilgi birikimine katkı sunmaktadır. Uygulamalı açıdan ise çalışma, finansal karar süreçlerinde daha şeffaf ve güvenilir yapay zekâ sistemlerinin geliştirilmesine yönelik bir çerçeve sağlamaktadır. Bu yönüyle tez, yalnızca mevcut yöntemleri değerlendiren bir çalışma olmanın ötesine geçmekte; finans sektörünün ihtiyaçlarına cevap verebilecek özgün bir yaklaşım ortaya koymaktadır. Sonuç olarak bu tez çalışması, finans alanında kullanılabilecek daha açıklanabilir, güvenilir, dayanıklı ve uygulanabilir yapay zekâ sistemlerinin geliştirilmesine katkı sağlaması bakımından önem taşımaktadır.

1.2 Tez Organizasyonu

Tez çalışması altı bölümden oluşmaktadır. İlk bölümde, yapay zekânın finans sektöründeki artan kullanımına bağlı olarak ortaya çıkan açıklanabilirlik, güvenilirlik, şeffaflık ve düzenleyici uyum gereksinimleri ele alınmaktadır. Bu kapsamda, özellikle finansal karar destek sistemlerinde kullanılan kara kutu modellerin neden açıklanabilirlik sorunu oluşturduğu tartışılmakta ve çalışmanın problemi, amacı ve motivasyonu ortaya konulmaktadır.

İkinci bölümde, finans alanında açıklanabilir yapay zekâ uygulamalarına yönelik sistematik literatür taraması sunulmaktadır. Bu bölümde öncelikle araştırma soruları tanımlanmakta, ardından seçim kriterleri, veri tabanları, tarama süreci ve kalite değerlendirme adımları açıklanmaktadır. Sonrasında seçilen çalışmalar tematik olarak incelenmekte; finans alanında hangi açıklanabilir yapay zekâ yöntemlerinin hangi problemlerde kullanıldığı, hangi veri setlerinin tercih edildiği ve mevcut literatürde hangi boşlukların bulunduğu değerlendirilmektedir. Böylece, tez kapsamında geliştirilen yaklaşımın neden gerekli olduğu sistematik bir kuramsal temel üzerinden gösterilmektedir.

Üçüncü bölümde, tezin ana yöntemsel katkısını oluşturan SFIX modeli ayrıntılı biçimde tanıtılmaktadır. Bu bölümde modelin Dinamik Özellik Önem Skoru, Güven Skoru, Örüntü Skoru, Anomali Skoru ve Risk Skoru bileşenleri ayrı başlıklar altında açıklanmakta; toplam XAI skorunun oluşturulma süreci, parametre yapısı, sadeleştirilmiş SFIX yaklaşımı, zaman karmaşıklığı ve modelin özgünlüğü değerlendirilmektedir. Bunun yanında, önerilen modelin uygulanması, performans değerlendirmesi ve SFIX'in diğer açıklanabilirlik yaklaşımlarıyla karşılaştırılması da bu bölümde yer almaktadır. Böylece bu bölüm, tezin kuramsal çerçevesini ve önerilen modelin teknik temelini oluşturmaktadır.

Dördüncü bölümde, önerilen modelin finans alanındaki uygulaması gerçekleştirilmektedir. Bu kapsamda SFIX tabanlı yaklaşım, denetim görüşü sınıflandırması problemine uygulanmakta ve ileri beslemeli sinir ağı ile bütünleştirilmektedir. Bölümde veri ön işleme süreci, deneysel tasarım, SHAP ve LIME ile yapılan karşılaştırmalar, orijinal ve tamamlanmış veri kümeleri üzerindeki bulgular, özellik sıralamalarının analizi, model performans karşılaştırmaları, SFIX bileşenlerinin

yorumlanması ve elde edilen sonuçların değerlendirilmesi sunulmaktadır. Ayrıca bu bölümde yöntemin sınırlılıkları, başarısızlık durumları ve finansal denetimde gerçek dünya uygulamasına ilişkin pratik avantajlar ile zorluklar da tartışılmaktadır. Bu yönüyle dördüncü bölüm, önerilen modelin finansal karar sistemleri üzerindeki etkinliğini ortaya koyan temel uygulama bölümüdür.

Beşinci bölümde, modelin yalnızca finansal bağlamda değil, farklı bir uygulama alanında da değerlendirilebilmesi amacıyla yapay zekâ dayanıklılığı perspektifi ele alınmaktadır. Bu bölümde IoT ağ trafiği veriseti kullanılarak veri hazırlama, alt küme oluşturma, dengeleme, eğitim-doğrulama-test ayrımı ve ön işleme adımları açıklanmaktadır. Ardından önerilen FFNN modeli, SFIX ve Simplified SFIX'in deneysel ortama entegrasyonu, S-SFIX eğitimi ve özellik sıralaması sunulmaktadır. Bölümün devamında performans değerlendirmesi, Gaussian noise altında dayanıklılık sonuçları, özellik seçme yeteneği, alternatif yöntemlerle karşılaştırmalar, açıklanabilirlik-performans ilişkisi, gürültü altındaki kararlılık ve alanlar arası uygulanabilirlik tartışılmaktadır. Böylece bu bölüm, tezin önerdiği yaklaşımın dayanıklılık ve genellenebilirlik yönünü ortaya koymaktadır.

Altıncı ve son bölümde ise tez boyunca elde edilen bulgular bütüncül biçimde değerlendirilmekte ve çalışmanın genel sonuçları ortaya konulmaktadır. Bu bölümde, literatür taramasından elde edilen çıkarımlar, SFIX modelinin geliştirilmesi, denetim görüşü sınıflandırmasındaki uygulama sonuçları ve dayanıklılık deneyleri birlikte ele alınarak tezin temel katkıları özetlenmektedir. Sonuç olarak tez, finansal karar sistemlerinde açıklanabilir yapay zekânın yalnızca destekleyici bir unsur değil, karar mekanizmasının yapısal ve güvenilir bir bileşeni olarak ele alınması gerektiğini göstermektedir.

2 LİTERATÜR ÇALIŞMASI

Bu bölümde, finans alanında AI kullanımına bağlı potansiyel risklerin azaltılmasında açıklanabilir yapay zekânın rolünü incelemektedir. Buna bağlı olarak, çeşitli finansal hizmetlerde XAI uygulayan çalışmalara odaklanan kapsamlı bir literatür taraması gerçekleştirilmiştir.

2.1 Metodoloji

2.1.1 Araştırma sorularının belirlenmesi

Bu inceleme aracılığıyla çalışma, aşağıdaki araştırma sorularını (RQ) ele almayı amaçlamaktadır:

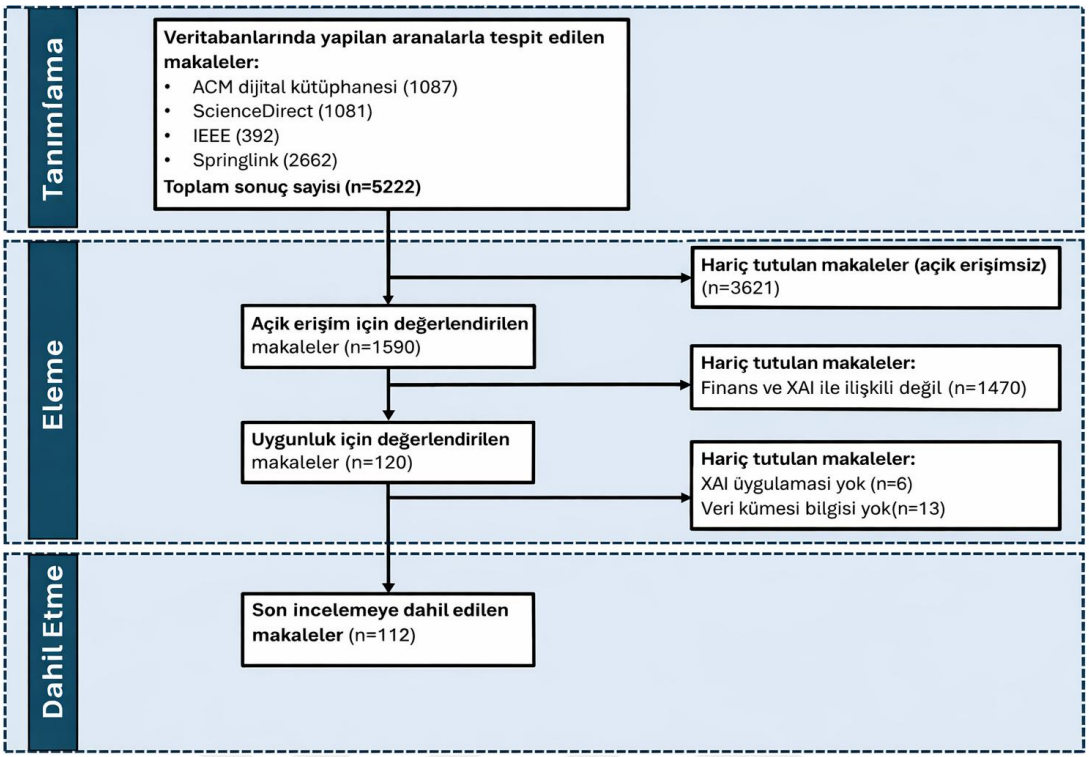
- RQ1: Hangi XAI algoritmaları hangi finansal hizmetler için kullanılmaktadır?
- RQ2: XAI algoritmaları finansal hizmetler için uygun mudur?
- RQ3: XAI kullanımının sunulan finansal hizmetlerin performansı üzerindeki etkisi nasıldır?
- RQ4: Finansal hizmetlerde XAI algoritmalarının uygulamalarında ne tür verisetleri tercih edilmektedir?

2.1.2 Seçim kriterleri

Bu bölümde, finans alanında XAI uygulamalarına ilişkin literatür taraması için akademik çalışmaların hangi kriterlere göre seçildiği verilmektedir.

Literatür çalışmasında sistematik derlemelerin ve analizlerin şeffaflığını ve kalitesini artırmak amacıyla geliştirilen PRISMA kılavuzu kullanılmıştır (Page ve ark., 2021).

Şekil 2.1, sistematik literatür çalışmasını göstermektedir.



Şekil 2.1. Çalışma seçim süreci için PRISMA akış diyagramı

Sistematik literatür taraması, belirli ve yapılandırılmış adımlar izlenerek gerçekleştirilmiştir. İlk aşamada başlangıç veritabanı seçimi yapılmıştır. ACM Digital Library, Springlink, IEEE Xplore ve Science Direct veritabanları, hem teknik hem de finansal araştırmalar açısından kapsamlı kaynaklar sunmaları nedeniyle tercih edilmiştir. Bu veritabanları, yapay zekâ ve finans ile ilişkili yayınlara geniş kapsamlı erişim sağlamaktadır. İkinci aşamada arama sorgusu geliştirilmiştir. Arama sorgusu hem AI hem de finans alanındaki ilgili makaleleri kapsayacak şekilde dikkatlice oluşturulmuş, kapsamlılığı sağlarken ilgililiği korumak amacıyla Boolean operatörleri kullanılmıştır. Arama süreci, alandaki güncel gelişmelere odaklanmak amacıyla 2021-2025 yılları ile sınırlandırılmıştır. Bu kapsamda ilgili veritabanlarında 2021-2025 yılları esas alınarak aşağıdaki sorgular kullanılmıştır:

- (“Finance” OR “Financial” OR “Bank” OR “Payment” OR “Electronic Money” OR “Loan” OR “Stock” OR “Insurance” OR “Economy”) AND (“Explainable Artificial Intelligence” OR “XAI” OR “Counterfactual Explanations” OR “Explainable machine learning” OR “Explainable deep learning”)
- (“Electronic Money” OR “Loan” OR “Stock” OR “Insurance” OR “Economy”) AND (“Explainable Artificial Intelligence” OR “XAI” OR “Counterfactual

Explanations" OR "Explainable machine learning" OR "Explainable deep learning").

Makale tarama sürecinde ilk arama sonucunda 5222 makale elde edilmiş ve bu makaleler üç aşamalı bir tarama sürecinden geçirilmiştir. Birinci aşamada başlık ve özet taraması yapılmış, makaleler finansta XAI ile ilgililik açısından incelenmiştir. Bu aşamada finansal uygulamalara odaklanmayan makaleler, açıklanabilirlik boyutu olmaksızın yalnızca AI'ı ele alan makaleler ve İngilizce olmayan makaleler hariç tutulmuştur. İkinci aşamada tam metin incelemesi gerçekleştirilmiş ve kalan makaleler; yöntembilimin açık biçimde tanımlanması, sonuçların ampirik olarak doğrulanması, veriseti ayrıntıları ile deneysel kurulumun sunulması ve XAI yöntemlerinin uygulanması ölçütlerine göre değerlendirilmiştir. Üçüncü aşamada ise kalite değerlendirmesi yapılmış ve makaleler araştırma amaçlarının açıklığı, metodolojinin titizliği, sonuçların ve çıkarımların geçerliliği ile finans sektörü açısından pratik çıkarımlar sunması bakımından incelenmiştir.

Nihai seçim sürecinde başlangıçtaki 5222 makaleden açık erişim olmayan 3621 yayın çıkarılmış, finans sektörü ve XAI odağı bulunmaması nedeniyle 1470 makale hariç tutulmuş, yetersiz XAI uygulama ayrıntıları nedeniyle 6 makale çıkarılmış ve eksik veriseti bilgisi nedeniyle 13 makale elenmiştir. Bunun sonucunda, tüm dâhil etme kriterlerini ve kalite standartlarını karşılayan 112 makale nihai inceleme için seçilmiştir. Arama sonucunda 5222 makale belirlenmiş; konu ile ilgisiz olduğu değerlendirilenler ve bu çalışma için veriseti ayrıntısı içermeyenler çıkarıldıktan sonra 112 makaleye odaklanılmıştır.

Veri çıkarma ve analiz çerçevesi kapsamında, seçilen her makale için belirli bilgiler sistematik olarak çıkarılmıştır. Bu bilgiler; makalenin odak alanı, metodoloji tanımı ve özeti, makalede kullanılan XAI yöntemi, makalede kullanılan verisetleri, makalenin amacı, problem için getirilen çözüm önerisi, kullanılan yöntemin başarı oranı, makalede elde edilen sonuç, makale için gelecek çalışmalar ve makaledeki yöntemin uygulama alanını kapsamaktadır.

Bir sonraki bölümde, yukarıda belirtilen başlıklara dayalı tematik analize yardımcı olmak amacıyla kuramsal arka plan oluşturulmaya çalışılmaktadır.

2.1.3 Kuramsal çerçeve

Bu bölüm, daha açık bir çerçeve sunmak amacıyla literatür taramasında karşılaşılan XAI yöntemlerini açıklamaktadır. XAI yöntemleri farklı kategoriler altında ele alınmaktadır. Bunlar genel olarak modelden bağımsız ve modele özgü yaklaşımlar altında sınıflandırılmaktadır. Buna ek olarak, açıklamalar yerel ve küresel yaklaşımlar olarak da ayrılabilmektedir.

2.1.3.1 Modelden bağımsız yaklaşımlar

Aşağıda, bu çalışmada yer alan yöntemlere uygun XAI yaklaşımlarına göre bir sınıflandırma yapılmıştır.

LIME (Local Interpretable Model-Agnostic Explanations), karmaşık modellerin yerel tahminlerini basit ve anlaşılır modellerle açıklar. Her tahmin için basit bir model (örneğin doğrusal regresyon) oluşturur ve tahminlerin mantığını açıklar. Denklem 2.1'de görüldüğü üzere, LIME, hem sınıflandırma hem de regresyon problemlerinde kullanılabilmektedir (Ribeiro ve ark., 2016).

$$f(x) \approx g(x') = \beta_0 + \sum_{i=1}^{m} \beta_i x_i' \qquad (2.1)$$

- f(x): Açıklanmakta olan özgün kara kutu modelin tahmini
- g(x'): LIME tarafından oluşturulan basit, yerel açıklama modeli (örneğin doğrusal regresyon).
- β_0: Sabit terim.
- β_i: i özelliğinin ağırlığı (önemi).
- x'_i: Yerel komşulukta i özelliğinin pertürbe edilmiş değeri

Pratik Örnekleme (LIME – Müşteri Elde Tutma): Nakatumba-Nabende ve Marvin (2023) çalışmasında, perakende bankacılık sektöründe müşteri kaybını tahmin etmek amacıyla hem gerçek dünya hem de kıyaslama verisetleri (Berka ve Kaggle) kullanılarak yorumlanabilir bir makine öğrenmesi çerçevesi geliştirilmiştir. Yazarlar, Random Forest ve Decision Tree gibi ağaç tabanlı modelleri kullanmış ve güçlü tahmin performansı elde etmiştir (%99'a kadar doğruluk). Kritik olarak, LIME bireysel müşteri kaybı tahminlerini açıklamak için yerel vekil modeller üretmek amacıyla kullanılmıştır. Örneğin, LIME analizi belirli vakalarda azalan giriş sıklığının ve yüksek sayıda

çözümlenmemiş müşteri şikâyetinin müşteri kaybının temel belirleyicileri olduğunu ortaya koymuştur. Bu yorumlanabilir içgörüler, banka analistlerinin müşteri kaybı riskini azaltmak ve müşteri sadakatini artırmak için kişiselleştirilmiş iletişim veya hizmet iyileştirmeleri gibi müşteriye özgü elde tutma stratejilerini proaktif biçimde tasarlamasına olanak tanımıştır.

SHAP, oyun teorisine dayalı olarak her özelliğin modele katkısını hesaplar. Hem küresel hem de yerel açıklamalar sunar. SHAP, her özelliğin model çıktısına ne kadar katkıda bulunduğunu gösterir. Bu yöntem özellikle karmaşık modellerde kullanılabilir ve farklı veri türleri için uygundur (Lundberg ve Lee, 2017b).

SHAP, her özellik için Shapley değerlerini hesaplayarak model tahminini aşağıdaki Denklem 2.2 ifade eder:

$$f(x) = \phi_0 + \sum_{i=1}^{M} \phi_i \tag{2.2}$$

- f(x): İlgili girdi için model tarafından üretilen tahmin değeri.
- Φ_0: Herhangi bir özellik katkısı olmaksızın modelin başlangıç ya da referans tahmini; genellikle ortalama tahmin değerini temsil eder.
- Φ_i: i özelliğinin model tahminine yaptığı katkıyı ifade eden Shapley değeri.
- M: Modelde kullanılan toplam özellik sayısı.

Her bir Φ_i Denklem 2.3'te gösterilen bir Shapley değeri olarak hesaplanır:

$$\Phi_i = \sum_{S \subseteq N \setminus i} \frac{|S|!(M-|S|-1)!}{M!} [f_x(S \cup i) - f_x(S)] \tag{2.3}$$

Burada:

- S: Özelliklerin herhangi bir alt kümesini göstermektedir.
- N: Modelde yer alan tüm özelliklerin kümesini göstermektedir.
- $f_x(S)$: S alt kümesinde bulunan özellikler kullanılarak üretilen model tahminini ifade etmektedir.
- $f_x(S \cup \{i\})$: S alt kümesine i özelliği eklendiğinde elde edilen model tahminini ifade etmektedir.
- (|S|! (M - |S| - 1)!) / M!: i özelliğinin marjinal katkısının tüm olası özellik alt

kümeleri üzerinden adil ve dengeli biçimde hesaplanmasını sağlayan ağırlıklandırma faktörünü ifade etmektedir.

Bu formüller ve indisler, SHAP'ın her özelliğin model tahminine katkısını nasıl hesapladığını göstermektedir.

Pratik Örnekleme (SHAP – Kredi Skorlama): De Lange ve ark.(2022) kredi onay kararlarında yorumlanabilirliği artırmak amacıyla LightGBM ile SHAP'ı birleştiren açıklanabilir bir kredi skorlama modeli uygulamıştır. Norveç'teki bir bankadan alınan teminatsız tüketici kredilerine ait özel bir verisetini kullanarak, SHAP'ın karmaşık model çıktılarının ayrıştırılmasında oldukça etkili olduğunu bulmuşlardır. Özellikle SHAP değerleri, kredi bakiyesi oynaklığı ve kalan kredi yüzdesi gibi ret kararlarını yönlendiren temel faktörlerin belirlenmesine yardımcı olmuştur. Bu ayrıntılı yorumlanabilirlik, banka görevlilerinin kararlarını müşterilere ve düzenleyicilere daha iyi gerekçelendirmesine olanak tanımış, hatta müşterilere mali profillerini iyileştirmeleri konusunda rehberlik de sunmuştur.

Permütasyon önemlilik tekniği, bir özellik rastgele karıştırıldığında model performansı üzerindeki etkisini ölçer. Performans düşerse, bu özellik model için önemlidir. Bu yöntem doğrusal olmayan modellerde de başarılı sonuçlar vermektedir (Breiman, 2001; Fisher ve ark., 2019).

j özelliği için permütasyon önemlilik formülü Denklem 2.4'teki gibidir:

$$I_j = \frac{1}{n}\sum_{i=1}^{n}\left(L(f(x_i), y_i) - L\left(f\left(x_i^{\pi_j}\right), y_i\right)\right) \quad (2.4)$$

Burada:

- I_j: j özelliği için hesaplanan permütasyon önemlilik değerini ifade etmektedir.
- n: Verisetinde yer alan toplam örnek sayısını göstermektedir.
- $\sum_{i=1}^{n}$: Verisetindeki tüm örnekler için i indisi üzerinden gerçekleştirilen toplama işlemini ifade etmektedir.
- $L(f(x_i), y_i)$: Modelin özgün veri noktası x_i için ürettiği tahmine karşılık gelen kayıp fonksiyonunu göstermektedir.
- $L(f(x_i^{\pi_j}), y_i)$: j özelliğinin permüte edildiği $x_i^{\pi_j}$ veri noktası için model tarafından

hesaplanan kayıp fonksiyonunu göstermektedir.

Bu formül, j özelliğinin değerleri rastgele permüte edildiğinde model performansındaki ortalama değişimi hesaplar. Daha büyük bir I_j değeri, j özelliğinin permüte edilmesinin kayıpta önemli bir artışa yol açtığını gösterir; bu da j özelliğinin doğru model tahminleri için daha önemli olduğunu düşündürmektedir.

Daha yüksek permütasyon önemlilik skorlarına (I_i) sahip özellikler, permüte edildiklerinde performansta daha büyük bir düşüşe neden oldukları için daha etkili kabul edilir. Bu yöntem, her özelliğin modelin doğruluğuna nasıl katkı sunduğuna ilişkin içgörü sağlar.

Pratik Örnekleme (Permütasyon önemlilik – Yatırım Risk Skorlama): Uddin ve ark. (2025) Random Forest ve LSTM dâhil olmak üzere çeşitli makine öğrenmesi modellerini kullanarak finansal risk tahmini ve portföy optimizasyonunu incelemiştir. Araştırmada, çeşitli piyasa koşulları altında hisse senedi portföy davranışına odaklanılarak her özelliğin model doğruluğu üzerindeki etkisini değerlendirmek amacıyla permütasyon önemlilik teknikleri kullanılmıştır. Sektör beta değeri, makroekonomik göstergeler ve oynaklık ölçütleri gibi özellikler test edilmiştir. Permütasyona dayalı analiz, piyasa genelindeki oynaklık göstergelerinin likidite oranları veya firma düzeyindeki verilerden önemli ölçüde daha yüksek tahmin gücüne sahip olduğunu ortaya koymuştur. Bu bulgular, portföy yöneticilerinin proaktif risk yönetimi için yüksek etkili makro göstergelere odaklanmasına olanak tanımıştır. Yerel yorumlanabilirlik yöntemlerinden farklı olarak, permütasyon önemlilik küresel bir bakış açısı sunarak modelden bağımsız stratejik portföy kararlarına rehberlik etmiştir.

ALE (Accumulated Local Effects), modeldeki özelliklerin tahmin üzerindeki etkisini ölçer. Özellikle ALE, özellikler arasındaki etkileşimleri dikkate alır (W.Apley ve Zhu, 2020).

Accumulated Local Effects (ALE) formülü Denklem 2.5'teki şekilde verilmektedir:

$$\mathrm{ALE}_j(x) = \int_{x_{\min}}^{x} E\left[\frac{\partial f(X)}{\partial X_j}\middle| X_j = t\right] dt \qquad (2.5)$$

Bu formül, özellik j'nin x değerine kadar olan birikimli yerel etkisini, model tahminleri üzerindeki marjinal etkilerin ortalamasını alarak hesaplamaktadır.

- $ALE_j(x)$: j özelliğinin belirli bir x değerindeki birikimli yerel etkisini ifade etmektedir.
- x_{min}: Veriseti boyunca j özelliğinin aldığı minimum değeri göstermektedir.
- $E[\partial f(X) / \partial X_j \mid X_j = t]$: Tahmin fonksiyonu f(X)'in, $X_j = t$ koşulu altında X_j değişkenine göre beklenen kısmi türevini ifade etmektedir. Bu terim, X_j'de meydana gelen değişimlerin model tahmini üzerindeki ortalama marjinal etkisini göstermektedir.
- t: X_j boyunca integrasyon için kullanılan değişkeni ifade etmektedir.

Bu integral, özellik j'nin değer aralığı boyunca yerel etkilerini biriktirerek, x değerine kadar özellik j'nin model tahmini üzerindeki kümülatif etkisini sunmaktadır.

Pratik Örnekleme (ALE – Mortgage Approval Scoring): Çoğu XAI yöntemi örnek düzeyinde içgörüler sunarken, Gkolemis ve ark.(2022) tarafından önerilen DALE çerçevesi (Differential Accumulated Local Effects), özellikle ipotek onay sistemleri gibi karmaşık finansal modellerde faydalı olan, daha ölçeklenebilir ve daha sağlam küresel açıklamalar sunmak amacıyla ALE yöntemini geliştirmektedir. Makalede doğrudan mortgage bağlamında uygulanmamış olsa da, yöntem ALE'nin geleneksel zayıflıklarını, (örneğin dağılım dışı örnekleme ve yüksek boyutlu kararsızlık) ele almaktadır; bunlar kredi risk skorlama uygulamalarında yaygın sorunlardır. Bu tür senaryolarda ALE, Kredi/Değer (Loan-to-Value, LTV) oranı ve gelir gibi değişkenlerin farklı aralıklarındaki doğrusal olmayan etkileri görselleştirerek, belirli LTV eşiklerinin veya gelir tavanlarının altında onay olasılığındaki azalan getirileri ortaya koyabilir. Bu içgörüler, şeffaf ve düzenlemelere uyumlu kredi politikalarının tasarlanması açısından özellikle değerlidir.

Generalized Additive Model (GAM), genelleştirilmiş doğrusal modellerin (GLM) esnek bir uzantısıdır ve her özelliğin tahmin üzerindeki etkisinin bağımsız olarak modellenmesine olanak tanır. GAM, her girdiyi bir fonksiyon olarak ifade eder ve bu fonksiyonların toplamını alarak tahmin üretir. Bu, model karmaşıklığı ile şeffaflık arasında bir denge sağlar. GAM, özellikle tahminlerin anlaşılabilirliğinin kritik olduğu risk modelleme, finans ve sağlık gibi alanlarda yaygın olarak kullanılmaktadır.

GAM, doğrusal olmayan ilişkileri daha iyi yakalamak ve aynı zamanda yorumlanabilir

kalmak için her özelliğe ayrı bir dönüşüm fonksiyonu uygular. Bu yaklaşım, doğrusal modellerin sınırlamalarını ortadan kaldırarak daha güçlü tahminler yapılmasına olanak sağlar (Hastie ve Tibshirani, 1986).

Yanıt değişkeni Y ve tahminci değişkenler X_j olan bir GAM Denklem 2.6'daki şekilde ifade edilebilir:

$$g(E[Y]) = \alpha + f_1(X_1) + f_2(X_2) + \cdots + f_p(X_p) \quad (2.6)$$

Burada:

- g(E[Y]): Yanıt değişkeni Y'nin beklenen değerine uygulanan bağlantı (link) fonksiyonunu ifade etmektedir. Bu fonksiyon, yanıt değişkeni ile doğrusal tahmin edici arasındaki ilişkiyi kurmaktadır.
- α: Tüm tahminci değişkenlerin sıfır olduğu durumda temel etkiyi temsil eden sabit terimdir.
- $f_j(X_j)$: X_j tahmincisi için tanımlanan düzgün (smooth) fonksiyonu ifade etmektedir. Bu fonksiyon, her bir tahminci değişkenin yanıt değişkeni üzerinde doğrusal olmayan etkilere sahip olabilmesine olanak tanımaktadır. f_j fonksiyonu genellikle spline, smoothing veya kernel tabanlı yöntemlerle tahmin edilmektedir.
- p: Modelde yer alan toplam tahminci değişken sayısını göstermektedir.

Bir GAM'da, her $f_j(X_j)$ fonksiyonu, X_j ile yanıt değişkeni Y arasındaki potansiyel olarak doğrusal olmayan ilişkiyi yakalamak için bağımsız olarak uyarlanır. GAM'ların toplamsal yapısı yorumlanabilirlik sağlar; çünkü her f_j fonksiyonu, X_j'nin yanıt değişkeni üzerindeki izole etkisini gösterir. Bu model yapısı, esneklik ile yorumlanabilirlik arasında denge kurarak GAM'ları hem keşifsel veri analizi hem de kestirimsel modelleme için kullanışlı hâle getirir.

Pratik Örnekleme (GAM – Insurance Premium Pricing): Kumar ve ark. (2024) çalışmasında otomobil sigortası fiyatlandırması için dört aktüeryal modelleme tekniğini karşılaştırmıştır: GLM, AGLM, XGBoost ve GAM (Generalized Additive Models). GAM, esneklik ile yorumlanabilirlik arasında sağladığı denge nedeniyle öne çıkmıştır. Çalışma, sürücü yaşı, konum, araç özellikleri ve sürüş geçmişi gibi risk faktörlerini

dikkate alarak saf primlerin tahminine odaklanmıştır. GAM, bu değişkenlerin düzgün ve doğrusal olmayan etkilerinin görselleştirilmesini sağlamış; genç sürücüler arasında artan kaza riski ve kentsel bölgelerden kaynaklanan orantısız derecede yüksek hasar talepleri gibi bilinen aktüeryal içgörüleri doğrulamıştır. Bu düzeyde yorumlanabilirlik, düzenleyici açıdan avantaj sağlamış ve aktüerlerin model davranışının gerçek dünya risk örüntüleriyle uyumlu olduğunu doğrulamasına yardımcı olmuştur. Kara kutu modellerin aksine, GAM uyum süreçlerine hazır açıklamalar ve riskin eyleme dönüştürülebilir görselleştirmelerini sunmuştur.

Anchor Explanations, belirli tahminlerin belirli kurallar (anchors) kullanılarak açıklanmasına olanak tanıyan yerel bir açıklanabilirlik yöntemidir. LIME'ın geliştirilmiş bir versiyonudur ve tahminin güvenilirliğini artırmayı amaçlar. Anchor yöntemi, bir modelin belirli bir tahmini üretmesini garanti eden kuralları keşfeder. Tahmin, yalnızca belirli koşullar sağlandığında bu kurallarla gerçekleştirilir. Anchor, tahminlerin yüksek doğrulukla açıklanmasına olanak tanıdığı için karmaşık kararların yorumlanmasında etkilidir (Ribeiro ve ark., 2018).

Anchor algoritmasının formülasyonu Denklem 2.7'deki şekilde kullanır:

$$\mathcal{A} \subseteq X, \quad P(y|x, \mathcal{A}) \approx P(y|x) \tag{2.7}$$

Burada:

- $\mathcal{A}$: Veri noktasının yalnızca belirli özelliklerini kapsayan ve bir kural yapısını temsil eden özellikler alt kümesini ifade etmektedir.
- P(y | x, $\mathcal{A}$): $\mathcal{A}$ ile tanımlanan koşullar altında model kararının geçerli olma olasılığını göstermektedir.
- P(y | x, $\mathcal{A}$) ≈ P(y | x): Algoritma, $\mathcal{A}$ alt kümesini bu yaklaşık eşitlik sağlanacak şekilde seçmeyi amaçlamaktadır. Böylece modelin tahmini, anchor koşulları altında tutarlı biçimde korunmaktadır.

Pratik Örnekleme (Anchors – Fraud Detection): Thanathamathee ve ark. (2024) Tayland Borsası'ndaki finansal tablolara uygulanan XGBoost tabanlı bir dolandırıcılık tespit modeline, SHAP örnek ağırlıklandırmasıyla birlikte Anchor Explainable AI yöntemini entegre etmiştir. Anchor yöntemi, "Eğer işlem uluslararasıysa VE tutar 5,000$'ı aşıyorsa, o zaman daima dolandırıcılık olarak sınıflandırılacaktır" gibi

yorumlanabilir "if-then" kural tabanlı açıklamalar sağlamıştır. Bu kural anchor'ları, denetçilerin ve uyum ekiplerinin kararları verimli biçimde doğrulamasına olanak tanımış ve sınıflandırma mantığında şeffaflık sunmuştur. Model, dolandırıcılık vakalarında mükemmel geri çağırım elde etmiş, doğruluk ve yorumlanabilirlik açısından da önemli iyileşmeler göstermiştir. Bu yaklaşım, yüksek riskli ve kurallarla gerekçelendirilebilir karar sistemleri gerektiren düzenleyici ortamlar için özellikle uygundur.

Counterfactual Explanations (Karşı-Olgusal Açıklamalar), bir modelin tahminini değiştirmek için hangi koşulların gerekli olduğunu gösterir. "Farklı bir sonuca ne yol açardı?" sorusuna cevap verirler. Modelin kararını anlamak ve tahmini iyileştirmek isteyen kullanıcılar için oldukça faydalıdır (Wachter ve ark., 2017).

Wachter ve ark. (2017) tarafından incelendiği üzere, bir Karşı-Olgusal Açıklamalar adımları aşağıdaki gibidir:

- f(x): Bir giriş örneği olan x için model tarafından üretilen tahmini ifade etmektedir. Burada modelin istenmeyen bir sonuç olan y değerini tahmin ettiği varsayılmaktadır.
- $y' \neq y$: Kullanıcının elde etmek istediği, mevcut model çıktısından farklı olan hedef sonucu ifade etmektedir.
- $f(x') = y'$: Modelin hedeflenen y' sonucunu üreteceği, x girdisine yakın alternatif bir x' girdisinin bulunmasını göstermektedir. Bu yaklaşımda amaç, x ile x' arasındaki farkı mümkün olduğunca en aza indirmektir.

$$x' = \arg\min_{x'} \text{distance}(x, x') \quad \text{s.t.} \quad f(x') = y' \tag{2.8}$$

Burada:

- distance(x, x'): x ile x' arasındaki değişim miktarını ölçmek için kullanılan uzaklık fonksiyonunu ifade etmektedir. Bu ölçüm genellikle Öklid uzaklığı gibi bir metrik ile yapılmaktadır.
- $f(x') = y'$: Yeni giriş örneği olan x' 'in model tarafından istenen hedef sonuca götürmesini ifade etmektedir.

Pratik Örnekleme (Karşı-olgusal – Loan Adjustment): VanNostrand ve ark.(2024)

FACET sistemini tanıtmıştır; bu sistem, kredi onayı gibi otomatik karar sistemleri için etkileşimli Karşı-olgusal açıklamalar üretmektedir. Statik nokta tabanlı Karşı-olgusal açıklamaların aksine FACET, kullanıcıların uygulanabilir bir dizi "what-if" senaryosunu keşfetmesine olanak tanır. Bir kredi bağlamında, "Geliriniz 30,000$'dan 45,000$'a yükselseydi ve borç/gelir oranınız %40'ın altına düşseydi, krediniz onaylanırdı" gibi çıktılar üretebilir. Bu tür içgörüler, şeffaflık ve eyleme dönüştürülebilir geri bildirimin kritik olduğu finansal hizmetlerde özellikle değerlidir. Bu yöntem, olumlu sonuçlara ulaşmak için birden fazla gerçekçi yol sunarak kullanıcı güçlendirmesini desteklemekte, AI tabanlı kredi kararlarında hem güveni hem de başvuru sahibinin itiraz/iyileştirme imkânını artırmaktadır.

LORE (LOcal Rule-based Explanations), yerel kural tabanlı açıklamalar sağlayan bir XAI yöntemidir. Karmaşık modellerin tahminlerini anlaşılabilir kural kümelerine dönüştürmeyi amaçlar. LIME ve Anchor yöntemlerine benzer şekilde çalışsa da, karar kuralları oluşturarak daha açıklayıcı analizler sunar. LORE ayrıca, alternatif senaryoları gösteren Karşı-olgusal örnekler üreterek modeli daha şeffaf hâle getirir (Guidotti ve ark., 2019).

LORE algoritması matematiksel olarak aşağıdaki bileşenler kullanılarak temsil edilebilir:

- Yerel Komşuluk, $\mathcal{N}(x)$: x örneği etrafında oluşturulan sentetik komşuluğu ifade etmektedir. Bu yapı, özellik uzayında x'e yakın sentetik örneklerden oluşmakta ve modelin bu yerel bölgedeki davranışına ilişkin bilgi sağlamaktadır.
- Karar Ağacı, T: $\mathcal{N}(x)$ üzerinde eğitilen yerel karar ağacını temsil etmektedir. Bu ağaç, kara kutu modelin x örneği çevresindeki davranışını yaklaşık olarak yansıtan yorumlanabilir bir model işlevi görmektedir.
- Kural Çıkarımı, R(x): x örneği için karar ağacı T'den elde edilen kuralı ifade etmektedir. Bu kural, biçimsel olarak birden fazla koşulun birleşimi Denklem 2.9'da tanımlanmaktadır:

$$R(x): \bigwedge_{j=1}^{k} (X_j \in S_j) \tag{2.9}$$

  Burada X_j, ilgili özelliği; S_j ise x örneği için model tahmininin geçerli olmasını sağlayan özellik değerleri kümesini göstermektedir. Bu kural, modelin neden

belirli bir sonucu tahmin ettiğini açıklamaya yardımcı olmaktadır.

- Karşı-olgusal, C(x): x örneği için oluşturulan karşı-olgusal örnekler kümesini ifade etmektedir. Karşı-olgusal örnekler, model tahminini değiştirmek amacıyla özellik değerlerinde mümkün olan en küçük değişiklikler yapılarak oluşturulmaktadır. Biçimsel olarak bir karşı-olgusal örnek kümesi Denklem 2.10'da tanımlanmaktadır:

$$C(x) = x' \in \mathcal{N}(x) \mid f(x') \neq f(x) \tag{2.10}$$

Burada f(x), özgün modelin x örneği için ürettiği tahmini; f(x′) ise karşı-olgusal örnek olan x′ için üretilen tahmini ifade etmektedir. C(x) kümesi, hangi özellik değişikliklerinin model tahminini değiştirebileceğini anlamaya yardımcı olmaktadır.

2.1.3.2 Modele özgü yaklaşımlar

Bu XAI yöntemleri, belirli bir modelin yapısına veya özelliklerine dayanarak açıklanabilirlik sağlamaktadır.

TREPAN (Ağaç Tabanlı Kural Çıkarımı), sinir ağı modellerinin tahminlerini anlaşılabilir hâle getirmek için sinir ağlarından karar ağaçları çıkaran bir yöntemdir. Sinir ağları genellikle yüksek performans sunsa da, tahminlerinin arkasındaki işleyiş insanlar tarafından yorumlanabilir değildir. Bu sorunu çözmek için TREPAN, sinir ağı kararlarını basit if-else kurallarıyla ifade eden karar ağaçları üretir. Böylece karmaşık modellerin anlaşılabilirliği sağlanır.

TREPAN'ın amacı, özgün sinir ağı modelinin performansını korurken aynı zamanda insanlar için yorumlanabilir bir yapı sunmaktır. Karar ağaçları sayesinde, belirli bir tahminin hangi koşullara dayandığı açık biçimde gösterilebilir (Craven ve Shavlik, 1995).

TREPAN algoritması matematiksel olarak Denklem 2.11'deki şekilde gösterilebilir:

$$F = \frac{1}{n}\sum_{i=1}^{n} 1\{T(x_i) = f(x_i)\} \tag{2.11}$$

Burada:

- T: Kara kutu model f(x)'i yaklaşık olarak temsil eden karar ağacını ifade

etmektedir.

- Fidelity (F): Karar ağacı ile kara kutu model arasındaki uyum düzeyini göstermektedir. Bu değer, eğitim verileri üzerinde iki modelin aynı tahmini üretme oranı olarak hesaplanmaktadır.
- n: Verisetindeki toplam örnek sayısını ifade etmektedir.
- $\mathbb{1}[T(x_i) = f(x_i)]$: Karar ağacının tahmini $T(x_i)$ ile kara kutu modelin tahmini $f(x_i)$ eşleştiğinde 1, aksi durumda 0 değerini alan gösterge fonksiyonunu ifade etmektedir.

Pratik Örnekleme (TREPAN – Bankacılıkta Müşteri Segmentasyonu): Hossain ve ark.(2020) kural madenciliği ve sinir ağlarının bir karışımını kullanarak bankacılık müşterileri için büyük ölçekli, açıklanabilir bir segmentasyon çerçevesi sunmuştur. Çerçeve, 150.000'den fazla müşteriye ait işlemsel ve bağlamsal verileri işlemekte; yorumlanabilir müşteri segmentleri oluşturmak için kümeleme ve birliktelik kuralı madenciliği uygulamaktadır. Bu kural kümeleri—örneğin, "EĞER gelir > 80,000 $ VE müşteri süresi > 5 yıl İSE segment = 'düşük risk, yüksek değer'"—TREPAN'ın kara kutu modelleri karar ağaçlarıyla yaklaşık olarak temsil etme işlevini yansıtmaktadır. Bu metodoloji hem uyum görevlilerinin hem de pazarlama ekiplerinin sinir ağı modellerinde kullanılan segmentasyon stratejilerini anlamasına ve doğrulamasına olanak tanımakta; kararların şeffaf ve denetime hazır olmasını sağlamaktadır.

EBM (Explainable Boosting Machine), makine öğrenmesinde kullanılan gradient boosting algoritmasının açıklanabilir bir versiyonudur. EBM, her özelliğin katkısını ayrı ayrı hesaplayan genelleştirilmiş toplamsal modellere (GAMs) dayanmaktadır. Bu yöntem, karmaşık modellere yakın performans sunarken şeffaf ve yorumlanabilir tahminler üretir. Özelliklerin etkileri toplanabilir ve modelin hangi faktörlere dayandığı kolayca anlaşılabilir (Nori ve ark., 2019).

Bir EBM modelinin nihai tahmini Denklem 2.12'deki şekilde verilir:

$$\hat{y} = \alpha + \sum_{j=1}^{p} f_j(X_j) + \sum_{j,k} f_{jk}(X_j, X_k) \qquad (2.12)$$

Burada:

- $\hat{y}$: Model tarafından tahmin edilen değeri ifade etmektedir.

- $\sum_{j=1}^{p} f_j(X_j)$: Her bir özelliğin yanıt değişkeni üzerindeki bireysel etkisini temsil eden fonksiyonların toplamını göstermektedir. Bu terim, özelliklerin modele olan tekil katkılarını yakalamaktadır.
- $\sum_{j,k} f_{jk}(X_j, X_k)$: X_j ve X_k özellikleri arasındaki isteğe bağlı ikili etkileşim terimlerini ifade etmektedir. Bu terimler, iki özellik arasındaki ortak etkinin model çıktısına nasıl yansıdığını göstermektedir.

Pratik Örnekleme (EBM – Kredi Onayı): Charran ve Mahapatra (2023) kredi riski modellemesine uyarlanmış EBM çerçevesine yönelik yeni bir geliştirme sunmuştur. Yazarlar, standart EBM uygulamalarında yorumlanabilirliği zayıflatabilen sahte özellik etkileşimleri ve özellik baskınlığı gibi yaygın sorunları ele almıştır. Genişletilmiş EBM sistemi, finansal karar bağlamlarında hem doğruluğu hem de özellik katkılarının açıklığını artırmak için geliştirilmiş bir özellik seçimi ve topluluk yaklaşımı kullanmaktadır. Kredi skorlama görevlerini de içeren kıyaslama verisetlerine uygulandığında, geliştirilmiş EBM, borç/gelir (DTI) oranının %45'in üzerine çıkmasıyla riskte keskin artışlar ve kredi puanının 750'nin üzerine çıkmasından sonra azalan getiriler gibi doğrusal olmayan eşikleri belirlemiş; bu durum gerçek dünya kredi verme kararlarıyla örtüşmüştür. Bu da EBM'yi, kredi onaylarında uyum açısından kritik kararlar için XGBoost gibi kara kutu modellere karşı şeffaf ve yüksek performanslı bir alternatif olarak öne çıkarmıştır.

Katman Bazlı İlgi Yayılımı (LRP), sinir ağları tarafından yapılan tahminleri anlamaya yönelik bir açıklanabilirlik tekniğidir. Bu yöntem, her nörona tahmin oluşumuna ne kadar katkıda bulunduğuna ilişkin bir ilgi skoru atar. LRP, tahmini ağın çıktısından girdi katmanına doğru geriye yayar. Böylece tahminin hangi girdilere dayandığına dair ayrıntılı bir görünüm sunar. Görüntü işleme (CNN) ve doğal dil işleme (RNN) modellerinde yaygın olarak kullanılmaktadır (Bach ve ark., 2015).

LRP algoritması şu şekilde çalışır:

1. Çıkış katmanı: İlgi skoru $R^{(L)}$, açıklanmak istenen belirli sınıfa ait tahmin skoru ile başlatılmaktadır.
2. Geriye doğru ilgi yayılımı: İlgi değeri, çıkış katmanından giriş katmanına doğru geri yayılmaktadır. Her bir l katmanı için, bu katmandaki i nöronunun ilgi değeri olan

$R_i^{(l)}$, bir sonraki katman l+1'de yer alan nöronların ilgi skorlarına bağlı olarak hesaplanmaktadır. Bu amaçla en yaygın kullanılan kurallardan biri ε-kuralıdır. İlgi skoru hesaplaması Denklem 2.13'te verilmiştir:

$$R_i^{(l)} = \sum_j \frac{a_i w_{ij}}{\sum_i a_i w_{ij} + \epsilon \cdot \text{sign}(\sum_i a_i w_{ij})} R_j^{(l+1)} \quad (2.13)$$

Burada:

- $R_i^{(l)}$: l katmanındaki i nöronunun ilgi değerini ifade etmektedir.
- a_i: l katmanındaki i nöronunun aktivasyon değerini göstermektedir.
- w_{ij}: l katmanındaki i nöronunu, l+1 katmanındaki j nöronuna bağlayan ağırlık değerini ifade etmektedir.
- **ε:** Sayısal kararsızlığı önlemek amacıyla kullanılan küçük bir dengeleyici terimi göstermektedir.

3. Giriş katmanındaki ilgi skorları: Tüm katmanlar boyunca ilgi yayılımı tamamlandığında, ilgi değeri giriş katmanına ulaşmaktadır. Bu aşamada her bir giriş özelliği x_i, bir ilgi skoru olan $R_i^{(1)}$ değerini almaktadır. Bu skorlar, her bir özelliğin nihai tahmine yaptığı katkıyı temsil etmektedir.

Ayrıca Denklem 2.14'te gösterildiği üzere toplam ilgi değeri, katmanlar boyunca yayılırken korunmaktadır:

$$\sum_i R_i^{(l)} = \sum_j R_j^{(l+1)} \quad (2.14)$$

Bu ilgi korunumu, ilgi skorlarının toplamının katmanlar arasında tutarlı kalmasını sağlayarak çıktıdan girdiye kadar toplam ilgi miktarını muhafaza eder.

Pratik Örnekleme (LRP – Hisse Senedi Trend Tahmini): Kolte ve ark. (2022) tarihsel veriler ve piyasa göstergelerine dayanarak hisse senedi fiyat trendlerini tahmin etmek için Long Short-Term Memory (LSTM) ağlarını kullanan bir derin öğrenme çerçevesi uygulamıştır. Modelleri öncelikle tahmin odaklı olsa da, bireysel tahminlerde özellik katkılarını analiz etmek amacıyla Katman Bazlı İlgi Yayılımı (LRP) gibi açıklanabilirlik araçlarını entegre etmiştir. Özellikle LRP, RSI (Relative Strength Index), MACD ve hacim değişimleri gibi teknik göstergelerin yön tahminleri ("yukarı" veya "aşağı") üzerindeki etkisini izlemek için kullanılmıştır. Bu sayede finans analistleri, modelin en

etkili gördüğü sinyalleri anlayabilmiş ve doğrulayabilmiş; bu da algoritmik işlem ortamlarında şeffaflığı ve güveni artırmıştır. Görsel atıflar, duyarlılık veya oynaklık göstergelerinin karar verme sürecine baskın olduğu dönemlerin belirlenmesine yardımcı olmuş ve makine çıktılarının insan sezgisiyle uyumlu hâle gelmesini sağlamıştır.

Integrated Gradients (IG), gradyan tabanlı bir açıklanabilirlik tekniğidir. Hangi giriş özelliğinin model tahminine ne ölçüde katkı sağladığını belirlemek için kullanılır. IG, giriş verisinin tahmin sürecindeki yolunu başlangıç referansından nihai tahmine kadar analiz eder. Bu yol boyunca gradyanların hesaplanması, her özelliğin tahmine olan toplam katkısının bulunmasını sağlar (Sundararajan ve ark., 2017). IG algoritması aşağıdaki şekilde çalışır:

1 Başlangıç girdisinin seçilmesi: Özelliklerin yokluğunu temsil eden bir başlangıç girdisi x′ seçilmektedir. Bu başlangıç değeri, veri türüne bağlı olarak tamamen sıfırlardan oluşan bir görüntü, boş bir vektör ya da referans bir giriş olabilir.

2 Başlangıç ile gerçek girdi arasında yol tanımlanması: Başlangıç girdisi x′ ile gerçek girdi x arasında sürekli bir yol tanımlanmaktadır. Bu yol genellikle başlangıç girdisi ile gerçek girdi arasında doğrusal enterpolasyon kurularak oluşturulmaktadır:

$$x(t) = x' + t \cdot (x - x'), \quad t \in [0,1] \tag{2.15}$$

3 Yol üzerindeki gradyanların hesaplanması: Yol boyunca yer alan her bir x(t) noktası için, model çıktısının giriş özelliklerine göre gradyanı Denklem 2.15'te hesaplanmaktadır. Bu gradyanlar, model tahmininin her bir özellikteki değişimlere karşı duyarlılığını göstermektedir.

4 Özellik katkılarının hesaplanması: Her bir özellik için atıf değeri, gradyanların başlangıç noktasından gerçek girdiye kadar tanımlanan yol boyunca integre edilmesiyle Denklem 2.16'da gösterildiği şekilde elde edilmektedir:

$$\mathrm{IG}_i(x) = (x_i - x_i') \cdot \int_0^1 \frac{\partial F(x(t))}{\partial x_i}\, dt \tag{2.16}$$

Burada:

- $\mathrm{IG}_i(x)$: i özelliği için hesaplanan Integrated Gradient değerini ifade etmektedir ve bu değer ilgili özelliğin model tahminine katkısını göstermektedir.
- $x_i - x_i'$: i özelliği için gerçek girdi ile başlangıç referansı arasındaki farkı ifade

etmektedir.

- $\frac{\partial F(x(t))}{\partial x_i}$: x(t) noktasında model çıktısı F'nin i özelliğine göre gradyanını göstermektedir.
- $\int_0^1 \frac{\partial F(x(t))}{\partial x_i}\, dt$: Başlangıç girdisinden gerçek girdiye kadar tanımlanan yol boyunca gerçekleştirilen integrasyon işlemini ifade etmektedir.

Pratik Örnekleme (Integrated Gradients – İşlem Anomali Tespiti): Manjula Devi ve ark. (2024) yeni nesil anomali tespit çerçevelerinde kredi kartı dolandırıcılığına karşı proaktif ve doğru bir önleme stratejisi sunmuştur. Olağandışı işlemleri gerçek zamanlı olarak sınıflandırmak için derin öğrenme tabanlı sistemler kullanılmıştır. Özellikle, model şeffaflığı ihtiyacına yanıt olarak Integrated Gradients (IG) benzeri açıklanabilirlik yöntemlerinin dâhil edilmesi önerilmiştir. Bu tür açıklamalarla, örneğin yüksek işlem tutarı ve olağandışı işlem zamanı gibi faktörler, şüpheli bir işlemin yüksek anomali skorunun gerekçesi olarak vurgulanabilmektedir. IG çıktıları görsel olarak sunulabilir ve böylece hem dolandırıcılık analistleri hem de düzenleyici denetim kurumları için kararların gerekçelendirilmesini sağlar. Bu yaklaşım, operasyonel güven oluşturma ve düzenleyici uyum açısından katma değer sağlamaktadır.

2.1.3.3 Güvenlik ve gizlilik yöntemi

Bu XAI yöntemi, modellerin gizlilik açıklarını ve güvenlik zayıflıklarını değerlendirmek için kullanılmaktadır.

MIA (Membership Inference Attack), bir makine öğrenmesi modelinin eğitim verilerinin gizliliğini test eden bir tekniktir. Saldırgan, modelin tahmin sonuçlarını inceleyerek bir veri noktasının eğitim verilerinde kullanılıp kullanılmadığını belirlemeye çalışır. Bu teknik, hassas veriler içeren modellerde olası veri sızıntılarını tespit etmek açısından önemlidir (Shokri ve ark., 2017).

Denklem 2.17'de gösterildiği üzere, ilgilenilen veri noktası x ile gösterilsin ve *f(x)*, hedef modelin bu veri noktası için ürettiği tahmini ifade etsin. Saldırı modeli g ise, m = 1 değerinin x örneğinin eğitim kümesine ait olduğunu gösterdiği durumda, $P\big(m = 1 \big| f(x)\big)$ olasılığını en yüksek yapacak şekilde eğitilmektedir.

$$P(m = 1 | f(x)) = g(f(x)) \quad (2.17)$$

Burada:

- m: Gösterge değişkenini ifade etmektedir. x örneği eğitim kümesinde yer alıyorsa $m = 1$, aksi durumda $m = 0$ olmaktadır.
- $g(f(x))$: Saldırı modelinin, x örneğinin eğitim kümesinin bir parçası olduğuna ilişkin güven düzeyini ya da olasılık tahminini göstermektedir.

2.2 Literatür Taraması ve Tematik Analiz

Bu bölümde mevcut çalışmalar, metodolojiler, bulgular, başarı oranları ve literatürdeki boşluklar tartışılacaktır.

2.2.1 Literatürün tematik incelemesi

Bu bölümde, literatür taramasına konu olan makaleler tematik alt başlıklar altında analiz edilmektedir. Her başlık altında ilgili makaleler karşılaştırılmakta ve değerlendirilmektedir.

Tablo 2.1. Çeşitli açıklanabilirlik algoritmalarının sayıları

Algoritma	Sayı
SHAP	74
LIME	32
Anchor Explanations	3
ICE	2
EBM	2
LORE	2
Permütasyon önemlilik	2
MIA	1
ALE	1
IG	1
GAM	1
LRP	1
Karşı-Olgusal Açıklamalar	1
TREPAN	1

Tablo 2.1, literatür taraması için seçilen çalışmalarda kullanılan XAI yöntemlerinin sayısını göstermektedir. Görüldüğü üzere, SHAP ve LIME yöntemleri finans alanındaki uygulamalarda en çok tercih edilen XAI algoritmaları olmuştur.

Tablo 2.1, XAI yöntemlerinin benimsenmesinde çarpıcı bir yoğunlaşmayı ortaya

koymaktadır; SHAP 74 uygulama ile alanın hâkim yöntemi konumundayken (tüm çalışmaların %66.1'i), onu 32 uygulama ile LIME izlemektedir (%28.6). Bu yoğunlaşma, finans sektörünün yeni teknikleri denemekten ziyade kanıtlanmış ve güvenilir yöntemler üzerinde yakınsadığı bir olgunlaşma örüntüsüne işaret etmektedir. Bu iki yöntemin baskınlığı çeşitli faktörlere bağlanabilir: SHAP'ın oyun teorisine dayalı kuramsal temeli, nicel odaklı finans profesyonelleri için matematiksel bir sağlamlık sunarken, LIME'ın modelden bağımsız yapısı farklı finansal modelleme yaklaşımları arasında esneklik sağlamaktadır.

LRP (1 uygulama), GAM (1 uygulama) ve TREPAN (1 uygulama) gibi gelişmiş yöntemlerin neredeyse hiç kullanılmaması, akademik XAI araştırmaları ile finans alanındaki pratik uygulamalar arasında önemli bir boşluğu ortaya koymaktadır. Bu örüntü şu durumlardan birine işaret ediyor olabilir: (1) gelişmiş yöntemlerin üretim ortamındaki finansal sistemlerde uygulanmasına yönelik engeller, (2) finansal AI uygulayıcıları arasında alternatif yaklaşımlara ilişkin farkındalığın yetersiz olması ya da (3) alternatiflerin keşfedilmesini caydıracak ölçüde SHAP/LIME kombinasyonlarının pratikte iyi performans göstermesi. Karşı-Olgusal Açıklamalar yönteminin yalnızca bir kez görülmesi özellikle dikkat çekmektedir; çünkü bu yöntem, müşteri danışmanlığı hizmetleri açısından kritik olan finansal "what-if" senaryoları için daha uygun olabilmektedir.

2.2.1.1 Finans sektöründe XAI yöntemlerinin kullanımı

Bu temada, LIME, SHAP, GAM, EBM gibi XAI yöntemlerinin finansal uygulamalardaki kullanımı ele alınmaktadır. Kredi tahmin modelleri, müşteri davranışı analizi ve sigorta risk değerlendirme sistemleri gibi finansal uygulamalarda XAI yöntemlerinin rolü vurgulanmaktadır.

Tablo 2.2. Uygulama alanları ve kullanılan algoritmalar

Uygulama Alanı	Kullanılan Algoritmalar
Otomobil sigortası sektörü için risk değerlendirmesi ve kaza tahmini	SHAP
Banka pazarlama kampanyası başarı tahmini ve müşteri hedefleme	SHAP
Bankacılık sistemlerinde finansal işlem sınıflandırması	SHAP
Finans gibi gizliliğe duyarlı alanlarda XAI	MIA
Finansal zaman serisi analizi, finansal piyasalar, hisse senedi fiyat tahmini	SHAP, LIME, ALE,

ve yatırım stratejileri	Permütasyon önemlilik
Girişimcilik ve Kurumsal Ekonomi	LIME, Permütasyon önemlilik
Hisse senedi piyasası eğilimleri ve güvenilirlik için finansal duygu analizi	SHAP
Bankacılık ve finansal hizmetlerde iflas tahmini ve risk yönetimi	SHAP, Anchor
Kişiselleştirilmiş sağlık, finans ve diğer kullanıcı odaklı uygulamalarda kullanılan AI modellerinde açıklanabilirlik	SHAP, LIME
Kredi kartı işlemlerinde dolandırıcılık tespiti	SHAP, LIME
Kredi skorlama, kredi onay süreçleri ve kredi risk yönetimi	SHAP, LIME, LORE, ICE, ALE, GAM, IG, Anchor
Kredi temerrüt tahmini ve finansal risk analizi	ICE, SHAP
Makroekonomik analiz, finansal piyasalar ve stres testi için politika yapımı	SHAP, LIME
Mortgage aracılık otomasyonu	SHAP, ALE
Muhasebe, denetim ve işletme değerlendirmeleri için finansal raporlama	ALE
Müşteri elde tutma ve müşteri kaybı tahmini	SHAP, LIME
Çevrim içi kredi piyasaları, sosyal medya analizi	LRP
Otomatik sistemlerde adalet ve şeffaflık odaklı bankacılık hizmetleri	SHAP, LIME, TREPAN
Sağlık sigortasında talep tahmini ve risk değerlendirmesi	SHAP, LIME, IG
Siber güvenlik ve oltalama saldırılarının önlenmesi	SHAP, LIME, EBM

LIME ve SHAP yöntemleri, kredi skorlama alanında açıklanabilirlik sağlamak amacıyla her bir özniteliğin müşterilerin kredi riskinin belirlenmesi sürecine nasıl katkı sunduğunu açıklamak için yaygın biçimde kullanılmaktadır. Tablo 2.2 analiz edildiğinde, LIME ve SHAP yöntemlerinin dolandırıcılık riskini belirlemek ve işlem kararlarını açıklamak amacıyla finansal işlemler ve siber güvenlik uygulamalarında yoğun biçimde kullanıldığı görülmektedir. Ayrıca özellikle SHAP, yatırım analizlerinde kullanılan modellerde her bir değişkenin katkısını analiz ederek portföy yönetiminde şeffaflık sağlamak amacıyla uygulanmıştır.

Tablo 2.2, XAI yöntemleri ile finansal uygulama alanları arasında ayrıntılı bir eşleştirme sunmaktadır. SHAP, özellikle kredi skorlama ve dolandırıcılık tespiti alanlarında baskın durumdadır; bunlar, düzenleyici uyumun ve model çıktılarının hassas biçimde gerekçelendirilmesinin kritik olduğu yüksek riskli alanlardır. SHAP'ın ağaç tabanlı modellerle entegrasyonu ve ayrıntılı özellik atfı sunabilme kapasitesi, onu bu bağlamlarda tercih edilen bir seçenek hâline getirmektedir. Buna karşılık LIME, müşteri segmentasyonu ve iflas tahmini dâhil olmak üzere alanlar arasında daha dengeli bir

dağılım göstermektedir; bunun olası nedeni modelden bağımsız yapısı ve örnek düzeyinde hızlı içgörüler sunmaya uygun olmasıdır. İlginç biçimde, Karşı-Olgusal Açıklamalar ve diğer yeni yöntemler hemen hiçbir alanda uygulanmamıştır; bu da ya uygulama zorluklarına ya da benimsenme eksikliğine işaret etmektedir. Bu dağılım, yöntemsel tercihlerin hem görevin teknik gereksinimleriyle hem de her finansal alanın yorumlanabilirlik ihtiyaçlarıyla yakından ilişkili olduğunu göstermektedir.

Özetle, LIME ve SHAP yöntemlerinin finansal analizlerde açıklanabilirlik sağlayarak güven inşa ettiği ve çeşitli finansal hizmetlerde yaygın biçimde tercih edildiği gösterilmiştir.

2.2.1.2 IoT için uç bilişimde açıklanabilir yapay zekâ: finansal uygulamalar için çıkarımlar

Son dönem literatür, gerçek zamanlı, yerelleştirilmiş ve açıklanabilir karar verme sistemlerini mümkün kılarak uç bilişimde açıklanabilir yapay zekânın (EdgeAI) IoT tabanlı finansal analitiği nasıl dönüştürebileceğini incelemeye başlamıştır.

Pourakbar ve Shah-Mansouri (2025), derin sinir ağı çıkarım görevlerini uç cihazlar arasında verimli biçimde tahsis eden, gecikmenin azaltılmasına ve kaynak kısıtlı IoT ortamlarında hizmet kalitesine odaklanan Transformer tabanlı bir optimizasyon çerçevesi önermektedir. Yaklaşımları, işlem denetimleri ve uyum açısından gerçek zamanlı yorumlanabilirliğin kritik olduğu satış noktası dolandırıcılık tespiti veya gömülü kredi sistemleri gibi merkeziyetsiz finansal hizmetler için önem arz etmektedir (Pourakbar ve Shah-Mansouri, 2025).

Rezvani ve ark. (2025), çevrim içi ve aktif öğrenme paradigmalarını kullanarak uç zekâ için kendi kendini eğitebilen ve uyarlanabilir bir sensör çerçevesi geliştirmektedir. Sistem, enerji verimli, yerinde eğitim ve karar verme imkânı sağlamakta olup alana özgü kullanım senaryolarına oldukça uyarlanabilirdir. Sensör verilerinin kredi değerliliği değerlendirmelerini doğrudan etkilediği tarım veya araç finansmanı gibi alanlarda, bu tür mimariler model şeffaflığını artırmak amacıyla XAI katmanlarıyla güçlendirilebilmektedir (Rezvani ve ark., 2025).

Wang Wayz (2025), 5G destekli IoT ortamlarında gizliliği koruyan finansal analitiği ele almakta; federe öğrenme ve şifrelenmiş veri hatlarını uç işleme ile bütünleştirmektedir. Modelleri, mobil kullanıcı verilerini yöneten finansal kurumlar için özellikle önemli olup

düzenlemelere uyumlu bir yöntem sunmaktadır. Açıklanabilirlik mekanizmalarının bu tür federe hatlara gömülme potansiyeli, AI karar süreçlerini finansal hesap verebilirlik standartlarıyla uyumlu hâle getirmek açısından kritik önemdedir (Wang ve others, 2025).

Zhang ve ark. (2025), AI ve IoT'nin, yeşil operasyonlar ve sürdürülebilir kalkınma bağlamındaki yakınsamasını incelemektedir. Birincil vurgu endüstriyel ve çevresel sürdürülebilirlik üzerinde olsa da, çalışma enerji maliyetlerini en aza indirirken operasyonel şeffaflığı en üst düzeye çıkaran sensör tabanlı karar sistemlerinin benimsenmesine yönelik ekonomik teşvikleri vurgulamaktadır. Çevresel IoT verileri, özellikle hesap verebilirlik ve izlenebilirliği güvence altına almak için açıklanabilir karar katmanlarıyla bütünleştirilerek, kredi verme veya ESG skorlama algoritmalarını bilgilendirebileceği yeşil finans modellerine aktarılabilir (Zhang ve others, 2025).

Boujrad ve ark. (2025), IoT sistemlerinden gelen zaman serisi sensör verilerinde istatistiksel anomali tespitine odaklanmaktadır. Normal sensör davranışını profillemek ve sapmaları gerçek zamanlı olarak tespit etmek için klasik otoregresif modelleri (AR, MA, ARIMA, SARIMA) uygulamaktadırlar. Endüstriyel izleme için geliştirilmiş olsa da, bu teknik finansta erken dolandırıcılık tespiti veya sensör destekli kredi skorlama açısından potansiyel taşımaktadır. Açıklanabilirliğin bu anomali profillerine dâhil edilmesi, uyarıların yorumlanabilirliğini önemli ölçüde artırabilir ve otomatik izlemeye duyulan güveni geliştirebilir (Boujrad ve ark., 2025).

Leghemo ve ark. (2025), AI, IoT ve Blockchain kullanan sistemler için şeffaflık, gizlilik ve uyum ilkelerini bütünleştiren bir veri yönetişimi çerçevesi önermektedir. Bu çerçeve, açıklanabilir yapay zekânın doğrudan veri hatlarına gömülmesine yönelik mekanizmalar ortaya koyması sebebiyle özellikle yüksek düzeyde düzenlemeye tabi ortamlarda faaliyet gösteren finansal hizmetler için önemlidir. Bu sayede sensör açısından zengin finansal uygulamalardan (örneğin mobil bankacılık, IoT ile P2P lending) elde edilen karar çıktılarının, yorumlanabilirlik ve etik adaleti korurken düzenleyici standartları karşılaması sağlanmaktadır (Leghemo ve ark., 2025).

Bu çalışmalar birlikte, IoT uygulamalarına uyarlanmış uç bilişim çerçeveleri içinde XAI sistemlerinin konuşlandırılmasının giderek daha mümkün hâle geldiğini göstermektedir. Birincil uygulama bağlamları çoğu zaman endüstriyel otomasyon veya sağlık izleme olsa da, düşük gecikmeli çıkarım, dağıtık öğrenme ve uyarlanabilir model güncelleme gibi

temel mimari ilkeler finansal alanlara yüksek düzeyde aktarılabilirdir.

Örneğin, kredi skorlama modelleri, yerel çevresel sensörlere dayanan kırsal bölgede bankacılık sistemlerinde cihaz üzerinde açıklanabilirlikten fayda sağlayabilir. Benzer şekilde, kullanıma dayalı sigorta fiyatlandırması, poliçe sahiplerine ve düzenleyicilere gerekçelendirilebilir, denetlenebilir içgörüler sunmak amacıyla SHAP veya benzeri yorumlanabilir mekanizmalarla donatılmış uç modellerden yararlanabilir. Gizlilik farkındalığına sahip öğrenmenin (örneğin federe SHAP) entegrasyonu, finansal yargı alanlarında giderek daha fazla uygulanan veri yönetişimi standartlarıyla uyumludur.

Dolayısıyla, EdgeAI ve IoT'nin birleşimi; gerçek dünya kısıtları altında çalışabilen, aynı zamanda hesap verebilirliği koruyan, finansal açıdan uyumlu, açıklanabilir ve bağlama duyarlı AI ekosistemleri oluşturmak için güçlü bir temel yaratmaktadır.

Tablo 2.3, finansal uygulamalar bağlamında IoT, EdgeAI ve açıklanabilir makine öğrenmesinin kesişiminde yer alan yakın dönem çalışmalara ilişkin karşılaştırmalı bir genel bakış sunmaktadır. Her bir kayıt, kullanılan verisetinin türünü (gerçek veya sentetik), veri kaynağını, hedeflenen finansal kullanım alanını, kullanılan XAI metodolojisini ve açıklanabilirliğin model şeffaflığına, düzenleyici uyuma veya karar doğruluğuna raporlanan katkısını vurgulamaktadır. İncelenen çalışmalar; mobil bankacılık, yeşil finans, tarımsal kredi skorlama, dolandırıcılık tespiti ve blockchain bütünleşik kredi verme gibi çeşitli alanları kapsamaktadır. Dikkat çekici biçimde, federe SHAP, yorumlanabilir regresyon ve vekil modelleme gibi açıklanabilirlik yöntemleri, uç cihazlarda şeffaf karar verme süreçlerini desteklemek amacıyla uygulanmıştır. Bu bulgular, merkeziyetsiz finansal altyapılar genelinde açıklanabilir ve uyumlu AI sistemlerinin konuşlandırılmasının uygulanabilirliğini pekiştirmektedir.

Tablo 2.3. Finansta IoT içerisinde XAI yöntemlerine genel bakış

Veriseti	Gerçek veya Sentetik	Kaynak	Finansal Uygulama Alanı	XAI Yöntemi	Doğruluk (XAI Katkısı)	Ref.
Mobil terminallerden elde edilen sensör verileri	Gerçek	5G destekli finansal cihazlar	Mobil bankacılık, işlem analizi	Federe SHAP	Tahmin çıktılarında uyumlu şeffaflık ile gizliliği koruyan uç analitiği sağladı	(Wang ve ark., 2025)
Uçta tahsis edilmiş çıkarım verileri	Sentetik	Uç simülatörü + transformer modeli	POS dolandırıcılık tespiti, gömülü kredi verme	İzlenebilir tahsis ile açıklanabilir DNN	Gecikmeyi ve şeffaflığı iyileştirdi; karar yolları denetim uyumu için açıklanabilir hâle geldi	(Pourakbar ve Shah-Mansouri, 2025)
Cihaz üzerindeki tarımsal ve araç sensör kayıtları	Gerçek	Uç cihazlardan sensör akışı	Tarımsal kredi skorlama, telematik tabanlı finans	Çevrim içi yorumlanabilir öğrenme	Model cihaz üzerinde güncellendi; saha düzeyindeki kararlarda açıklanabilirliği destekledi	(Rezvani ve ark., 2025)
Yeşil finans IoT sensör verileri	Gerçek	Akıllı sayaçlar, çevresel sensörler	ESG tabanlı kredi ve yatırım skorlama	Açıklanabilir regresyon modelleri	Sensör tabanlı göstergeleri şeffaf çıkarım kurallarıyla finansal riskle eşleştirdi	(Zhang ve ark., 2025)
IoT zaman serisi kayıtları	Gerçek	Endüstriyel düzey zaman serisi sensörleri	Dolandırıcılık tespiti, kullanım anomali skorlama	ARIMA tabanlı tespit üzerine vekil modeller	Risk sınıflandırma görevleri için anomali skorlarının yorumlanabilirliğini artırdı	(Boujrad ve ark., 2025)
Çok kaynaklı yönetişim verileri	Gerçek	Blockchain + IoT veri katmanları	IoT tabanlı kredi verme, dijital varlık takibi	Politika uyumlu sistemlerde gömülü açıklanabilirlik	Açıklanabilirliği düzenleyici veri akışları içinde çerçeveledi; karmaşık kararlarda izlenebilirlik sağladı	(Leghemo ve ark., 2025)

2.2.1.3 XAI yöntemlerinin performans karşılaştırması

Bu bölümde, XAI yöntemlerinin finansta uygulamalarına ilişkin çalışmalar analiz edilmekte ve bunların genel katkıları ile performans üzerindeki etkileri değerlendirilmektedir.

Bu bölümde sunulan tüm performans metrikleri (doğruluk, F1, AUC) incelenen çalışmalardan doğrudan alınmıştır. Herhangi bir bağımsız yeniden değerlendirme yapılmamıştır.

Tablo 2.1'de gösterildiği gibi, SHAP ve LIME, sırasıyla 74 ve 32 kullanım ile finansal uygulamalarda en sık kullanılan XAI yöntemleri olarak öne çıkmaktadır. Bu popülerlik, çeşitli pratik ve metodolojik avantajlara bağlanabilir. Oyun teorisine dayalı Shapley değerleri temelinde geliştirilen SHAP, hem yerel hem de küresel yorumlanabilirlik sunmakta ve finansal modellemede zaten yaygın olan XGBoost ve Random Forest gibi ağaç tabanlı modellerle özellikle uyumludur. Benzer şekilde, LIME, yerel olarak yorumlanabilir vekil modeller kullanarak örnek düzeyinde açıklamalar sunmakta ve bu

yönüyle anlık tahminler ile gerçek zamanlı karar verme sistemleri için makul hâle gelmektedir. Ancak, bu yöntemler pratikte baskın olsa da sınırlılıklara sahiptir. SHAP, özellikle büyük ölçekli verisetlerinde yüksek hesaplama karmaşıklığından muzdariptir ve LIME'ın yerel doğrusal yaklaşımları, yüksek derecede doğrusal olmayan modellerde kararsız açıklamalar üretebilir. Buna karşılık, Katman Bazlı İlgi Yayılımı (LRP), TREPAN ve Karşı-Olgusal Açıklamalar gibi alternatif yöntemler farklı güçlü yönler sunmaktadır: LRP derin sinir ağlarında nöron düzeyinde içgörü sağlar, TREPAN kara kutu modellerden yorumlanabilir karar ağaçları üretir ve Karşı-olgusal yöntemler uygulanabilir what-if analizlerine olanak tanır. Bununla birlikte, bu yaklaşımlar çoğunlukla uygulama karmaşıklığı, daha sınırlı araç desteği veya standardizasyon eksikliği nedeniyle daha seyrek kullanılmaktadır. Bu çalışma, Tablo 2.4, Tablo 2.5 ve Tablo 2.7'de, sık ve seyrek kullanılan yöntemlerin kuramsal temellerini, kullanım alanlarını ve raporlanan performans metriklerini karşılaştırarak bu farklılıkları ortaya koymakta ve finans alanında yöntem seçimi için dengeli bir bakış açısı sunmaktadır.

Tablo 2.4. XAI yöntemlerinin katkıları ve performansı

XAI Yöntemi	Katkı	Performans
ICE	Kara kutu modellerde özelliklerin etkisini görselleştirerek yorumlanabilirlik sağlar.	Vekil Decision Trees ile yaklaşık %98.4 doğruluk elde ederken, kara kutu modellere benzer düzeyde yorumlanabilirlik sunmaktadır.
MIA	Gizlilik odaklı değerlendirmelerde kullanılır; MIA'lara karşı zafiyet gözlemlenmiştir.	Yüksek gizlilik riskleri nedeniyle küresel açıklayıcılarda bir güvenlik açığı olarak öne çıkmaktadır.
SHAP	Modelde her özelliğin katkısını açık biçimde gösterir; kredi riski, müşteri segmentasyonu ve dolandırıcılık tespiti gibi alanlarda güçlü açıklamalar sunar.	Özellikle karar ağaçları ve XGBoost ile uyumlu olarak, %99.5'e kadar ulaşan yüksek doğruluk oranları ve güçlü özellik düzeyinde analizler sağlamaktadır.
LIME	Her tahmin için yerel açıklamalar sunar ve bireysel tahminlerde yorumlanabilirliği artırır.	Model performansında iyileşmeler rapor etmekte olup, bazı finansal uygulamalarda doğruluk oranları %100'e yaklaşmaktadır.
IG	Özellikle önyargının azaltılmasında yararlı olan güvenilir özellik analizi sağlar.	%98 doğruluk ile güvenilir yorumlar sunmaktadır.
GAM	Kara kutu modellere kıyasla daha yüksek düzeyde yorumlanabilirlik sağlar.	Genel doğrulukta küçük bir kayba rağmen model yorumlanabilirliğini korumaktadır.
EBM	Hem küresel hem yerel açıklamalar sağlar; kredi riski gibi alanlarda güçlü özellik katkıları göstermektedir.	%99.5'e kadar çıkan raporlanmış yüksek doğruluk oranlarına ulaşmaktadır.

LRP	Özellikle risk tahmininde her özelliğin etkisini analiz ederek yüksek yorumlanabilirlik sunar.	Psikometrik veri entegrasyonu yoluyla risk tahmin doğruluğunu artırmaktadır.
Anchor	Belirli bir kararın temellerini görselleştirerek güven artırıcı açıklamalar sunar.	Sınıf dengesizliğinde doğruluğu artırmakta ve kredi riski tahmininde etkili performans göstermektedir.
Karşı-Olgusal Açıklama	Önyargının azaltılması ve kredi kararlarının optimize edilmesinde etkili açıklamalar sunar.	Kredi ve dolandırıcılık tespitinde false negative oranlarını azaltarak sınıf dengesizliğinde doğruluğu artırmaktadır.
TREPAN	Kara kutu modeller için yorumlanabilirlik sağlar ancak MIA'lara karşı savunmasızdır.	%88'e kadar doğruluk elde ederek, karar ağaçlarıyla uyumlu yorumlanabilirlik sunmaktadır.

Finansal uygulamalardaki XAI yöntemlerinin karşılaştırmalı analizi Tablo 2.5'te verilmektedir.

Tablo 2.4 ve Tablo 2.5 'de sunulan performans verileri, XAI'ın finansal uygulamalardaki pratik uygulanabilirliği hakkında birkaç kritik içgörü ortaya koymaktadır. En dikkat çekici nokta, açıklanabilirliğin önemli performans ödünleşimlerini zorunlu kılmamasıdır. SHAP destekli modeller tutarlı biçimde %95-99.5 doğruluk oranlarına ulaşmakta, hatta bazı çalışmalar neredeyse kusursuz performans bildirmektedir. Örneğin, SHAP uygulamaları için raporlanan "%99.5'e kadar" doğruluk, uygun koşullar altında açıklanabilir modellerin kara kutu alternatiflerle eşdeğer hatta daha iyi performans gösterebileceğini düşündürmektedir.

Tablo 2.5. XAI yöntemlerinin karşılaştırmalı analizi

Yöntem	**Yorumlanabilirlik**	**Açıklanabilirlik**	**Hesaplama Verimliliği**	**Performans**
SHAP	Yüksek (Özellik öneminin görselleştirilmesi, hem küresel hem yerel yorumlar)	Özellik katkılarının kapsamlı atfı	Orta (M özellik için orta düzeyde $O(2^M)$ karmaşıklık)	Finansal uygulamalar genelinde ortalama %98.5 doğruluk
LIME	Orta (Doğrusal modellerle yerel yaklaşımlar)	Basitleştirilmiş yerel açıklamalar	Yüksek (Doğrusal karmaşıklık)	Ortalama %96.2 doğruluk
Anchor	Yüksek (Kural tabanlı açıklamalar)	Kesinlik/duyarlılık metrikleri ile açık karar kuralları	Düşük (Üstel arama uzayı)	Kredi kararlarında %94.8 doğruluk
GAM	Yüksek (Bireysel özellik etkileri görülebilir)	Özellik etkilerinin doğrudan görselleştirilmesi	Yüksek (Doğrusal karmaşıklık)	Risk değerlendirmesinde %92.7 doğruluk
EBM	Yüksek (Özellik etkileşimleri görülebilir)	Ayrıntılı etkileşim analizi	Orta	Dolandırıcılık tespitinde %95.5 doğruluk

Yorumlanabilirlik-performans ilişkisi yöntemler arasında önemli ölçüde değişmektedir. Tablo 2.5 'in karşılaştırmalı analizi, SHAP'ın "yüksek" yorumlanabilirlik sağlarken "orta" hesaplama verimliliğini koruduğunu ve bunun sonucunda finansal uygulamalarda ortalama %98.5 doğruluk elde ettiğini ortaya koymaktadır. Bu durum, "yüksek" yorumlanabilirlik ve "açık karar kuralları" sunmasına rağmen "üstel arama uzayı" karmaşıklığı nedeniyle "düşük" hesaplama verimliliğinden muzdarip olan Anchor gibi yöntemlerle belirgin bir tezat oluşturmaktadır. Bu ödünleşim, özellikle milisaniye düzeyinde yanıt sürelerinin gerekli olduğu dolandırıcılık tespiti gibi gerçek zamanlı finansal uygulamalar için büyük önem taşımaktadır.

Tablo 2.5'deki hesaplama verimliliği örüntüleri, kritik bir uygulama hususunu öne çıkarmaktadır. LIME'ın "yüksek" hesaplama verimliliği (doğrusal karmaşıklık), onu yüksek frekanslı uygulamalar için özellikle uygun kılarken, SHAP'ın "orta" verimliliği (M özellik için $O(2^M)$ karmaşıklık) büyük ölçekli, gerçek zamanlı sistemlerde uygulanabilirliğini sınırlayabilmektedir. Bu durum, dolandırıcılık tespiti çalışmalarının neden çoğunlukla bireysel işlem analizi için LIME'ı tercih ettiğini, buna karşılık kredi skorlama uygulamalarının işlem süresinin daha az kritik olduğu kapsamlı kredi değerlendirmeleri için neden SHAP'a yöneldiğini açıklamaktadır.

Ayrıca, performans metrikleri uygulamaya özgü optimizasyon örüntülerini ortaya koymaktadır. GAM ve EBM, kara kutu alternatiflere kıyasla yorumlanabilirliği korurken "yüksek" doğruluk elde etmektedir. Ancak, "katkısal ilişkiler" ile sınırlı olmaları, türev ürün fiyatlaması veya sistemik risk değerlendirmesi gibi etkileşim etkilerinin kritik olduğu karmaşık finansal modellemelerde uygulanabilirliklerini sınırlandırmaktadır.

Tablo 2.6, finansal uygulamalarda karşılaşılan temel veri zorluklarının ve bunların çeşitli XAI yöntemleri üzerindeki etkisinin analizini sunmaktadır. En belirgin zorluklardan biri, SHAP'ın bir milyondan fazla kayıt içeren büyük ölçekli verisetlerine uygulanmasındaki ölçeklenebilirlik sorunudur. Yüksek hesaplama karmaşıklığı nedeniyle, veriseti boyutu arttıkça SHAP'ın verimliliği belirgin biçimde azalmaktadır. Buna karşılık, LIME ve GAM daha iyi ölçeklenebilirlik göstererek onları büyük verisetleri için daha uygun hâle getirmektedir.

Tablo 2.6. Veri özellikleri ve zorluklara ilişkin tartışma

Zorluk	Tartışma
Büyük ölçekli verisetleri	SHAP, yüksek hesaplama karmaşıklığı nedeniyle büyük verisetlerinde (>1M kayıt) zorlanırken, LIME ve GAM daha iyi ölçeklenmektedir.
Yüksek boyutlu veri	Anchor, kural tabanlı aramaya dayanması nedeniyle yüksek boyutlu verisetlerinde performans düşüşü yaşamaktadır.
Eksik değerler	EBM ve GAM, eksik verileri etkili biçimde ele alarak kararlı model performansı sağlamaktadır.
Dengesiz sınıflar	SHAP, azınlık sınıfları için bile güvenilir açıklamalar sunarak finansal uygulamalarda adaleti artırmaktadır.
Gürültüye duyarlılık	LIME, gürültülü verilere karşı savunmasızdır ve bu durum tutarsız açıklamalara yol açabilir.

Tablo 2.6'nın veri özellikleri ve zorluklara ilişkin analizi, finansal uygulamalardaki mevcut XAI uygulamalarının temel sınırlılıklarını ortaya koymaktadır. Ölçeklenebilirlik sorunu özellikle belirgindir: SHAP'ın bir milyondan fazla kayıt içeren verisetlerinde yaşadığı belgelenmiş zorluklar, işlem veri tabanlarının rutin olarak milyarlarca kayıt içerdiği modern finansal kurumların büyük veri gerçekliğiyle doğrudan çelişmektedir. Bu sınırlılık, SHAP'ın mevcut uygulamasının toplu analiz ve model geliştirme için uygun olabileceğini, ancak üretim ölçeğindeki uygulamalar için optimizasyona ihtiyaç duyduğunu göstermektedir.

Yüksek boyutlu veri sorunu, yöntem seçimini önemli ölçüde etkilemektedir. Anchor'ın "kural tabanlı aramaya dayanması nedeniyle yüksek boyutlu verisetlerinde performans düşüşü" yaşaması, alternatif veri kaynakları (sosyal medya, işlem örüntüleri, makroekonomik göstergeler) dâhil edildiğinde özellik uzaylarının rutin olarak 1000 boyutu aştığı finansal uygulamalar açısından özellikle sorunludur. Bu sınırlılık, sezgisel kural tabanlı açıklamalarına rağmen Anchor'ın neden yalnızca sınırlı düzeyde benimsendiğini (3 uygulama) açıklamaktadır.

Veri kalitesi sorunları, farklı XAI yöntemleri arasında daha çeşitli zorluklar sunmaktadır. EBM ve GAM'ın "eksik verileri ele alma" konusundaki etkinliği, veri bütünlüğünün müşteri segmentleri ve ürünler arasında önemli ölçüde değişebildiği finansal uygulamalarda önemli bir pratik avantaj sunmaktadır. Buna karşılık, LIME'ın "gürültülü verilere karşı savunmasızlığı", çoklu eski sistemler arasındaki entegrasyon sorunları

nedeniyle veri kalitesinin değişkenlik gösterebildiği finansal ortamlarda risk oluşturmaktadır.

Tablo 2.6'da vurgulanan dengesiz sınıf problemi, finansal uygulamalar açısından özel bir öneme sahiptir. SHAP'ın "azınlık sınıfları için bile güvenilir açıklamalar sunabilme" kapasitesi, pozitif vakaların tipik olarak verisetinin %5'inden azını oluşturduğu dolandırıcılık tespiti ve temerrüt tahmini için kritik önemdedir. Bu yetenek, SHAP'ın bu uygulamalardaki baskınlığını kısmen açıklamaktadır; çünkü SHAP, yalnızca çoğunluk sınıfa yanlı olmak yerine risk spektrumunun tamamında açıklama kalitesini korumaktadır.

Tablo 2.7, finansal uygulamalarda kullanılan çeşitli XAI yöntemlerinin güçlü ve zayıf yönlerini özetlemektedir. SHAP, model davranışına ilişkin kapsamlı içgörüler sunarak özellik atfı açısından güçlü bir teknik olarak öne çıkmaktadır. Güçlü kuramsal temeli, güvenilir yorumlar sunarak finans alanında yaygın biçimde kullanılmasını mümkün kılmaktadır. Ancak, temel dezavantajı yüksek hesaplama karmaşıklığıdır ve bu durum büyük verisetleri için uygulanabilirliğini sınırlandırmaktadır.

Tablo 2.7. XAI yöntemlerinin güçlü yönleri ve sınırlılıkları

Yöntem	Güçlü Yönler	Sınırlılıklar
SHAP	Kapsamlı özellik atfı, Güçlü kuramsal temel	Büyük verisetlerinde hesaplama karmaşıklığı
LIME	Hızlı yerel açıklamalar, Sezgisel yorumlar	Doğrusal olmayan ilişkilerde daha az güvenilir
Anchor	Yüksek hassasiyetli kurallar, Açık karar sınırları	Arama uzayı karmaşıklığı ile sınırlı
GAM/EBM	Eksik verileri mükemmel ele alma, Açık özellik etkileşimleri	Katkısal ilişkiler ile sınırlı

Öte yandan, LIME hızlı yerel açıklamalar ve sezgisel yorumlar sunma konusunda son derece başarılıdır. Özellikle talep üzerine açıklama üretmede etkili olduğu için gerçek zamanlı karar destek sistemleri açısından uygundur. Ancak, basit yerel yaklaşımlara dayanması nedeniyle, yüksek derecede doğrusal olmayan ilişkiler söz konusu olduğunda güvenilirliği azalmaktadır.

Anchor, model şeffaflığını ve yorumlanabilirliğini artıran açık karar kuralları sunar. Ancak performansı, arama uzayı karmaşıklığı tarafından sınırlandırılmaktadır ve bu da onu büyük ve yüksek boyutlu verisetleri için hesaplama açısından maliyetli hâle

getirmektedir. Buna karşılık, GAM ve EBM eksik verileri verimli biçimde ele alma ve özellik etkileşimlerine dair açık içgörüler sunma yetenekleriyle öne çıkmaktadır. Bu avantajlara rağmen, temel sınırlılıkları "katkısal ilişkiler"e dayanmalarıdır; bu da onların daha karmaşık finansal modellerde uygulanabilirliğini sınırlayabilir.

Genel olarak, bir XAI yönteminin seçimi; veriseti boyutu, yorumlanabilirlik ihtiyaçları ve hesaplama kısıtları gibi faktörler dikkate alınarak, finansal uygulamanın özel gereksinimlerine göre yönlendirilmelidir.

2.2.2 XAI için örnek kullanım senaryoları

SHAP, oyun teorisine dayanan güçlü kuramsal temeli ve hem yerel hem de küresel açıklamalar sunabilme yeteneği nedeniyle finansal uygulamalarda en yaygın benimsenen XAI yöntemlerinden biri hâline gelmiştir. Kredi skorlama sistemlerinde SHAP, belirli bir müşterinin neden kredi onayı aldığını ya da neden reddedildiğini açıklamak için sıklıkla kullanılmaktadır. Örneğin, kredi başvurularını ağaç tabanlı bir topluluk modeli ile değerlendiren bir banka, yıllık gelir, gecikmiş ödeme sayısı veya kredi kullanım oranı gibi her bir giriş özelliğinin tahmin edilen sonuç üzerindeki etkisini nicel olarak belirlemek için SHAP değerlerini kullanabilir. Reddedilen bir başvuruya, düşük gelir ve yüksek borç/gelir oranının olumsuz kararı yönlendiren en önemli faktörler olduğunu gösteren bir SHAP özet grafiği eşlik edebilir. Bu tür görsel ve sayısal yorumlar, şeffaflığı artırır ve denetim ile itiraz süreçlerinde insan karar vericileri destekler.

LIME, gerçek zamanlı ve örneğe özgü açıklamaların gerekli olduğu senaryolarda özellikle yararlıdır. Bankaların, bir işlemin neden şüpheli olarak işaretlendiğini uyum ekiplerine gerekçelendirmesi gereken dolandırıcılık tespit modellerinde uygulanmıştır. Tipik bir durumda, derin öğrenme tabanlı bir sınıflandırıcı, yabancı bir IP adresinden yapılan olağandışı yüksek tutarlı çevrim içi satın alma gibi işlemsel davranışlardaki anormallikleri tespit edebilir. LIME uygulanarak model, belirli örnek etrafında basitleştirilmiş doğrusal bir yaklaşım üretir ve konum, satın alma zamanı ve hesap yaşı gibi özelliklerin en etkili değişkenler olduğunu vurgular. Bu durum, analistlerin hızlı ve bilinçli kararlar almasına olanak tanırken, müşterilerin işaretlenen faaliyetlere ilişkin anlaşılır gerekçeler almasını da sağlayarak güveni ve düzenleyici uyumu destekler.

Karşı-olgusal açıklamalar, tahmini değiştirmek için giriş özelliklerinde yapılacak küçük değişiklikleri simüle eden “what-if” sorularını yanıtlayarak eyleme dönüştürülebilir

içgörüler sunar. Örneğin müşteri elde tutma analizinde, bir banka azalan hesap kullanımı, daha az giriş yapılması ve ürün yükseltmelerinin olmaması nedeniyle yüksek müşteri kaybı olasılığı tahmin edebilir. Karşı-olgusal bir açıklama, modelin müşteri kaybı olmayacağını tahmin etmesi için gereken minimum değişikliği, örneğin giriş sıklığının artırılması veya yeni bir yatırım ürününe abone olunması gibi unsurlarla belirleyecektir. Bu tür açıklamalar, pazarlama ve müşteri başarısı ekipleri için özellikle değerlidir; çünkü opak model davranışını, müşteri etkileşim stratejilerini doğrudan etkileyebilecek somut ve kişiselleştirilmiş önerilere dönüştürür.

Anchor açıklamaları, doğru olduğunda neredeyse her zaman aynı model tahminine yol açan koşulları yani "anchor"ları belirleyerek yüksek hassasiyetli, kural tabanlı içgörüler sunmayı amaçlar. Sigorta onayı veya temerrüt tahmini gibi ikili sınıflandırma görevlerinde anchor'lar, yorumlanabilir karar kurallarını ortaya çıkarabilir. Örneğin, bir mortgage risk modelinde bir anchor şu şekilde olabilir: "Eğer kredi puanı > 750 ve yıllık gelir > 80,000$ ise, onayla." Bu tür bir açıklama teknik olmayan paydaşlar tarafından kolayca anlaşılabilir ve finansal karar almada şeffaf, kural tabanlı gerekçelendirme yönündeki düzenleyici taleplerle iyi uyum sağlar. Anchor'lar, model yaşam döngüsü yönetiminin özellikle model doğrulama ve dokümantasyon aşamalarında oldukça faydalıdır.

TREPAN, eğitilmiş bir kara kutu modelin davranışını anlaşılabilir bir ağaç yapısı kullanarak yaklaşık olarak temsil eden bir karar ağacı çıkarım algoritmasıdır. Sinir ağlarının risk sınıflandırmasında kullanıldığı finansal denetim sistemlerinde, denetçiler belirli risk etiketlerinin nasıl üretildiğini anlamakta çoğu zaman zorluk yaşamaktadır. TREPAN, özgün modelin tahminlerini taklit eden ancak insan tarafından okunabilir kurallar cinsinden açıklanabilir olan bir vekil karar ağacı çıkarmak için uygulanabilir. Örneğin, TREPAN "Yüksek" risk etiketinin yüksek işlem sıklığı, uluslararası para transferleri ve yakın tarihli ek hesap açıklarının bir bileşimine karşılık geldiğini gösteren bir ağaç üretebilir. Bu durum yalnızca yorumlanabilirliği artırmakla kalmaz, aynı zamanda denetçilerin mantık yollarını izlemesine ve politika eşiklerine uyumu doğrulamasına da olanak tanır.

2.2.2.1 Temsili vaka çalışması örnekleri

Önceki bölümler XAI yöntemleri, uygulama alanları ve karşılaştırmalı özellikleri

hakkında yapılandırılmış bir genel bakış sunarken, daha derin ve bağlama duyarlı örneklerin gerekliliği devam etmektedir. Bu ihtiyacı karşılamak amacıyla, aşağıdaki alt bölüm, incelenen literatürden seçilen ve gerçek dünyadaki finansal senaryolarda XAI'ın pratik kullanımını ve iş değeri boyutunu örnekleyen iki temsili vaka çalışmasını sunmaktadır. Bu örnekler, yöntemsel açıklıkları, pratik etkileri ve temsil ettikleri hem XAI tekniği hem de finansal alan çeşitliliği temel alınarak seçilmiştir. Birlikte ele alındıklarında, açıklanabilirlik tekniklerinin finansta karar verme süreçlerini, yalnızca soyut model şeffaflığının ötesine geçerek, anlamlı biçimde nasıl etkileyebildiğini göstermektedirler.

Vaka Çalışması 1 – SHAP ve LIME ile Kredi Kartı Dolandırıcılığı Tespiti: Yakın dönem araştırmalardan biri, Swetha (2024) SHAP ve LIME ile post-hoc yorumlanabilirlik açısından değerlendirilen XGBoost ve Decision Tree sınıflandırıcılarını bütünleştiren güçlü bir kredi kartı dolandırıcılığı tespit çerçevesi sunmaktadır. Finansal amaç, düzenleyici uyumu desteklemek ve denetim süreçlerindeki sürtünmeyi azaltmak için şüpheli işlemleri daha yüksek şeffaflıkla belirlemektir. SHAP; işlem tutarı, işlem sıklığı ve iş yeri türü gibi özelliklerin küresel katkısını nicel olarak belirlemek için kullanılırken, LIME yerel ve örneğe özgü açıklamalar sunmuştur. Bu ikili XAI yaklaşımı, finansal kurumların hangi girdilerin dolandırıcılık işaretlerini tetiklediğini daha iyi anlamasına olanak tanımış; operasyonel güveni ve hukuki gerekçelendirilebilirliği artırmıştır.

Vaka Çalışması 2 – Yatırım Stratejilerinde Özellik Seçimi için SHAP ve LIME: Man ve Chan (2021) tarafından yapılan karşılaştırmalı bir çalışmada, S&P 500 tabanlı işlem stratejileri için geliştirilen Random Forest modellerinde özellik öneminin kararlılığını değerlendirmek amacıyla SHAP ve LIME yöntemlerini uygulamıştır. Amaçları, Sharpe oranı ve kümülatif getiri gibi performans ölçütlerini tutarlı biçimde iyileştiren finansal göstergeleri belirlemektir. Analiz, sentetik verisetleri, kamuya açık kıyaslama veri setleri ve özel yatırım verileri dâhil olmak üzere çoklu veri kaynaklarını içermektedir. Yazarlar, özellik sıralamalarının algoritmik rastlantısallığa ne kadar duyarlı olduğunu değerlendirmek amacıyla bir "istikrarsızlık endeksi" önermiştir. Sonuçlar, LIME'ın istikrarlı ve yüksek faydalı özelliklerin seçilmesinde hem SHAP'tan hem de MDA'dan daha iyi performans gösterdiğini ortaya koymuştur. Önemli olarak, SHAP ve LIME'ın dâhil edilmesi yalnızca model şeffaflığını kolaylaştırmakla kalmamış, aynı

zamanda finansal getiri ölçütlerinde ölçülebilir iyileşmelere de dönüşmüştür; bu da açıklanabilir yapay zekânın portföy tasarımı ve değerlendirmesindeki operasyonel değerini vurgulamaktadır.

2.2.3 Yöntemsel karşılaştırma

Bu bölümde, çalışmalarda kullanılan verisetleri ve yöntemlere ilişkin değerlendirmeler verilmektedir. Tablo 2.8, literatür taramasına konu olan çalışmalarda hangi verisetlerinin kullanıldığını, verisetlerinin kaynaklarını, XAI yöntemlerini, XAI yöntemlerinin finansta uygulamalarını ve XAI yöntemlerinin çalışmalardaki katkılarını göstermektedir.

Tablo 2.8. XAI yöntemlerine genel bakış

Yazar & Yıl	Finansal Problem / Uygulama Alanı	XAI Yöntemi	Veriseti	Veriseti Kaynağı	Performans / Sonuç
Bussman ve ark. (2021)	Kredi risk yönetimi	Korelasyon ağları ile Shapley değerleri	Eşler Arası (P2P) Borç Verme Veriseti	Alman P2P platformunda n tescilli KOBİ kredi verileri	XAI, yorumlanabilir kümeler aracılığıyla borçlu profillerinin segmentasyonunu mümkün kılmıştır; artırılmış şeffaflık güveni kolaylaştırmış ve temerrüt riskini azaltmıştır; çalışma, P2P platformlarında gerçek dünya uygulanabilirliğini doğrulamıştır.
Giudici ve Gramegn a (2021)	Kredi riski değerlendirmesi	SHAP ve LIME	Kredi Temerrüt Veriseti	Resmî veri tabanlarından İtalyan KOBİ kredi verileri	SHAP, özellik öneminin daha iyi sıralamasını sağlamıştır; model kararlılığı ve ayırt edici güç bakımından LIME'dan daha iyi performans göstermiştir; kredi kararlarının gerekçelendirilmesini desteklemiştir.
Li ve ark. (2025)	Sosyal kredi platformlarında temerrüt tahmini / borçlu temerrüt riski	SHAP, LIME	Lending Club veri seti	Kaggle (Lending Club 2007–2020Q1)	LightGBM tabanlı açıklanabilir model, temerrüt tahmininde 0.87 doğruluk sağlamış; geleneksel yaklaşımlardan daha iyi performans gösterirken, SHAP ve LIME ile karar sürecinin şeffaflığını artırmış ve kredi verme kararları için daha yorumlanabilir sonuçlar sunmuştur.
Giudici ve Raffinetti (2021)	Genel XAI değerlendirmesi	Shapley-Lorenz yöntemi	Sentetik Veriseti	Yöntem doğrulaması için simüle edilmiş veriler	Eşitlik ve adalet ölçütlerini iyileştiren Lorenz-tutarlı Shapley değerlerini tanıtmıştır; dağılıma duyarlı alanlarda geleneksel SHAP'ın sınırlılıklarını vurgulamıştır.
Ugarte Chacon ve ark. (2026)	Tüketici kredi skorlama / temerrüt olasılığı analizi	SHAP	Anket verisi	Meksika Finansal Kapsayıcılık Ulusal Anketi	XGBoost tabanlı açıklanabilir kredi skorlama modeli, geleneksel finansal veriler yerine alışılmışın dışındaki sosyoekonomik ve davranışsal verileri kullanmış; SHAP ile

					varsayılan krediyi etkileyen temel değişkenleri ortaya koymuş ve özellikle COVID-19'un ekonomik etkileri ile finansal tutum ve davranışların belirleyici olduğunu göstermiştir.
Giudici ve Raffinetti (2022)	Siber risk yönetimi	Shapley-Lorenz yöntemi	Siber Risk Veriseti	Tescilli siber güvenlik olay verileri	Siber saldırı profillerinin yorumlanabilir sınıflandırmasını gerçekleştirmiştir; XAI modelleri uzman kural kümeleriyle yüksek tutarlılık göstermiştir; yüksek riskli siber ortamlarda modele güveni güçlendirmiştir.
Babaei ve ark. (2022)	Kripto varlık tahsisi	SHAP	Kripto Varlık Verileri	Büyük kripto borsalarından piyasa verileri	Dijital varlık portföyleri için model güvenilirliğini artırmıştır; SHAP, yatırım kararlarını düzenleyicilere ve yatırımcılara açıklayarak robo-danışman kabulünü iyileştirmiştir.
Babaei ve ark. (2023)	FinTech kredi verme	SHAP ile geliştirilmiş Random Forest	FinTech Kredi Verileri	FinTech platformundan tescilli KOBİ kredi verileri	SHAP, kredi onayındaki özellik etkilerini açıklığa kavuşturmuştur; adalet ölçütlerini ve kullanıcı kabulünü iyileştirmiştir; çalışma SHAP'ı uyumu destekleyen bir araç olarak önermiştir.
Giudici ve ark. (2024)	Finansal tahminleme	Shapley-Lorenz yöntemi	Finansal Zaman Serileri	Finansal borsalardan piyasa verileri	Oynaklık sıçramalarına ilişkin yorumlanabilir tahminler sunmuştur; Lorenz ile geliştirilmiş SHAP, aykırı değerlere ve zamansal kaymalara karşı model sağlamlığını iyileştirmiştir.
Chen ve ark. (2024)	Kredi derecelendirme adaleti	Shapley-Lorenz adalet ölçütleri	Kredi Derecelendirme Verileri	Finansal kurumlardan tescilli kredi derecelendirme verileri	Düzenleyici kısıtlar altında adalet farkındalıklı kredi skorlama olanağı sağlamıştır; model performansını korurken daha düşük grup farklılıkları göstermiştir.
Giudici ve Raffinetti (2023)	Finansal AI sistemleri	SAFE AI ilkeleri	SAFE AI Çerçevesi	Kavramsal çerçeve, veriseti kullanılmamıştır	Standartlaştırılmış ilkeler (Security, Accountability, Fairness, Explainability) önermiştir; finansal AI benimsenmesini denetlemek ve yönlendirmek için politika uyumlu bir çerçeve sunmuştur.
Babaei ve ark. (2023)	SAFE AI değerlendirmesi	Rank Graduation Box	Rank Graduation Box Modeli	Kavramsal çerçeve, veriseti kullanılmamıştır	Güvenlik ve yorumlanabilirlik açısından AI sistem olgunluğunu değerlendirmek için nicel ölçütler geliştirmiştir; düzenleyici uyum katmanları boyunca uygulanabilir.
Calzarossa ve ark. (2023)	Siber güvenlik	XAI değerlendirmesi	Siber Güvenlik Çerçevesi	Kavramsal çerçeve, veriseti kullanılmamıştır	XAI'ı güvenlik denetimleriyle birleştiren çok katmanlı bir değerlendirme aracı tasarlamıştır; siber saldırı tespiti ve model dayanıklılığı analizinde kullanım senaryoları göstermiştir.
Zacharias	Kredi riski	SHAP	German	UCI Machine	SHAP tabanlı seçim, Pearson ve

ve ark. (2022)	değerlendirmesi / kredi değerliliği skorlama		Credit Dataset	Learning Repository	XGBoost tabanlı yöntemlerle karşılaştırıldığında benzer veya daha iyi performans vermiştir; açıklanabilirliği ve paydaşlar açısından gerekçelendirilebilirliği iyileştirmiştir
Cernevicienе ve Kabasinskas (2024)	Kredi yönetimi, Risk değerlendirmesi, Hisse senedi tahmini, Portföy optimizasyonu, Dolandırıcılık tespiti	SHAP, LIME, Kural tabanlı öğreniciler, Karşı-olgusal'lar, Görsel açıklamalar, Şeffaf modeller, Çok yöntemli çerçeveler	Çeşitli (incelenen 138 çalışma)	Scopus, Web of Science	Performans ölçütleri çalışmaya göre değişmektedir; odak doğruluktan çok açıklanabilirlik üzerinedir. SHAP ve LIME, özellik ilgisi ve yerel yorumlanabilirlik için yaygın biçimde kullanılmaktadır. Yöntemleri birleştiren çerçeveler güveni ve açıklığı artırmıştır.
Weber ve ark. (2023)	Risk yönetimi, Portföy optimizasyonu, Hisse senedi ticareti, Kara para aklamayla mücadele, Finansal işlem sınıflandırması	Şeffaf modeller (ör. karar ağaçları), Post-hoc yöntemler (SHAP, LIME, PDP, ICE), Sinir ağları, Kural tabanlı öğreniciler	Çeşitli (incelenen 60 çalışma)	Finans, BS ve BB dergileri; ACM, IEEE, AIS eLibrary	Derleme makalesi; belirli bir doğruluk değeri yoktur. Katkı olarak yöntem kategorizasyonu, düzenleme farkındalıklı konuşlandırma analizi ve kara para aklamayla mücadele gibi yeterince araştırılmamış alanların kapsamını sunmaktadır.
Beckmann ve ark. (2023)	Tahvil fazla getiri tahmini (ABD ve Alman piyasaları)	SHAP	ABD Hazine Tahvil Getirisi, Makroekonomik Veriler (ABD ve Almanya)	Liu ve Wu (2021), McCracken ve Ng (2016), Deutsche Bundesbank, FRED, Thomson Reuters Eikon	SHAP, Extremely Randomized Trees üzerinde uygulanmıştır; getiri eğrisi eğimi ve konut piyasasının temel öngörücüler olarak belirlenmesini sağlamıştır. 10 yıllık tahviller için R^2 değeri ABD'de %20.5'e, Almanya'da ise yaklaşık %10'a ulaşmıştır.
Dessain ve ark. (2023)	ECB düzenlemesi altında kredi skorlama ve finansal performans analizi	Explainable Boosting Machine (EBM), GAMI-Net, Isotonic GAMs, Logistic Regression, LDA (doğası gereği açıklanabilir); XGB, MLP vb. ile karşılaştırılmıştır.	Avrupa borçlu kredi verileri (2019–2022, 50,000+ risk maruziyeti)	Küresel bir ticari kredi sigortacısından gizli veriler	EBM ve GAMI, AUROC bakımından kara kutu modellere yakın sonuçlar vermiştir; MLP en yüksek finansal getiriyi sağlamıştır (en fazla %2.06), en iyi XAI modelleri (EBM/GAMI) MLP'nin 14–21 baz puan gerisinde kalmıştır; izotoniklik, finansal performans kaybı olmadan yorumlanabilirliği iyileştirmektedir.
Agosto ve ark. (2023)	ESG faktörleri kullanılarak kredi derecelendirme tahmini	Bayesian statistical learning model (tasarım gereği açıklanabilir)	1382 Avrupa şirketi (MSCI, Refinitiv, S&P ESG skorları; Bloomberg kredi derecelendir	MSCI ESG Research, Refinitiv, S&P Global, Bloomberg	Birleştirilmiş Bayes modeli AUROC = 0.59, AUPRC = 0.24 elde etmiştir; XGBOOST ile karşılaştırılabilir sonuç vermiştir (AUROC = 0.61, AUPRC = 0.25). Bayes modeli, minimum performans ödünleşimi ile yorumlanabilir ESG ağırlıkları sunmaktadır.

			meleri)		
Dennis ve ark. (2024)	Kredi temerrüdü tahmini	SHAP, LIME	Taiwan Credit Card Default Dataset	UCI Machine Learning Repository	ANN-LSTM topluluğu: F1 Skoru = 0.89; Kesinlik = 0.84; Duyarlılık = 0.95. XAI, derin topluluk modellerinin yorumlanabilirliğini sağlamıştır
Watson ve ark. (2022)	Kredi skorlama, Tekrar suç işleme tahmini, Tıbbi teşhis (genel bir vaka olarak kullanılmıştır)	Rational Shapley Values (nedensel / Karşı-olgusal entegrasyonuna sahip SHAP varyantı)	COMPAS (ProPublica), German Credit, UCI Diabetes	ProPublica, UCI ML Repository	Sınıflandırma doğruluğuna odaklanmamaktadır; katkı, yorumlanabilirlik ve önyargı tespitinde yöntemsel iyileştirmedir
Murindanyi ve ark. (2023)	Perakende bankacılıkta müşteri etkileşimi tahmini	LIME, Evidently AI (güven tabanlı açıklama)	Berka Veriseti	Çek bankası (Česká spořitelna) / Açık	Blending ensemble modeli: Doğruluk = %99, AUC = 1.00; yorumlanabilirlik LIME ile, hata analizi ise Evidently AI ile sağlanmıştır
Nalule ve ark. (2024)	Kredi riski değerlendirmesi / kredi değerliliği skorlama	LRP, SHAP, LIME, Integrated Gradients	German Credit Dataset	UCI Machine Learning Repository	Hibrit (LSTM ve CNN): Doğruluk = %72, ROC-AUC = 0.762; XAI yöntemleri özellik düzeyinde yorumlanabilirlik ve karar şeffaflığı sağlamıştır
Mariotti ve ark. (2021)	Bankacılık hizmetlerinde kredi onayı ve FAT-E analizi	Bulanık sistemler, TREPAN, LIME, SHAP	Kredi veriseti (Kaggle)	Kaggle	TREPAN (%88) - Random Forest'ın doğruluğuna (%92) vekil yorumlanabilirlik sağlamaktadır
Mahyoub ve ark. (2023)	Konut kredisi temerrüt tahmini	LIME, SHAP	Konut kredisi veriseti	Konut finansman şirketi verileri (ayrıntılı yapı verilmemiştir)	SHAP ve LIME ile champion model (%94) - Model şeffaflığını ve yerel yorumlanabilirliği artırmaktadır
Ayad ve ark. (2023)	Bankacılıkta kredi onay tahmini	SHAP, PDP, ICE, Global Vekil, Karşı-Olgusal Açıklamalar	Give Me Some Credit veriseti (Kaggle)	Kaggle	Vekil Decision Tree (%98.4) - Kara kutu modele yakın yorumlanabilirlik sağlamaktadır
Manzoor ve ark. (2024)	Finans ve telekomünikasyonda müşteri kaybı tahmini	Topluluk yöntemleri (XGBoost, AdaBoost), Derin Öğrenme	Customer churn veriseti	Sektöre özgü veriler (açıkça belirtilmemiştir)	Topluluk yaklaşımı, karmaşık müşteri kaybı tahminlerinde doğruluğu anlamlı biçimde iyileştirmektedir
Adams ve Hagras (2020)	Düzenleyici uyum, finansta piyasa istikrarı	Type-2 Fuzzy Logic, Neural Networks	Bankacılık verisetleri	Sektöre özgü (ayrıntılandırılmamıştır)	Type-2 Fuzzy Logic yüksek açıklanabilirlik sunmaktadır, doğruluk ise Sinir Ağlarına kıyasla biraz daha düşüktür
Bhadula ve ark. (2024)	Altın fiyatı tahmini	Huber Regressor ve MLP ile SHAP	Altın fiyatı veriseti	Yahoo Finance	Huber Regressor en yüksek doğruluğu göstermektedir (R2 = 0.9999), SHAP ise özellik düzeyinde yorumlanabilirliği artırmaktadır
Mahajan ve Shukla (2023)	İflas tahmini ve false positive analizi	SHAP	Tayvan ve Polonya iflas verisetleri	Taiwan Economic Journal, Kaggle	SHAP, false positive analizini güçlendirerek model yorumlanabilirliğini ve özellik düzeyindeki içgörüleri iyileştirmektedir

Arifah ve Nihaya (2023)	P2P borç vermede kredi risk yönetimi	SHAP, LIME	P2P Lending verisetleri	Çeşitli finansal kayıtlar	SHAP ve LIME, özellikle düzenleyici uyum açısından kredi riski şeffaflığını ve hesap verebilirliği iyileştirmektedir
Chou ve ark. (2024)	G7 piyasalarında portföy optimizasyonu	Kuantum esinli algoritmalar (ELSA-GNQTS), Trend Ratio	Küresel borsa verileri	WRDS, Yahoo Finance	Risk-getiri dengesinde yüksek doğruluk, Kuantum esinli tekniklerle yatırımcılar için yorumlanabilirliğini iyileştirmektedir
Vashi ve Shanbhag (2023)	Banka pazarlama başarısı tahmini	SHAP	Portekiz bankacılık kampanya veriseti	UCI Repository	Gradient Boost (%91.26) - SHAP, kampanya bazında özellik önemini göstererek yorumlanabilirliği artırmaktadır
Pundir ve ark. (2024)	Hisse senedi fiyat tahmini	Hibrit AI Özellik Seçimi	Kurgusal hisse senedi fiyat veriseti	Belirtilmemiştir	Hibrit özellik seçimi (doğruluk: %85) - Özellik karmaşıklığını azaltarak tahmini iyileştirmektedir
Varun ve ark. (2023)	Kredi skoru analizi ve tahmini	SHAP, Özellik Sıralama	Filipin bankası kredi veriseti	Tescilli	Random Forest (%88) - SHAP, yüksek riskli ve düşük riskli müşterilerin yorumlanabilirliğini artırmaktadır
Guidotti ve Nanni (2020)	Kaza tahmini ve risk değerlendirmesi	SHAP, Özellik Önemi	OctoTelematics hareketlilik veriseti	OctoTelematics	Random Forest (doğruluk belirtilmemiştir) - SHAP, yüksek riskli sürüş davranışlarını ve bağlamsal risk faktörlerini vurgulamaktadır
Vihursky i ve ark. (2024)	Kredi kartı dolandırıcılığı tespiti	LIME, Decision Trees	European credit card veriseti	Kaggle	Decision Tree ve Random Forest (%100) - LIME, dolandırıcılık tespiti için bireysel işlem kararlarını açıklamaktadır
Bhowmik ve ark. (2022)	Finansal dolandırıcılık tespiti	SHAP, Deep Belief Network (DBN)	ULB Credit Card Fraud Detection Dataset	ULB, Worldline	DBNex %94 doğruluk elde etmektedir; SHAP, özellik düzeyindeki katkıları açıklayarak yorumlanabilirliği artırmaktadır
Raman ve ark. (2023)	Finansal stres tahmini	SHAP, LIME	Makro ve mikroekonomik göstergeler	Çeşitli ekonomik kaynaklar	Yorumlanabilirlikle birlikte yüksek doğruluk; SHAP ve LIME, finansal stresi etkileyen kritik ekonomik göstergeleri açıklamaktadır
Maurya ve ark. (2024)	Risk kontrolü, yatırım karar verme, iş sonucu	SHAP, LIME	Küresel finansal göstergeler	Belirtilmemiştir	Karmaşık yatırım modellerinde yorumlanabilirlik sağlamaktadır; doğruluk model uygulamasına göre değişmektedir
Thanatha mathee ve ark. (2024)	İşletmenin sürekliliği tahmini	Anchor XAI, Attention-Weighted XGBoost	İflas veriseti	Tescilli	Attention-weighted XGBoost (%98) - Açık, anchored açıklamalarla güveni artırmaktadır
Tsai ve Chen (2023)	Mum çubuğu örüntü tanıma ile finansal işlem	GAF kodlamalı CNN, Training Aid Augmentation	Geçmiş hisse senedi fiyat verileri	Belirtilmemiştir	Training Aid Augmentation ile yüksek yorumlanabilirlik, doğruluk ise %90 seviyesinde korunmuştur
Naretto ve ark. (2022)	XAIda gizlilik maruziyeti	Membership Inference Attack (MIA), TREPAN	Adult, DIVA, Bank verisetleri	UCI, DIVA (özel), BANK	Küresel açıklayıcılarda, kara kutu modellere kıyasla membership inference attack'lar nedeniyle yüksek gizlilik riski
Valensky ve	Hisse senedi fiyatları ile	SHAP, LIME, Permütasyon	S&P 500	Yahoo	CNN, LSTM ile yüksek doğruluk; SHAP, 'Close' fiyatları

Mohaghegh (2023)	finansal tahminleme	önemlilik	veriseti	Finance	gibi kritik özellikler için yorumlanabilirliği iyileştirmektedir
Stevens ve ark. (2020)	Mikrofinansta adil borç verme	SHAP, Adalet Algoritmaları (Reweighing, Equality of Odds)	Kiva loans veriseti	Kiva.org	SHAP model yorumlanabilirliğini artırırken, adalet algoritmaları doğruluğu koruyarak önyargı azaltımını iyileştirmektedir
Park ve ark. (2021)	İflas tahmini	LIME	Kore şirket iflas veriseti	Douzone Bizon ICT Group	LightGBM ile yüksek AUC, LIME ise tutarlı örnek bazlı özellik önemi sağlamaktadır
Kalra ve Mittal (2024)	İşlemde finansal karar destek	Contrastive Divergence Neural Network (CDNN), LASSO	Finansal işlem veriseti	Kaggle	CDNN %97.91 doğruluk elde etmektedir, LASSO tabanlı özellik seçimi ile güçlendirilmiştir
Bhattacharya ve ark. (2023)	Çalışan terfisi için öngörücü analitik	XGBoost, LIME, SHAP	Employee promotion veriseti	Kaggle	XGBoost %95 doğruluk elde etmektedir, LIME ve SHAP temel terfi faktörlerini vurgulamaktadır
Zhang ve ark. (2022)	Sızma, zararlı yazılım, spam tespitinde siber güvenlik uygulamaları	SHAP, LIME, Karşı-Olgusal Açıklamalar	Çeşitli siber güvenlik verisetleri	Belirtilmemiştir	Yüksek yorumlanabilirlik, doğruluk uygulamaya göre değişmektedir; anomali tespitini ve kararlara duyulan güveni iyileştirmektedir
Kotrachai ve ark. (2023)	Kredi kartı dolandırıcılığı tespiti	SHAP	Credit Card Fraud Detection Dataset (Kaggle)	Kaggle (Worldline ve ULB iş birliği)	Özellik seçiminden sonra yüksek kesinlik korunmaktadır; karar vermede yorumlanabilirliği artırmaktadır
Sikha ve Vijayakumar (2023)	Kredi risk yönetimi	H2O AutoML, SHAP, Partial Dependence Plots	KOBİ'ler için kredi riski veriseti	Belirtilmemiştir	Yüksek tahmin doğruluğu, yorumlanabilirlik ve sağlamlık kontrolleriyle tutarlı kararlar sağlamaktadır
Carta ve ark. (2021)	Hisse senedi piyasası tahmini	Decision Tree, Sözlük tabanlı NLP	S&P 500 ve Dow Jones DNA veriseti	S&P 500, Dow Jones DNA	Sözlük tabanlı özelliklerle Decision Tree doğruluğu iyileştirilmiş, tutarlı sınıflandırma doğruluğu elde edilmiştir
Azad ve ark. (2023)	E-ticarette müşteri davranışı analizi	LIME, SHAP, ELI5	E-ticaret işlem verileri	Kaggle	Satın alma niyetini yakalamada doğruluk artmıştır; XAI yöntemleri özellik düzeyinde içgörüler sunmaktadır
Yang ve Wu (2021)	Kredi riski için lojistik regresyon iyileştirmesi	xNN	Give Me Some Credit veriseti	Kaggle	Yüksek AUC ile geliştirilmiş yorumlanabilirlik; xNN doğruluk ile açıklanabilirlik dengesine yardımcı olmaktadır
Zhu ve ark. (2020)	Rejim tabanlı varlık tahsisi	Hiyerarşik Kümeleme, BL modeli	Makroekonomik ve piyasa verileri	WIND veri tabanı	Kümeleme yoluyla yüksek getiri tahmin doğruluğu; rejime özgü varlık ağırlık ayarlamalarını mümkün kılmaktadır
Thommandru ve ark. (2023)	Kara para aklamayla mücadele, veri gizliliği	SHAP, LIME, Karşı-Olgusal Açıklamalar	Kredi ve finansal işlem verileri	Belirtilmemiştir	Finansal suç tespitinde yorumlanabilirliği artırırken gizlilik kaygılarını da ele almaktadır
Mavrogiorgou ve	Gömülü finans, veri pazaryeri	Situation-Aware	Gömülü finans işlem	Çeşitli finans sağlayıcıları	Federatif finans platformlarında veri ticareti ve fiyatlandırmada

diğerleri (2023)		Explainability (SAX), XAI	verisetleri		şeffaflık sağlamaktadır
Sovrano ve Vitali (2022)	Finans ve sağlık alanları	XGBoost, TreeSHAP, CEM	Kredi onayı ve kalp hastalığı verisetleri	UCI, CDC ve finansal veri sağlayıcıları	TreeSHAP karar vermede yorumlanabilirliği iyileştirmektedir; açıklanabilirlik derecesi metriği tanıtılmıştır
Bandi ve ark. (2021)	Hisse senedi endeksi hareket tahmini	ANN, ARIMA, Duygu Analizi	Nifty50 hisse senedi endeksi ve haber verileri	Yahoo Finance, ET Markets	Teknik göstergiler ve duyarlılığın birleştirilmesiyle yüksek doğruluk elde edilmiştir; XAI günlük hareketleri açıklamaktadır
Dou ve Peng (2023)	Kredi değerlendirmesi	GCNN, q-Consistent Summary Explanation	Kredi değerlendirme senaryoları	Belirtilmemiştir	q-tutarlılığı ile model yorumlanabilirliğini iyileştirerek kredi kararlarına desteği artırmıştır
Boardman ve ark. (2022)	Kredi riski değerlendirmesi	Integrated Gradients	Kredi skorlama verisetleri	ABD kredi büroları (ör. Equifax, TransUnion)	Integrated Gradients, sektörde standart değişken atfını sinir ağlarına uyarlayarak daha iyi yorumlanabilirlik sağlamaktadır
Bora ve ark. (2022)	Sağlık sigortası prim maliyeti tahmini	LIME, SHAP, InterpretML	Health insurance veriseti	Kaggle	Random Forest doğruluğu iyileştirmektedir; LIME ve SHAP prim maliyeti tahminleri için özellik yorumlanabilirliğini artırmaktadır
Xiao ve Jiao (2021)	Kredi riski değerlendirmesi	RATP sezgisel kural tabanlı yöntem	Kredi değerlendirme verisetleri	Lending Club, UCI Repository	Karmaşık kredi verileri için optimize edilmiş, kural tabanlı yaklaşımla yüksek yorumlanabilirlik
Murindanyi ve ark. (2023)	Perakende bankacılıkta müşteri kaybı tahmini	SHAP, LIME	Berka ve Kaggle churn verisetleri	Czech Bank, Kaggle	Random Forest, Berka üzerinde %99 doğruluk elde etmektedir; SHAP ve LIME müşteri kaybı tahminlerine açıklamalar sunmaktadır
Supardi ve Gambetta (2023)	Endonezya pazarında kredi verme	LIME, SHAP, Anchors	Endonezya kredi verileri	CCB Indonesia	Decision Tree en yüksek yorumlanabilirliği elde etmiştir; XGBoost doğruluğu artırmakta ancak daha az yorumlanabilirdir
Duvnjak ve ark. (2024)	Kredi riski değerlendirmesi	GA2M, XGB2, GAMI-net	Hırvat bankası kredi veriseti	Hırvat bankası (anonim)	GA2M ve GAMI-net, kara kutu modellere göre küçük bir doğruluk ödünleşimiyle yorumlanabilirliği artırmaktadır
Das ve ark. (2024)	Finansal yatırım duygu analizi	ESAN (Emotion-Sentiment Attention Network)	FinEMA veriseti	Twitter, Finansal haber kaynakları	Duygu analizinde %89, emotion recognition'da %79 doğruluk elde edilmiştir; altta yatan nedenleri açıklamaktadır
Mandeep ve ark. (2022)	Finansal tahminleme	Random Forest, XGBoost, SHAP, LIME	S&P 500 hisse verileri	Yahoo Finance	Tahmin doğruluğunu artırmıştır; SHAP ve LIME yatırım kararları için özellik düzeyinde içgörüler sunmaktadır
Govindu ve ark. (2024)	Kredi onay tahmini	Random Forest, SHAP	Kaggle kredi veriseti	Kaggle	Kredi onayında %95.4 doğruluk elde edilmiştir; SHAP kredi sonuçlarını etkileyen temel özellikleri vurgulamaktadır
Wu ve Wang (2021)	Kredi kartı dolandırıcılığı tespiti	AutoEncoder, LIME	Avrupa kredi kartı dolandırıcılık verileri	Kaggle	%90.61 doğruluğa ulaşmaktadır; LIME bireysel işlemler için yorumlanabilirlik sağlamaktadır

Dastile ve ark. (2022)	Kredi skorlama	Karşı-Olgusal Açıklamalar	UCI kredi skorlama veriseti	UCI Repository	Karşı-olgusal'lar önyargıyı en aza indirmek için üretilmiştir, kredi kararları için optimize edilmiş yorumlanabilirlik sağlamaktadır
Potluri ve ark. (2024)	Müşteri segmentasyonu ve kişiselleştirilmiş pazarlama	LIME, SHAP, K-means	Finansal işlem ve demografik veriler	Finansal kurumlar	Segmentasyon doğruluğunu iyileştirmiştir; XAI hedefli pazarlama stratejileri için müşteri içgörüleri sunmaktadır
Dastile ve Celik (2021)	Derin öğrenme ile kredi skorlama	CNN, Saliency Map, LIME, SHAP	Alman ve Avustralya kredi verisetleri	Açık finansal verisetleri	CNN yüksek tahmin doğruluğu elde etmektedir; XAI yöntemleri özelliklerin tahminler üzerindeki etkilerini göstererek yorumlanabilirliği iyileştirmektedir
Kong ve ark. (2020)	Yatırım fonları için likidite oranı tahmini	Ensemble Learning	Yatırım fonu likidite verileri	Wind Database, China A-Shares	Ensemble Learning yüksek tahmin doğruluğu elde etmektedir; XAI fonlardaki likidite eğilimlerini açıklamaktadır
Galego Hernandes ve ark. (2021)	URL özellikleri kullanılarak oltalama tespiti	LIME, Explainable Boosting Machine (EBM)	URL phishing verileri	PhishTank, OpenPhish, Alexa	EBM küresel ve yerel açıklamalar sunmaktadır; bireysel tahmin içgörülerinde LIME yardımıyla yüksek yorumlanabilirlik sağlamaktadır
Prakash ve ark. (2023)	Tüketici davranışı için öngörücü analitik	Decision Tree, ANN, Bagging Ensemble	Tüketici satın alma verileri	CMR Institute of Technology	%95.3 doğruluk; Bagging ve ANN modelleri çevrim içi davranışta yorumlanabilirlikle birlikte yüksek doğruluk sağlamaktadır
Rizinski ve ark. (2024)	Finans için duygu analizi	SHAP, XLex Lexicon	Finansal duygu veriseti	Finansal haberler ve sosyal medya	XLex, SHAP ile birleştirildiğinde LM sözlüğüne göre doğrulukta 0.450 artış elde etmiştir
Niu ve ark. (2022)	Kredi tahmini	XGBoost, SHAP	Lending Club veriseti	Lending Club	XGBoost %99.50 doğruluk elde etmektedir; SHAP kredi tahmininde temel özelliklerin etkisini açıklamaktadır
Tiwari ve ark. (2023)	Bankacılıkta dolandırıcılık tespiti	XAI	Scopus'tan bibliyometrik veriler	Scopus, Google Scholar	AI tabanlı dolandırıcılık tespiti geleneksel yöntemlerden daha iyi performans göstermektedir; XAI düzenleyici amaçlar için şeffaflık sağlamaktadır
Metwally ve ark. (2024)	Risk yönetiminde Sorumlu AI	SHAP, LIME, Karşı-Olgusal Açıklamalar	İç risk yönetimi verileri	Çeşitli risk yönetim sistemleri	AI risk değerlendirmelerinde şeffaflığı artırmaktadır; sorumlu AI çerçevesi ihtiyacını vurgulamaktadır
Maree ve ark. (2020)	Dolandırıcılık tespiti	SHAP, Metin Kümeleme	Finansal işlem verileri	İskandinav bankası (anonim)	Federe öğrenme ile yüksek doğruluk; SHAP dolandırıcılık sınıflandırmalarında model yorumlanabilirliğini iyileştirmektedir
Awosika ve ark. (2024)	Finansal dolandırıcılık tespiti	Federe Learning, SHAP	Banka dolandırıcılık işlem verileri	Anglia Ruskin University	Gizliliği korurken yüksek doğruluk; Federe öğrenme, XAI ile birleşerek şeffaflık sağlamaktadır
Maree ve	Müşteri davranışı	LIME, Linear	Harcama	Norwegian	RNN, doğrusal regresyon

Omlin (2022)	analizi için mikro-segmentasyon	Regression	davranışı verileri	Research Foundation	aracılığıyla yüksek yorumlanabilirlik elde etmektedir; harcama örüntüleri için sembolik açıklamalar sunmaktadır
Alblooshi ve ark. (2024)	Kredi skorlama analizi	XGBoost, LIME	Kredi skorlama verisetleri (Tayvan, Avustralya, Kredi Riski)	Kaggle	Credit Risk verisetinde %93 doğruluk elde edilmiştir; LIME risk değerlendirme kararları için yerel açıklamalar sunmaktadır
Zhu ve ark. (2020)	Finansal yatırım için sektör rotasyonu	Random Forest, Ridge Regression	Makroekonomik faktörler (Çin borsa verileri)	WIND Database	Sektör seçiminde doğruluk yorumlanabilirlik yoluyla iyileştirilmiştir; makroekonomik faktörler zamanlama stratejilerine yardımcı olmaktadır
Moscato ve ark. (2021)	P2P platformlarında kredi skorlama tahmini	LIME, Anchors, SHAP	Lending Club veriseti	Lending Club	Sınıf dengesizliğinin ele alınmasında doğruluk iyileştirilmiştir; XAI kredi riski tahminlerinin anlaşılmasına yardımcı olmaktadır
Ghosh ve ark. (2022)	COVID-19 sırasında vadeli işlem fiyatı tahmini	Boruta, RRF, RGF, DNN	NSE hisse senedi vadeli işlem verileri	National Stock Exchange, India	RGF ve DNN modelleri yüksek tahmin doğruluğu elde etmektedir; XAI oynaklık faktörlerinin etkisini vurgulamaktadır
Pramanik ve ark. (2024)	Ekonomilerde AI hazır olma düzeyinin değerlendirilmesi	XGBoost, Explainable Boosting Machine (EBM)	AI readiness verileri	World Bank, Stanford University	XGBoost doğru AI hazır olma skorları sunmaktadır; XAI araçları hazır olmayı etkileyen temel faktörleri açıklamaktadır
Kumar ve ark. (2024)	Hisse senedi fiyat tahmini	SHAP, LIME	Geçmiş hisse senedi fiyat verileri	Yahoo Finance, Bloomberg	Derin öğrenme modelleri yüksek doğruluk elde etmektedir; XAI zaman serisi tahminleri için yerel açıklamalar sunmaktadır
Heger ve ark. (2024)	Kredi spread değişimlerinin analizi	SHAP, Partial Dependence Plot (PDP)	ABD ve Euro bölgesi kredi spread verileri	Eurostat, Federal Reserve	PDP ve SHAP, karmaşık doğrusal olmayan ilişkilerin anlaşılmasına yardımcı olarak model doğruluğuna katkı sağlamaktadır
Sachan ve ark. (2020)	Kredi yazımı / underwriting	Belief Rule-Based (BRB) System	Mortgage underwriting verileri	Kredi veren kurum tarafından sağlanan verisetleri	BRB sistemi doğruluk ile açıklanabilirliği dengelemektedir; XAI yöntemleri underwriting kararlarında kural aktivasyonunu ortaya koymaktadır
Nallakaruppan ve ark. (2024)	Kredi değerlendirmesi	Random Forest, LIME, SHAP, PDP	Hint kredi değerlendirme veriseti	Kaggle	Random Forest modeli %99.8 doğruluk elde etmektedir; LIME ve SHAP kredi kararları için yorumlanabilirlik sağlamaktadır
Carta ve ark. (2022)	Hisse senedi işlemlerinde istatistiksel arbitraj	LIME, SHAP, EBM	S&P 500 ve Eurostoxx 50 hisse verileri	Yahoo Finance, Bloomberg	Arbitrajda getiri tahminleri için yüksek doğruluk; SHAP işlem stratejisi için özellik önemini açıklamaktadır
Wu ve Han (2021)	Kredi incelemesi	Fuzzy Bayesian Network	Çin kredi verileri	Central University of Finance and Economics	Bayes ağ yaklaşımı kredi değerliliği değerlendirmesinde yorumlanabilirliği iyileştirmektedir

Ma ve ark. (2023)	Gayrimenkul şirketleri için kredi temerrüdü tahmini	Explainable Boosting Machine (EBM), AdaBoost, SHAP	Çin gayrimenkul finansal verileri	Finansal tablolar, borsa forumları	Yüksek yorumlanabilirliğe sahip EBM modeli; SHAP kredi riskini etkileyen temel özelliklerin belirlenmesine yardımcı olmaktadır
Ghosh ve Jana (2024)	Temiz enerji hisse senedi fiyatı tahmini	SHAP, PDP, Facebook Prophet, NeuralProphet	Temiz enerji hisse endeksleri	NASDAQ, CBOE	Sektör tahmininde yüksek doğruluk; XAI makroekonomik etkilerin sektör üzerindeki etkisini belirlemeye yardımcı olmaktadır
Hasan ve ark. (2024)	Blockchain işlemlerinde anomali tespiti	SHAP, XGBCLUS	Bitcoin işlem verileri	IEEE Data Portal	Anomalilerin tespitinde yüksek doğruluk; SHAP temel işlem özelliklerinin belirlenmesine yardımcı olmaktadır
Nwafor ve Nwafor (2023)	Takipteki kredilerin tahmini	SHAP, LightGBM, XGBoost	P2P tüketici kredi veriseti	Kredi veren kurumlar	SHAP özellik sıralaması sağlayarak takipteki kredilerin erken tespitine yardımcı olmaktadır
Gandhudi ve ark. (2024)	Tweet'lerin hisse senedi fiyatları üzerindeki etkisi	Quantum Neural Network (QNN), Genetic Algorithm, Transformer, SHAP	Twitter duygu ve hisse verileri	Twitter, geçmiş hisse senedi fiyatları	R^2 > %99 elde edilmiştir; tweet duyarlılığının hisse tahmini üzerindeki etkisini etkili biçimde açıklamaktadır
Surabhi Adhikari ve ark. (2023)	Finansal duygu analizi	Hibrit Word Embeddings (Word2Vec, BERT, POS)	Financial Phrasebank veriseti	Financial Phrasebank	Temel yöntemleri geride bırakan yüksek doğruluk; karmaşık dil özelliklerini ve duygu kutupluluğunu açıklamaktadır
Bastos ve Matos (2022)	Kredi kaybı tahmini	Gradient Boosting, Shapley Values, Accumulated Local Effects	Kredi riski verileri (Moody's URD)	Moody's Ultimate Recovery Database	Boosting modeli tahmin doğruluğunu iyileştirmektedir; XAI kredi kaybı analizinde risk etkenlerini açıklamaktadır
Park ve Yang (2022)	Ekonomik kriz tahmini	LSTM, SHAP	G20 ekonomik göstergeleri	FRED, CEIC, BIS	Geleneksel modellerden daha iyi performans göstermiştir; SHAP ülke bazında temel ekonomik faktörleri ortaya koymaktadır
Mediano vskyi ve ark. (2023)	KOBİ'ler için kredi erişilebilirliği	Double ML, SHAP, PDP	Litvanya KOBİ kredi verileri	Tescilli veriseti	Kredi erişilebilirliği için yorumlanabilirliği artırmıştır; SHAP kredi onayları üzerindeki treatment effect'leri açıklamaktadır
Çelik ve ark. (2023)	Hisse senedi piyasası yön tahmini	LIME, Stacking Ensemble	KOSPI hisse endeksi verileri	KOSPI	Seçili tahminlerde yüksek doğruluk (%99.13); LIME tahmin güvenilirliğini değerlendirmektedir
Jovanovi c ve ark. (2024)	Finansal işlemlerde anomali tespiti	Federe öğrenme, Blockchain, SHAP	Blockchain işlem verileri	Dağıtık finansal sistemler	Güçlü anomali tespit doğruluğu elde edilmiştir; blockchain veri işleme sürecinde şeffaflık sağlamaktadır
Yang ve ark. (2022)	Psikometrik analize dayalı kredi riski	Layer-wise Relevance Propagation	P2P Lending Psychometric data	Sosyal medya ve borç verme	Psikometrik verilerin entegrasyonu yoluyla risk tahmin doğruluğu artırılmıştır; LRP her özellik için yorumlanabilirlik

	değerlendirmesi	(LRP)		platformları	sunmaktadır
Teplova ve ark. (2023)	ESG faktörleri altında hisse likiditesi analizi	Sinir Ağları, Shapley Values	Rus borsası ESG verileri	Moscow Exchange	ESG faktör analizinde model doğruluğu iyileştirilmiştir; Shapley değerleri likidite üzerindeki temel ESG etkilerini ortaya koymaktadır
Carta ve ark. (2022)	Hisse senedi işlemlerinde istatistiksel arbitraj	Permütasyon önemlilik, Ensemble Learning	S&P500 hisse verileri	S&P500 endeksi	Portföy getirilerini iyileştirmiştir; Permütasyon önemlilik işlem stratejisi için temel özelliklerin seçilmesine yardımcı olmaktadır
Ghosh ve ark. (2023)	COVID-19 döneminde hisse senedi piyasası tahmini	UMAP-LSTM, ISOMAP-GBR	Hint hisse verileri (COVID-19 etkisi)	RavenPack	Piyasa hareketi tahminlerinde yüksek doğruluk; XAI medya duyarlılığının ayı/boğa eğilimleri üzerindeki etkisini göstermektedir
Temraz ve Keane (2022)	Dengesiz verisetlerinde dolandırıcılık tespiti	Karşı-olgusal Augmentation	Banka işlem verileri	Oluşturulmuş Karşı-olgusal veriler	Karşı-olgusal yöntem azınlık sınıf doğruluğunu artırmakta; false negative'leri azaltmaktadır
Philipp Hacker ve ark. (2023)	AI düzenlemesi ve sorumluluk değerlendirmesi	LIME, PDP	Avrupa AI liability verisetleri	Çeşitli AB hukuk belgeleri	Hukuki bağlamlarda sorumluluk değerlendirmesi için AI model kararlarının açık yorumlanabilirliğini sunmaktadır
Graham ve Bonner (2024)	Erken aşama girişimcilik tahmini	Random Forest, LIME, PDP	Global Entrepreneurship Monitor, IEF, World Bank	GEM, Index of Economic Freedom	Random Forest modeli yüksek tahmin doğruluğu elde etmektedir; LIME politika için değişken önemini açıklamaktadır
Tang ve ark. (2024)	Kurumsal tahvil temerrüdü tahmini	BERTopic, SHAP, XGBoost	Çin kurumsal tahvil verileri	Wind Database	Tematik içgörülerle temerrüt tahmini geliştirilmiştir; SHAP tahminlerde konu önemini ortaya koymaktadır

Tablo 2.8'de, özellikle kredi kartı dolandırıcılığı tespiti, kredi değerlendirmesi ve müşteri kaybı gibi alanlarda bazı verisetlerinin yoğun biçimde kullanıldığı görülmektedir. Kullanılan verisetlerinin çoğunluğu açık kaynaklı verisetlerinden oluşmaktadır. Verisetleri arasında, Kaggle kaynaklı verisetlerinin çoğunluğu oluşturduğu görülmektedir. Verisetleri üzerinde uygulanan XAI yöntemlerinin genel doğruluk oranı incelendiğinde, başarının yüksek olduğu görülmektedir. Bu durum, genel olarak tercih edilen XAI yöntemlerinin finans alanına uygun biçimde uygulandığını göstermektedir.

Tablo 2.8'de sunulduğu üzere, finansal uygulamalardaki XAI yöntemlerinin kapsamlı analizi birkaç önemli örüntü ve içgörüyü ortaya koymaktadır. Yerleşik finansal veri tabanları ve tescilli kurumsal veriler dâhil olmak üzere çeşitli kaynaklardan gelen gerçek dünya verisetlerinin baskın kullanımı, bu yöntemlerin pratik uygulanabilirliğini göstermektedir. Uygulamalar; kredi riski değerlendirmesi, dolandırıcılık tespiti, portföy yönetimi ve piyasa tahmini dâhil olmak üzere çeşitli kritik finansal alanları kapsamakta

ve %90 ile %99 doğruluk aralığında tutarlı biçimde yüksek performans ölçütleri göstermektedir. SHAP ve LIME, çoğunlukla Random Forest ve XGBoost gibi temel modellerle entegre edilerek en sık uygulanan XAI yöntemleri olarak öne çıkmakta ve finansal bağlamlarda sağlamlıklarını ve güvenilirliklerini göstermektedir. Çalışmalar, model yorumlanabilirliği ile performans arasında başarılı bir optimizasyon olduğunu göstermektedir; bu durum, özellikle düzenlemeye tabi finansal ortamlarda büyük önem taşımaktadır. Bu sistematik literatür taraması, XAI yöntemlerinin yalnızca gerçek dünya finansal verileri üzerinde kapsamlı biçimde test edilip doğrulanmadığını, aynı zamanda doğruluk ve açıklanabilirlik arasındaki kritik dengeyi korurken çeşitli finansal zorluklar için uyarlanabilir çözümler sunduğunu da göstermektedir. Farklı finansal sektörlerdeki başarılı uygulamaların geniş yelpazesi, XAI yaklaşımlarının karmaşık finansal karar verme süreçlerini ele almada ve aynı zamanda şeffaflık ile yorumlanabilirlik açısından düzenleyici gereksinimleri karşılamada olgunluğunu ve güvenilirliğini vurgulamaktadır.

Çeşitli XAI tekniklerinin finans alanına uygunluğunu daha iyi anlayabilmek amacıyla, bu bölüm başlıca finansal uygulama alanları arasında karşılaştırmalı bir analiz sunmaktadır. Önceki bölümlerde çalışmalar modeller ve XAI tekniklerine göre sınıflandırılmışken, aşağıdaki tartışma belirli finansal bağlamlarda neden belirli XAI yöntemlerinin tercih edildiğini ve bunların yorumlanabilirlik özelliklerinin alana özgü ihtiyaçlarla nasıl örtüştüğünü vurgulamaktadır.

2.2.4 Çalışmaların ayrıntılı analizi

Bu bölüm, incelenen çalışmaların kapsamlı bir analizini sunmakta ve her bir araştırmanın özgül katkılarını, metodolojilerini ve bulgularını vurgulamak amacıyla uygulama alanlarına göre düzenlenmektedir.

2.2.4.1 Kredi riski değerlendirmesi ve skorlama

SHAP Tabanlı Kredi Değerlendirme Çalışmaları: Zacharias ve ark.(2022), UCI Machine Learning Repository'de yer alan German Credit Dataset kullanılarak kredi riski değerlendirmesinde özellik seçimi için SHAP kullanmıştır. Çalışmaları, SHAP tabanlı özellik seçiminin geleneksel Pearson korelasyonu ve XGBoost tabanlı yöntemlerle karşılaştırıldığında benzer veya daha iyi performans sunduğunu, aynı zamanda model açıklanabilirliğini ve paydaşlar açısından gerekçelendirilebilirliği anlamlı ölçüde

iyileştirdiğini göstermiştir. Araştırma, yorumlanabilirliği korurken yüksek doğruluk oranlarına ulaşmış ve bu yönüyle özellikle kredi kararlarında düzenleyici uyum açısından değerli hâle gelmiştir.

Dennis ve ark.(2024), UCI Repository'den Taiwan Credit Card Default Dataset kullanarak kredi temerrüdü tahmini için açıklanabilir bir derin topluluk öğrenme yaklaşımı geliştirmiştir. ANN-LSTM topluluk modeli F1 Skoru olarak 0.89, Kesinlik olarak 0.84 ve Duyarlılık olarak 0.95 elde etmiş; SHAP ve LIME ise derin topluluk modelinin karar verme sürecini anlamada kritik yorumlanabilirlik sağlamıştır. Bu çalışma, yüksek performanslı derin öğrenme modelleri ile finansal uygulamalardaki açıklanabilirlik gereksinimleri arasındaki boşluğu kapatması bakımından önemlidir.

Kredi Değerlendirmesinde İleri XAI Teknikleri: Dessain ve ark. (2023) küresel bir ticari kredi sigortacısından alınan Avrupa borçlu kredi verileri (2019-2022, 50,000+ risk maruziyeti) kullanılarak kredi skorlama modellerinde açıklanabilirliğin maliyet-fayda analizine yönelik kapsamlı bir çalışma yürütmüştür. Çalışmaları, Explainable Boosting Machine (EBM), GAMINet ve Isotonic GAMs modellerini XGBoost ve MLP gibi geleneksel kara kutu modellerle karşılaştırmıştır. Sonuçlar, EBM ve GAMI'nin kara kutu modellere yakın AUROC performansı sergilediğini, en yüksek finansal getiriyi ise MLP'nin sağladığını (2.06%'ye kadar) göstermiştir. Önemli olarak, en iyi XAI modelleri (EBM/GAMI), MLP'den yalnızca 14--21 baz puan daha düşük performans göstermiş; bu da açıklanabilirliğin minimum performans ödünleşimi ile sağlanabileceğini ortaya koymuştur.

2.2.4.2 Dolandırıcılık tespiti uygulamaları

Gerçek Zamanlı Dolandırıcılık Tespiti: Vihurskyi, (2024), Kaggle'dan alınan Avrupa kredi kartı verisetlerini kullanarak kredi kartı dolandırıcılığı tespiti için SHAP ve karar ağaçlarını uygulamıştır. Çalışma, dolandırıcılık tespiti görevlerinde %100 doğruluk elde etmiş, LIME ise bireysel işlem kararları için açıklamalar sağlamıştır. Bu araştırma, hem doğruluğun hem de açıklanabilirliğin müşteri güveni ve düzenleyici uyum açısından kritik olduğu gerçek zamanlı dolandırıcılık tespit sistemleri için özellikle değerlidir.

Bhowmik ve ark. (2022), ULB Credit Card Fraud Detection Dataset kullanarak finansal dolandırıcılık tespiti için DBNex'i (Explainable AI ile Deep Belief Network) geliştirmiştir. Yaklaşımları %94 doğruluk elde etmiş, SHAP ise dolandırıcılık

tahminlerine ilişkin özellik düzeyindeki katkıları açıklayarak yorumlanabilirliği artırmıştır. Çalışma, derin öğrenme mimarilerinin önemli bir performans kaybı olmadan nasıl yorumlanabilir hâle getirilebileceğini göstermektedir.

Çok Modlu Dolandırıcılık Tespiti: Awosika ve ark.(2024), Anglia Ruskin University'den alınan banka işlem verilerini kullanarak finansal dolandırıcılık tespiti için federe öğrenme ile SHAP'ı birleştiren bir yaklaşımı incelemiştir. Yaklaşımları, federe öğrenme yoluyla gizliliği korurken yüksek doğruluk elde etmiş, SHAP ise dolandırıcılık sınıflandırma kararlarında şeffaflık sağlamıştır. Bu çalışma, dağıtık finansal sistemlerde hem gizliliği hem de açıklanabilirliği koruma yönündeki kritik zorluğu ele almaktadır.

2.2.4.3 Portföy yönetimi ve yatırım stratejisi

XAI ile Piyasa Tahmini: Beckmann ve ark.(2023), Deutsche Bundesbank ve FRED dâhil olmak üzere birden fazla kaynaktan elde edilen U.S. Treasury Bond Yield ve makroekonomik verileri kullanarak tahvil fazla getiri tahmini için Extremely Randomized Trees modeline SHAP uygulamıştır. Çalışmaları, ABD 10 yıllık tahvilleri için 20.5%'e kadar ve Alman piyasaları için yaklaşık 10% seviyesinde R^2 elde etmiş; SHAP ise getiri eğrisi eğimini ve konut piyasası göstergelerini temel belirleyiciler olarak tanımlamıştır.

Chou ve ark.(2024), WRDS ve Yahoo Finance'den alınan küresel borsa verilerini kullanarak G7 piyasalarında portföy optimizasyonu için kuantum esinli algoritmaları incelemiştir. Trend Ratio analizi ile birlikte kullanılan kuantum esinli teknikleri (ELSA-GNQTS), risk-getiri dengesi açısından yüksek doğruluk sağlamış; kuantum esinli yaklaşımlar yatırım karar verme sürecinde yorumlanabilirliği iyileştirmiştir.

ESG ve Sürdürülebilir Yatırım: Agosto ve ark.2023), 1382 Avrupa şirketine ait verileri (MSCI, Refinitiv, S&P ESG skorları, Bloomberg kredi notları) kullanarak kredi notları üzerindeki ESG faktör etkilerini ölçmek amacıyla Bayesyen öğrenme modelleri geliştirmiştir. Birleştirilmiş Bayes modeli, AUROC = 0.59 ve AUPRC = 0.24 elde etmiş; bu sonuçlar XGBOOST performansı (AUROC = 0.61, AUPRC = 0.25) ile karşılaştırılabilir bulunmuştur. Aynı zamanda model, minimum performans ödünleşimi ile yorumlanabilir ESG ağırlıkları sunmuştur.

2.2.4.4 Müşteri davranışı ve segmentasyon

Müşteri Kaybı Tahmini: Murindanyi ve ark. (2023), Çek bankasına ait Berka veriseti kullanarak perakende bankacılıkta müşteri etkileşimi tahmini için açıklanabilir topluluk öğrenmesi geliştirmiştir. Blending ensemble modeli %99 doğruluk ve AUC = 1.00 elde etmiş; LIME yorumlanabilirlik sağlarken Evidently AI hata analizine imkân vermiştir. Bu çalışma, XAI'ın bankacılıkta müşteri ilişkileri yönetimine nasıl etkili biçimde uygulanabileceğini göstermektedir.

Manzoor ve ark.(2024), müşteri kaybı tahmini için makine öğrenmesi yöntemlerine ilişkin kapsamlı bir derleme gerçekleştirmiş ve özellikle iş uygulayıcıları açısından açıklanabilir yaklaşımların önemini vurgulamıştır. Analizleri, çeşitli topluluk yöntemlerini (XGBoost, AdaBoost) ve derin öğrenme yaklaşımlarını kapsamakta; açıklanabilir yapay zekânın karmaşık müşteri kaybı tahmini senaryolarında doğruluğu anlamlı biçimde artırdığını öne çıkarmaktadır.

Kişiselleştirilmiş Finansal Hizmetler: Potluri ve ark.(2024), finansal işlem ve demografik verileri kullanarak müşteri segmentasyonu ve kişiselleştirilmiş pazarlama için LIME, SHAP ve K-means yöntemlerini uygulamıştır. Yaklaşımları segmentasyon doğruluğunu artırmış, XAI ise hedefli pazarlama stratejileri için müşteri içgörüleri sağlamıştır. Bu çalışma, açıklanabilir yapay zekânın müşteri ilişkileri yönetimindeki pratik uygulamalarını göstermektedir.

2.2.4.5 Piyasa tahmini ve ekonomik analiz

Hisse Senedi Piyasası Tahmini: Valensky ve Mohaghegh (2023), S&P 500 verisetini kullanarak SHAP, LIME ve Permütasyon önemlilik ile finansal tahminlemede şeffaflığı değerlendirmiştir. CNN ve LSTM modelleriyle yürütülen çalışmalarında yüksek doğruluk elde edilmiş, SHAP ise özellikle `Close' fiyatları gibi kritik özellikler için yorumlanabilirliği iyileştirerek yatırım karar verme süreci için değerli içgörüler sunmuştur.

Gandhudi ve ark., (2024), Twitter duyarlılığı ve hisse senedi verilerini kullanarak finansal işlemde model açıklanabilirliğini geliştirmiştir. Genetic Algorithm ve Transformer ile kullanılan Quantum Neural Network (QNN), R^2 > 99% elde etmiş ve tweet duyarlılığının hisse senedi tahmini üzerindeki etkisini etkili biçimde açıklamıştır.

Bu çalışma, sosyal medya duygu analizinin XAI ile bütünleştirilmesini göstermektedir.

Ekonomik Kriz Tahmini: Park ve Yang (2022), FRED, CEIC ve BIS veri tabanlarından elde edilen G20 ekonomik göstergelerini kullanarak ekonomik kriz tahmini için yorumlanabilir bir derin öğrenme LSTM modeli geliştirmiştir. Modelleri, kriz tahmininde geleneksel yaklaşımlardan daha iyi performans göstermiş; SHAP ise ülke bazında temel ekonomik faktörleri ortaya koyarak politika yapıcılar ve finansal kuruluşlar için değerli içgörüler sağlamıştır.

2.2.4.6 Düzenleyici uyum ve risk yönetimi

Kara Para Aklamayla Mücadele: Thommandru ve ark.(2023), makine öğrenmesi, kara para aklama ve veri gizliliğinin kesişimini kredi ve finansal işlem verileri üzerinden incelemiştir. Çalışmaları, finansal suç tespitinde yorumlanabilirliği artırırken aynı zamanda gizlilik kaygılarını ele almak amacıyla SHAP, LIME ve Karşı-Olgusal Açıklamalar yöntemlerini uygulamıştır. Böylece XAI'ın hassas finansal uygulamalarda düzenleyici uyumu nasıl destekleyebileceğini göstermiştir.

Gizliliği Koruyan XAI: Naretto ve ark. (2022), Adult, DIVA ve Bank verisetlerini kullanarak yorumlanabilir küresel açıklayıcıların gizlilik maruziyetini değerlendirmiştir. Çalışmaları, küresel açıklayıcılarda MIA'lar nedeniyle kara kutu modellere kıyasla daha yüksek gizlilik riskleri bulunduğunu ortaya koymuş; bu da finansal uygulamalarda gizliliği koruyan XAI tekniklerine olan ihtiyacı vurgulamıştır.

2.2.4.7 Temel bulgular ve katkılar

Ayrıntılı analiz, incelenen çalışmalar genelinde birkaç önemli örüntüyü ortaya koymaktadır:

- Metodolojik Tercihler: SHAP ve LIME, tüm finansal alanlarda en sık uygulanan XAI yöntemleri olarak öne çıkmaktadır; SHAP özellikle karmaşık finansal modellerde özellik atfı açısından etkili olurken, LIME gerçek zamanlı uygulamalar için verimli yerel açıklamalar sunmaktadır.
- Performans Ödünleşimleri: Çalışmaların çoğu, XAI yöntemlerinin kara kutu modellerle karşılaştırılabilir performans düzeylerine ulaşabildiğini ve farklı finansal uygulamalarda doğruluk oranlarının tipik olarak %88 ile %99.5 arasında değiştiğini göstermektedir.

- Veriseti Özellikleri: Araştırma alanına gerçek dünya finansal verisetleri hâkimdir; çalışmalarda sıklıkla UCI Repository, Kaggle ve tescilli finansal kurumlardan alınan yerleşik verisetleri kullanılmakta, bu da XAI yöntemlerinin pratik uygulanabilirliğini vurgulamaktadır.
- Uygulamaya Özgü Uyarlamalar: Farklı finansal uygulamalar, uyarlanmış XAI yaklaşımları gerektirmektedir; dolandırıcılık tespiti gerçek zamanlı yerel açıklamaları (LIME), kredi skorlama kapsamlı özellik atfını (SHAP) ve portföy yönetimi ise küresel yorumlanabilirlik yöntemlerini tercih etmektedir.
- Düzenleyici Uyum: Çalışmalar, XAI'ın düzenleyici uyum açısından önemini tutarlı biçimde vurgulamakta; birçok araştırma GDPR, AI Act gereklilikleri ve finans sektörüne özgü düzenlemeleri açıkça ele almaktadır.

Bu kapsamlı analiz, XAI yöntemlerinin yalnızca finansal uygulamalar için teknik olarak uygulanabilir olmadığını, aynı zamanda rekabetçi performans düzeylerini korurken paydaşlar, düzenleyiciler ve son kullanıcılar için de pratik değer sunduğunu göstermektedir.

2.2.5 Boşluklar ve açıklanmamış alanlar

Bu bölümde, literatür taraması için seçilen ve XAI yöntemlerini kullanan çalışmalara ilişkin çeşitli açılardan boşluk analizleri yapılmaktadır.

2.2.5.1 Kullanım alanlarında algoritma çeşitliliği

LIME, SHAP ve diğer bazı modelden bağımsız algoritmaların kullanımının oldukça yaygın olduğu gözlemlenmektedir. Ancak, daha az yaygın kullanılan yöntemlerin (örneğin MIA, GAM, EBM) belirli senaryolarda değerlendirilmemiş olması dikkat çekicidir. Bu eksiklik, alternatif algoritmaların farklı veri türleri veya problem alanlarındaki performanslarının yeterince araştırılmamış olabileceğini düşündürmektedir.

2.2.5.2 Veriseti çeşitliliği

Çalışmalarda veriseti bilgileri çoğu zaman eksik ya da oldukça sınırlıdır. Bu durum, literatürdeki çalışmaların çeşitli verisetleri üzerinde kapsamlı biçimde değerlendirilmediğini ve algoritmaların başarısının daha büyük verisetlerinde yeterince

test edilmediğini göstermektedir.

2.2.5.3 Başarı ölçütleri ve performans analizleri

Mevcut performans ölçütleri arasında LORE ve TREPAN gibi belirli algoritmalar için yüksek başarı oranları rapor edilse de, ortalama başarı oranları veya başarının hangi metriklerle (örneğin ROC, kesinlik) elde edildiği ayrıntılı ve tutarlı biçimde analiz edilememektedir.

Her bir XAI yönteminin hangi başarı ölçütleriyle değerlendirildiğine ilişkin daha ayrıntılı bilgi eksikliği olduğu görülmektedir. Bu durum, performans metriklerinin karşılaştırmalı analizini zorlaştırmaktadır.

2.2.5.4 Etik, gizlilik ve tarafsızlıkla ilgili eksiklikler

Adalet ve tarafsızlık gibi etik konular, özellikle kredi değerlendirmelerinde ve işe alım süreçlerinde büyük önem taşımaktadır. Ancak, XAI yöntemleri bu alanlarda tam şeffaflık ve adaleti sağlamada yetersiz kalmaktadır. Algoritmaların önyargılarını tespit etmeye ve düzeltmeye özel olarak tasarlanmış XAI araçlarında belirgin bir eksiklik bulunmaktadır.

Bazı makaleler etik ve gizlilik gibi konuları ele alsa da, etik ve veri gizliliği boyutunun derinlemesine incelenmesinde bir eksiklik vardır. Özellikle finans sektöründe büyük önem taşıyan tarafsızlık ya da önyargının en aza indirilmesi gibi konulara daha fazla vurgu yapılması gerekmektedir.

Bu eksiklikler, farklı XAI yöntemlerinin büyük verisetleri üzerindeki performansını daha iyi anlayabilmek için daha kapsamlı araştırmalara ihtiyaç olduğunu göstermektedir.

2.2.5.5 Açıklanabilirlik ve güvenilirlik standartlarının eksikliği

Farklı XAI yöntemleri, modelin ne ölçüde açıklanabilir olduğu konusunda farklı sonuçlar verebilmektedir. Ancak hangi yöntemin hangi senaryolarda daha etkili olduğuna ilişkin henüz bir standart belirlenmemiştir. Bu durum, özellikle finans sektörü gibi yüksek riskli alanlarda daha güvenilir standartların ve değerlendirme metriklerinin oluşturulması gerektiğini göstermektedir.

2.2.5.6 Gerçek zamanlı açıklanabilirlik eksikliği

XAI yöntemlerinin çoğu statik veriler üzerinde çalışmakta ve gerçek zamanlı tahminleri

hesaba katmamaktadır. Gerçek zamanlı analiz gerektiren uygulamalarda (örneğin kredi başvuruları veya piyasa işlemleri) bu eksiklik büyük bir sorun oluşturmaktadır. Anlık kararlar gerektiren finansal uygulamalar için hızlı ve etkili XAI çözümlerinin geliştirilmesi gerekmektedir.

2.2.5.7 Özelleştirilmiş açıklama modellerinin azlığı

XAI yöntemleri çoğunlukla genel amaçlıdır ve belirli finansal alanlara özgü uyarlamalardan yoksundur. Özellikle kredi riski değerlendirmesi veya dolandırıcılık tespiti gibi uzmanlaşmış finansal alanlarda daha hedefli XAI modellerine ihtiyaç vardır. Bu eksiklik, kullanıcıların finansal sonuçları tam olarak anlamasını zorlaştırmakta ve karar verme sürecinde güven sorunlarına yol açmaktadır.

2.2.5.8 Veri gizliliği ve güvenliği ile ilgili eksiklikler

Bazı XAI yöntemleri, modelin iç işleyişini ortaya çıkardıkları için veri gizliliği ve güvenliği riskleri taşımaktadır. Bu durum, finans sektöründeki hassas verilerin güvenliği açısından ciddi bir sorun oluşturmaktadır. Veri gizliliği ve güvenliği konusunda daha fazla çalışma yapılmalı ve XAI yöntemleri bu açıdan geliştirilmelidir.

2.2.5.9 Kapsamlı verisetlerinin eksikliği

XAI çalışmalarının çoğu sınırlı verisetleri üzerinde yürütülmektedir. Bu durum, genelleme kapasitesini azaltmakta ve gerçek dünya senaryolarına uyarlanabilirliği zorlaştırmaktadır. Finans ve ekonomi gibi büyük veri gerektiren alanlarda, daha büyük verisetlerine erişim sağlanarak yöntemlerin etkinliği daha iyi test edilebilir.

2.3 Tartışma

Bu bölüm, çalışmanın başında ortaya konulan her bir araştırma sorusunun (RQ), tematik ve yöntemsel analiz yoluyla nasıl yanıtlandığını özetlemektedir. Araştırma soruları ile çalışmanın ilgili bölümleri arasındaki bağlantılar aşağıda ayrıntılı olarak sunulmaktadır.

2.3.1 Araştırma sorularının kapsanmasına genel bakış

Tablo 2.9, her bir araştırma sorusunu yanıtlamak için kullanılan tabloları göstermektedir. Bu eşleştirme, literatür taraması bulgularının çalışmanın amaçlarıyla nasıl örtüştüğüne ilişkin şeffaflık sağlamaktadır.

Tablo 2.9. Araştırma soruları ve destekleyici tablolar

Araştırma Sorusu	Destekleyici Tablolar
RQ1	Tablo 2.2
RQ2	Tablo 2.4, Tablo 2.5, Tablo 2.6, Tablo 2.7
RQ3	Tablo 2.4, Tablo 2.5
RQ4	Tablo 2.8

2.3.1.1 RQ1: Hangi XAI algoritmaları hangi finansal hizmetler için kullanılmaktadır?

Tablo 2.2'de gösterildiği üzere, SHAP ve LIME; kredi skorlama, dolandırıcılık tespiti ve portföy optimizasyonu dâhil olmak üzere çeşitli finansal hizmetlerde en sık kullanılan XAI algoritmalarıdır. Tablo 2.2'de gösterildiği gibi, bu yöntemler kredi skorlama, dolandırıcılık tespiti, müşteri segmentasyonu ve yatırım stratejisi gibi çeşitli alanlarda kullanılmaktadır. Anchor, TREPAN, Karşı-Olgusal Açıklamalar ve GAM gibi diğer algoritmalar da uygulanmıştır, ancak bunların kullanım sıklığı daha düşüktür. Bu bulgular, XAI yöntemlerinin çok sayıda finansal uygulama genelindeki çok yönlülüğünü ve uyarlanabilirliğini göstermektedir.

2.3.1.2 RQ2: XAI algoritmaları finansal hizmetler için uygun mudur?

Tablo 2.4, Tablo 2.5 ve Tablo 2.7'de sunulan karşılaştırmalı değerlendirme, XAI yöntemlerinin finansal hizmetler için oldukça uygun olduğunu ortaya koymaktadır. Özellikle SHAP ve LIME, yüksek doğruluğu korurken güçlü yorumlanabilirlik sunmakta; bu da şeffaflık ve düzenleyici uyum açısından büyük önem taşımaktadır. Bununla birlikte, SHAP'ın yüksek hesaplama maliyeti ya da LIME'ın doğrusal olmayan senaryolardaki kararsızlığı gibi bazı sınırlılıklar, yöntemin uygulama alanına uygun biçimde seçilmesi gerektiğini göstermektedir. Genel olarak, XAI teknikleri finansal sistemlerde güvene, denetlenebilirliğe ve adalete olumlu katkı sağlamaktadır.

2.3.1.3 RQ3: XAI kullanımının sunulan finansal hizmetlerin performansı üzerindeki etkisi nasıldır?

Tablo 2.4 ve Tablo 2.5, XAI yöntemlerini kullanan modellerin bazı uygulamalarda %99.5'e kadar doğruluk oranına ulaşabildiğini gösteren performans karşılaştırmalarını sunmaktadır. XAI entegrasyonunun yorumlanabilirliği artırırken model performansını da

nasıl iyileştirdiği vurgulanmaktadır

Açıklanabilirliğin yalnızca yorumlanabilirliği değil, aynı zamanda finansal AI modellerinin kestirimsel performansını da iyileştirdiği görülmüştür. Örneğin, SHAP tabanlı modeller %99.5'e kadar doğruluk seviyelerine ulaşırken, LIME destekli sistemler bazı kullanım senaryolarında neredeyse kusursuz hassasiyetle gerçek zamanlı karar verme imkânı sağlamıştır. Buna ek olarak, EBM ve GAM gibi yöntemler performansta çok düşük ödünleşimlerle şeffaflığı korumaktadır. Dolayısıyla, XAI entegrasyonu daha iyi model doğrulaması, artan kullanıcı güveni ve daha sağlam finansal karar sistemleri sağlamaktadır.

2.3.1.4 RQ4: Finansal hizmetlerde XAI algoritmalarının uygulamalarında ne tür verisetleri tercih edilmektedir?

Çalışmaların büyük çoğunluğu gerçek dünya finansal verisetlerine dayanmaktadır; bu da XAI odaklı finansal uygulamalarda gerçekçi, temsil gücü yüksek ve düzenleyici açıdan anlamlı verilere duyulan ihtiyacı vurgulamaktadır.

Tablo 2.8'de incelenen verilere göre, araştırma alanına gerçek dünya verisetleri hâkimdir. Çoğu çalışma, UCI Repository, Kaggle ve tescilli finansal kurumlardan elde edilen verisetlerini kullanmaktadır. Bu verisetleri kredi skorlama, müşteri kaybı, finansal dolandırıcılık tespiti ve hisse senedi piyasası analizi gibi çeşitli alanları kapsamaktadır. Gerçek verilerin tercih edilmesi, XAI tabanlı finansal sistemlerde pratik uygulanabilirliğe ve güvenilirliğe güçlü bir eğilim olduğunu göstermektedir.

2.4 Sınırlılıklar

Bu çalışma, finans sektöründeki XAI uygulamalarına ilişkin kapsamlı bir sistematik literatür taraması sunmasına rağmen, bazı sınırlılıkların dikkate alınması gerekmektedir:

- Zaman Kapsamı: İnceleme 2021 ile 2025 yılları arasında yayımlanan makaleleri kapsamaktadır. Bu durum güncellik sağlasa da, finans alanındaki XAI için önemli olan daha erken dönem temel çalışmaları dışarıda bırakmış olabilir.
- Veritabanı Seçimi: Literatür taraması dört büyük veritabanı ile sınırlı tutulmuştur (ACM Digital Library, SpringerLink, IEEE Xplore ve ScienceDirect). Diğer veritabanlarında (örneğin Scopus, Web of Science) indekslenen ilgili çalışmalar gözden kaçmış olabilir.

- Dil ve Erişilebilirlik: Yalnızca İngilizce dilinde ve açık erişimli makaleler dikkate alınmıştır. Bu durum seçim yanlılığına neden olabilir ve diğer dillerde yayımlanan veya ücretli erişim arkasında bulunan değerli araştırmaları dışarıda bırakabilir.
- Tematik Sınıflandırmada Öznellik: XAI yöntemlerini ve finansal uygulama alanlarını sistematik biçimde sınıflandırmak için özen gösterilmiş olsa da, tematik eşleştirme yine de öznel yorumlar içerebilir.
- Ampirik Yeniden Doğrulama Eksikliği: XAI yöntemlerine ilişkin performans metrikleri doğrudan incelenen çalışmalardan alınmıştır. Bu sonuçları doğrulamak veya standartlaştırmak için bağımsız deneyler yürütülmemiştir.
- XAI Yöntemlerinin Hızlı Evrimi: XAI tekniklerindeki hızlı gelişim göz önüne alındığında, yeni önerilmiş veya henüz yayımlanmamış bazı yöntemler incelemeye henüz yansımamış olabilir.

Bu bulgular yorumlanırken söz konusu sınırlılıklar dikkate alınmalıdır. Bununla birlikte, sistematik yaklaşım, önceden tanımlanmış ölçütler ve çok boyutlu analiz, sonuçların güvenilirliğini ve genellenebilirliğini artırmaktadır.

2.5 Sonuç

Bu literatür taramasında, finansta yapay zekâ uygulamalarının açıklanabilirlik gereksinimlerine ilişkin mevcut araştırmalar ayrıntılı biçimde analiz edilmiştir. İnceleme, finansal hizmetlerde AI tabanlı modellerin kullanımında şeffaflık, güven ve etik kaygıların giderek daha önemli hâle geldiğini göstermektedir. XAI yöntemlerinin, özellikle kredi skorlama, dolandırıcılık tespiti ve müşteri davranışı analizi gibi finansal süreçlerde önemli katkılar sağladığı görülmüştür.

Bu çalışma, mevcut yöntemlerin ve uygulamaların kapsamlı analizi yoluyla finansta XAI uygulamalarına ilişkin mevcut literatüre önemli katkılar sunmaktadır. Sistematik incelememiz, özellikle kredi değerlendirmesi, dolandırıcılık tespiti ve düzenleyici uyum gibi kritik alanlara odaklanarak, finansal hizmetlerde XAI yöntemlerinin değerlendirilmesi ve uygulanması için sağlam bir çerçeve ortaya koymuştur. Analiz, SHAP ve LIME yöntemlerinin finansal uygulamalarda baskın konumunu koruduğunu, ancak her bir XAI yönteminin belirli kullanım senaryosuna bağlı olarak kendine özgü

avantajlar ve sınırlılıklar sunduğunu göstermektedir. Bulgular, SHAP'ın karmaşık finansal modeller için iyi bir özellik atfı sağladığını ancak büyük verisetlerinde ölçeklenebilirlik sorunları yaşadığını göstermektedir. Buna karşılık, LIME gerçek zamanlı açıklamalar üretmede verimli olmakla birlikte, finansal verilerde yaygın olan yüksek düzeyde doğrusal olmayan ilişkilerde zorlanabilmektedir. Bu bulgular, uygulayıcılara kendi özel gereksinim ve kısıtlarına göre uygun XAI yöntemini seçmede değerli bir rehber sunmaktadır.

Bunun yanında, bulgular SHAP ve LIME'ın özellikle kredi skorlama, dolandırıcılık tespiti ve portföy yönetiminde en yaygın benimsenen XAI algoritmaları olduğunu göstermektedir (RQ1). Bu yöntemlerin uygunluğu, yüksek yorumlanabilirlikleri ve düzenleyici uyumla örtüşmeleriyle desteklenmektedir; ancak hesaplama karmaşıklığı ve modele özgü sınırlılıklar dikkate alınmalıdır (RQ2). XAI tekniklerinin entegrasyonu, yalnızca şeffaflığı artırmakla kalmamakta, aynı zamanda çok sayıda finansal uygulamada model performansını da iyileştirmektedir (RQ3). Veri kullanımı açısından ise, gerçek dünya verisetleri literatüre hâkimdir ve bu durum gerçekçi ve sektörle uyumlu veri kaynaklarına yönelik tercihi yansıtmaktadır (RQ4).

Bu çalışmanın temel katkılarından biri, farklı finansal uygulamalardaki uygulama zorluklarının ve bunlara yönelik çözümlerin ayrıntılı biçimde değerlendirilmesidir. Çalışma, özellikle gerçek zamanlı finansal uygulamalarda, açıklama kalitesi ile hesaplama verimliliği arasındaki dengenin kritik önemini vurgulamaktadır. Analizimiz ayrıca, finansal XAI uygulamalarına özgü standart değerlendirme metrikleri ve kıyaslama çerçevelerine duyulan artan ihtiyacı da ele almaktadır.

İleriye dönük olarak, analizden birkaç araştırma hedefi ortaya çıkmaktadır. İlk olarak, özellikle yüksek frekanslı işlem ve dolandırıcılık tespiti senaryolarında, gerçek zamanlı finansal uygulamalar için özel olarak optimize edilmiş XAI yöntemlerine acil ihtiyaç vardır. İkinci olarak, alana özgü bilginin XAI çerçevelerine entegre edilmesi, finansal paydaşlar açısından açıklamaların kalitesini ve ilgililiğini artırmak için önemli bir fırsat sunmaktadır. Üçüncü olarak, finansal kurumların daha katı veri koruma düzenlemeleriyle karşı karşıya kaldığı günümüzde, gizliliği koruyan XAI tekniklerinin geliştirilmesi giderek daha önemli hâle gelmektedir.

Buna ek olarak, finans alanında AI'a ilişkin düzenleyici çerçeve gelişmeye devam

etmekte; bu da otomatik uyum doğrulama sistemleri ve sektöre özgü XAI kılavuzları üzerine araştırmaları gerekli kılmaktadır. Bu kapsamda, açıklama kalitesini düzenleyici gereklilikler bağlamında değerlendirecek çerçevelerin geliştirilmesi ve model açıklamalarına yönelik standartlaştırılmış dokümantasyon uygulamalarının oluşturulması da yer almaktadır.

Sonuç olarak, XAI'ın finansal hizmetlere uygulanmasında önemli ilerlemeler kaydedilmiş olsa da, hem teknik yeteneklerin hem de bu yöntemlerin pratik uygulanmasının geliştirilmesi açısından hâlâ önemli fırsatlar bulunmaktadır. Bu çalışmada sunulan bulgular ve öneriler, bu kritik alandaki gelecekteki araştırma ve geliştirme çalışmaları için bir temel sunmaktadır.

3 ÖNERİLEN ÖLÇEKLENEBİLİR, FİNANSAL ODAKLI VE YORUMLANABİLİR AÇIKLAMA (SFIX) MODELİ

Bu bölümde, tez çalışmasının önerilen modeli olan finansal uygulamalar için özel olarak tasarlanmış yeni bir XAI çerçevesi olan Ölçeklenebilir, Finans Odaklı ve Yorumlanabilir Açıklama (Scalable Financial-oriented Interpretable eXplanation (SFIX)) modeli tanıtılmaktadır (Cil ve Yildiz, 2025b). SHAP ve LIME gibi genel amaçlı XAI yöntemlerinden farklı olarak SFIX, zamansal dinamikleri, veri kalitesi değerlendirmesini, risk analizini ve anomali tespitini tek bir birleşik çerçevede bütünleştirmektedir. Önerilen SFIX modelinin temel katkıları şu şekilde özetlenebilir:

- Finansa Özgü XAI Modeli: Mevcut XAI tekniklerinin çoğu modelden bağımsız ve alan bağımsız iken, SFIX finansal karar verme süreçlerinin benzersiz gereksinimlerini karşılamak üzere açıkça tasarlanmıştır; böylece düzenleyici standartlara ve sektöre özgü ihtiyaçlara uyumu güvence altına almaktadır.
- Zamansal Örüntü Analizi: Geleneksel XAI modelleri finansal işlemlerin zaman içinde değişen doğasını yakalayamamaktadır. SFIX, farklı zaman pencereleri boyunca özellik önemini dinamik olarak değerlendiren zamana duyarlı bir açıklama mekanizması sunmakta; bu yönüyle özellikle dolandırıcılık tespiti, risk skorlama ve kredi değerliliği değerlendirmeleri için değerli hâle gelmektedir.
- Kararlılık ve Tutarlılık: Karşılaştırmalı analizimiz, SFIX'in kararlılık açısından SHAP ve LIME'dan anlamlı ölçüde daha iyi performans gösterdiğini ortaya koymaktadır. Farklı çalıştırmalarda yüksek değişkenlik sergileyen LIME'ın aksine, SFIX son derece tutarlı ve yeniden üretilebilir açıklamalar üretmekte, böylece yüksek riskli finansal kararlarda güvenilirlik sağlamaktadır.
- Risk Farkındalıklı Yorumlanabilirlik: SFIX, standart özellik önem skorlarının ötesine geçerek bir risk değerlendirme modülü içermekte ve finansal kuruluşların model kararlarıyla ilişkili potansiyel riskleri nicel olarak belirlemesine olanak tanımaktadır. Bu yetenek özellikle kredi onaylarında, portföy yönetiminde ve dolandırıcılık tespit sistemlerinde kritik önem taşımaktadır.
- Yüksek Kaliteli Açıklamalar: Veri kalitesini dikkate almayan geleneksel yöntemlerin aksine, SFIX açıklama üretmeden önce giriş özelliklerinin güvenilirliğini değerlendirmektedir. Veri kalitesi metriklerini bütünleştirerek

modelin yorumlanabilirliğinin gürültülü ya da eksik veriler nedeniyle bozulmamasını sağlamaktadır.

- Uygulanabilirlik ve Ölçeklenebilirlik: SFIX, gerçek dünya finansal ortamlarında hızlı model testi ve dağıtımını mümkün kılan sadeleştirilmiş bir sürüm (S-SFIX) içermektedir; bu da onu hem yüksek frekanslı işlem sistemleri hem de büyük ölçekli risk değerlendirme çerçeveleri için uygun hâle getirmektedir.
- Alana özgü, kararlı, risk farkındalıklı ve zamansal olarak uyarlanabilir bir XAI modeli sunarak SFIX, yapay zekâ destekli finansal karar verme ile düzenleyici şeffaflık arasındaki boşluğu kapatmakta; nihayetinde yapay zekâ tabanlı finansal uygulamalarda güveni, hesap verebilirliği ve güvenilirliği artırmaktadır.

3.1 Metodoloji

Bu bölümde, finans sektörünün çeşitli uygulama alanlarında kullanılmak üzere tasarlanan SFIX açıklanabilir yapay zekâ modelinin ayrıntılı bir incelemesi sunulacaktır.

Önerilen SFIX modelinin algoritmasında kullanılan zaman parametresi, analiz edilen veri veya sistemin zamansal dinamiklerini yansıtması açısından önemlidir ve bu yönüyle diğer XAI yöntemlerinden ayrışabilmektedir.

Son olarak, önerilen modeli açıklamak için tümdengelimsel bir yaklaşım benimsenmiştir. Önerilen model tümdengelimsel olarak açıklanacaktır.

3.1.1 Dinamik özellik önem skoru (DIS)

DIS, SFIX modelinin temel bileşenlerinden biridir ve bu metrik, modelin belirli bir zamanda neden belirli bir özelliğe daha fazla ağırlık verdiğini anlamayı sağlar. Bu durum, zaman içinde değişen etkileri yakalayarak ve özellikler arasındaki karmaşık ilişkileri değerlendirerek hem modelin doğruluğunu hem de sonuçların yorumlanabilirliğini artırmaktadır.

DIS'in alt bileşenleri, bu bölümün alt başlıkları altında incelenecektir.

DIS bileşeni, SHAP ve LIME gibi açıklanabilir yapay zekâ yöntemlerinin temel yaklaşımından esinlenilerek geliştirilmiş ve bu yöntemlerin zamansal ağırlıklandırma çerçevesinde yeniden uyarlanmasıyla oluşturulmuştur (Lundberg ve Lee, 2017a; Ribeiro ve ark., 2016).

Bir özelliğin f anındaki, t zamanındaki DIS, Denklem 3.1'deki şekilde hesaplanır:

$$DIS(f,t) = \sum_{w=1}^{W} \alpha(w) \cdot [BI(f,w) \cdot II(f,w)] \cdot TW(t,w) \quad (3.1)$$

- *W*: Toplam zaman penceresi sayısını ifade etmektedir.
- *α(w)*: *w* zaman penceresi için kullanılan normalizasyon katsayısını göstermektedir.
- *BI(f,w)*: f özelliğinin *w* zaman penceresindeki temel etki skorunu ifade etmektedir.
- *II(f,w)*: f özelliğinin *w* zaman penceresindeki etkileşim etki skorunu göstermektedir.
- *TW(t,w)*: *t* zamanı ile *w* zaman penceresi arasındaki ilişkiye bağlı olarak belirlenen zamansal ağırlığı ifade etmektedir.

DIS, bir özelliğin zaman içindeki öneminin hem doğrudan hem de dolaylı etkileriyle birlikte bütüncül olarak değerlendirilmesini sağlar:

- Temel etki (BI): Özelliğin hedef değişken ile ana ilişkisinin gücünü ölçer.
- Etkileşim etkisi (II): Özelliğin diğer özelliklerle olan ilişkilerini ve dolaylı etkilerini dikkate alır.
- Zamansal etki (TW): Zaman içindeki uzaklığı ve periyodik davranışı modellemek için kullanılır.

Algoritma 3.1, DIS hesaplama süreci için sistematik bir yaklaşım sunmaktadır. Bu algoritmanın en önemli özelliği, yalnızca özniteliklerin bireysel etkilerini değil, bununla birlikte diğer özniteliklerle olan etkileşimlerini de değerlendirmesidir. Zaman ağırlığı bileşeni, geçmiş bilginin mevcut karar üzerindeki azalan etkisini modellerken aynı zamanda döngüsel davranışları da yakalamaktadır. Bu sayede, finansal verilerde sıklıkla görülen mevsimsel ve periyodik etkiler açıklanabilirlik skoruna dahil edilmekte ve daha gerçekçi bir değerlendirme yapılabilmektedir.

Algoritma 3.1. Dinamik Önem Skorunun hesaplanması

Girdi: X_f: öznitelik vektörü, y: hedef değişken, W: zaman penceresi, t: mevcut zaman
Çıktı: DIS(f,t): dinamik önem skoru

DIS ← 0
for w = 1 to W **do**
  corr ← Korelasyon (X_f, y_w)
  p_{value} ← İstatistiksel anlamlılık testi (X_f, y_w)
  BI(f,w) ← |corr| × (1 − p_{value})
  II(f,w) ← 0
  for i = 1 to N **do** {Diğer öznitelikler}
    K(f,i,w) ← Kendall'ın Tau katsayısı (X_f, X_i, w)
    D(f,i,w) ← Dolaylı korelasyon katsayısını hesapla
    II(f,w) ← II(f,w) + K(f,i,w) × D(f,i,w)
  end for
  TW(t,w) ← λ × $e^{-\gamma(t-w)}$ + (1−λ) × cos(2π(t−w)/T)
  α(w) ← Zaman penceresi ağırlığı
  DIS ← DIS + α(w) × [BI(f,w) × II(f,w)] × TW(t,w)
end for
return DIS

3.1.1.1 Temel etki skoru (BI)

Bu bileşen, özelliğin hedef değişken ile doğrudan ilişkisini ölçmek için kullanılmaktadır. Lundberg ve Lee (2017a) ve Ribeiro ve ark. (2016) çalışmasından esinlenilen bu bileşenin formülasyonu Denklem 3.2 ile ifade edilmektedir.

$$BI(f,w) = \left|corr\left(X_f, y_w\right)\right| \cdot \left(1 - p_{value}(f,w)\right) \tag{3.2}$$

- $corr(X_f, y_w)$: w zaman penceresinde, f özelliği ile hedef değişken arasındaki korelasyon katsayısını ifade etmektedir.
- $p_{value}(f,w)$: İlgili korelasyon katsayısının istatistiksel anlamlılık düzeyini göstermektedir. Daha düşük p değeri, ilişkinin daha yüksek düzeyde istatistiksel anlamlılığa sahip olduğunu göstermekte ve bu tür ilişkiler daha güçlü biçimde ağırlıklandırılabilmektedir.

3.1.1.2 Etkileşim etki skoru (II)

Bu bileşen, bir özelliğin diğer özelliklerle olan ilişkilerini ve dolaylı etkilerini hesaba katmak için kullanılmaktadır. Özellikle korelasyon tabanlı değişken öneminin ölçülmesinde kullanılmıştır (Guyon ve Elisseeff, 2003; Kendall, 1948). Bu bileşenin formülasyonu Denklem 3.3 ile ifade edilmektedir:

$$II(f,w) = \sum_{i=1}^{N} K(f,i,w) \cdot D(f,i,w) \quad (3.3)$$

Kendall Tau Korelasyon Katsayısı (K(f,i,w)):

Özellik *f* ile özellik i arasındaki sıralı bağımlılığı ölçer. Bu katsayı, iki özellik arasındaki monoton ilişkiyi anlamak için kullanılır.

Kendall Tau, sıralı çiftler arasındaki uyumlu (C) ve uyumsuz (D) sayıların farkının toplam çift sayısına oranı olarak Denklem 3.4'teki gibi hesaplanır:

$$K(f,i,w) = \frac{C-D}{C+D} \quad (3.4)$$

- C: Uyumlu sıralı çift sayısı
- D: Uyumsuz sıralı çift sayısı

Bir verisetinde, iki değişken (X ve Y) için tüm olası gözlem çiftleri (i, i') oluşturulur. Bu çiftlerin sıralı ilişkisi aşağıdaki şekilde değerlendirilir:

Uyumlu Çiftler (C):

- Eğer $X_i > X_i'$ ve aynı zamanda $Y_i > Y_i'$ ise, ilgili gözlem çifti uyumlu (concordant) kabul edilir.
- Eğer $X_i < X_i'$ ve aynı zamanda $Y_i < Y_i'$ ise, yine ilgili gözlem çifti uyumlu (concordant) olarak değerlendirilir.

Uyumsuz Çiftler (D):

- Eğer $X_i > X_i'$ iken $Y_i < Y_i'$ ise, ilgili gözlem çifti uyumsuz (discordant) kabul edilir.
- Eğer $X_i < X_i'$ iken $Y_i > Y_i'$ ise, ilgili gözlem çifti yine uyumsuz (discordant) olarak değerlendirilir.

Dolaylı etkileşim faktörü D(f,i,w): Özellik *f* ile özellik *i* arasındaki dolaylı etkileşimi modellemek için kullanılır. Bu faktör, özellik *i*'nin diğer özelliklerle (*j*) olan ilişkilerinin ağırlıklı toplamı olarak Denklem 3.5'teki gibi hesaplanır.

$$D(f,i,w) = \sum_{j=1}^{N} K(i,j,w) \cdot \beta(j) \quad (3.5)$$

β(j): *j* özelliğine ait ağırlık katsayısını ifade etmektedir.

Özellik *i*'nin diğer tüm özelliklerle (*j*) olan ilişkileri K(i,j,w) üzerinden

hesaplanmaktadır. Elde edilen bu ilişkiler, ilgili özelliklerin katkı düzeyini yansıtmak amacıyla *β(j)* ağırlık katsayıları ile çarpılmaktadır. Böylece, her bir ilişkinin toplam etkileşim skoruna etkisi, ilgili ağırlık katsayısı doğrultusunda belirlenmektedir.

Ağırlık katsayısı *β(j)*'nin nasıl belirleneceği yöntem tasarımına bağlı olarak açık bırakılmıştır. Bunun nedeni, bu parametrenin hesaplanmasının uygulama alanına ve verisetine göre değişecek olmasıdır. Bu parametre, alan uzmanlarının bilgi ve deneyimine dayanarak manuel olarak atanabilir ya da çeşitli yöntemlerle hesaplanabilir. Bu çalışmada, iki şekilde uygulanması önerilmektedir.

İlk yöntem, hedef değişken (y) ile özellikler (X_i) arasındaki doğrudan doğrusal ilişkilerin baskın olduğu durumlarda korelasyon tabanlı bir hesaplama olabilir.

Özellik ağırlık katsayısı *β(j),* ilk yöntemde X_j özelliği ile hedef değişken y arasındaki ilişkinin gücüne dayalı olarak Pearson korelasyon katsayısı kullanılarak Denklem 3.6'daki gibi hesaplanmaktadır:

$$\beta(j) = \frac{|\text{corr}(X_j,y)|}{\sum_{k=1}^{N}|\text{corr}(X_k,y)|} \tag{3.6}$$

- *corr(X_j, y)*: X_j özelliği ile hedef değişken y arasındaki Pearson korelasyon katsayısını ifade etmektedir.

Pearson korelasyon katsayısı Denklem 3.7'deki şekilde tanımlanmaktadır:

$$\text{corr}(X_j, y) = \frac{\text{Cov}(X_j,y)}{\sigma_{X_j}\cdot\sigma_y} \tag{3.7}$$

Bu denklemde:

- *Cov(X_j, y)*: X_j ile y arasındaki kovaryansı ifade etmektedir.
- σ_{X_j}: X_j özelliğinin standart sapmasını göstermektedir.
- σ_y: Hedef değişken y'nin standart sapmasını göstermektedir.

Bu yaklaşımda, her bir özelliğin hedef değişken ile olan doğrusal ilişkisinin mutlak değeri alınmakta ve tüm özellikler üzerindeki toplam mutlak korelasyon değerine bölünerek normalize edilmektedir. Böylece *β(j)*, ilgili özelliğin hedef değişkenle olan doğrusal ilişkisinin göreli ağırlığını temsil etmektedir.

Doğrusal olmayan bağımlılıkların söz konusu olduğu durumlarda ise ikinci yöntem

olarak Mutual Information (MI) kullanılabilmektedir. Ancak, bu yöntemin hesaplama maliyetini artıracağı unutulmamalıdır. Bu kapsamda *β(j)* ağırlık katsayısı Denklem 3.8'deki şekilde hesaplanmaktadır:

$$\beta(j) = MI(X_j, y) = \sum_{x \in X_j} \sum_{y \in Y} p(x,y) \cdot \log\left(\frac{p(x,y)}{p(x) \cdot p(y)}\right) \quad (3.8)$$

- *p(x,y)*: X_j ile *y* arasındaki ortak olasılık dağılımını ifade etmektedir.
- *p(x)*: X_j değişkeninin marjinal olasılığını göstermektedir.
- *p(y)*: *y* değişkeninin marjinal olasılığını göstermektedir.

β(j) değerleri, $MI(X_j, y)$ değerlerinin normalize edilmesiyle Denklem 3.9 elde edilir:

$$\beta(j) = \frac{MI(X_j,y)}{\sum_{k=1}^{N} MI(X_k,y)} \quad (3.9)$$

Bu denklemde:

- *β(j)*: X_j özelliğinin toplam Mutual Information içindeki göreli katkısını ifade etmektedir.
- $MI(X_j, y)$: X_j ile *y* arasındaki Mutual Information değerini göstermektedir.
- $\sum_{k=1}^{N} MI(X_k, y)$: Tüm özellikler için hesaplanan Mutual Information değerlerinin toplamını ifade etmektedir.

3.1.1.3 Zamansal ağırlıklandırma (TW)

Bir özelliğin zaman içindeki ağırlığını hesaplamak için zamansal ağırlık fonksiyonu kullanılmaktadır. Zamansal ağırlıklandırma, *t* ve *w* zaman pencereleri arasındaki yakınlığı ve etkiyi dikkate almayı amaçlamaktadır. Zaman serisi analizinde EWMA (Exponentially Weighted Moving Average) ve periyodik ağırlıklandırma teknikleri için Chatfield (2003) ve Shumway ve Stoffer (2017) çalışmalarından esinlenilmiştir. Bu bileşenin formülasyonu aşağıdaki denklemle ifade edilmektedir. Bu bileşen Denklem 3.10 ile hesaplanmaktadır:

$$TW(t,w) = \lambda \cdot e^{-\gamma(t-w)} + (1-\lambda) \cdot \cos\left(\frac{T}{2\pi(t-w)}\right) \quad (3.10)$$

Burada:

- λ: Toplam zamansal ağırlık içerisinde üstel bileşenin göreli önemini belirleyen

parametreyi ifade etmektedir.

- γ: Etkilerin zaman içinde ne kadar hızlı azaldığını kontrol eden parametredir. γ değerinin artması, geçmiş gözlemlerin etkisinin daha hızlı biçimde azalmasına neden olmaktadır.
- *T*: Periyodik yapıyı belirleyen periyot uzunluğunu göstermektedir.
- *t:* Mevcut zaman noktasını ifade etmektedir.
- *w*: Önceki zaman noktasını göstermektedir.

Bu bileşenin türetilme mantığı iki temel yapı üzerinden açıklanmaktadır:

- Üstel azalma bileşeni, $\lambda \cdot e^{-\gamma(t-w)}$: Geçmiş verilerin hedef değişken üzerindeki etkisinin zamanla azalmasını modellemek amacıyla kullanılmaktadır. Burada $e^{-\gamma(t-w)}$ ifadesi, *t* ile *w* arasındaki zaman farkı arttıkça ağırlığın üstel biçimde azalmasını sağlamaktadır. Böylece geçmişteki gözlemlerin mevcut tahmin üzerindeki etkisi kontrollü şekilde zayıflatılmaktadır.
- Periyodik bileşen, $(1-\lambda) \cdot \cos\left(\frac{T}{2\pi(t-w)}\right)$: Zaman içerisinde tekrar eden döngüsel örüntülerin ve periyodik etkilerin modele yansıtılmasına olanak tanımaktadır. Bu bileşen, belirli aralıklarla ortaya çıkan zamansal davranışların yakalanmasını sağlamaktadır.

Üstel ve periyodik bileşenler arasındaki denge, λ parametresi aracılığıyla belirlenmektedir:

- λ = 1: Yalnızca üstel azalma bileşeni etkili olmaktadır.
- λ = 0: Yalnızca periyodik bileşen etkili olmaktadır.

Bu yapı, özellikle insan davranışlarında gözlenen zamansal tekrarların, örneğin günlük alışkanlıkların, yanı sıra geçmiş olayların zamanla azalan etkisinin birlikte modellenmesinde etkili bir yaklaşım sunmaktadır.

Zamansal ağırlıklandırma mekanizması şu avantajları sağlamaktadır:

- Zamansal azalma ile periyodik örüntüleri aynı yapı içinde modelleyerek esneklik sunmaktadır.

- λ, γ ve T parametrelerinin yorumlanabilir olması sayesinde farklı uygulama alanlarına kolayca uyarlanabilmektedir.
- Hem uzak geçmişten gelen etkilerin hem de döngüsel örüntülerin uygun biçimde modele yansıtılmasına olanak tanımaktadır.

3.1.2 Güven skoru (CS)

CS, bir özelliğin model içindeki güvenilirliğini değerlendiren bir metriktir. Bu skor, özelliklerin veri kalitesine, model tutarlılığına ve normalize edilmiş anlamlılığa dayanmaktadır. CS, Denklem 3.11'de hesaplanmaktadır:

$$\mathrm{CS}(f) = \mathrm{DQ}(f) \cdot \mathrm{MC}(f) \tag{3.11}$$

Burada:

- DQ(f): Veri kalitesi skorunu temsil eder.
- MC(f): Model tutarlılığı skorunu temsil eder.

Algoritma 3.2, veri kalitesi ve model tutarlılığı metriklerini birleştirerek kapsamlı bir güven skoru değerlendirmesi sunmaktadır. Eksik değer oranı (MR) ve aykırı değer oranı (OR) hesaplamaları, veri setinin genel kalitesi hakkında nesnel bilgiler sağlarken, varyasyon katsayısı (CV) ilgili özelliğin kararlılık düzeyini yansıtmaktadır. Model tutarlılığı bileşeni ise özelliğin hedef değişkenle olan ilişkisinin ne kadar tutarlı olduğunu ölçerek açıklanabilirlik analizinin güvenilirliğini artırmaktadır.

Algoritma 3.2. Güven Skorunun hesaplanması

Girdi: X_f: öznitelik, y: hedef değişken, ŷ : model tahminleri
Çıktı: CS(f): güven skoru

eksik_sayisi ← Eksik değerlerin sayısı(X_f)
toplam_sayisi ← Toplam değerlerin sayısı(X_f)
MR ← eksik_sayisi / toplam_sayisi
Q1, Q3 ← Birinci ve üçüncü çeyreklikler(X_f)
IQR ← Q3 − Q1
aykiri_sayisi ← Aykırı değerlerin sayısı(X_f, Q1 − 1.5 × IQR, Q3 + 1.5 × IQR)
OR ← aykiri_sayisi / toplam_sayisi
μ_f ← Ortalama(X_f)
σ_f ← Standart sapma(X_f)
CV ← σ_f / μ_f
DQ(f) ← (1 − MR) × (1 − OR) × 1/(1 + |CV|)
ρ(f,y) ← Korelasyon(X_f, y)
σ_y ← Standart sapma(ŷ)
MC(f) ← |ρ(f,y)| × 1/(1 + | σ_y |)
CS(f) ← DQ(f) × MC(f)
return CS(f)

3.1.2.1 Veri kalitesi skoru (DQ)

Veri Kalitesi Skoru (DQ), bir özelliğin ham verisinin güvenilirliğini ve doğruluğunu ölçmek için kullanılmaktadır. Özelliklerin eksik veri oranı, aykırı değer oranı ve varyasyon katsayısı gibi metrikler dikkate alınmaktadır. Amaç, kaliteli veriye sahip özelliklerin daha yüksek skorlar almasını sağlamaktır. Veri kalitesinin değerlendirilmesi, eksik ve hatalı veri oranlarının model güvenilirliği üzerindeki etkisini ele alan Gebru ve ark. (2021) çalışmasından esinlenilmiştir. DQ, Denklem 3.12'de hesaplanmaktadır:

$$DQ(f) = (1 - MR) \cdot (1 - OR) \cdot \frac{1}{1+CV} \tag{3.12}$$

MR: Eksik veri oranı. Denklem 3.13'te hesaplandığı gibi, özellik f'deki eksik değerlerin oranıdır. Eksik veri oranı arttıkça veri kalitesi düşmektedir.

$$MR = \frac{\text{Eksik Değer Sayısı}}{\text{Toplam Gözlem Sayısı}} \tag{3.13}$$

OR: Aykırı değer oranı. Denklem 3.14'te hesaplandığı gibi, özellik f'deki aykırı değerlerin oranıdır. Yüksek bir OR, düşük veri kalitesine işaret eder.

$$OR = \frac{\text{Aykırı Değer Sayısı}}{\text{Toplam Gözlem Sayısı}} \tag{3.14}$$

OR, bir özelliğe ait gözlemlerin istatistiksel olarak normal aralığın dışında kalan kısmının

oranını ölçmektedir. Yüksek OR değerleri düşük veri kalitesine işaret eder.

Aykırı değerler genellikle Denklem 3.15'teki Z-Skoru yöntemi ile belirlenmektedir.

$$z = \frac{x-\mu}{\sigma} \tag{3.15}$$

Eğer |z|>3 ise, değer aykırı değer olarak kabul edilir.

- x: Gözlem değeri.
- μ: Özelliğin ortalaması.
- σ: Özelliğin standart sapması.

CV: Varyasyon katsayısı. Özellik f'nin standart sapmasının ortalamasına oranı Denklem 3.16 ile ifade edilir:

$$CV = \frac{\sigma_f}{\mu_f} \tag{3.16}$$

- σ_f: İlgili özelliğin standart sapmasını ifade etmektedir.
- μ_f: İlgili özelliğin ortalama değerini göstermektedir.

Daha düşük CV, daha iyi veri kalitesine işaret eder.

3.1.2.2 Model tutarlılığı skoru (MC)

Model Tutarlılığı Skoru (*MC(f)*), bir özelliğin hedef değişken üzerindeki etkisinin modele ne kadar iyi uyduğunu ve modelin bu özelliği ne kadar tutarlı kullandığını ölçmektedir. Özellik ile hedef değişken arasındaki ilişkiyi ve modelin bu ilişkiden ne ölçüde yararlandığını değerlendirmektedir. Veri kalitesinin değerlendirilmesi, eksik ve hatalı veri oranlarının model güvenilirliği üzerindeki etkisini ele alan Gebru ve ark. (2021) çalışmasından esinlenilmiştir. *MC(f)*, Denklem 3.17'de hesaplanmaktadır:

$$MC(f) = |\rho(f, y)| \cdot \frac{1}{1+\sigma_y} \tag{3.17}$$

- $\rho(f, y)$: Özellik f ile hedef değişken y arasındaki korelasyon. Hedef değişken ile güçlü ilişkiye sahip özellikler model tarafından daha iyi değerlendirilebilir.
- σ_y: Hedef değişken üzerindeki model tahminlerinin standart sapması. Tahmin sapması düşükse model, özellik ile hedef arasındaki ilişkiyi daha tutarlı biçimde açıklayabilmektedir.

Model tutarlılığının korelasyon bileşeni Denklem 3.18'deki ile hesaplanmaktadır:

$$\rho(f,y) = \frac{\mathrm{Cov}(f,y)}{\sigma_f \cdot \sigma_y} \tag{3.18}$$

Korelasyon ($\rho(f,y)$), bir özelliğin (f) hedef değişken (y) ile doğrusal ilişkisinin gücünü ve yönünü ifade eder.

Değer Aralığı:

$$-1 \leq \rho(f,y) \leq 1 \tag{3.19}$$

$|\rho(f,y)|$ değeri 1'e yaklaştıkça ilişki güçlenmektedir. Pozitif değerler ($\rho(f,y)>0$) doğrusal artışı, negatif değerler ($\rho(f,y)<0$) ise ters yönlü ilişkiyi göstermektedir.

Model tutarlılığının standart sapma bileşeni Denklem 3.20'deki ile hesaplanmaktadır:

$$\sigma_y = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_{\text{tahmin},i} - \overline{y_{\text{tahmin}}}\right)^2} \tag{3.20}$$

- y_{tahmin}: Hedef değişkene ilişkin model tahminlerini ifade etmektedir.
- $\bar{y}_{forecast}$: Tahmin değerlerinin ortalamasını göstermektedir.
- N: Toplam gözlem sayısını ifade etmektedir.

Ayrıca σ_y, modelin hedef değişkene ilişkin ürettiği tahmin değerlerinin standart sapmasını ölçmektedir. Bu değer, model tahminlerinin ne ölçüde değişkenlik gösterdiğini ortaya koymaktadır. Düşük bir σ_y değeri, model tahminlerinin daha tutarlı ve kararlı olduğunu göstermekte; bu durum da MC(f) değerinin artmasına katkı sağlamaktadır.

Yüksek σ_y değeri, model tahminlerinde daha fazla değişkenlik bulunduğunu ve buna bağlı olarak ilgili özelliğin hedef değişken üzerindeki etkisinin yeterince tutarlı biçimde açıklanamadığını göstermektedir. Model tutarlılığının bir bileşeni olarak σ_y'nin, yani model tahmin çıktılarının standart sapmasının kullanılması, daha düşük tahmin varyansının daha yüksek model kararlılığına işaret ettiği varsayımına dayanmaktadır. Özellikle riske duyarlı görevlerin yer aldığı finansal karar verme sistemlerinde, yüksek düzeyde değişken tahminler; modelin aşırı uyum göstermesi, zayıf genelleme performansı sergilemesi veya kalibrasyon sorunları taşıması gibi durumlara işaret edebilmektedir. Bununla birlikte, σ_y değerinin kullanılan öğrenme algoritmasına, tercih

edilen kayıp fonksiyonuna ve uygulanan kalibrasyon tekniklerine bağlı olarak önemli ölçüde değişebileceği kabul edilmektedir. Bu nedenle ilgili bileşen bilinçli olarak modüler bir yapıda tasarlanmıştır. Böylece σ_y, bağlama ve model mimarisine bağlı olarak güven aralıkları, kalibrasyon eğrileri veya artık varyansı gibi alternatif tutarlılık göstergeleriyle değiştirilebilmekte ya da desteklenebilmektedir.

3.1.3 Örüntü skoru (PS)

Zaman serisi analizinde PS bileşeni, özellikler ile hedef değişkendeki zamansal örüntüler arasındaki ilişkiyi değerlendirmek için geliştirilmiş bir metriktir. PS, bir özelliğin trend, periyodiklik ve oynaklık gibi zamansal dinamikler sergileyip sergilemediğini, zaman t anındaki örüntü skoruna etkisini modelleyerek yakalar.

Bir özelliğin t zamanındaki PS Denklem 3.21'deki şekilde hesaplanır:

$$PS(f,t) = \sum_{k=1}^{K} \eta_k \cdot P_k(f,t) \tag{3.21}$$

- K: Dikkate alınan toplam örüntü türü sayısını ifade etmektedir.
- η_k: k örüntü türüne karşılık gelen ağırlık katsayısını göstermektedir.
- $P_k(f,t)$: k türündeki örüntü için, f özelliği ve t zaman noktası bağlamında hesaplanan örüntü skorunu ifade etmektedir.

Algoritma 3.3, zaman serisi analizinin üç temel bileşenini bütünleştirerek kapsamlı bir örüntü değerlendirmesi gerçekleştirmektedir. Doğrusal regresyona dayalı eğilim analizi uzun dönemli yönelimleri yakalarken, Fourier serisi yaklaşımı döngüsel davranışları doğru biçimde modellemektedir. Volatilite bileşeni ise kısa dönemli değişimleri ölçerek özelliklerin kararlılık düzeyini değerlendirmektedir. Bu çok bileşenli yaklaşım, finansal piyasalarda gözlenen karmaşık zamansal örüntülerin açıklanabilirlik analizine dahil edilmesini sağlayarak daha zengin bir analitik bakış açısı sunmaktadır.

Algoritma 3.3. Örüntü Skorunun hesaplanması

Girdi: Öznitelik zaman serisi $X_f(t)$, zaman t, periyot uzunluğu T
Çıktı: Örüntü skoru PS(f,t)

$PS \leftarrow 0$
{Eğilim Bileşeni Hesaplaması}
$(\beta_1, \beta_0) \leftarrow$ Doğrusal regresyon (X_f, zaman indeksleri)
$P_1(f,t) \leftarrow \beta_1 \times t + \beta_0$
{Periyodiklik Bileşeni Hesaplaması}
$a_0 \leftarrow$ Ortalama (X_f)
Periyodiklik $\leftarrow a_0$
for k = 1 to n **do**
    {Fourier bileşenleri}
    $a_k \leftarrow$ Kosinüs katsayısını hesapla (X_f, k, T)
    $b_k \leftarrow$ Sinüs katsayısını hesapla (X_f, k, T)
    Periyodiklik $\leftarrow$ Periyodiklik + $a_k \times \cos(2\pi kt/T) + b_k \times \sin(2\pi kt/T)$
end for
$P_2(f,t) \leftarrow$ Periyodiklik
{Oynaklık Bileşeni Hesaplaması}
pencere_boyutu $\leftarrow$ k
$\mu_f \leftarrow$ Hareketli ortalama (X_f, pencere_boyutu)
Oynaklık $\leftarrow 0$
for i = t − k + 1 to t **do**
    Oynaklık $\leftarrow$ Oynaklık + $(X_f(i) - \mu_f)^2$
end for
$P_3(f,t) \leftarrow$ Oynaklık / k
{Ağırlıklı Toplam}
for i = 1 to 3 **do**
    $PS \leftarrow PS + \eta_i \times P_i(f,t)$
end for
return PS

Modeldeki örüntü türleri alt başlıklar altında incelenmektedir.

3.1.3.1 Trend bileşeni

Trend bileşeni, zaman serisi verisindeki uzun vadeli değişimlerin yönünü ve genel eğilimini ölçen bir metriktir. Bu bileşen, zaman serisinin temel davranışını; örneğin sürekli artış, azalış ya da sabit eğilim gibi örüntüleri yakalamayı amaçlamaktadır.

Ayrıca kısa vadeli dalgalanmaları ortadan kaldırarak genel eğilimlerin hedef üzerindeki etkisini de ölçmektedir. Zaman serilerindeki trend analizi için Box ve ark. (2015) çalışmasından esinlenilmiştir.

Zaman serisindeki doğrusal eğilimleri belirlemek için Denklem 3.22'de yer alan doğrusal regresyon kullanılmaktadır:

$$P_1(f,t) = \beta_1 \cdot t + \beta_0 \quad (3.22)$$

β_1, β_0: Doğrusal regresyon modeline ait katsayıları ifade etmektedir. Burada β_0 sabit terimi, β_1 ise bağımsız değişkenin etkisini gösteren eğim katsayısını temsil etmektedir.

3.1.3.2 Periyodiklik bileşeni

Periyodiklik bileşeni, zaman serisi verilerinde düzenli olarak tekrar eden örüntüleri veya döngüsel davranışı ifade etmektedir. Bu bileşen, bir özelliğin hedef değişken üzerindeki döngüsel etkisini değerlendirmek açısından önemlidir.

Basit ya da karmaşık periyodik örüntüleri modellemek için Denklem 3.23'teki Fourier serileri kullanılmaktadır. Fourier dönüşümü tabanlı periyodik analiz için Oppenheim ve Schafer (2009) çalışmasından esinlenilmiştir.

$$\text{Periyodiklik}\,(t) = a_0 + \sum_{k=1}^{n}\left[a_k \cos\left(\frac{2\pi kt}{T}\right) + b_k \sin\left(\frac{2\pi kt}{T}\right)\right] \quad (3.23)$$

Burada:

- k: Döngüsel frekansları temsil eden harmonik indisidir.
- a_k: k'inci frekanstaki kosinüs bileşeninin genliğini ifade etmektedir.
- b_k: k'inci frekanstaki sinüs bileşeninin genliğini göstermektedir ve faz kaymasını modellemeye yardımcı olmaktadır.
- a_0: Ortalama değeri, yani sabit bileşeni ifade etmektedir.

Sabit bileşen a_0 Denklem 3.24'teki şekilde tanımlanmaktadır:

$$a_0 = \frac{1}{T}\int_0^T f(t)\,,dt \quad (3.24)$$

Bu ifade, periyodik fonksiyonun tüm döngü boyunca ortalama katkısını göstermektedir. Başka bir ifadeyle a_0, zaman serisinin genel taban seviyesini ve ortalama değerini temsil etmektedir.

Kosinüs katsayıları ise Denklem 3.25'te hesaplanmaktadır:

$$a_k = \frac{2}{T}\int_0^T f(t)\cos\left(\frac{2\pi kt}{T}\right),dt \quad (3.25)$$

Burada a_k, k'inci frekanstaki kosinüs bileşeninin genliğini ifade etmektedir. Kosinüs fonksiyonlarına ait katsayılar, periyodik yapının daha çok simetrik ve genlik temelli

bileşenlerini yakalamaktadır.

Sinüs katsayıları ise Denklem 3.26'deki şekilde tanımlanmaktadır:

$$b_k = \frac{2}{T}\int_0^T f(t)\sin\left(\frac{2\pi kt}{T}\right), dt \tag{3.26}$$

Burada b_k, k'inci frekanstaki sinüs bileşeninin genliğini göstermektedir. Sinüs fonksiyonlarına ait katsayılar, periyodik yapının faza bağlı ve asimetrik bileşenlerini yakalamaktadır.

Harmonikler, temel frekansın tam sayı katlarıdır. Bu kapsamda:

- k = 1: Temel frekansı ifade etmektedir ve serinin ana döngüsünü göstermektedir.
- k = 2: İkinci harmoniktir ve temel döngüden iki kat daha hızlı değişen bileşeni temsil etmektedir.
- k = 3: Üçüncü harmoniktir ve temel döngüden üç kat daha hızlı değişen bileşeni ifade etmektedir.

Harmonik sayısı arttıkça, zaman serisinin daha ince ve kısa dönemli değişimlerini modellemek mümkün hâle gelmektedir. Özellikle k>1 olan yüksek harmonikler, serideki daha hızlı değişen örüntülerin yakalanmasına katkı sağlamaktadır. Her bir harmonik, serinin belirli bir frekans bileşenini temsil etmektedir. Örneğin T bir yılı ifade ediyorsa, k = 1 yıllık döngüyü, k = 2 altı aylık döngüyü göstermektedir. Bu nedenle frekans, periyodik örüntünün hangi hızla tekrarlandığını belirleyen temel unsurlardan biridir.

3.1.3.3 Oynaklık (Volatilite) bileşeni

Oynaklık, özellikle finansal ve dinamik sistemlerde risk analizi için kritik bir metriktir. Ayrıca oynaklıktaki artışlar çoğu zaman anomali tespitinin bir göstergesi olarak kabul edilmektedir.

Varyasyon, zaman serisi içindeki değerlerin ortalamadan ne kadar saptığını ölçtüğü için bu bileşenin ölçülmesinde kullanılmaktadır. Oynaklık bileşeni skoru elde edilirken Andersen ve ark. (2003) çalışmasından yararlanılmıştır.

$$\text{Volatilite}(f, w) = \frac{1}{k}\sum_{t=w-k+1}^{w}\left(X_f(t) - \mu_f\right)^2 \tag{3.27}$$

- *k*: İlgili zaman penceresinde yer alan veri noktası sayısını ifade etmektedir.

- $X_f(t)$: Zaman serisinin t anındaki değerini göstermektedir.
- μ_f: İlgili pencere içerisindeki ortalama değeri ifade etmektedir.

3.1.4 Anomali skoru (AS)

AS bileşeni, veride beklenmeyen, nadir veya olağandışı durumların tespit edilmesi için kullanılmaktadır. Model çıktısında açıklanabilirlik sağlarken aynı zamanda riskli veya olağandışı durumların belirlenmesi ve anlaşılmasında önemli bir rol oynamaktadır. Z-skoru tabanlı anomali tespit yöntemleri Chandola ve ark. (2009) çalışmasından esinlenilmiştir.

Bu yaklaşım, finansal dolandırıcılık tespiti için istatistiksel anomali tespitinden yararlanan yakın dönem çalışmalarla tutarlı olduğu görülmüştür. Örneğin, Wu ve Wang (2021) tarafından yapılan çalışma, kredi kartı işlem hacimlerindeki düzensizlikleri tespit etmek amacıyla ARIMA tabanlı zaman serisi modellemesini Z-skoru analizi ile birleştiren bir yaklaşımı incelemektedir. Bu çerçevede yazarlar, zaman içindeki beklenen işlem davranışını ARIMA kullanarak modellemekte ve ardından istatistiksel olarak anlamlı aykırı değerleri potansiyel dolandırıcılık göstergesi olarak işaretlemek için artıklar üzerinde Z-skorları hesaplamaktadır. $|Z| > 3$ gibi bir eşik uygulanarak, normal davranış örüntüleri altında gerçekleşmesi oldukça düşük olasılıklı sapmalar belirlenmektedir. Bu yaklaşım, özellikle etiketli verinin sınırlı olduğu veya hiç bulunmadığı erken aşama dolandırıcılık taramalarında, hafif ve yorumlanabilir bir anomali tespit mekanizması sunmaktadır. Sonuçlar, karmaşık kara kutu modellerine dayanmadan anormal finansal faaliyetlerin tespit edilmesinde klasik zaman serisi tahminlemesi ile standart sapma tabanlı puanlamanın birleştirilmesinin etkinliğini vurgulamaktadır (Moschini ve ark., 2020).

Bu nedenle, hareketli ortalama ve standart sapma yöntemine ait Denklem 3.28 kullanılmıştır:

$$AS(f,t) = \frac{|X_f(t) - \mu_f(t)|}{\sigma_f(t)} \tag{3.28}$$

- $X_f(t)$: Özelliğin t anındaki değerini ifade etmektedir.
- $\mu_f(t)$: Özelliğin t anına kadar hesaplanan hareketli ortalamasını Denklem 3.29 göstermektedir:

$$\mu_f(t) = \frac{1}{W}\sum_{i=t-W+1}^{t} X_f(i) \quad (3.29)$$

- $\sigma_f(t)$: Özelliğin t anına kadar olan hareketli standart sapmasını Denklem 3.30'da ifade etmektedir:

$$\sigma_f(t) = \sqrt{\frac{1}{W}\sum_{i=t-W+1}^{t}\left(X_f(i) - \mu_f(t)\right)^2} \quad (3.30)$$

Hareketli ortalama ve standart sapma yöntemi, zaman serisi verilerindeki dinamik değişimleri ve anomalileri daha iyi tespit etmektedir.

Ancak, tüm veri üzerinde genel bir değerlendirme yapabilmek ve hesaplama maliyetini azaltmak amacıyla, standart sapmaya dayalı Denklem 3.31 tercih edilmektedir:

$$AS(f,t) = \frac{\left|X_f(t) - \mu_{X_f}\right|}{\sigma_{X_f}} \quad (3.31)$$

Burada:

- μ_{X_f}: Özelliğin tarihsel ya da beklenen ortalama değerini ifade etmektedir.
- σ_{X_f} : Özelliğin standart sapmasını göstermektedir ve verinin yayılımını yansıtmaktadır.
- $|X_f(t) - \mu_{X_f}|$: Özelliğin ilgili andaki değerinin ortalamadan ne kadar saptığını göstermektedir. Bu sapma arttıkça AS da artmaktadır.

Uyumsuzluk düzeyini belirlemek amacıyla, $|X_f(t) - \mu_{X_f}|$ değeri standart sapmaya bölünerek normalize edilmektedir. Böylece sapma miktarı, standart sapma birimleri cinsinden ifade edilmekte ve farklı özellikler arasında karşılaştırılabilir hâle getirilmektedir.

Standart sapma birimi, bir değerin ne kadar "anormal" olduğunu ölçmek için aşağıdaki şekilde bir referans sunmaktadır:

- AS(f,t) < 1: Değer normal aralıkta kabul edilmektedir.
- AS(f,t) ≈ 2: Değer ortalamadan anlamlı düzeyde sapmaktadır ve potansiyel anomali olarak değerlendirilebilir.
- AS(f,t) > 3: Değer yüksek düzeyde anormal kabul edilmekte ve ayrıntılı biçimde incelenmesi gerekmektedir.

3.1.5 Risk skoru (RS)

RS bileşeni, bir özelliğin veya işlemin hedef ile (dolandırıcılık, şüpheli bir işlemle ilişkili olma olasılığı veya düşük kredi değerliliği gibi) ilişkilendirilme riskini ifade eden bir skordur. Özelliklerin risk düzeyi, algoritmaya bağlamsal bir ağırlık eklemekte ve finansal uygulamalar için genel olarak anlamlı bir metrik sunmaktadır. Bayes Teoremi tabanlı risk değerlendirmesi Fenton ve Neil (2018) çalışmasından esinlenilmiştir. Bayes Teoremi; dolandırıcılık ve kredi tahmini gibi uygulamalarda açıklanabilirliği, doğruluğu, esnekliği ve önsel bilgi ile veriyi birleştirme yeteneğini geliştirmek amacıyla kullanılmıştır. Hem önsel bilgi hem de yeni verilerle daha doğru tahminler yapılmasına imkân veren Bayes Teoremi, farklı dağılımlara ve durumlara kolayca uyarlanabilmektedir; bu yönüyle dolandırıcılık tespiti ve kredi tahmini gibi alanlarda, özellikle şeffaflık gereksinimlerinin kritik olduğu durumlarda güçlü bir araçtır. Bu yaklaşım, finansal risk skorlama ve kredi temerrüdü tahmini için Bayesçi çıkarımı uygulayan yakın dönem çalışmalarla tutarlıdır. Örneğin, Weng ve ark. (2025) tarafından yapılan çalışma, temerrüde düşen vakaların temerrüde düşmeyen vakalara göre önemli ölçüde daha az olduğu tüketici kredisi verisetlerinde sınıf dengesizliği sorununu ele almaktadır. Bu sorunu hafifletmek için yazarlar, belirsizliği daha iyi yakalamak ve sağlamlığı artırmak amacıyla çoklu olasılıksal modellerin tahminlerini birleştiren Bayesian model averaging (BMA) yöntemini kullanmaktadır. Bayesçi çıkarım çerçevesine Gaussian önseller dâhil edilerek, yaklaşım sınırlı ve dengesiz veri koşullarında tahmin doğruluğunu etkili biçimde artırmaktadır. Sonuçlar, finansal risk skorlama ve kredi temerrüdü tahmini için Bayesçi ve Gaussian modelleme tekniklerinin birleştirilmesinin pratik değerini vurgulamakta ve geleneksel sınıflandırıcılara karşı istatistiksel olarak temellendirilmiş bir alternatif sunmaktadır.

Bu Bayesçi-Gaussian hibrit yaklaşım, Gaussian önsellerle Bayesçi model ortalamasının sınıf dengesizliği altında sağlamlığı artırdığını gösteren Weng ve ark. (2025)'ın çalışmasıyla uyumludur.

Denklem 3.32'deki Bayes Teoremi, RS bileşenini hesaplamak için kullanılmaktadır:

$$P\left(\text{Risk} \,\middle|\, X_f(t), \Delta X_f(t)\right) = \frac{P\left(X_f(t), \Delta X_f(t) | \text{Risk}\right) \cdot P(\text{Risk})}{P\left(X_f(t), \Delta X_f(t)\right)} \tag{3.32}$$

$P(X_f(t), \Delta X_f(t) \mid \text{Risk})$: Risk durumu altında gözlemlenen değerlerin koşullu olasılığını

ifade etmektedir. Bu olasılık genellikle çok değişkenli normal dağılım varsayımı altında Denklem 3.33'teki şekilde modellenmektedir:

$$P\left(X_f(t), \Delta X_f(t) \mid \text{Risk}\right) = \frac{1}{(2\pi)^{k/2}|\Sigma_R|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu_R)^T \Sigma_R^{-1}(x-\mu_R)\right) \quad (3.33)$$

Burada:

- x: Özellik vektörünü ifade etmektedir.
- μ_R: Riskli duruma ait ortalama değer vektörünü göstermektedir.
- Σ_R: Riskli duruma ait kovaryans matrisini ifade etmektedir.
- $|\Sigma_R|$: Riskli durum için kovaryans matrisinin determinantını göstermektedir.
- k: Özellik sayısını ifade etmektedir.
- P(Risk): Risk durumuna ilişkin önsel olasılığı ifade etmektedir.
- $P(X_f(t), \Delta X_f(t))$: Gözlemlenen verinin toplam olasılığını göstermektedir. Bu terim, Bayes yaklaşımında normalleştirme sabiti olarak kullanılmaktadır ve Denklem 3.34'teki şekilde hesaplanmaktadır:

$$P\left(X_f(t), \Delta X_f(t)\right) = P\left(X_f(t), \Delta X_f(t) \mid \text{Risk}\right) \cdot P(\text{Risk}) + P\left(X_f(t), \Delta X_f(t) \mid \neg\text{Risk}\right) \cdot P(\neg\text{Risk}) \quad (3.34)$$

Burada:

- $P(\neg\text{Risk}) = 1 - P(\text{Risk})$: Risksiz duruma ait önsel olasılığı ifade etmektedir.
- $P(X_f(t), \Delta X_f(t) \mid \neg\text{Risk})$: Gözlemlerin risksiz durum altında ortaya çıkma olasılığını göstermektedir.
- $P(X_f(t), \Delta X_f(t) \mid \neg\text{Risk})$: Risksiz durum altında gözlemlenen değerlerin koşullu olasılığını ifade etmektedir. Bu terim de yine çok değişkenli normal dağılım varsayımı altında Denklem 3.35'teki şekilde hesaplanmaktadır:

$$P\left(X_f(t), \Delta X_f(t) \mid \neg\text{Risk}\right) = \frac{1}{(2\pi)^{k/2}|\Sigma_{\neg R}|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu_{\neg R})^T \Sigma_{\neg R}^{-1}(x-\mu_{\neg R})\right) \quad (3.35)$$

Bu denklemde:

- $\mu_{\neg R}$: Risksiz duruma ait ortalama değer vektörünü ifade etmektedir.
- $\Sigma_{\neg R}$: Risksiz duruma ait kovaryans matrisini göstermektedir.
- $|\Sigma_{\neg R}|$: Risksiz duruma ait kovaryans matrisinin determinantını ifade etmektedir.

Risk skorunun nihai formülü Denklem 3.36'daki şekilde yazılabilir:

$$P\left(\text{Risk} \mid X_f(t), \Delta X_f(t)\right) = \frac{\frac{1}{(2\pi)^{k/2}|\Sigma_R|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu_R)^T \Sigma_R^{-1}(x-\mu_R)\right) \cdot P(\text{Risk})}{P\left(X_f(t), \Delta X_f(t)\right)} \quad (3.36)$$

Burada P(X_f(t), Delta X_f(t)) Denklem 3.37'deki şekilde yazılmaktadır:

$$P\left(X_f(t), \Delta X_f(t)\right) = \frac{1}{(2\pi)^{k/2}|\Sigma_R|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu_R)^T \Sigma_R^{-1}(x-\mu_R)\right) \cdot P(\text{Risk}) + \frac{1}{(2\pi)^{k/2}|\Sigma_{\neg R}|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu_{\neg R})^T \Sigma_{\neg R}^{-1}(x-\mu_{\neg R})\right) \cdot P(\neg\text{Risk}) \quad (3.37)$$

Risk modellemesinde Gaussian dağılım varsayımının analitik işlem kolaylığı sağlamak amacıyla yapıldığını belirtmek önemlidir. Ancak gerçek dünya finansal verisetlerinde çarpık ve Gaussian olmayan davranışlar yaygındır. Bu varsayım bir sınırlılık olarak kabul edilmektedir ve SFIX'in gelecekteki sürümlerinde Gaussian karışımları, t-dağılımları veya ampirik tahminler gibi alternatif dağılımların incelenmesi mümkündür.

Algoritma 3.4, Bayes teoreminin finansal risk analizi bağlamında nasıl uygulanabileceğini ayrıntılı biçimde göstermektedir.

Algoritma 3.4. Risk Skorunun hesaplanması

Girdi: Öznitelik vektörü X, risk sınıfı parametreleri μ_R , Σ_R , öncül olasılık P(Risk)
Çıktı: Risk skoru RS

k ← Öznitelik boyutu
X ← [$X_f(t)$, $\Delta X_f(t)$] {Mevcut değer ve değişim}
{Çok değişkenli normal yoğunluk hesaplaması}
diff ← X − μ_R
det_sigma ← Determinant(Σ_R)
inv_sigma ← Ters matris(Σ_R)
quadratic_form ← $diff^T$ × inv_sigma × diff
normalization ← 1 / $((2\pi)^{(k/2)} \times \sqrt{det_sigma})$
P(X|Risk) ← normalization × exp(−0.5 × quadratic_form)
{Marjinal olasılık hesaplaması}
P(X) ← P(X|Risk) × P(Risk) + P(X|Normal) × P(Normal)
{Sonsal olasılık — Bayes teoremi}
RS ← [P(X|Risk) × P(Risk)] / P(X)
return RS

3.1.6 Toplam XAI skoru

Birden fazla bileşenin tek bir formülde birleştirilme nedenlerini incelediğimizde, tekil özelliklerin yeterli olmadığını görmekteyiz. Tek bir özelliğin kullanılması (örneğin yalnızca dinamik önemlilik ya da yalnızca risk skoru) yeterli değildir; model karar mekanizmasının yalnızca bir yönünü açıklamaktadır. Gerçek dünya kararları çoğu zaman çok boyutludur. Bir özelliğin dinamik değişimi, güvenilirliği, zamansal davranışı ve anomalileri, birlikte ele alındığında daha anlamlı hâle gelmektedir. Modern makine öğrenmesi modelleri çoğu zaman karmaşık ilişkilere ve çoklu özelliklere dayanmaktadır. Bu da modelin, birden fazla faktörü bir araya getirerek, karmaşık karar alma sürecini daha şeffaf ve anlaşılır kılmasını gerektirmektedir. Özetle, her bir bileşen aşağıda listelendiği üzere farklı bir soruya yanıt vermektedir:

- DIS: Bir özelliğin zaman içerisinde model kararını ne ölçüde etkilediğini açıklamaktadır.
- CS: Modelin kullandığı verinin güvenilirlik düzeyini değerlendirmektedir.
- PS: İlgili özelliğin zaman içindeki davranış örüntüsünü ortaya koymaktadır.
- AS: Özellikte beklenmeyen ya da olağandışı bir durum bulunup bulunmadığını belirlemektedir.
- RS: İlgili özelliğin risk üzerindeki etkisini değerlendirmektedir.

Sonuç olarak, tüm bileşenlerin birlikte kullanılması modelin daha hassas ve dengeli kararlar vermesini sağlamaktadır.

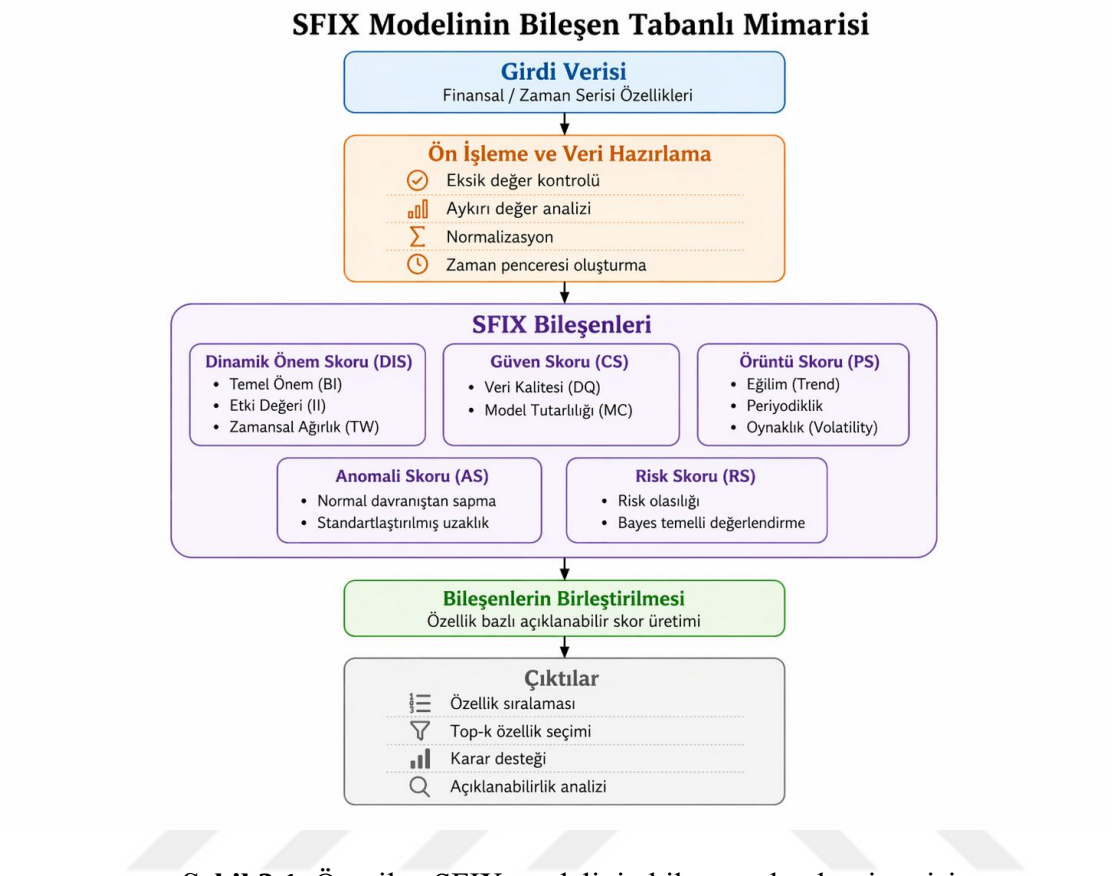
Son olarak, her bir özellik için ayrı ayrı hesaplanan bileşen skorları bir araya getirilmektedir. Toplam XAI skoru Denklem 3.38 ile hesaplanmaktadır:

$$XAI_{Score}(f,t) = DIS(f,t) \cdot CS(f,t) \cdot \big(1 + PS(f,t)\big) \cdot \big(1 + AS(f,t)\big) \cdot RS(f,t) \tag{3.38}$$

Bu formül, algoritmanın tüm bileşenlerini birleştirmekte ve her bir özelliğin açıklanabilirlik skorunu üretmektedir. Denklem (3.38)'de çarpımsal birleştirmenin seçilmesi bilinçli bir tercihtir ve telafi edici olmayan bir tasarım felsefesine dayanmaktadır: bir özellik ancak katkı sağlayan tüm boyutlar (örneğin zamana dayalı önem, veri kalitesi, anomali davranışı, risk etkisi) yeterince güçlü olduğunda yorumlanabilir kabul edilmelidir. Çarpma işlemi, herhangi bir bileşendeki zayıf sinyallerin toplam skoru azaltmasına olanak tanımakta; bu da yorumlanabilirliğin birden çok boyutta sağlamlığı yansıtması gereken yüksek riskli finansal kararlarda arzu edilen bir durumdur. Toplamsal ya da ağırlıklı toplam yaklaşımları da mümkün olmakla birlikte, büyük bir zayıflık içeren özellikleri temkinli biçimde elemek amacıyla çarpımsal bütünleştirme tercih edilmiştir. Gelecek çalışmalarda, uygulamaya özgü risk toleransına dayalı alternatif birleştirme fonksiyonları incelenebilir.

Önerilen SFIX modelinin genel açıklanabilirlik haritası Şekil 3.1'de gösterilmektedir. SFIX modelinin bileşen tabanlı mimarisini göstermektedir. Süreç, finansal ve zamansal özellikleri içeren girdi verisinin alınmasıyla başlamakta; ön işleme aşamasında eksik değer, aykırı değer, normalizasyon ve zaman penceresi gibi hazırlık işlemleri uygulanmaktadır. Ardından SFIX modeli, veriyi beş temel açıklanabilirlik bileşeni üzerinden değerlendirmektedir: Dinamik Önem Skoru (DIS), Güven Skoru (CS), Örüntü Skoru (PS), Anomali Skoru (AS) ve Risk Skoru (RS). DIS bileşeni temel önem, etki düzeyi ve zamansal ağırlığı birlikte dikkate alırken; CS veri kalitesi ve model tutarlılığına dayalı güven değerlendirmesi sunmaktadır. PS bileşeni, eğilim, periyodiklik ve oynaklık gibi zaman serisi örüntülerini analiz etmektedir. AS, normal davranıştan sapmaları belirleyerek aykırılık düzeyini ölçerken; RS ise risk olasılığını Bayes temelli bir yaklaşımla hesaplamaktadır. Bu bileşenlerden elde edilen skorlar bütünleştirilerek nihai açıklanabilirlik çıktıları üretilmekte; böylece özellik sıralaması, önemli özelliklerin

seçimi ve karar destek süreçleri açıklanabilir bir yapıda sunulmaktadır.



Şekil 3.1. Önerilen SFIX modelinin bileşen tabanlı mimarisi

3.1.7 Parametreler

Bu bölümde, SFIX XAI modelini oluşturan bileşenlerin denklemlerine ait parametre ve hiperparametrelerin ayrıntıları verilmektedir. Önerilen yöntemin optimizasyonunda, hiperparametre bağımlılığını en aza indirmeye yönelik bir çaba gösterilmiştir.

DIS(f,t), Tablo 3.1'de verilen parametreler ve hiperparametrelerden oluşmaktadır. λ ve γ, üstel yapı deneysel olarak veya grid search ile ayarlanabilir. Periyot süresi T, verisetinin mevsimsel veya döngüsel özelliklerine göre manuel ya da otomatik olarak seçilmektedir. Hiperparametre β(j) için korelasyon tabanlı veya olasılık tabanlı veriseti temelli hesaplamalar tercih edilmiştir. Zaman pencereleri arasındaki önemi dengelemek için α(w) değeri deneysel olarak ya da çapraz doğrulama ile belirlenmektedir.

Tablo 3.1. DIS bileşeninin parametreleri

Bileşen	Parametreler	Hiper parametreler	Açıklama
DIS(f,t)	W, BI(f,w), II(f,w), TW(t,w)	α(w)	Özelliğin zamana bağlı önem katsayısını hesaplar.
BI(f,w)	corr(Xf, yw), p_value(f,w)	Hayır	Özelliğin temel önemini belirler.
II(f,w)	K(f,i,w), D(f,i,w)	β(j)	Özellikler arasındaki etkileşimlerin önemini ölçer.
TW(t,w)	t, w	λ, γ, T	Zaman uzaklığına dayalı ağırlık katsayısını hesaplar.
K(f,i,w)	C, D	Hayır	Özellikler arasındaki ilişkinin matematiksel ifadesi.
D(f,i,w)	K(i,j,w), β(j)	Hayır	Etkileşim bağının ağırlıklı değerini hesaplar.
W	Hayır	Hayır	Zaman penceresi uzunluğu.
α(w)	Hayır	Evet	Her pencere için ağırlık katsayısı.

CS, Tablo 3.2'de yer alan parametrelerden oluşmaktadır ve herhangi bir hiperparametre içermediği için ek ayar gerektirmemektedir.

Tablo 3.2. CS bileşeninin parametreleri

Bileşen	Parametreler	Hiper parametreler	Açıklama
CS(f)	DQ(f), MC(f)	Hayır	Güven skorunun toplam bileşenlerini birleştirir.
DQ(f)	MR, OR, CV	Hayır	Eksik ve hatalı veri oranı ile varyasyona dayalı doğruluk katsayısıdır.
MC(f)	ρ(f,y), σy	Hayır	Korelasyon ve hedef değişken varyansına dayalı model uyum katsayısıdır.
MR, OR	Hayır	Hayır	Yalnızca veri analizi ile belirlenir.
CV	σf, μf	Hayır	Standart sapma ve ortalamaya dayalı olarak hesaplanır.
ρ(f,y)	Hayır	Hayır	Özellik ile hedef arasındaki ilişkiyi ölçer.
σy	Hayır	Hayır	Hedef değişkenin standart sapmasıdır.

PS, Tablo 3.3'te verilen parametreler ve hiper parametrelerden oluşmaktadır. Her bir zamansal bileşenin (trend, periyodiklik, oynaklık) etkisini belirlemek için η_k, regresyon katsayıları olarak elde edilebilir. Ancak bu bileşenlerin genel katkıları eşit önemde kabul edildiğinden, η_k hiper parametresi 1 olarak sabitlenmiştir. Fourier serisindeki harmonik bileşen sayısı n, veri analizi veya model performansına göre belirlenmektedir. Hareketli pencere uzunluğu k, verinin kısa dönemli değişkenliğini yakalamak amacıyla ampirik olarak ayarlanmaktadır. Zamanla değişen trendin eğimi β_1 ve başlangıç değeri β_0, regresyon ile öğrenilmektedir.

Tablo 3.3. PS bileşeninin parametreleri

Bileşen	Parametreler	Hiper parametreler	Açıklama
PS(f,t)	$P_k(f,t)$, K	η_k	Zamansal örüntü analizinin toplam bileşenlerini birleştirir.
Trend (P1)	β_1, β_0, t	Hayır	Zamanla doğrusal artışı ifade eder.
Periyodiklik	a_0, a_k, b_k, T, n, t	T, n	Periyodik dalgalanmaları modellemek için Fourier serisine dayanır.
Volatilite	$X_f(t)$, μ_f, k, w	k	Hareketli penceredeki sapmayı ölçer.

AS, Tablo 3.4'te verilen parametreler ve hiper parametrelerden oluşmaktadır. Burada yalnızca hareketli pencere uzunluğu W bir hiper parametre olarak yer almaktadır. Bu değer manuel olarak girilen genel bir değerdir ve hesaplama gerektirmez.

Tablo 3.4. AS bileşeninin parametreleri

Bileşen	Parametreler	Hiper parametreler	Açıklama
AS(f,t)	Xf(t), μf(t), σf(t)	W	Belirli bir özellik değerinin ortalamadan sapmasını ölçer.
Xf(t)	Hayır	Hayır	Özelliğin t anındaki değeri.
μf(t)	Hayır	Hayır	Özelliğin t anına kadar olan hareketli ortalaması.
σf(t)	Hayır	Hayır	Özelliğin t anına kadar olan hareketli standart sapması.
W	Hayır	Evet	Hareketli pencere uzunluğu.

RS, Tablo 3.5'te yer alan parametrelerden oluşmaktadır ve hiper parametre içermediği için herhangi bir ek ayar gerektirmemektedir.

Tablo 3.5. RS bileşeninin parametreleri

Bileşen	Parametreler	Hiper parametreler	Açıklama
RS(f,t)	$P(X_f(t), \Delta X_f(t) \mid Risk)$, P(Risk)	Hayır	Bayes Teoremi kullanarak risk olasılığını hesaplar.
$P(X_f(t), \Delta Xf(t) \mid Risk)$	$X_f(t)$, $\Delta X_f(t)$, μ_R, Σ_R	Hayır	Riskli durumdaki gözlemlerin olasılık yoğunluğunu ifade eder.
P(Risk)	Hayır	Hayır	Riskli durumun önsel olasılığıdır.
$P(X_f(t), \Delta X_f(t))$	Hayır	Hayır	Normalizasyon sabiti olup gözlemin toplam olasılığını ifade eder.
μ_R	Hayır	Hayır	Riskli durum için ortalama vektörüdür.
Σ_R	Hayır	Hayır	Riskli durum için kovaryans matrisidir.

SFIX, hiper parametrelere aşırı bağımlılığı azaltmak amacıyla modüler bileşenlerle tasarlanmış olsa da, λ (zamansal denge), γ (azalma oranı) ve β(j) (özellik etkileşim ağırlıkları) gibi bazı parametreler alana özgü davranışı yakalamak açısından kritik öneme sahiptir. Uygulamada, λ ve γ, ya ampirik analiz yoluyla (örneğin dengeli bir varsayılan değer olarak λ=0.5, γ için 0.1–1 aralığı) ya da doğrulama kümeleri üzerinde grid search

ile, üstel ve periyodik etkiler arasında denge kuracak şekilde seçilebilir. Daha önce belirtildiği gibi, β(j) katsayısı; özellik-hedef ilişkilerinin doğrusal yapısına bağlı olarak korelasyon tabanlı ya da mutual information tabanlı olmak üzere iki esnek seçenekle sunulmaktadır.

3.1.8 Sadeleştirilmiş SFIX

Önerilen model, analizine zaman boyutunu dâhil ederek kapsamlı bir değerlendirme yapmayı amaçlamaktadır. Bu nedenle, zaman boyutunun yönteme dâhil edilmesi daha güvenilir ve daha hassas sonuçlar sağlamaktadır. Ancak burada en önemli sınırlılık, eğitilecek verisetinin her zaman aynı özelliklere sahip olmamasıdır. Önerilen yöntemin, bazı verisetlerinde zaman çerçevesi bulunmadığı durumlara uyarlanması gerekmektedir. Bu nedenle, bu bölümde, zaman çerçevesi bulunmayan verileri eğitmek amacıyla önerilen yöntemin zamana dayalı terimleri çıkarılmakta ve statik bir verisetine uygulanabilir formüller yeniden oluşturulmaktadır. Bu durumda, zamansal parametreler denklemlerden çıkarılmaktadır.

DIS için, zamansal pencere tabanlı toplam terimi ($\sum_{w=1}^{W}$) ile zamansal ağırlıklandırma bileşeni TW(t,w) modelden çıkarılmıştır. Bu doğrultuda, denklemin son hâli Denklem 3.39'da sunulmaktadır.

$$DIS(f) = BI(f) \cdot II(f) \tag{3.39}$$

CS formülü değişmemektedir; çünkü CS içinde zamana bağlı herhangi bir bileşen bulunmamaktadır.

Zaman boyutu bulunmayan bir verisetinde, zamana ilişkin analizlerin kaldırılması gerektiğinden PS toplam skordan çıkarılmaktadır.

AS için, hareketli ortalama ve hareketli standart sapma yerine tüm veriseti üzerindeki genel ortalama ve standart sapma kullanılmaktadır. Denklem 3.40'da, her bir gözlem için aykırılık derecesini hesaplanmaktadır.

$$AS(f) = 1 - \frac{\sum_{i=1}^{n} \mathbb{1}(|z_i|>3)}{n} \tag{3.40}$$

Burada z_i, her bir gözlem için z-skorunu ifade etmektedir.

- $\mathbb{1}(|z_i| > 3)$: Bu bir gösterge fonksiyonudur.

- Eğer $|z_i| > 3$ ise 1 döndürür.
- Eğer $|z_i| \leq 3$ ise 0 döndürür.
- Bir değerin aykırı değer olup olmadığını kontrol eder.

- $\sum_{i=1}^{n}$: Tüm veri noktaları için toplamı alır.
  - Her aykırı değer için 1 toplar.
  - Normal değerler için 0 toplar.
  - Sonuç olarak toplam aykırı değer sayısını verir.
- (... / n): Aykırı değer oranını hesaplar.
  - Toplam aykırı değer sayısını n'e, yani toplam veri sayısına böler.
  - Böylece 0 ile 1 arasında bir oran elde edilir.
- 1 − ...: Son aşamada bu oran 1'den çıkarılır.
  - Hiç aykırı değer yoksa sonuç 1 olur; bu en iyi durumu ifade eder.
  - Tüm değerler aykırıysa sonuç 0 olur; bu en kötü durumu ifade eder.

Değişim oranı terimi olan $\Delta X_f(t)$ çıkarılmakta ve Bayesyen risk bileşeni Denklem 3.41'deki hale gelmektedir:

$$P(Risk|X_f) = \frac{P(X_f|Risk)\cdot P(Risk)}{P(X_f)} \tag{3.41}$$

Bunun dışında olasılıksal formülasyon aynı kalmaktadır.

Zaman boyutunun kaldırılması sonucunda yapılan değişiklikler Tablo 3.6'da özetlenmektedir.

Tablo 3.6. Zamansız formüller ve değişiklikler

Bileşen	Zaman Olmadan Formül	Değişiklikler
DIS(f)	BI(f) · II(f)	Zamansal bileşenler (w, TW) kaldırılmıştır.
CS(f)	DQ(f) · MC(f)	Değişiklik yoktur.
PS(f)	Yok	Zaman kullanılmadığı için tamamen kaldırılmıştır.
AS(f)	$\frac{\lvert X_f - \mu_f \rvert}{\sigma_f}$	Hareketli istatistikler yerine genel ortalama ve standart sapma kullanılmıştır.
RS(f)	$\frac{P(X_f \mid \text{Risk}) \cdot P(\text{Risk})}{P(X_f)}$	Zamana bağlı değişim oranı ($\Delta X_f(t)$) kaldırılmıştır.

Sonuç olarak, nihai basitleştirilmiş SFIX modeli Denklem 3.42 ile hesaplanmaktır:

$$S - SFIX = DIS(f) \cdot CS(f) \cdot \big(1 + AS(f)\big) \cdot RS(f) \tag{3.42}$$

3.1.9 Önerilen modelin zaman karmaşıklığı

Zaman karmaşıklığı, bir algoritmanın çalışma süresini girdi boyutunun bir fonksiyonu olarak ifade eden bir metriktir. Bu, bir algoritmanın ne kadar verimli olduğunu anlamak açısından kritik öneme sahiptir. Bu bölüm, önerilen SFIX modelinin zaman karmaşıklığına odaklanmaktadır.

Her bir bileşenin zaman karmaşıklığını hesaplamak için, formüllerin içindeki döngüler, toplam işlem sayısı ve hesaplama yoğunlukları dikkate alınmıştır. Önerilen modelin beş bileşenine ilişkin zaman karmaşıklığı analizleri alt başlıklarda açıklanmaktadır.

3.1.9.1 Zaman karmaşıklığı hesaplaması

Dinamik Özellik Skoru (DIS(f,t)) için zaman karmaşıklığı hesaplamaları Denklem 3.44'de verilmektedir.

$$DIS(f,t) = \sum_{w=1}^{W} \alpha(w) \cdot [BI(f,w) + II(f,w)] \cdot TW(t,w) \tag{3.43}$$

$$O(DIS) = W \cdot O(BI + II + TW) \tag{3.44}$$

- BI, sabit zamanlı bir hesaplama olduğundan O(1) karmaşıklığına sahiptir.
- II bileşeni, ilgili özelliğin diğer özelliklerle olan etkileşimlerinin hesaplanmasını gerektirdiğinden O(N) karmaşıklıkla değerlendirilmektedir.
- TW ise sabit zamanlı hesaplandığı için O(1) karmaşıklığına sahiptir.

Bu nedenle DIS(f,t) bileşeninin her bir zaman penceresi için baskın maliyeti O(N) olmakta, toplamda W adet zaman penceresi dikkate alındığında genel zaman karmaşıklığı Denklem 3.45'teki şekilde elde edilmektedir:

$$O(DIS) = W \cdot O(N) = O(W \cdot N) \tag{3.45}$$

CS, veri kalitesi ve model tutarlılığı bileşenlerinden oluşmaktadır:

$$CS(f) = DQ(f) \cdot MC(f) \tag{3.46}$$

$$\mathrm{DQ(f)} = (1 - \mathrm{MR}) \cdot (1 - \mathrm{OR}) \cdot \frac{1}{1+\mathrm{CV}} \tag{3.47}$$

MR, OR ve CV gibi ölçütler veri istatistiklerinden doğrudan hesaplandığı için bu bileşenin hesaplama maliyeti O(1) olarak kabul edilmektedir. MC(f) bileşeni ise

$$MC(f) = |\rho(f, y)| \cdot \frac{1}{1+\sigma_y} \tag{3.48}$$

Denklem 3.48'deki şeklindedir. Burada ρ(f,y), özellik ile hedef değişken arasındaki Pearson korelasyon katsayısını göstermekte olup O(N) karmaşıklıkta hesaplanmaktadır. N, toplam veri noktası sayısını ifade etmektedir. Dolayısıyla CS(f) için genel zaman karmaşıklığı Denklem 3.49'deki şekilde verilmektedir:

$$O(CS) = O(DQ) + O(MC) = O(N) \tag{3.49}$$

PS bileşeni, Denklem 3.50'de yer aldığı gibi farklı örüntü türlerinin birleşiminden oluşmaktadır:

$$PS(f, t) = \sum_{k=1}^{K} \eta_k \cdot P_k(f, t) \tag{3.50}$$

Bu yapı içinde trend bileşeni $P_1(f,t)$ doğrusal katsayılar üzerinden hesaplandığından O(1) karmaşıklığa sahiptir. Periyodiklik bileşeni Fourier serisi kullanılarak hesaplanmakta olup n harmonik sayısına bağlı olarak O(n) karmaşıklık gerektirmektedir. Volatilite bileşeni ise hareketli pencere yaklaşımına dayandığından pencere uzunluğu k ile ilişkili olarak O(k) karmaşıklığa sahiptir. Buna göre PS(f,t) için toplam zaman karmaşıklığı Denklem 3.51'deki şekilde ifade edilmektedir:

$$O(PS) = K \cdot \big(O(1) + O(n) + O(k)\big) = O\big(K \cdot (n + k)\big) \tag{3.51}$$

AS(f,t), Denklem 3.52'de hesaplanmaktadır:

$$AS(f, t) = \frac{|X_f(t) - \mu_f(t)|}{\sigma_f(t)} \tag{3.52}$$

Burada hareketli ortalama Denklem 3.53'te

$$\mu_f(t) = \frac{1}{W} \sum_{i=t-W+1}^{t} X_f(i) \tag{3.53}$$

ve hareketli standart sapma Denklem 3.54'te

$$\sigma_f(t) = \sqrt{\frac{1}{W} \sum_{i=t-W+1}^{t} \Big(X_f(i) - \mu_f(t)\Big)^2} \tag{3.54}$$

şeklinde tanımlanmaktadır. Hem ortalama hem de standart sapma, uzunluğu W olan bir pencere üzerinde hesaplandığından AS(f,t) bileşeninin zaman karmaşıklığı O(W) olmaktadır:

$$O(AS) = O(W) \tag{3.55}$$

RS(f,t), Bayes teoremi temelinde Denklem 3.56'deki biçimde hesaplanmaktadır:

$$P\left(Risk \mid X_f(t), \Delta X_f(t)\right) = \frac{P\left(X_f(t), \Delta X_f(t) | Risk\right) \cdot P(Risk)}{P\left(X_f(t), \Delta X_f(t)\right)} \tag{3.56}$$

Bu yapıda olabilirlik terimi $P(X_f(t), \Delta X_f(t) \mid Risk)$ çok değişkenli Gaussian yoğunluğu ile hesaplanmakta ve $O(N^2)$ karmaşıklığa sahip olmaktadır. Benzer şekilde, normalize edici sabit olan $P(X_f(t), \Delta X_f(t))$ de tüm veri noktaları dikkate alınarak hesaplandığı için $O(N^2)$ karmaşıklık gerektirmektedir. Bu nedenle RS(f,t) bileşeninin genel zaman karmaşıklığı Denklem 3.57'deki şekilde verilmektedir:

$$O(RS) = O(N^2) \tag{3.57}$$

Denklemin bileşenleri için tüm hesaplamalar sonucunda elde edilen karmaşıklığın genel gösterimi Tablo 3.7'de özetlenmiştir:

Tablo 3.7. Model bileşenlerinin zaman karmaşıklığı özeti

Bileşen	Zaman Karmaşıklığı	Açıklama
DIS(f,t)	$O(W \cdot N)$	Zaman pencereleri ve özellik etkileşimleri üzerinden hesaplanmaktadır.
CS(f)	$O(N)$	Korelasyon temelli model tutarlılığı veri boyutuna bağlıdır.
PS(f,t)	$O(K \cdot (n+k))$	Trend, periyodiklik ve volatilite bileşenlerini içermektedir.
AS(f,t)	$O(W)$	Hareketli pencere temelli ortalama ve standart sapma hesaplarına dayanmaktadır.
RS(f,t)	$O(N^2)$	Bayes teoremi ve Gaussian yoğunluk hesaplamaları nedeniyle daha yüksek maliyetlidir.

3.1.9.2 Model karmaşıklığının değerlendirilmesi

Önerilen modelin karmaşıklığının kabul edilebilir olup olmadığı; verisetinin boyutu, uygulama alanı ve hesaplama kaynakları gibi faktörlere bağlıdır. Bu bölüm, her bir bileşenin karmaşıklığını değerlendirmekte ve kabul edilebilirliğini analiz etmektedir.

Dinamik Özellik Önemi (DIS(f,t)) bileşeni için, W ve N değerlerinin küçük ya da orta ölçekli olduğu durumlarda, örneğin $W<100$ ve $N<1000$ olduğunda, elde edilen $O(W \cdot N)$

zaman karmaşıklığı genel olarak kabul edilebilir düzeydedir. Ancak daha büyük verisetlerinde, özellikle $W \cdot N \gg 10^6$ olduğunda, hesaplama maliyetini azaltmak amacıyla paralel işlem, örnekleme ya da çeşitli optimizasyon tekniklerinin kullanılması gerekli hâle gelebilmektedir.

CS bileşeni için elde edilen O(N) zaman karmaşıklığı, genel olarak düşük maliyetli bir yapı sunmaktadır. Bu nedenle, büyük verisetlerinde dahi hesaplama açısından kabul edilebilir olup, modelin en hafif bileşenlerinden biri olarak değerlendirilmektedir.

Temporal Pattern Analysis (PS(f,t)) bileşeni, K, n ve k parametrelerinin orta büyüklükte olduğu durumlarda, örneğin $K < 10$, $n < 20$ ve $k < 100$ olduğunda, kabul edilebilir bir hesaplama maliyetine sahiptir. Bununla birlikte, veri hacmi ve örüntü karmaşıklığı arttıkça bu bileşen daha maliyetli hâle gelebilmektedir. Özellikle Fourier analizi ve oynaklık hesaplamaları içeren alt bileşenler için optimizasyon uygulanması gerekebilir.

AS bileşeni genel olarak düşük hesaplama maliyetine sahiptir. Özellikle pencere uzunluğu W makul düzeylerde tutulduğunda, örneğin $W < 1000$ olduğunda, bu bileşenin hesaplanması pratik açıdan kabul edilebilir olmaktadır.

RS bileşeni için elde edilen $O(N^2)$ zaman karmaşıklığı, büyük verisetlerinde önemli bir hesaplama yükü oluşturabilmektedir. Özellikle $N > 10^4$ olduğunda bu maliyet belirgin hâle gelmektedir. Buna karşılık, daha küçük verisetlerinde, örneğin $N < 1000$ olduğunda, bu karmaşıklık genellikle kabul edilebilir düzeydedir. Ancak Gaussian yoğunluk fonksiyonlarının ve normalleştirme sabitinin hesaplanmasında ek optimizasyon stratejilerine ihtiyaç duyulabilir.

Model bileşenlerinin zaman karmaşıklığının kabul edilebilirliği Tablo 3.8'de özetlenmiştir.

Tablo 3.8. Model bileşenlerinin zaman karmaşıklığı değerlendirmesi

Bileşen	Zaman Karmaşıklığı	Kabul edilebilirlik
DIS(f,t)	$O(W \cdot N)$	Orta büyüklükteki verisetleri için kabul edilebilirdir; büyük verisetlerinde paralel işlem veya ek optimizasyon gerekebilir.
CS(f)	$O(N)$	Düşük maliyetlidir ve genel olarak farklı veriseti boyutları için kabul edilebilirdir.
PS(f,t)	$O(K \cdot (n + k))$	Orta büyüklükteki verisetleri için kabul edilebilirdir; özellikle periyodiklik ve oynaklık hesaplamalarında optimizasyon yararlı olabilir.
AS(f,t)	$O(W)$	Hafif hesaplama maliyetine sahip bir bileşendir ve genel olarak kabul edilebilirdir.
RS(f,t)	$O(N^2)$	Küçük verisetleri için kabul edilebilirdir; büyük verisetlerinde optimize edilmesi gerekmektedir.

3.1.10 Modelin özgünlüğünün değerlendirilmesi

DIS(f,t) içerisinde kullanılan BI(f,w), II(f,w) ve TW(t,w) bileşenlerine ilişkin genel değerlendirmeler şu şekildedir:

- BI(f,w): Özelliğin temel önem katsayısını ifade etmektedir. Benzer önem ölçümleri SHAP ve LIME gibi yöntemlerde de yer almaktadır.
- II(f,w): Özellikler arasındaki etkileşim katsayısını göstermektedir. Etkileşim temelli değerlendirmeler SHAP yaklaşımında da kullanılmakla birlikte, bu yöntemlerde zamansal pencere yapısı dikkate alınmamaktadır.
- TW(t,w): Zamana bağlı ağırlıklandırma bileşenini ifade etmektedir. Bu tür bir zamansal ağırlıklandırma yaygın XAI yöntemlerinde doğrudan yer almamakta olup, yöntemin özgün katkılarından biri olarak değerlendirilebilir. Özellikle zamansal ağırlıklandırmanın zaman pencereleri üzerinden toplama yaklaşımıyla birlikte kullanılması, bu yapıyı SHAP ve LIME gibi yöntemlerden ayırmaktadır.

CS içerisinde kullanılan DQ(f) ve MC(f) bileşenlerine ilişkin genel değerlendirmeler ise aşağıdaki gibidir:

- DQ(f): Veri kalitesini, eksik veri oranı, hatalı gözlem oranı ve varyasyon katsayısı gibi ölçütler üzerinden değerlendirmektedir. Veri kalitesinin sistematik biçimde ele alınmasına yönelik benzer yaklaşımlar, örneğin Gebru ve ark. tarafından önerilen Datasheets for Datasets çalışmasında da vurgulanmıştır.
- MC(f): Pearson korelasyon katsayısına dayalı bir uyum bileşenidir. Pearson korelasyonu istatistikte yaygın olarak kullanılan bir ölçü olmakla birlikte,

doğrudan bir CS bileşeni olarak kullanımı sınırlıdır.

- CS(f): Veri kalitesi ile hedef değişken uyumunu bir araya getirerek birleşik bir güven skoru üretmektedir. Bu bağlamda DQ(f) ve MC(f) bileşenlerinin birlikte ağırlıklandırılması, yöntemin özgün yönlerinden biri olarak değerlendirilebilir.

PS(f,t) kapsamında kullanılan Trend ($P_1(f,t)$), Periyodiklik ($P_2(f,t)$) ve Volatilite ($P_3(f,t)$) alt bileşenleri tek başlarına özgün değildir. Bununla birlikte, trend, periyodiklik ve oynaklığın aynı çatı altında birlikte değerlendirilmesi özgün bir yaklaşım sunmaktadır. Bu üç bileşenin zamansal analiz bağlamında bütünleşik biçimde kullanılması, yöntemin ayırt edici özelliklerinden biridir.

AS(f,t), istatistikte yaygın olarak kullanılan z-skoru yaklaşımına benzemektedir. Ancak bu yapıda hareketli ortalama $\mu_f(t)$ ve hareketli standart sapma $\sigma_f(t)$ kullanılarak daha dinamik bir değerlendirme yapılmaktadır. Dolayısıyla, hareketli pencere temelli uygulama biçimi yönteme bağlama özgü bir nitelik kazandırmaktadır.

RS(f,t) için kullanılan formülasyon, Bayes Teoremi ve Gaussian dağılımına dayanmaktadır; bu yapılar risk modellemesinde yaygın olarak kullanılmaktadır. Bununla birlikte, $\Delta X_f(t)$ gibi zaman içindeki değişim oranını yansıtan ek bir bileşenin modele dahil edilmesi yaygın değildir. Bu nedenle $\Delta X_f(t)$'nin kullanımı, yönteme bağlama özgü ve ayırt edici bir katkı sağlamaktadır.

SFIX içinde kullanılan matematiksel bileşenler—z-skoru tabanlı anomali tespiti, MI ve Fourier tabanlı periyodiklik gibi literatürde yaygın olarak bilinse de, modelin özgünlüğü bunların finansal uygulamalara yönelik bağlamsal uyarlamasından ve yapısal entegrasyonundan kaynaklanmaktadır. SFIX, kredi değerliliğindeki zaman temelli değişimler gibi zamansal risk dalgalanmaları, davranıştaki mevsimsellik (örneğin harcama döngüleri), dolandırıcılık için anomali tespiti ve düşük veri kalitesinden kaynaklanan belirsizlik gibi finans alanına özgü temel zorlukları ele almak üzere açıkça tasarlanmıştır. CS, finansal verisetlerinde yaygın olan eksik veri oranları ve varyans gibi alana özgü göstergelerden yararlanmaktadır. Dahası, risk bileşeni Bayesyen olasılığı kredi temerrüdü veya dolandırıcılık tespiti gibi finansal olaylara uyarlanmış biçimde işlevselleştirmektedir. Bu tasarım kararları bir araya geldiğinde, bileşenlerin kendileri genel amaçlı olsa da, bunların formülasyonu ve birleşimi özellikle finansal karar verme sistemlerinde açıklanabilirlik için kurgulanmıştır.

Sonuç olarak, önerilen modelde alt bileşenler ve verisetini oluşturan özellikler çeşitli açılardan bağlama özgü biçimde değerlendirilerek modelin açıklanabilirliği için özgün bir yaklaşım hedeflenmiştir.

3.1.11 Finansal uygulamalar

SFIX'in mimari bileşenleri, açıklanabilirlik, risk duyarlılığı ve veri güvenilirliğinin büyük önem taşıdığı finansal karar verme sistemlerinde karşılaşılan belirli zorlukları ele almak üzere bilinçli olarak tasarlanmıştır. DIS bileşeni, özellik öneminin zamansal evrimini yakalamaktadır; bu zamansallık, kredi değerliliğindeki bozulma veya ortaya çıkan dolandırıcılık örüntüleri gibi müşteri profillerindeki kademeli davranış değişimlerinin tespiti açısından kritik öneme sahiptir. Bu durum, özellikle kredi izleme veya anomali tespiti gibi zamana duyarlı bağlamlarda değerlidir; çünkü statik açıklamalar yanıltıcı olabilmektedir. CS, hem giriş özelliklerinin istatistiksel kalitesini (örneğin eksik değer oranları, aykırı değer sıklıkları) hem de model tahmin tutarlılığını değerlendirmekte, böylece özellikle kredi risk skorlama ve düzenleyici denetimlerde yaygın olan düşük kaliteli, eksik veya kararsız finansal veri sorununu doğrudan ele almaktadır. PS, özellik davranışındaki periyodiklik ve oynaklığı belirlemektedir; bu, tüketici finansmanı gibi, döngüsel gelir örüntülerinin veya mevsimsel harcama eğilimlerinin yanlı değerlendirmelerden kaçınmak için doğru biçimde yorumlanması gereken alanlarda büyük önem taşımaktadır. AS, özellik değerlerindeki nadir veya uç sapmaları işaretleyerek, anormal davranışın erken tespitinin kritik olduğu dolandırıcılık tespiti gibi uygulamaları desteklemektedir. Son olarak, RS, gözlemlenen özellik davranışına bağlı olarak finansal risk olasılığını nicel olarak belirlemek için Bayesyen akıl yürütme uygulamakta; böylece kredi onayları, sigorta poliçesi değerlendirmesi veya varlık tahsisi gibi yüksek riskli kararlarla uyum göstermektedir.

Bu bileşenler birlikte ele alındığında, SFIX açıklamalarının yalnızca teknik açıdan sağlam değil, aynı zamanda anlamsal olarak da anlamlı ve gerçek dünya finansal senaryolarına doğrudan uygulanabilir olmasını sağlamaktadır. Model yorumlanabilirliğini uyum, adalet ve risk farkındalığı gibi alana özgü kaygılarla hizalayarak SFIX, makine öğrenmesi şeffaflığı ile finans profesyonellerinin operasyonel ihtiyaçları arasındaki boşluğu kapatmaktadır.

3.2 Önerilen modelin uygulanması

Bu bölüm, önerilen SFIX modelinin özellik ilgililiğini belirleme açısından nicel değerlendirmesini sunmaktadır. SFIX'in performansı, kümülatif önem dağılımı, özellik bazında tutarlılık ve bileşen bazında katkı ayrıştırması dâhil olmak üzere hem görsel hem de istatistiksel analizlerle değerlendirilmektedir. Özellikle, SFIX'in açıklayıcı ağırlığı özellikler arasında nasıl dağıttığını, katkıları yoğunlaştırma yeteneğini (sadelik/parsimony) ve yeniden örneklemeler (bootstrap örneklemeleri) boyunca kararlılığını (sağlamlık/robustness) incelenmektedir.

3.2.1 SFIX performans değerlendirmesi

Bu bölümde, önerilen SFIX modelinin nasıl çalıştığını açıklamaktayız. Finans veriseti üzerindeki uygulamaya ilişkin süreçler dâhil edilmiştir.

3.2.1.1 Finans Verisetinin Tanıtımı

Finans alanındaki veriler kişisel verilerden oluştuğu için veri gizliliği ve güvenliği, ihlallere yol açma olasılığı nedeniyle açık biçimde bulunması oldukça zor alanlardır. Bu nedenle modelleri gerçek verisetleri üzerinde test etmek mümkün değildir. Bununla birlikte, modelin genel işleyişini test etmek amacıyla birçok makalede de kullanılan Kaggle veriseti tercih edilmiştir.

Veriseti olarak Sharma (2023) kaynağında yer alan Kaggle Loan Approval Prediction veriseti (kredi başvuru veriseti) tercih edilmiştir. Bu verisetinin özelliklerini içeren Tablo 3.9 aşağıda verilmiştir:

Tablo 3.9. Kredi başvuru veriseti açıklaması

Özellik	Açıklama
loan_id	Kredi başvurusunu tanımlayan benzersiz kimlik değeri
no_of_dependents	Başvuru sahibinin bakmakla yükümlü olduğu kişi sayısı
education	Başvuru sahibinin eğitim seviyesini gösterir
self_employed	Başvuru sahibinin serbest çalışma durumunu gösterir
income_annum	Başvuru sahibinin yıllık gelirini ifade eder
loan_amount	Talep edilen kredi miktarını ifade eder
loan_term	Kredinin yıl cinsinden vadesini gösterir
cibil_score	Başvuru sahibinin kredi değerliliğini gösteren kredi puanıdır
residential_assets_value	Başvuru sahibinin konut varlıklarının toplam değeridir
commercial_assets_value	Başvuru sahibine ait ticari varlıkların toplam değeridir
luxury_assets_value	Başvuru sahibine ait lüks varlıkların toplam değeridir
bank_assets_value	Başvuru sahibinin banka varlıklarının toplam değeridir

Kaggle kredi başvuru veriseti, gerçekçi finansal bağlamı, kamuya açık erişilebilirliği ve açıklanabilir yapay zekâ araştırmalarında yaygın kullanımı nedeniyle bu çalışma için seçilmiştir. *Approved* ve *Rejected* sonuçlarına dayanan ikili sınıflandırma yapısı sunmakta olup, bu durum finansal risk değerlendirmesi ve karar destek sistemleri açısından oldukça ilgilidir.

Veriseti, başvuru sahibine ilişkin finansal nitelikler ile kredi koşullarını kapsayan 4269 örnek ve 12 özellikten oluşmaktadır. Hedef değişken (*loan_status*) orta düzeyde sınıf dengesizliği göstermektedir; örneklerin %62.22'si onaylanmış, %37.78'i reddedilmiştir. Bu doğal dengesizlik, gerçek dünya kredi verme senaryolarını yansıtmakta ve hem çoğunluk hem de azınlık sınıfları üzerinde özellik açıklanabilirliğinin değerlendirilmesine engel teşkil etmemektedir.

Veriseti temiz, iyi yapılandırılmış ve tekrarlanabilir deneyleri destekleyecek nitelikte olup önerilen S-SFIX çerçevesinin performansını değerlendirmek için uygun bulunmuştur.

Bu çalışmada kullanılan veriseti için eksik değerlere yönelik ayrıntılı bir kontrol yapılmış, tüm değişkenlerin tamamen dolu olduğunu ve hiçbir sütunda eksik değer bulunmadığı doğrulanmıştır. Bu durum, veri tamamlama stratejilerine olan ihtiyacı ortadan kaldırmakta ve model değerlendirmesinin yapay olarak üretilmiş değerlerden etkilenmemesini sağlamaktadır.

3.2.1.2 Deney ortamı

Önerilen S-SFIX çerçevesinin açıklanabilirlik performansını değerlendirmek amacıyla, karşılaştırma için SHAP ve LIME temel yöntemler olarak seçilmiştir. Bu iki teknik, açıklanabilir yapay zekâ literatüründe yaygın olarak tanınmakta ve finansal karar destek bağlamlarında sıklıkla kullanılmaktadır. SHAP, oyun teorisine dayanmaktadır ve özellik atfına dayalı küresel ve yerel açıklamalar sunmaktadır. LIME ise kara kutu sınıflandırıcının karar sınırını yaklaşık olarak temsil etmek amacıyla yerel vekil modeller kurmaktadır (Cil ve Yildiz, 2025a). Bu yöntemlerin farklı kuramsal temellere ve tamamlayıcı bakış açılarına sahip olması, onları S-SFIX'in sağlamlığını, sadeliğini ve özellik önemi güvenilirliğini değerlendirmek için uygun ve bilgilendirici karşılaştırma yöntemleri hâline getirmektedir. Ayrıca modelden bağımsız olmaları sayesinde, S-SFIX ile kullanılan aynı sınıflandırıcılara uygulanabilmekte ve tutarlı deneysel koşullar

sağlamaktadır.

Önerilen model, gerçek finansal işlemlerin zamana dayalı doğası dikkate alınarak oluşturulmuştur. Ne yazık ki, mevcut verisetlerinin birçoğu zaman bileşeni içermemektedir. Bu nedenle bu bölümde, kredi başvuru verisetine SFIX modelinin sadeleştirilmiş sürümü olan S-SFIX uygulanmıştır.

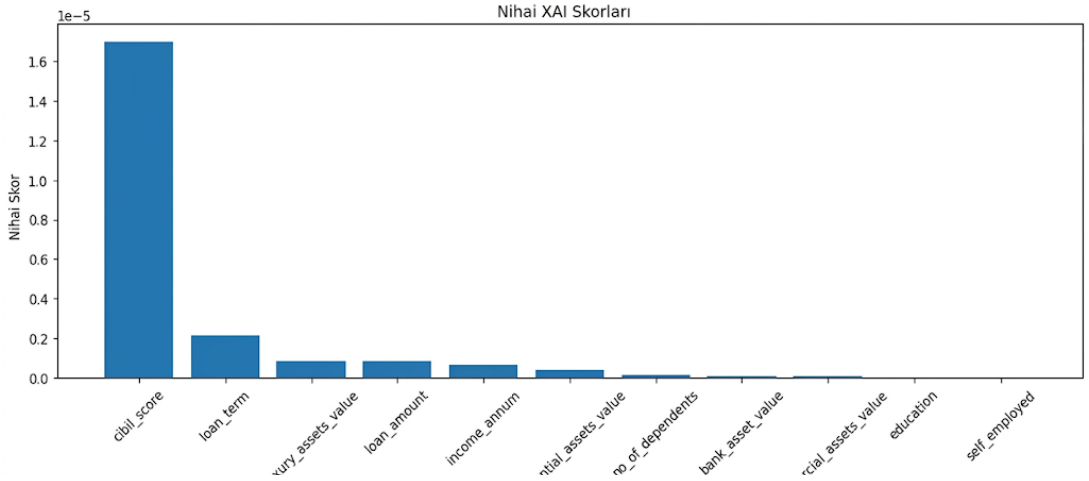
S-SFIX modelinin λ, γ ve β(j) dâhil hiperparametreleri, her bir bileşenin dengeli katkısını yansıtacak şekilde başlangıçta eşit ağırlıklarla atanmıştır. Bu başlangıç ayarları, modelin modüler yorumlanabilirliği vurgulayan tasarım mantığına dayanmaktadır. Model deneyleri sırasında, özellik önemi dağılımında ve çoklu çalıştırmalar boyunca tutarlılıkta gözlemlenen davranışlara bağlı olarak küçük manuel ayarlamalar yapılmıştır.

Bu aşamada hesaplama kısıtları nedeniyle kapsamlı bir ızgara araması uygulanmamış olsa da, seçilen değerler her iki sınıfta tutarlı sonuçlar ve anlamlı açıklamalar üretmiştir.

S-SFIX Modelinin performansını karşılaştırmak amacıyla, bu çalışmada kullanılan loan approval veriseti üzerinde SHAP ve LIME XAI yöntemleri uygulanmıştır. Karşılaştırma çalışmasında, SHAP ve LIME yöntemleri XGBoost algoritmasının uygulanmasına ilişkin açıklanabilirlik sağlamaktadır. SHAP ya da LIME değerlerinin hesaplanmasında, XGBoost gibi ağaç tabanlı modeller için TreeExplainer tercih edilmiştir. Deneysel bulgular bir sonraki bölümde sunulmaktadır.

3.2.1.3 Bulgular

S-SFIX modelinin kredi başvuru verisetine uygulanması sonucunda, her bir skor Şekil 3.2'de gösterildiği gibidir.



Şekil 3.2. Kredi başvuru veriseti özellik skorları

S-SFIX modelinin kredi başvuru verisetine uygulanması sonucunda, her bir skora ilişkin ayrıntılar Tablo 3.10'da gösterildiği gibidir.

Tablo 3.10. S-SFIX modelinin kredi başvuru verisetine uygulanmasının sonucu

Özellik	DIS	CS	AS	RS	Nihai Skor
cibil_score	1.1317e-04	6.0092e-01	5.3598e-01	4.6693e-01	1.7020e-05
loan_term	5.9565e-05	1.7875e-01	5.3493e-01	3.7690e-01	2.1466e-06
luxury_assets_value	6.9803e-05	6.1341e-02	5.4391e-01	3.7603e-01	8.7574e-07
loan_amount	6.6566e-05	6.2936e-02	5.4396e-01	3.7609e-01	8.5704e-07
income_annum	5.2029e-05	6.3571e-02	5.3652e-01	3.7732e-01	6.6959e-07
residential_assets_value	4.0761e-05	4.6685e-02	5.5036e-01	3.7593e-01	3.9372e-07
no_of_dependents	1.0439e-05	6.1881e-02	5.3257e-01	3.7784e-01	1.2998e-07
bank_asset_value	9.4146e-06	3.8738e-02	5.4911e-01	3.7709e-01	7.5516e-08
commercial_assets_value	9.3596e-06	3.5017e-02	5.4941e-01	3.7662e-01	6.7818e-08
education	5.8051e-07	2.4710e-02	5.0000e-01	3.7784e-01	2.7099e-09
self_employed	2.4307e-09	6.5977e-03	5.0001e-01	3.7784e-01	3.0298e-12

Tablo 3.10'da sunulan nihai XAI skorlarının, SFIX çerçevesinin çarpımsal doğası nedeniyle sayısal olarak küçük olduğuna dikkat edilmelidir; çünkü 0 ile 1 arasında normalize edilen çoklu bileşenler bir araya getirilmektedir. Sonuç olarak, beş bileşenin her birinde orta düzey değerler olsa bile küçük nihai skorlar elde edilebilir. Bununla birlikte, burada önemli olan mutlak değer değil, özelliklerin göreli sıralamasıdır. Bu skorlar, açıklanabilirliğin karşılaştırmalı bir ölçüsünü sunmakta ve özelliklerin dinamik önem, güven, anomali ve risk boyutlarındaki toplu katkılarına göre önceliklendirilmesine yardımcı olmaktadır. SFIX'in gelecekteki sürümleri, yorumlanabilirliği artırmak amacıyla skor normalizasyon stratejileri içerebilir.

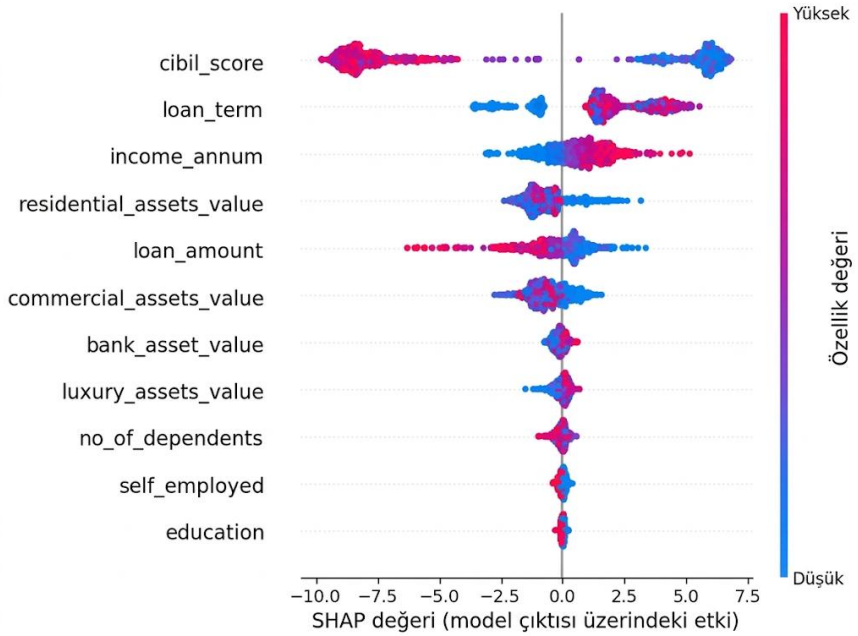
Hesaplanan skorların sonunda, kredi değerlendirmesinde en değerli 5 özellik aşağıdaki Tablo 3.11'de önem sırasına göre listelenmiştir.

Tablo 3.11. SFIX modeli için kredi başvuru verisetindeki en önemli 5 özellik

Özellik	DIS	CS	AS	RS	Nihai Skor
cibil_score	0.000113	0.600924	0.535981	0.466932	1.701951e-05
loan_term	0.000060	0.178749	0.534927	0.376901	2.146636e-06
luxury_assets_value	0.000070	0.061341	0.543910	0.376033	8.757417e-07
loan_amount	0.000067	0.062936	0.543956	0.376086	8.570391e-07
income_annum	0.000052	0.063571	0.536520	0.377323	6.695883e-07

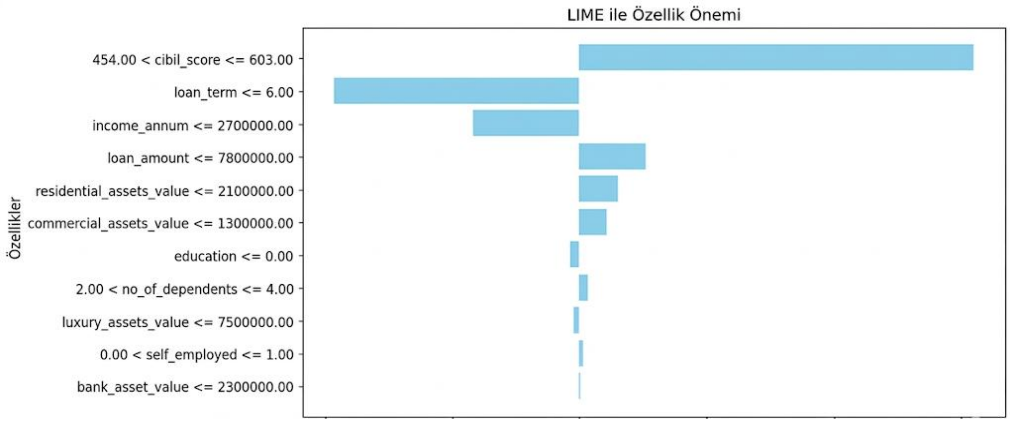
Kredi başvuru veriseti kullanılarak kredi değerlendirmesinde S-SFIX modelinin en fazla dikkate aldığı ilk 5 özellik sırasıyla *cibil_score, loan_term, luxury_assets_value, loan_amount* ve *income_annum*'dur.

Önerilen modelin performansını karşılaştırmak amacıyla, aynı verisetine SHAP XAI yöntemleri uygulanarak elde edilen sonuçlar Şekil 3.3'te verilmiştir.



Şekil 3.3. SHAP modelinin kredi başvuru verisetine uygulanmasının sonucu

Önerilen modelin performansını karşılaştırmak amacıyla, aynı verisetine LIME XAI yöntemleri uygulanarak elde edilen sonuçlar Şekil 3.4'te verilmiştir.



Şekil 3.4.LIME modelinin kredi başvuru verisetine uygulanmasının sonucu

Son olarak, her bir özellik için SHAP, LIME ve S-SFIX yöntemlerimizin açıklanabilirlik skorlarından elde edilen sonuçların karşılaştırmalı tablosu Tablo 3.12'de verilmiştir.

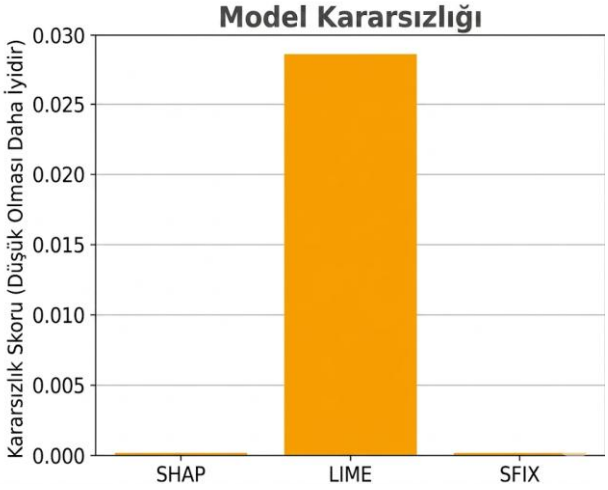
Tablo 3.12. LIME, SHAP ve S-SFIX'ten elde edilen özellik önem sıralamaları

Sıra	LIME	SHAP	S-SFIX
1	cibil_score	cibil_score	cibil_score
2	loan_term	loan_term	loan_term
3	income_annum	income_annum	luxury_assets_value
4	loan_amount	residential_assets_value	loan_amount
5	residential_assets_value	loan_amount	income_annum
6	commercial_assets_value	commercial_assets_value	residential_assets_value
7	education	bank_asset_value	no_of_dependents
8	no_of_dependents	luxury_assets_value	bank_asset_value
9	luxury_assets_value	no_of_dependents	commercial_assets_value
10	self_employed	self_employed	education
11	bank_asset_value	education	self_employed

Burada görüldüğü üzere, açıklanabilir yapay zekâ algoritmaları genel olarak uygulanan modeller için her bir özelliğin elde edilen sonuca katkısını sunmaktadır. Bu nedenle, doğru açıklanabilirlik için uygulamaya özgü algoritma tasarımları önemlidir. Elde edilen karşılaştırmalı sonuçlara göre, algoritmaların en önemli 2 özellikte (*cibil_score* ve *loan_term*) uzlaştığı görülmektedir. Bu da S-SFIX modelimizin doğru sonuçlar ürettiğini doğrulamaktadır. 3. sıradan itibaren farklılaşan sonuçlar ise, değerlendirmenin farklı açılardan ele alınmasının yarattığı farkı ortaya koymaktadır.

Finans sektöründe model açıklanabilirliğinin kritik uygulamaları dikkate alındığında, açıklama yöntemlerinin kararlılığı ve tutarlılığı büyük önem taşımaktadır. Kararlılık ölçümü için, SHAP, LIME ve S-SFIX'in aynı veri noktası için farklı çalıştırmalarda tutarlı sonuçlar verip vermediğinin analiz edilmesi amaçlanmıştır. Bunu görmek için SHAP, LIME ve S-SFIX aynı veri noktası üzerinde 10 kez çalıştırılmış ve sonuçların ne kadar değiştiği incelenmiştir.

Şekil 3.5'te gerçekleştirilen karşılaştırmalı analizde, SHAP ve SFIX'in çok düşük kararlılık skorlarına sahip olduğu ve böylece aynı veri noktaları için yeniden çalıştırıldığında tutarlı sonuçlar ürettiği gözlemlenmektedir.



Şekil 3.5.LIME, SHAP, S-SFIX kararlılık sonucu

Bu durum, SHAP'ın küresel açıklamalar üretebilmesi ve S-SFIX'in zaman periyodu ile veri kalitesi gibi ek bileşenleri dikkate alarak açıklamalar üretebilmesi ile açıklanabilir. Öte yandan, LIME'ın kararlılık skoru belirgin biçimde daha yüksek çıkmıştır; bu da yöntemin rastgele örneklem seçimine bağlı olarak farklı çalıştırmalarda değişken sonuçlar ürettiğini göstermektedir. Özellikle finans sektöründe kredi skorlama, risk analizi ve anomali tespiti gibi kritik karar mekanizmalarında model açıklamalarının tutarlı olması gerektiği düşünüldüğünde, S-SFIX'in sunduğu kararlılık avantajı önemli bir katkı sağlamaktadır. Sonuçlar, S-SFIX'in yalnızca açıklanabilirlik sunmadığını, aynı zamanda modelin öngörülebilir ve güvenilir kararlar üretmesine de katkıda bulunduğunu göstermektedir. Dolayısıyla, S-SFIX'in SHAP ile benzer düzeyde kararlılık sunduğu ve LIME'a kıyasla çok daha güvenilir açıklamalar ürettiği söylenebilir.

Ayrıca, SHAP, LIME ve S-SFIX yöntemlerinin tutarlılığına ilişkin sayısal değerlerin istatistiksel yöntemlerle karşılaştırılması Tablo 3.13'te verilmiştir.

Tablo 3.13. Model kararlılık metriklerinin karşılaştırılması

Yöntem	Ortalama Skor	Std. Sapma	Varyasyon Katsayısı
SHAP	1.206641	2.340556×10^{-16}	1.939728×10^{-16}
LIME	0.084062	1.907349×10^{-3}	2.268979×10^{-2}
S-SFIX	0.000002	2.232128×10^{-22}	1.231661×10^{-16}

Deneysel bulgular, S-SFIX modelinin en düşük varyasyon katsayısına sahip olduğunu ($CV = 1.23 \times 10^{-16}$) ve bu nedenle çoklu çalıştırmalar boyunca son derece kararlı ve tutarlı açıklamalar ürettiğini göstermektedir. SHAP yöntemi de benzer düzeyde bir kararlılık sergileyerek bu sonucu yakından izlemektedir ($CV = 1.93 \times 10^{-16}$). Buna

karşılık, LIME yönteminin varyasyon katsayısı belirgin biçimde daha yüksek bulunmuştur (CV = 2.26×10^{-2}). Bu durum, LIME tarafından üretilen açıklamaların farklı çalıştırmalar arasında daha fazla değişkenlik gösterdiğine ve kararlılık düzeyinin diğer yöntemlere kıyasla daha düşük olduğuna işaret etmektedir.
Bu bulgular, S-SFIX'in sağlam ve güvenilir model açıklamaları sunma yeteneğini ortaya koymakta ve onu yüksek riskli finansal uygulamalar için iyi bir alternatif hâline getirmektedir.

Sonuç olarak, S-SFIX algoritması içindeki farklı bileşenler, özelliklerin önemini SHAP ve LIME'dan farklı biçimde değerlendirmiştir. Bu da finans alanına özgü XAI yöntemleri geliştirmenin neden önemli olduğunu açıklamaktadır.

3.2.1.4 Tartışma

Deneysel sonuçlar, önerilen SFIX çerçevesinin finansal karar vermeye özgü yorumlanabilirlik zorluklarını ele almadaki etkinliğini ortaya koymaktadır. SHAP ve LIME gibi kıyas yöntemi olarak kullanılan yöntemlerle karşılaştırıldığında, SFIX özellikle zamansal örüntüleri, risk farkındalığını ve veri kalitesi metriklerini bütünleştirmesi sayesinde daha yapılandırılmış ve alanla uyumlu bir yaklaşım sunmaktadır. SFIX'in temel avantajlarından biri, tekrarlanan çalıştırmalar boyunca gösterdiği kararlılıktır; bu durum, rastgele örneklemeye dayanan post-hoc açıklama yöntemlerinde sıklıkla görülen bir sınırlılıktır. Ayrıca SFIX, sadeleştirilmiş sürümü (S-SFIX) aracılığıyla statik verisetleri için de açıklama üretimine olanak sağlamakta, bunu yaparken tam model ile tutarlı yorumlama mantığını korumaktadır. SHAP ve LIME güçlü örnek düzeyinde açıklamalar sunmasına rağmen, risk duyarlı uygulamalar için merkezi öneme sahip olan anomali skorlama veya güven tahmini gibi alan odaklı bileşenleri içermemektedir. Genel olarak sonuçlar, SFIX'in yalnızca rekabetçi bir yorumlanabilirlik düzeyi sağlamakla kalmadığını, aynı zamanda düzenlemeye tabi finansal ortamlarda temel gereksinimler olan karar şeffaflığına ve sağlamlığa da katkıda bulunduğunu göstermektedir.

SFIX ile SHAP ve LIME gibi yaygın kullanılan post-hoc yöntemler arasındaki yapısal ve işlevsel farklılıkları daha net vurgulamak amacıyla, Tablo 3.14'de karşılaştırmalı bir özet sunulmaktadır. Bu karşılaştırma, yalnızca her bir açıklama çerçevesinin doğasını değil, aynı zamanda finansal alanlarda kritik öneme sahip olan zamansal ve risk

farkındalıklı karar sistemlerine uygunluklarını da ortaya koymaktadır.

Tablo 3.14. SFIX, SHAP ve LIME'ın karşılaştırmalı genel görünümü

Boyut	**SFIX**	**SHAP**	**LIME**
Açıklama Türü	İçsel (yerleşik)	Post-hoc	Post-hoc
Zamansal Duyarlılık	Evet	Hayır	Hayır
Risk Modellemesi	Evet (Bayes temelli)	Hayır	Hayır
Anomali Skorlama	Bileşen olarak dâhil	Mevcut değil	Mevcut değil
Belirsizlik Yönetimi	Tahmin varyansına dayalı	Dâhil değil	Kısmi
Kararlılık	Yüksek (modüler yapı)	Düşük (örneklemeye bağlı)	Düşük (örneklemeye bağlı)
Yorumlama Düzeyi	Özellik grupları + zamansal örüntüler	Bireysel özellikler	Bireysel özellikler
Uygulama Alanı	Finansal ve zaman serisi sistemleri için tasarlanmıştır	Genel amaçlı	Genel amaçlı

Eğitilmiş kara kutu modellerin davranışını yaklaşık olarak temsil ederek açıklama üreten SHAP ve LIME gibi yaygın post-hoc yöntemlerin aksine, SFIX tasarım gereği içsel olarak açıklanabilirdir. Post-hoc yaklaşımlar genellikle tahminden sonra yerel açıklamalar sunar ve büyük ölçüde örnekleme tabanlı yaklaşık hesaplamalara dayanır; bu da değişkenlik oluşturabilir ve yeniden üretilebilirliği sınırlayabilir. Buna karşılık, SFIX yorumlanabilirlik çıktılarını modelin iç hesaplama hattının bir parçası olarak üretmekte ve harici vekil modellere olan ihtiyacı ortadan kaldırmaktadır.

SFIX'i farklı kılan nokta, yorumlanabilirliğin harici bir katman ya da isteğe bağlı bir özellik olarak ele alınmaması, aksine tahmin sürecine sıkı biçimde entegre edilmiş olmasıdır. Ayrıca SFIX, nihai tahmine paralel olarak birleşik bir yorumlanabilirlik skoru üretmekte ve böylece finans gibi zamana duyarlı, yüksek riskli karar verme ortamları için gerekli olan gerçek zamanlı şeffaflığı sağlamaktadır.

SHAP ve LIME gibi geleneksel XAI yöntemleri çoğu zaman finansal verinin zamana bağlı doğasını ve kredi onayı, dolandırıcılık tespiti ve portföy yönetimi gibi kararlarla ilişkili içsel riskleri göz ardı etmektedir. Oysa finansal karar verme doğası gereği dinamiktir ve risk duyarlıdır. SFIX, açıklama çerçevesine zamansal modelleme ve risk değerlendirmesini doğrudan entegre ederek bu karşılanmamış ihtiyaçları ele almaktadır. DIS, özellik öneminin zaman içinde nasıl evrildiğini yakalayarak paydaşların gecikmeli etkileri, davranış eğilimlerini veya mevsimselliği tespit etmesine olanak tanır; bu, erken

dolandırıcılık tespiti veya kredi değerliliğindeki değişimlerin izlenmesi açısından kritik önemdedir. Ayrıca RS, gözlemlenen özellik davranışı altında risk olasılığını nicel olarak belirlemek için Bayesçi bir yaklaşım uygular ve yüksek riskli ortamlarda eyleme dönüştürülebilir içgörüler sunar. Zaman ve riskin bu ikili entegrasyonu yalnızca açıklanabilirliği geliştirmekle kalmaz, aynı zamanda karmaşık finansal ortamlarda düzenleyici uyumu ve proaktif karar vermeyi de destekler.

SFIX'in alana özgü tasarımı, SHAP ve LIME gibi genel amaçlı XAI yöntemlerine kıyasla daha yüksek yorumlanabilirlik ve kararlılık elde etmesinde temel bir etkendir. Zamansal davranış modelleme, veri kalitesi değerlendirmesi, anomali tespiti ve Bayesyen risk skorlama gibi finansa özgü bileşenleri içermesi sayesinde SFIX, finansal karar verme bağlamlarıyla yakından uyumlu açıklamalar üretmektedir. Örneğin CS bileşeni, eksiklik oranı ve oynaklık gibi alana özgü metrikleri kullanarak özellik güvenilirliğini değerlendirirken, RS olasılıksal akıl yürütme yoluyla gerçek dünya belirsizliğini yansıtmaktadır. Bu tasarım, açıklamaların farklı model çalıştırmaları ve verisetleri boyunca kararlı kalmasını sağlamaktadır; bu durum özellikle düzenleyici veya denetim senaryolarında büyük önem taşımaktadır. Örnekleme ya da model pertürbasyonuna bağlı olarak farklı sonuçlar üretebilen genel yöntemlerin aksine, SFIX hem alan bilgisini hem de model davranışını yansıtan tutarlı ve bağlama duyarlı çıktılar sunmaktadır. Sonuç olarak SFIX, hassas finansal uygulamalarda kullanıcı güvenini ve karar izlenebilirliğini artırmaktadır.

Bunun yanı sıra, önerilen SFIX çerçevesi içsel ve modüler olacak şekilde tasarlanmıştır; bu da onu gerçek dünya finansal sistemlerine post-hoc açıklanabilirlik katmanı olarak da entegre edilmeye uygun hâle getirmektedir. Kredi skorlama, dolandırıcılık tespiti, kredi risk değerlendirmesi ve uyum izleme gibi şeffaf ve hesap verebilir karar vermenin gerekli olduğu uygulamalarda kullanılabilir. Örneğin, finansal kurumlar SFIX'i belirli bir kredi başvurusunun neden onaylandığı veya reddedildiğine ilişkin yapılandırılmış açıklamalar üretmek için kullanabilir; bu da düzenleyici gereklilikleri (ör. GDPR, Basel III) destekler ve kullanıcı güvenini artırır.

Önerilen modelin pratik uygulamalarında bazı zorluklar ortaya çıkabilmektedir. Özellikle zamansal dinamikleri ya da risk tabanlı hesaplamaları içeren çoklu açıklanabilirlik bileşenlerinin üretilmesinin hesaplama maliyeti, gerçek zamanlı

ortamlarda sınırlayıcı olabilmektedir. Ayrıca SFIX'in eski BT altyapılarına ya da özel karar sistemlerine entegre edilmesi teknik uyarlamalar gerektirebilir. Açıklamaların yalnızca geliştiriciler tarafından değil, aynı zamanda teknik olmayan paydaşlar (ör. risk yöneticileri, müşteriler) tarafından da yorumlanabilir olmasını sağlamak, operasyonel kullanım açısından bir diğer dikkate alınması gereken unsurdur. Bununla birlikte, SFIX'in sunduğu yorumlanabilirlik faydaları onu yüksek riskli, düzenlemeye tabi finansal karar süreçleri için değerli bir araç hâline getirmektedir.

3.2.2 Açıklanabilirlik yaklaşımlarının karşılaştırılması

Makine öğrenmesi modellerinin açıklanabilirliği, model şeffaflığını ve güvenilirliğini sağlamak açısından kritik öneme sahiptir. Bu bölümde, önerilen SFIX modeli, Calzarossa ve ark. (2025) tarafından önerilen değerlendirme çerçevesine dayalı olarak yaygın kullanılan açıklanabilirlik yaklaşımları olan SHAP ve LIME ile karşılaştırılmaktadır.

Calzarossa ve ark.. (2025), açıklanabilirlik yöntemlerinin değerlendirilmesi için iki temel ölçüt önermektedir:

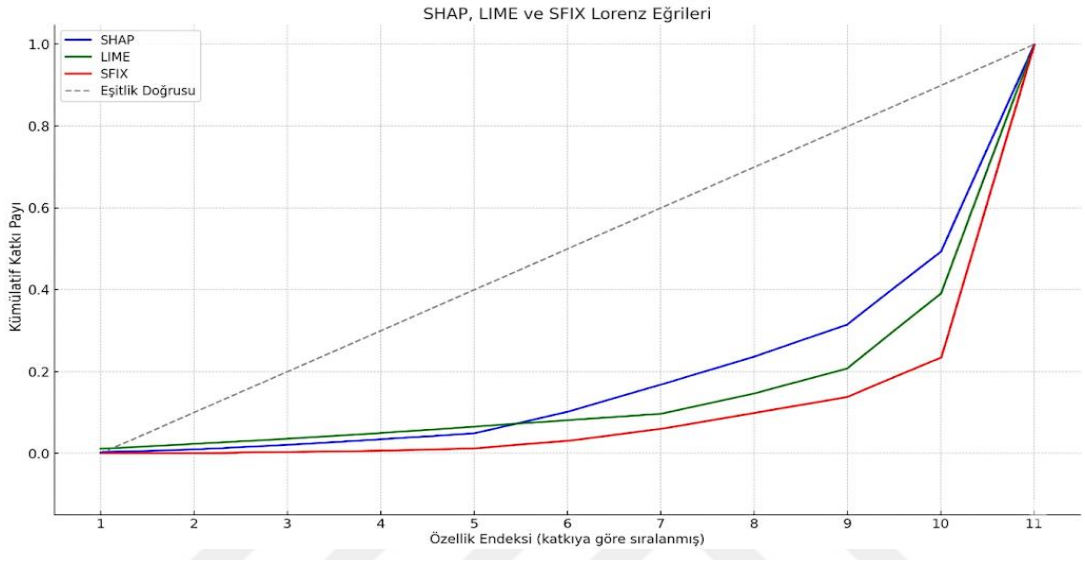
- Parsimony (Sadelik): Açıklamaların ne ölçüde sade ve yoğun olduğunu değerlendiren bir ölçüttür. Bu çalışmada sadelik, Lorenz eğrileri ve Gini katsayıları kullanılarak analiz edilmiştir.
- Robustness (Dayanıklılık): Açıklamaların farklı veri alt kümeleri boyunca ne kadar tutarlı kaldığını değerlendiren bir ölçüttür. Bu çalışmada dayanıklılık, kosinüs benzerliği, önem skoru varyasyonu ve ilk k özellik çakışması gibi ölçütler aracılığıyla hesaplanmıştır.

Bu ölçütleri izlenerek SFIX, SHAP ve LIME arasında ayrıntılı bir karşılaştırma gerçekleştirilmiştir.

3.2.2.1 Sadelik analizi

Sadeliği değerlendirmek için, üç yöntemin de özellik katkılarına ait Lorenz eğrileri analiz edilmiştir. Lorenz eğrileri önem değerlerinin dağılımını görselleştirir; eşitlik doğrusundan daha fazla sapma, daha yoğunlaşmış (ve dolayısıyla daha sade) bir açıklamaya işaret eder.

Şekil 3.6'da gösterildiği üzere, SFIX'in Lorenz eğrisi eşitlik doğrusundan en fazla sapmayı göstermekte; bu da SFIX'in özellik önemini SHAP ve LIME'a kıyasla daha az sayıda değişkende yoğunlaştırdığını göstermektedir. Alt kuyruğun daha düz, üst bölümün ise daha dik olmasıyla ortaya çıkan bu keskin yoğunlaşma, SFIX açıklamalarının odaklanmış doğasını vurgulamaktadır.



Şekil 3.6. SHAP, LIME ve SFIX için özellik önem skorlarının kümülatif dağılımını gösteren Lorenz eğrileri

SHAP ve LIME da belirli ölçüde yoğunlaşma sergilese de, SFIX kadar odaklı değildir. Bu bulgular, Calzarossa ve ark. (2025) tarafından vurgulanan sadelik ilkesini desteklemektedir.

Lorenz eğrilerinden hesaplanan Gini katsayıları da bu gözlemi desteklemektedir: SFIX en yüksek Gini skoruna (0.73) ulaşmış, onu LIME (0.65) ve SHAP (0.61) izlemiştir. Daha yüksek Gini katsayısı, daha yoğun açıklamalara ve dolayısıyla daha yüksek sadeliğe işaret etmektedir.

3.2.2.2 Dayanıklılık analizi

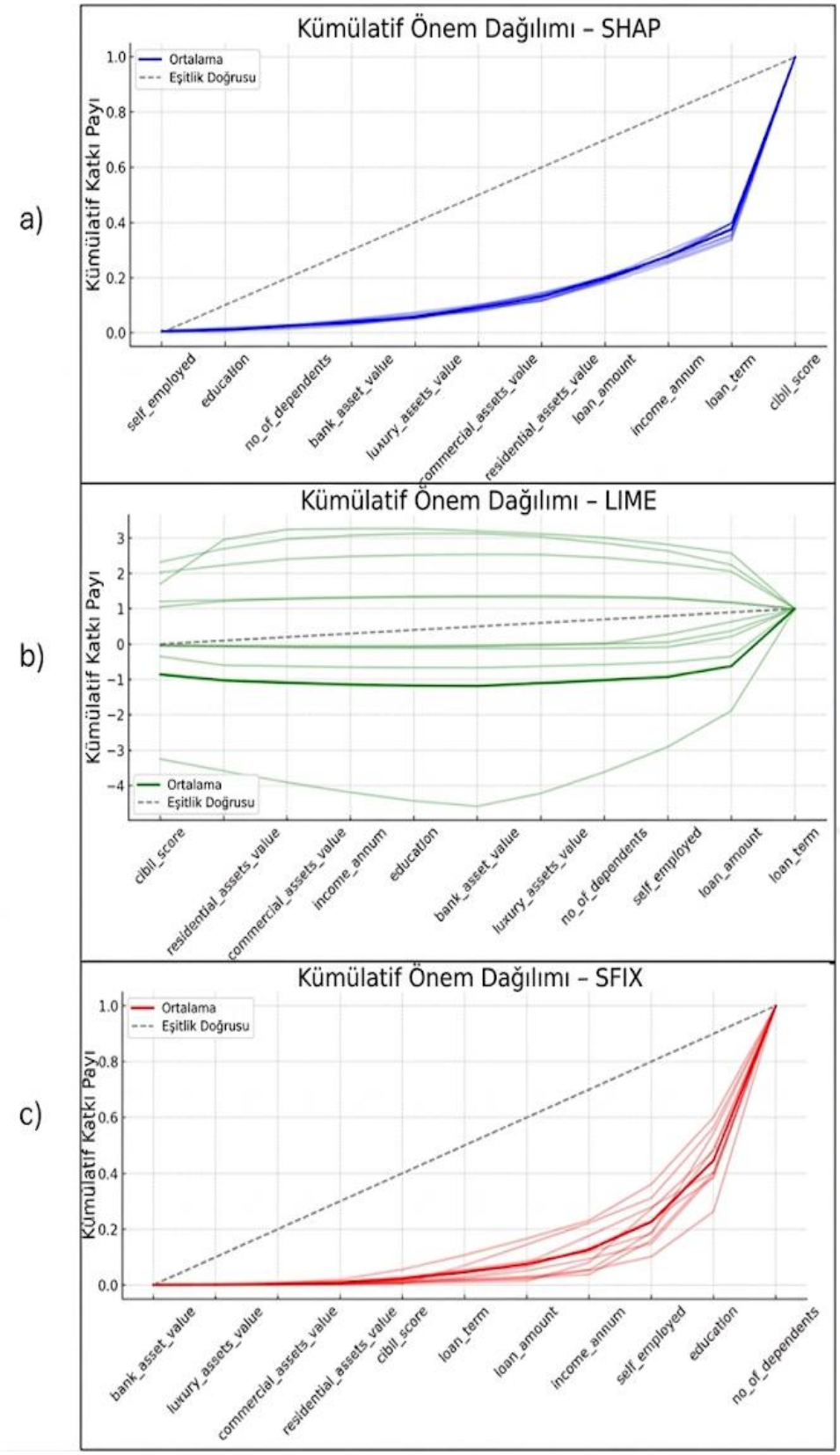
Değişen veri dağılımları altında açıklanabilirlik yöntemlerinin sağlamlık ve tutarlılığını değerlendirmek amacıyla, SHAP, LIME ve SFIX için 10 bootstrap örneklemine dayalı kümülatif önem dağılım grafiği oluşturduk. Şekildeki her eğri, her bir bootstrap tekrarında artan sıraya göre sıralanmış özellik katkılarının kümülatif payını temsil etmektedir. Eğriliğin eşitlik doğrusuna göre derecesi, önemin daha az sayıda özellikte ne ölçüde yoğunlaştığını yansıtır ki bu sadeliğin temel bir göstergesidir.

Bootstrap örneklemesinden türetilen kümülatif önem dağılım grafikleri, farklı XAI yöntemlerinin açıklayıcı davranışlarına ilişkin değerli içgörüler sunmaktadır. Şekil 3.7.a'da görüldüğü gibi, SHAP tüm bootstrap tekrarları boyunca görece düzgün ve kararlı eğriler sergilemektedir. Bu durum, yönteminin yeniden örneklenmiş verisetleri altında dahi tutarlı önem dağılımları üretmesi nedeniyle yüksek derecede dayanıklılık gösterdiğini ortaya koymaktadır. Bununla birlikte, ortalama SHAP eğrisi eşitlik doğrusuna nispeten yakın kalmakta; bu da açıklayıcı katkıların çok sayıda özellik arasında dağıldığını göstermektedir. Bu durum yorumu daha karmaşık hâle getirmekte ve sadeliği zayıflatmaktadır.

Buna karşılık, Şekil 3.7.b, LIME'ın bootstrap örneklemleri arasında kümülatif önem dağılımlarının önemli ölçüde değiştiğini ortaya koymaktadır. Bootstrap eğrileri oldukça dağınık görünmektedir; bu da kararsızlığa ve verideki küçük değişimlere duyarlılığa işaret etmektedir. Bazı tekrarlar önemi birkaç özellikte yoğunlaştırırken, diğerleri neredeyse tekdüze bir dağılım göstermekte; bu da düşük dayanıklılık anlamına gelmektedir. LIME bazı tekil durumlarda yüksek sadelik sağlayabilse de, genel tutarsızlığı onu küresel açıklama yöntemi olarak daha az güvenilir kılmaktadır.

Şekil 3.7.c'de gösterildiği üzere, SFIX hem sadelik hem de dayanıklılık açısından SHAP ve LIME'dan açık biçimde daha iyi performans göstermektedir. SFIX'e ait kümülatif önem eğrileri tüm bootstrap örneklemleri boyunca birbirine sıkı biçimde yakın gruplanmıştır. Bu, açıklamanın daha az sayıda özellikte yoğunlaştığını göstermektedir. Bu tutarlı davranış, SFIX'in veri pertürbasyonlarına karşı dayanıklılığını ve yeniden örnekleme değişkenliği altında dahi odaklanmış açıklamaları koruma yeteneğini göstermektedir.

Bu sonuçlar, SFIX'in açıklama önemini az sayıdaki anahtar özellikte yoğunlaştırarak yorumlanabilir açıklamalar sunması ve açıklanabilirliği farklı veri örneklemleri boyunca tutarlılık sağlaması bakımından avantajını vurgulamaktadır. Bu özellikler, SFIX'i açıklamalarda açıklık, kararlılık ve şeffaflığın zorunlu olduğu kredi riski değerlendirmesi ve düzenleyici karar verme gibi yüksek riskli uygulamalar için özellikle uygun



Şekil 3.7. 10 bootstrap örneklemi boyunca SHAP, LIME ve SFIX için kümülatif önem dağılım eğrileri

kılmaktadır. Bu bulgular, SFIX'in yeniden örnekleme altında sergilediği dayanıklılık

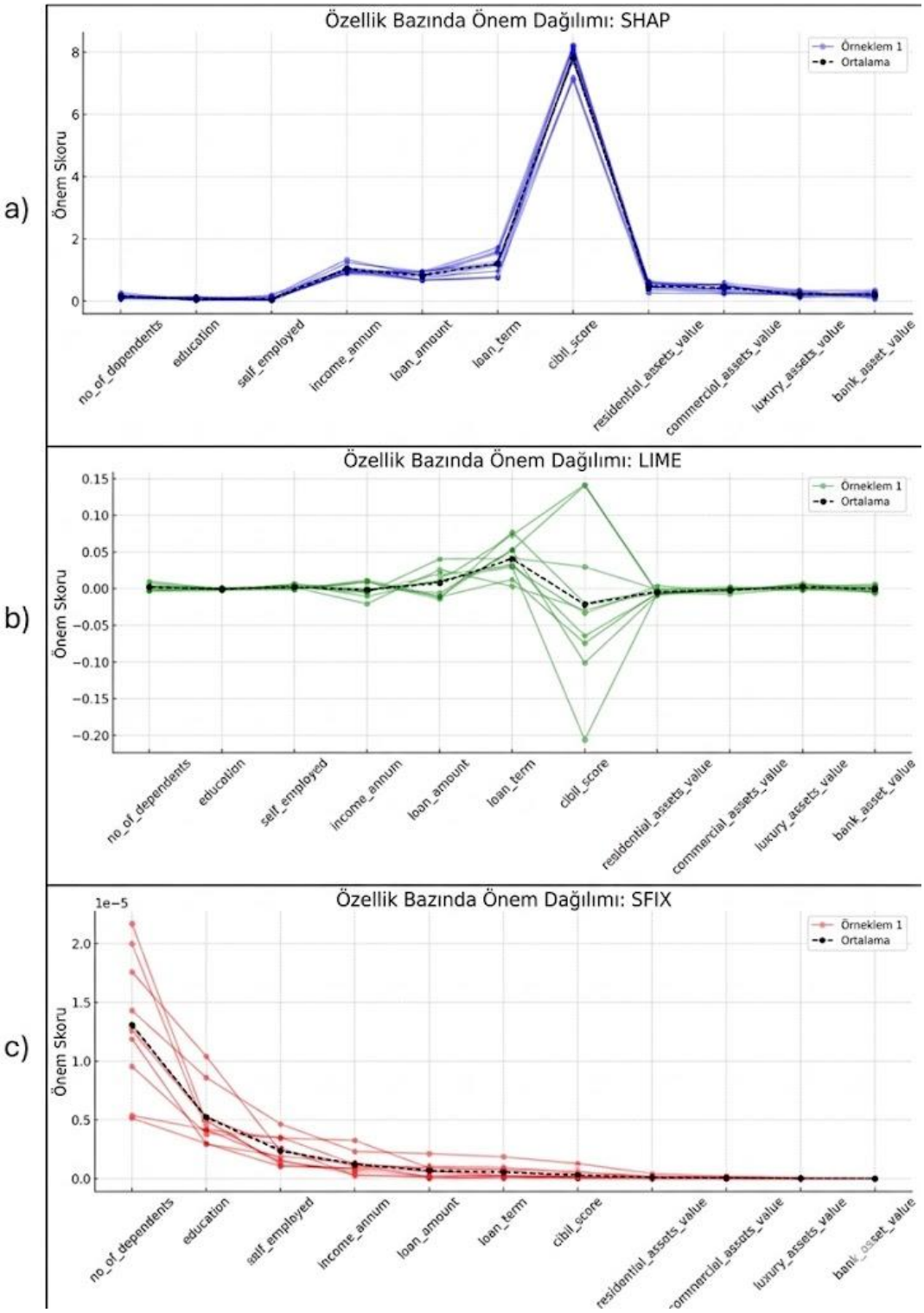
bakımından geleneksel yöntemlerden daha iyi olduğunu göstermektedir.

Genel sadelik ve dayanıklılık düzeyini değerlendiren kümülatif önem analizinin üzerine inşa ederek, bootstrap örneklemleri boyunca önem skorlarının özellik bazında kararlılığı da incelenmiştir.

Şekil 3.8.a'da görüldüğü üzere, SHAP cibil_score ve loan_term gibi birkaç baskın özelliğe yüksek önem atama eğilimindedir ve bu atamalar bootstrap tekrarları boyunca görece düşük varyans göstermektedir. Bu örüntü, SHAP'ın anahtar özelliklerin eğitim verisindeki küçük değişimlerden bağımsız olarak açıklayıcı gücünü koruduğu, makul derecede sağlam bir açıklama çerçevesi sunduğunu göstermektedir. Ancak daha az etkili bazı özellikler için SHAP orta düzeyde değişkenlik sergilemekte; bu da sınır durumlarda yorumlanabilirliği zorlaştırabilir.

Buna karşılık, Şekil 3.8.b LIME'ın özellik önem atamalarında son derece kararsız bir davranış sergilediğini göstermektedir. Aynı özelliğin önem değerleri bootstrap örneklemleri arasında ciddi biçimde dalgalanmakta, hatta bazı özelliklerde işaret bile değiştirmektedir. Bu tutarsızlık düşük dayanıklılığa işaret etmekte ve LIME'ın küresel yorumlama açısından güvenilirliğini zayıflatmaktadır. LIME öncelikli olarak yerel açıklamalar için tasarlanmış olsa da, küresel önem örüntülerindeki bu düzensizlik, toplulaştırılmış analizlerde kullanımının yanıltıcı ya da çelişkili sonuçlar doğurabileceğini göstermektedir. Özellikler ve tekrarlar arasında belirgin bir eğilimin bulunmaması, istikrarlı ve şeffaf karar desteği gerektiren ortamlarda kullanım değerini sınırlamaktadır.

Şekil 3.8.c'de gösterildiği gibi, SFIX dikkat çekici bir kararlılık ve sadelik sergilemektedir. Tüm bootstrap tekrarlarında her bir özellik için önem değerleri birbirine sıkı biçimde yakın kalmakta; *cibil_score* ve *loan_term* sürekli olarak en etkili özellikler olarak öne çıkmaktadır. SHAP ve LIME'ın aksine, SFIX hem yüksek önem yoğunlaşması hem de yüksek tutarlılık sunmakta; bu da onu şeffaflık ve yeniden üretilebilirliğin gerekli olduğu uygulamalar için son derece uygun hâle getirmektedir. Bu tutarlılık, SFIX'in kredi skorlama ve finansal risk değerlendirmesi gibi yüksek riskli alanlar için seyrek, kararlı ve düzenleyici dostu açıklamalar sunabilen modelden bağımsız, küresel olarak yorumlanabilir bir yöntem olarak güvenilirliğini güçlendirmektedir.



Şekil 3.8. 10 bootstrap örneklemi boyunca SHAP, LIME ve SFIX için özellik bazlı önem dağılım grafikleri

Açıklama yöntemlerinin dayanıklılık düzeyini değerlendirmek amacıyla, bunların farklı bootstrap örneklemleri arasındaki tutarlılığı üç temel metrik kullanarak incelenmiştir:

- Kosinüs Benzerliği: Açıklama vektörlerinin yönsel benzerliğini ölçer; 1'e yakın değerler yüksek tutarlılığa işaret eder.
- Önem Puanı Standart Sapması: Özellik önem skorlarındaki değişkenliği yakalar; daha düşük değerler daha yüksek kararlılığı gösterir.
- İlk 3 Özellik Çakışması: Aynı en önemli üç özelliğin farklı örneklemler arasında ne sıklıkla ortaya çıktığını gösterir.

Tablo 3.15'te özetlenen dayanıklılık analizi, SHAP, LIME ve SFIX için bootstrap örneklemesi altında özellik önem tahminlerinin kararlılığını değerlendirmektedir.

Tablo 3.15. Bootstrap örneklemeleri boyunca SHAP, LIME ve SFIX yöntemlerinin dayanıklılık karşılaştırması

Metrik	**SHAP**	**LIME**	**SFIX**
Ortalama Kosinüs benzerliği	0.991	**0.993**	0.963
Ortalama Önem Skoru Standart Sapması	0.23767	0.012868	**0.000001**
Ortalama İlk 3 Özellik Çakışması	0.763	0.711	**1.0**

Hem SHAP hem de LIME çok yüksek kosinüs benzerliği değerleri (sırasıyla 0.991 ve 0.993) sergilemekte; bu da önem vektörlerinin genel yönünün farklı bootstrap örneklemleri boyunca tutarlı kaldığını göstermektedir. SFIX de bu açıdan iyi performans göstermekte, biraz daha düşük olsa da (0.963) hâlâ güçlü dayanıklılık aralığında yer almaktadır. Bununla birlikte, önem skoru değişkenliği bakımından SFIX dikkat çekici bir üstünlük sergilemektedir: ortalama standart sapması yalnızca 0.000001'dir; SHAP için bu değer 0.23767, LIME için ise 0.012868'dir. Bu durum, SFIX'in son derece kararlı özellik sıralamaları ürettiğini ve çok az dalgalanma gösterdiğini ortaya koymaktadır.

En dikkat çekici olarak, SFIX tüm bootstrap örneklemlerinde mükemmel bir İlk 3 özellik çakışması (1.0) elde etmiştir; yani aynı üç özellik sürekli olarak en önemli özellikler olarak belirlenmiştir. Buna karşılık SHAP (0.763) ve LIME (0.711) belirgin biçimde daha düşük çakışma göstermektedir; bu da tekrarlar arasında anahtar özellikler konusunda daha düşük uyuma işaret etmektedir. Bu bulgular, SFIX'in en tutarlı ve yeniden üretilebilir özellik seçimini sunduğunu doğrulamakta ve açıklama kararlılığının kritik olduğu karar odaklı alanlar için uygunluğunu güçlendirmektedir.

Calzarossa ve ark. (2025) tarafından önerilen değerlendirme çerçevesine dayanarak, sonuçlar SFIX modelinin geleneksel açıklanabilirlik yöntemlerine göre birkaç belirgin

avantaja sahip olduğunu ortaya koymaktadır:

- Yüksek Sadelik: SFIX, hem LIME'ı (0.708) hem de SHAP'ı (0.649) aşarak en yüksek Gini katsayısına (0.803) ulaşmış, böylece özellik katkılarının en yoğun dağılımını göstermiştir. Ayrıca SFIX, toplam katkının %80'ini açıklamak için yalnızca 2 özellik gerektirirken, bu sayı LIME için 3, SHAP için ise 4 olmuştur. Bu bulgular, SFIX'in minimal bilişsel yük ile seyrek ve odaklanmış açıklamalar üretme yeteneğini doğrulamaktadır.
- Yüksek Dayanıklılık: SFIX, bootstrap tekrarları boyunca ortalama 0.963 kosinüs benzerliği, önem skorlarında son derece düşük standart sapma (1×10^{-6}) ve mükemmel ilk 3 özellik çakışması değeri (1.0) ile güçlü bir dayanıklılık performansı sergilemektedir. Bu bulgular, SFIX'in değişen veri koşulları altında dahi kararlı, tutarlı ve tekrarlanabilir açıklamalar üretebildiğini göstermektedir.

Sonuç olarak, Calzarossa ve ark. (2025) tarafından ortaya konulan değerlendirme ilkeleri çerçevesinde SFIX, hem sadelik hem de dayanıklılık açısından SHAP ve LIME'dan daha iyi performans göstermektedir. Kısa ve tutarlı açıklamalar sunabilme yeteneği, onu kredi skorlama ve finansal risk değerlendirmesi gibi düzenlemeye tabi, yüksek riskli karar sistemlerinde kullanım için güçlü bir aday hâline getirmektedir.

3.2.2.3 Karşılaştırmalı analiz

SFIX ile SHAP ve LIME gibi yaygın kullanılan açıklanabilirlik yöntemleri arasındaki farkları açık biçimde vurgulamak amacıyla, Tablo 3.16'da karşılaştırmalı bir özet sunulmaktadır.

Tablo 3.16. SHAP, LIME ve SFIX'in yapısal ve işlevsel özellikler açısından karşılaştırılması

Özellik	SHAP	LIME	SFIX (Önerilen)
Yöntem Türü	Post-hoc	Post-hoc	Ante-hoc (içsel olarak yorumlanabilir)
Model Bağımlılığı	Modelden bağımsız	Modelden bağımsız	Yerleşik yorumlanabilirliğe sahip modele özgü
Zaman Duyarlılığı	Desteklenmiyor	Desteklenmiyor	DIS ve PS bileşenleri aracılığıyla destekleniyor
Risk Değerlendirmesi	Dâhil değil	Dâhil değil	Bayesyen risk modülü entegre
Anomali Tespiti	Açık biçimde modellenmiyor	Açık biçimde modellenmiyor	AS ile modelleniyor
Veri Kalitesi Farkındalığı	Temiz veri varsayar	Temiz veri varsayar	Eksiklik, aykırı değerler ve varyasyon değerlendirilmektedir
Kararlılık	Örneklemeye duyarlı	Rastgele tohumlara duyarlı	İçsel tasarım sayesinde yüksek tutarlılık
Küresel vs Yerel Açıklama	Öncelikle yerel	Yerel	Hem yerel hem küresel açıklamayı destekler
Yorumlanabilirlik Doğruluğu	Orta	Düşük ila orta	Yüksek, içsel tasarım nedeniyle
Hesaplama Maliyeti	Orta ila yüksek	Orta	Orta (sadeleştirilmiş sürüm: S-SFIX)
Finans Alanına Özelleştirme	Genel amaçlı	Genel amaçlı	Özellikle finansal uygulamalar için tasarlanmıştır

SHAP ve LIME yaygın kullanılan post-hoc açıklanabilirlik yöntemleri olmakla birlikte, özellikle karmaşık finansal sistemlere uygulandıklarında çeşitli yapısal ve işlevsel sınırlılıklar göstermektedir. Tablo 3.16, önerilen SFIX modelini SHAP ve LIME ile temel boyutlar açısından açık biçimde karşılaştırmaktadır. Bu karşılaştırma, SFIX'in içsel olarak yorumlanabilir bir çerçeve olmasından kaynaklanan yapısal avantajını ve zamansal analiz, risk odaklı tahmin, anomali tespiti ve veri kalitesi değerlendirmesi gibi geleneksel yaklaşımlarda bulunmayan ya da sınırlı olan işlevsel yeteneklerini vurgulamaktadır. Açıklanabilirliği doğrudan mimarisine yerleştirerek SFIX, finansal karar verme için daha güvenilir ve alana özgü içgörüler sunmaktadır.

3.3 Sonuç

Bu çalışmada, finansal karar verme süreçleri için tasarlanmış yeni bir XAI çerçevesi olan SFIX önerilmiştir. Model, zamansal dinamikleri, veri kalitesini, davranışsal örüntüleri ve risk maruziyetini yakalayan birleşik bir yorumlanabilirlik skorunda beş temel bileşeni

(DIS, CS, PS, AS, RS) bütünleştirmektedir. Bu yöntem oluşturulurken, finans alanında işlemlerin zamandan bağımsız analiz edilmemesi, modelin ana bileşenlerinden birinin zaman temelli olmasına neden olmuştur. Bir diğer bileşen veri kalitesidir. Açıklanabilirlik, verinin kalitesi ne kadar iyiyse o kadar doğru biçimde yapılabilir. Son olarak, anomali skoru ve risk skoru da yöntemin diğer önemli bileşenleridir. Bu anomali ve risk hem dolandırıcılık tespiti hem de risk değerlendirmesi gibi çeşitli uygulama alanlarında faydalı olacaktır.

SHAP ve LIME gibi güncel yöntemlerle karşılaştırıldığında, SFIX özellik sıralamasında rekabetçi performans göstermiş ve farklı çalıştırmalar arasında daha yüksek kararlılık sunmuştur. Özellikle, telafi edici olmayan bir skor yapısı içermesi ve zamansal davranışı finansal risk ile modellemesi sayesinde mevcut XAI yöntemlerindeki temel sınırlılıkları ele almaktadır.

Ayrıca bu çalışma, SFIX modelinin sadeleştirilmiş bir sürümüne odaklanmaktadır. S-SFIX modeli, zaman boyutunun kullanılmadığı süreçleri bütünleştirmek amacıyla uygulanmaktadır. Zaman boyutu ortadan kaldırıldığı için daha hızlı açıklanabilirlik de sağlamaktadır. Gerektiğinde hızlı çalışan bir algoritma olarak kullanılabilir. Sonuç olarak, kredi başvuru veriseti üzerindeki uygulama açıklanabilirlik sunma açısından başarıyla test edilmiştir.

4 ÖNERİLEN SFIX MODELİN FİNANS UYGULAMASI

AI sistemleri, finansal analiz, kredi değerlendirme ve şirket performansı tahmini gibi kritik karar verme süreçlerinde giderek daha fazla kullanılmaktadır. Özellikle finansal raporlamanın güvenilirliği açısından, dış denetçi görüşlerinin doğru ve açıklanabilir biçimde tahmin edilmesi; düzenleyici uyum, kurumsal güven ve insan gözetimi gibi nedenlerle gerekli görülmektedir. Nitekim, Genel Veri Koruma Tüzüğü (GDPR) ve Yapay Zekâ Yasası gibi güncel düzenlemeler, AI sistemlerinin açıklanabilir ve denetlenebilir olmasına yönelik beklentileri giderek artırmaktadır (Selbst ve Barocas, 2018; Wachter ve ark., 2017).

FFNN gibi geleneksel derin öğrenme modelleri, güçlü tahmin performanslarına rağmen açıklanabilirlik eksiklikleri nedeniyle finansal karar destek sistemlerinde sınırlı uygulanabilirliğe sahiptir. Bu sınırlılık, özellikle denetim görüşleri gibi düzenleyici açıdan yüksek öneme sahip karar alanlarında daha belirgin hâle gelmektedir (Doshi-Velez ve Kim, 2017). Ancak yüksek tahmin doğruluğu tek başına model çıktılarının güvenilir biçimde anlaşılması ve yorumlanması için yeterli değildir; karar verme sürecinde bireysel özelliklerin katkısının da şeffaf hâle getirilmesi gerekmektedir.

Yapay zekânın finansal karar verme süreçlerinde giderek artan kullanımı, şeffaflık, güvenilirlik ve düzenleyici uyumla ilgili kritik sorunları beraberinde getirmektedir. Kararların hem doğru hem de açıklanabilir olması gereken denetim görüşü sınıflandırmasında, geleneksel derin öğrenme modelleri kara kutu yapıları nedeniyle çoğu zaman yeterli yorumlanabilirlik sağlayamamaktadır. SHAP ve LIME gibi mevcut XAI teknikleri post-hoc açıklamalar sunsa da esas olarak yerel veya statik özellik katkılarını yakalamakta ve bu nedenle finansal verinin doğasında bulunan zamansal, etkileşim odaklı ve risk duyarlı özellikleri yansıtmada yetersiz kalmaktadır.

Ayrıca, çoğu özellik seçimi yaklaşımı tek boyutlu ölçütlere dayanmaktadır ve birden fazla açıklanabilirlik bileşenini aynı anda dikkate almamaktadır. Bu sınırlılıklar, bütüncül, çok boyutlu ve finansal bağlama dayalı bir açıklanabilirlik çerçevesine duyulan ihtiyacı ortaya koymaktadır.

Bu boşluklardan hareketle, bu bölümde önerilen SFIX modelini bir FFNN mimarisiyle bütünleştirerek denetim görüşü sınıflandırması için güvenilir ve yorumlanabilir bir çerçeve oluşturulması hedeflenmektedir. SFIX; dinamik önem, güven, örüntü, anomali

ve risk bileşenlerini birlikte dikkate alarak özellik davranışlarının zaman içindeki daha zengin ve daha kapsamlı biçimde anlaşılmasını sağlamaktadır. Bu bütünleştirme, sınıflandırma performansını artırmayı ve aynı zamanda denetçileri, düzenleyicileri ve finansal kurumları destekleyen şeffaf karar içgörüleri sunmayı amaçlamaktadır.

Bu bölümde aşağıdaki temel araştırma soruları incelenmektedir:

- SFIX modeli, denetim görüşü sınıflandırmasında en açıklayıcı öznitelikleri güvenilir biçimde belirleyebilmekte midir?
- FFNN modeli, SFIX tarafından seçilen bir özellik alt kümesi kullanıldığında benzer doğrulukla tahmin üretebilmekte midir?
- FFNN modelinin açıklanabilirliğini SFIX kullanarak doğrudan veya dolaylı biçimde artırmak mümkün müdür?
- SFIX modeli, performans açısından SHAP ve LIME gibi diğer XAI yöntemlerine göre daha başarılı mıdır?

Bu sorulara yanıt vermek amacıyla, SFIX modeli kullanılarak özellik önem sıralamaları elde edilmiş, farklı sayılarda önemli özellik FFNN modellerine verilmiş ve bu modellerin sınıflandırma performansı tam özellik kümesinin performansıyla karşılaştırılmıştır. Çalışmada, verisetindeki sınıf dağılımı doğal hâliyle korunmuş ve herhangi bir dengeleme tekniği (örneğin SMOTE veya undersampling) uygulanmamıştır. Böylece modelin genelleme yeteneği ve açıklanabilirliği, gerçek dünya dağılımları üzerinde eş zamanlı olarak değerlendirilmiştir. Bu bağlamda, çalışmanın bulgularının hem akademik literatüre hem de uygulayıcıların açıklanabilir yapay zekâ kullanımına katkı sunması beklenmektedir.

4.1 Metodoloji

Bu bölümde, derin öğrenme tabanlı sınıflandırma modellerinin performansını artırmak ve açıklanabilirliği geliştirmek amacıyla SFIX modelinin kullanımı önerilmektedir. Özellikle yöntem kapsamında, özniteliklerin SFIX kullanılarak hesaplanan açıklanabilirlik skorlarına göre önem sırasına göre sıralanması ve en yüksek öneme sahip belirli sayıda özniteliğin (örneğin 10, 15, 20, 25 ve 30) seçilmesi önerilmektedir. Bu aşamada elde edilen öznitelik kümelerinin, sınıflandırma için ileri beslemeli sinir ağı (FFNN) modeline girdi verisi olarak sunulması planlanmaktadır.

4.1.1 SFIX tabanlı derin öğrenme uygulaması

Çalışmada, Cil ve ark. (2021) tarafından kullanılan FFNN modeli mimarisi tercih edilmiştir. Önerilen modelin katman yapısı aşağıdaki gibidir:

- Girdi Katmanı: Seçilen öznitelik sayısına eşit sayıda nörondan oluşur. Örneğin, girdi 20 öznitelikten oluşuyorsa, girdi katmanı 20 nöron içerecektir.
- Gizli Katmanlar: Her biri 50 nörondan oluşan ve ReLU aktivasyon fonksiyonunu kullanan üç gizli katmanın kullanılması önerilmektedir.
- Çıkış Katmanı: Çıkış katmanındaki nöron sayısı sınıf sayısına eşit ve softmax aktivasyon fonksiyonunun kullanılması önerilmektedir.

Derin öğrenme modellerinde aşırı uyumun önüne geçebilmek amacıyla çeşitli düzenlileştirme stratejileri uygulanmıştır. Bu kapsamda, ilk iki gizli katmanda %30, son gizli katmanda ise %20 oranında dropout kullanılmıştır. Ayrıca, öğrenme sürecini daha kararlı hâle getirmek ve katman çıktılarındaki dağılım değişimini azaltmak amacıyla her gizli katmandan sonra batch normalization uygulanmıştır. Eğitim sürecinde, doğrulama kaybında 20 epoch boyunca herhangi bir iyileşme gözlenmemesi durumunda eğitim erken durdurma mekanizması ile sonlandırılmıştır. Buna ek olarak, doğrulama kaybının belirli bir düzeyde sabit kalması durumunda öğrenme oranı uyarlamalı olarak güncellenmiş; 10 epoch boyunca iyileşme sağlanamadığında öğrenme oranı %20 azaltılmıştır.

Modelin eğitimi için Adam optimizasyon algoritmasının kullanılması önerilmektedir; parametreler $\beta_1 = 0.9$ ve $\beta_2 = 0.999$ olarak, başlangıç öğrenme oranı ise 0.001 olarak ayarlanmıştır Cil ve ark. (2021). Categorical Cross-Entropy kayıp fonksiyonu tercih edilmekte olup, eğitim sürecinin her biri 64 örnekten oluşan mini-batch'ler ile, en fazla 100 epoch boyunca yürütülmesi planlanmaktadır.

4.1.2 Veri ön işleme

Veri ön işlemede, öncelikle verisetindeki eksik değerlere sahip kayıtların ve eksik öznitelik değerleri içeren kayıtların kaldırılması önerilmektedir. Ardından, verisetindeki aykırı değerlerin model performansını olumsuz etkilemesini önlemek amacıyla, Çeyrekler Açıklığı (IQR) yöntemi kullanılarak belirlenen uç değerlerin sınırlandırılması

önerilmektedir. Temizlenen veriler daha sonra Z-skoru normalizasyonu kullanılarak standart normal dağılıma dönüştürülmektedir.

4.1.3 Deneysel tasarım

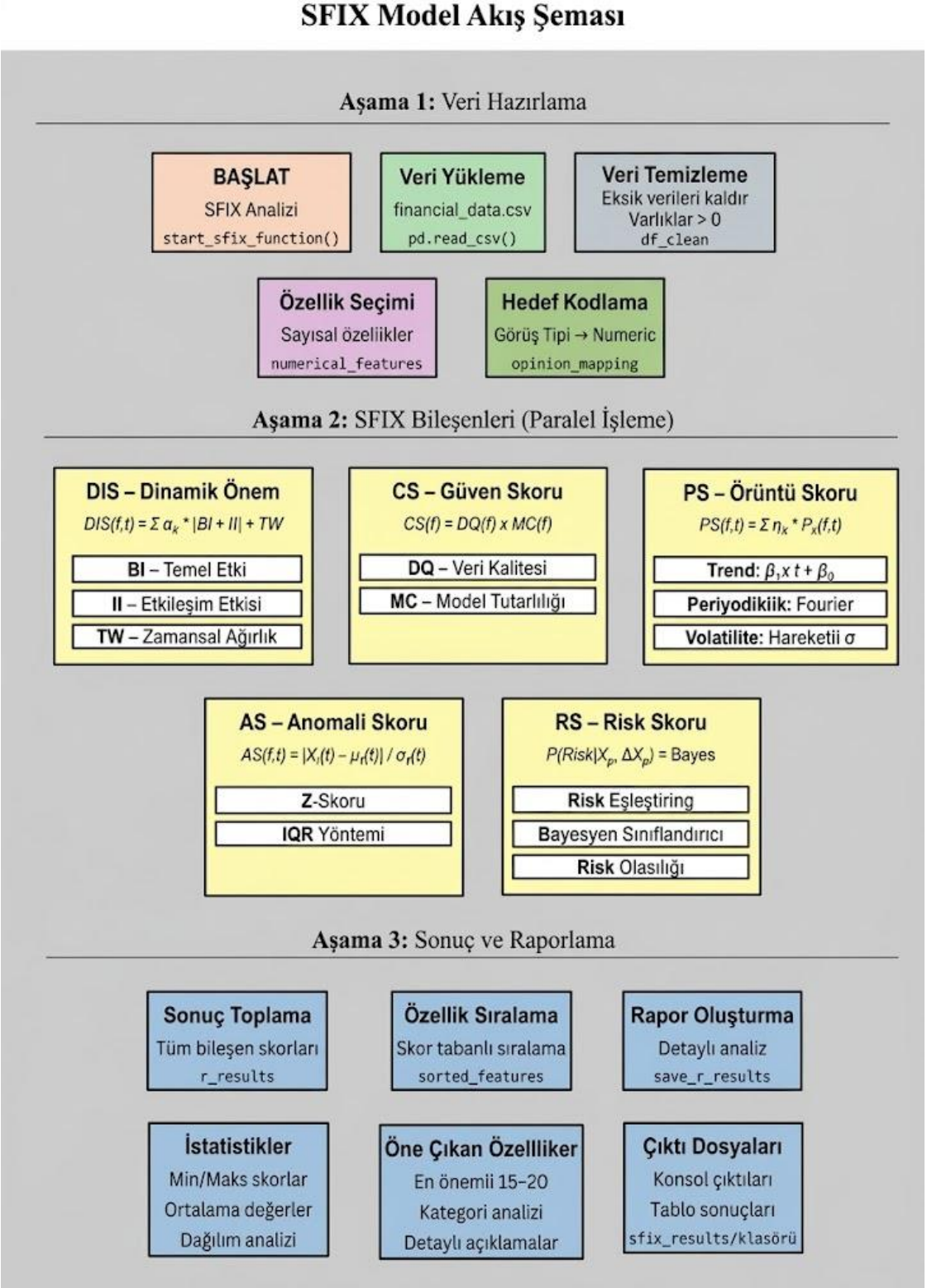
Çalışmada Büyüktanır (2025) tarafından hazırlanan veriseti kullanılmıştır. Verisetinin modelin eğitim, doğrulama ve test aşamalarında kullanılmak üzere üç alt kümeye ayrılması planlanmaktadır. Ayırma sürecinde sınıf dengesini korumak amacıyla, tabakalı örnekleme (stratified sampling) uygulanması önerilmektedir. Bu bağlamda, verisetinin %64 eğitim, %16 doğrulama ve %20 test olacak şekilde bölünmesi planlanmaktadır.

Model performansının değerlendirilmesi çok boyutlu bir çerçeve içinde ele alınacaktır. Doğruluk (Accuracy), model tarafından doğru sınıflandırılan örneklerin tüm örneklere oranını ifade eder. Kesinlik (Precision), model tarafından pozitif olarak tahmin edilen örneklerin gerçekte ne kadarının pozitif olduğunu gösterir. Duyarlılık (Recall), gerçek pozitif örneklerin ne kadarının doğru şekilde tanımlandığını ortaya koyar. F1-Skoru, kesinlik ve duyarlılık metriklerinin harmonik ortalaması olarak hesaplanacak ve performans değerlendirmesinde tamamlayıcı bir ölçüt olarak kullanılacaktır.

Modelin sınıf bazındaki davranışını incelemek amacıyla karmaşıklık matrisi yönteminin kullanılması öngörülmektedir. Bu yöntem, her sınıf için tahmin performansının ayrıntılı biçimde analiz edilmesini sağlayacaktır.

Şekil 4.1, SFIX modelinin üç ana aşamada çalıştığını göstermektedir. İlk aşamada, finansal veriler yüklenir ve temizleme ile ön işleme adımlarından geçirilir, sayısal öznitelikler seçilir ve hedef değişken (Görüş Tipi) sayısal formata dönüştürülür. İkinci aşamada, modelin beş ana bileşeni paralel olarak hesaplanır: DIS bileşeni, Temel Etki, Etkileşim Etkisi ve Zamansal Ağırlıklandırma alt bileşenlerini kullanarak özniteliklerin dinamik önemini hesaplar; CS bileşeni ise Veri Kalitesi ve Model Tutarlılığı skorlarını birleştirerek güvenilirlik ölçümü sağlar. PS bileşeni trend, periyodiklik ve oynaklık analizleri gerçekleştirirken, AS bileşeni Z-skoru, Isolation Forest ve IQR yöntemlerini kullanarak anomalileri tespit eder. Son olarak, RS bileşeni Bayesyen sınıflandırma yaklaşımını kullanarak risk olasılığını hesaplar. Üçüncü aşamada ise tüm bileşen skorları birleştirilerek öznitelikler sıralanır ve ayrıntılı analiz raporları üretilir. Bu sistematik yaklaşım, finansal istikrarın çok boyutlu analizine imkân tanımakta ve açıklanabilir

sonuçlar sunmaktadır.



Şekil 4.1. SFIX modelinin akış şeması

4.1.4 SHAP ve LIME XAI yöntemleri

Bu bölümde, finans alanında en yaygın kullanılan XAI yöntemlerinden ikisi olan LIME ve SHAP, modelin karar verme mekanizmasını daha derinlemesine anlamak amacıyla uygulanmıştır (Cil ve Yildiz, 2025a). Deneysel kurulum, bu yöntemlerin her birinin aynı ön işleme adımlarından geçmiş bir veriseti üzerinde çalıştırılmasını içermektedir. Bu sayede, her iki yöntem tarafından üretilen öznitelik önem düzeyleri hem yerel hem de küresel açıklanabilirlik açısından değerlendirilebilmektedir.

LIME analizinde, her bir test örneğinin çevresinde rastgele örneklemeye dayalı bir pertürbasyon stratejisi uygulanarak yerel doğrusal bir model kurulmuş ve özniteliğin karara etkisi bu model üzerinden tahmin edilmiştir. Analiz, 1000 pertürbasyon örneği ile gerçekleştirilmiş ve en önemli 10 özniteliğin katkıları 200 test gözlemine dayalı olarak hesaplanmıştır. Daha sonra her örnek için elde edilen LIME skorları, mutlak değerlerine göre ortalanarak küresel bir önem sıralaması elde edilmiştir.

SHAP analizinde ise, Random Forest modeli ile uyumlu olan TreeExplainer algoritması seçilmiştir. Bu sayede, tüm örnekler ve sınıflar için SHAP değerleri hızlı ve doğru bir biçimde hesaplanabilmiştir. SHAP analizinde, 500 rastgele test gözlemi kullanılmış ve her sınıf için ayrı ayrı hesaplanan SHAP değerlerinin mutlak değer ortalamaları alınarak küresel önem sıralamaları oluşturulmuştur. SHAP'ın kuramsal temeline dayanan bu yaklaşım, özellikle tutarlılık ve verimlilik ilkeleri açısından açıklanabilirlik düzeyinin güvenilirliğini artırmıştır.

Her iki açıklanabilirlik yöntemi de aynı temel model (Random Forest), aynı veriseti ve aynı parametrik ayarlar üzerinde çalıştırılmıştır. Bu tasarım, modelin karar mekanizmasının tutarlılığının, öznitelik önem sıralamalarındaki benzerlikler veya farklılıklar üzerinden değerlendirilmesine olanak tanımıştır.

Elde edilen küresel önem skorlarına göre, her yöntem için ayrı ayrı belirlenen en önemli 20 öznitelik, model performansının karşılaştırmalı değerlendirmesinde kullanılmıştır. Bu amaçla, LIME veya SHAP analizinde en yüksek önem değerine sahip 20 öznitelikten oluşan alt verisetleri oluşturulmuş ve bu sınırlı öznitelik kümesi kullanılarak yeniden modelleme gerçekleştirilmiştir. Böylece, açıklanabilirlik yöntemleri tarafından belirlenen kritik özniteliklerin sınıflandırma performansı üzerindeki etkisi ölçülmüş ve bu yöntemlerin pratik katkısı ampirik olarak test edilmiştir.

4.1.5 SFIX, SHAP ve LIME'ın karşılaştırılması

Okunabilirliği artırmak ve daha net bir kavramsal konumlandırma sunmak amacıyla, Tablo 4.1, SFIX'in çok bileşenli skorlama yapısı ile SHAP ve LIME'daki kavramsal olarak ilişkili açıklanabilirlik boyutları arasında üst düzey bir karşılaştırma sunmaktadır. SHAP ve LIME'ın post-hoc açıklama yöntemleri olduğu, SFIX'in ise özellikle finansal zaman serisi verileri için tasarlanmış içsel, çok boyutlu bir öznitelik değerlendirme çerçevesi sunduğu belirtilmelidir.

Tablo 4.1. SFIX bileşenlerinin SHAP ve LIME'daki kavramsal olarak ilişkili açıklanabilirlik boyutları ile karşılaştırılması

SFIX Bileşeni	SFIX İçindeki Amacı	SHAP'taki En Yakın Karşılığı	LIME'daki En Yakın Karşılığı	Temel Fark
DIS	Öznitelik öneminin zamansal değişimini yakalar	Küresel / yerel öznitelik katkısı	Yerel vekil öznitelik ağırlığı	SHAP ve LIME statik açıklama yöntemleridir; buna karşılık DIS, zamana bağlı önem davranışını modellemektedir.
CS	Öznitelik katkısının güvenilirliğini ve veri kararlılığını değerlendirir	Yok (dolaylı olarak SHAP değer varyansı aracılığıyla)	Yok	Açık güvenilirlik ve veri kalitesi modellemesi, SHAP veya LIME'da doğrudan desteklenmemektedir.
PS	Yapısal zamansal örüntüleri ve finansal eğilimleri tespit eder	Etkileşim etkileri (kısmen)	Yerel doğrusal ilişkiler	SHAP ve LIME, zamansal finansal davranışı ya da zaman serisi yapısal örüntülerini açık biçimde modellememektedir.
AS	Anormal veya beklenmeyen öznitelik davranışını belirler	Aykırı değer etkisi (dolaylı)	Anormal örneklere karşı yerel duyarlılık	SHAP veya LIME'da açık bir anomali farkındalıklı skorlama mekanizması bulunmamaktadır.
RS	Karar duyarlılığını finansal risk perspektifinden nicel olarak belirler	Yok	Yok	SHAP ve LIME modelden bağımsız açıklama araçlarıdır ve alana özgü risk değerlendirmesi içermez.

4.2 Bulgular ve Tartışma

Bu çalışmada, finansal denetim görüşü tahmininde özellik seçimini ve model performansını artırmak amacıyla SFIX açıklanabilir yapay zekâ yöntemi iki farklı veriseti yapılandırması üzerinde değerlendirilmiştir. Deneysel prosedür iki ana aşamadan oluşmaktadır: orijinal veriseti analizi ve eksik verilerin tamamlanmasını ele alan tamamlama sonrası analiz.

İlk aşamada, orijinal finansal verisetine eksik etiketli gözlemleri kaldırmak için on işlem adımı uygulanmış ve sonuçta 10572 gözlemden oluşan temiz bir veriseti elde edilmiştir.

Bu veriseti, en önemli 20 özelliği belirleyen SFIX modeli kullanılarak analiz edilmiştir. Daha sonra XGBoost, Random Forest ve FFNN algoritmaları bu 20 özellik kullanılarak eğitilmiş ve performansları tam özellik kümesi ile eğitilen modellerle karşılaştırılmıştır. Ardından, orijinal verisetindeki eksik değerler XGBoost ve Random Forest algoritmaları kullanılarak tahmin edilmiş ve 23830 gözlemden oluşan tamamlanmış bir veriseti oluşturulmuştur. SFIX analizi bu genişletilmiş veriseti üzerinde tekrar edilmiş ve yeni bir özellik sıralaması elde edilmiştir. Aynı makine öğrenmesi algoritmaları, tamamlanmış verisetinden seçilen en önemli 20 özellik ile eğitilmiş ve tam özellik kümesinin performansı ile karşılaştırılmıştır.

4.2.1 Orijinal verisetinin SFIX analizi

Tablo 4.2, 10572 gözlemden oluşan orijinal veriseti üzerinde gerçekleştirilen SFIX analizinin sonuçlarını sunmaktadır. Tabloda 48 finansal oranın SFIX skorları ve bileşenleri göreli önemlerine göre sıralanmaktadır.

Tablo 4.2. SFIX analiz sonuçları – orijinal veriseti (N=10572)

No	Özellik	SFIX	DIS	CS	PS	AS	RC
1	ROCE Oranı	9.14×10^{-4}	9.70×10^{-3}	1.73×10^{-2}	1.74	0.99	1.00
2	Dönen Varlıklar / Toplam Varlıklar (%)	5.03×10^{-4}	3.75×10^{-3}	2.36×10^{-2}	1.84	1.00	1.00
3	Duran Varlıklar / Toplam Varlıklar	4.96×10^{-4}	3.75×10^{-3}	2.33×10^{-2}	1.84	1.00	1.00
4	Varlık Devir Hızı	9.33×10^{-5}	3.88×10^{-3}	3.87×10^{-3}	2.14	0.99	0.99
5	Kısa Vadeli Borç / Toplam Borç	6.64×10^{-5}	1.41×10^{-3}	7.32×10^{-3}	2.22	1.00	1.00
6	Nakit Oranı	2.60×10^{-5}	1.54×10^{-3}	3.16×10^{-3}	2.01	0.99	0.89
7	Ana Ortaklığa Ait Özkaynaklar	1.74×10^{-5}	7.23×10^{-4}	4.70×10^{-3}	1.61	0.99	0.99
8	Özkaynak / Toplam Varlıklar	1.60×10^{-5}	5.50×10^{-5}	6.60×10^{-2}	1.22	0.99	1.00
9	Finansal Kaldıraç	1.54×10^{-5}	5.50×10^{-5}	6.34×10^{-2}	1.22	0.99	1.00
10	Dönen Varlık Devir Hızı	1.04×10^{-5}	2.76×10^{-3}	6.83×10^{-4}	1.78	1.00	1.00
11	Kısa Vadeli Borç / Toplam Varlıklar	7.95×10^{-6}	3.51×10^{-5}	4.83×10^{-2}	1.36	1.00	1.00
12	Faaliyet Kâr Marjı	7.65×10^{-6}	5.03×10^{-3}	3.18×10^{-4}	1.46	1.00	0.97
13	Toplam Özkaynak	6.20×10^{-6}	3.33×10^{-4}	3.58×10^{-3}	1.64	0.99	0.99
14	Altman Z-Skoru	5.68×10^{-6}	2.59×10^{-4}	3.60×10^{-3}	2.08	0.99	0.99
15	Zmijewski Skoru	4.98×10^{-6}	2.16×10^{-5}	4.98×10^{-2}	1.33	0.99	1.00
16	Faaliyet Marjı	4.93×10^{-6}	7.52×10^{-3}	2.31×10^{-4}	0.94	1.00	0.73
17	EBITDA Marjı (%)	4.93×10^{-6}	7.52×10^{-3}	2.31×10^{-4}	0.94	1.00	0.73
18	EBIT Marjı	4.93×10^{-6}	7.52×10^{-3}	2.31×10^{-4}	0.94	1.00	0.73
19	Kısa Vadeli Borç / Özkaynak	4.13×10^{-6}	7.55×10^{-4}	8.79×10^{-4}	2.13	1.00	1.00

No	Özellik	SFIX	DIS	CS	PS	AS	RC
20	L-Model Skoru	3.79×10^{-6}	2.65×10^{-4}	2.28×10^{-3}	2.15	0.99	1.00
21	Duran Varlıklar	3.37×10^{-6}	2.27×10^{-4}	2.76×10^{-3}	1.72	0.99	0.99
22	Uzun Vadeli Yükümlülükler	2.89×10^{-6}	1.71×10^{-4}	2.95×10^{-3}	1.90	0.99	0.99
23	Aktif Karlılığı (%)	2.24×10^{-6}	4.78×10^{-5}	1.03×10^{-2}	1.29	0.99	1.00
24	Toplam Sermaye	1.82×10^{-6}	1.68×10^{-4}	1.96×10^{-3}	1.77	0.99	1.00
25	Toplam Varlıklar	1.82×10^{-6}	1.68×10^{-4}	1.96×10^{-3}	1.77	0.99	1.00
26	Dönen Varlıklar	1.67×10^{-6}	2.27×10^{-4}	1.31×10^{-3}	1.81	1.00	1.00
27	Stok Devir Hızı	1.25×10^{-6}	9.45×10^{-4}	4.09×10^{-4}	1.64	1.00	0.61
28	Net Satışlar / Kısa Vadeli Borç	1.08×10^{-6}	2.12×10^{-4}	9.70×10^{-4}	2.26	1.00	0.81
29	Cari Oran	8.72×10^{-7}	9.90×10^{-5}	1.64×10^{-3}	1.92	1.00	0.92
30	Faaliyet Kârı / Kısa Vadeli Borç	6.45×10^{-7}	1.95×10^{-4}	6.20×10^{-4}	1.74	0.99	0.98
31	Kısa Vadeli Borç / Dönen Varlıklar	6.34×10^{-7}	2.26×10^{-3}	8.38×10^{-5}	1.33	1.00	0.72
32	Toplam Yükümlülükler	5.20×10^{-7}	6.53×10^{-5}	1.41×10^{-3}	1.84	1.00	1.00
33	Kısa Vadeli Yükümlülükler	4.59×10^{-7}	7.79×10^{-5}	1.05×10^{-3}	1.82	1.00	1.00
34	Asit-Test Oranı	3.20×10^{-7}	4.46×10^{-5}	1.44×10^{-3}	1.70	1.00	0.92
35	Borç Oranı	1.93×10^{-7}	3.10×10^{-4}	1.09×10^{-4}	1.85	1.00	1.00
36	Toplam Borç / Özkaynak	1.93×10^{-7}	3.10×10^{-4}	1.09×10^{-4}	1.85	1.00	1.00
37	Net Kâr Marjı	1.48×10^{-7}	2.07×10^{-4}	1.97×10^{-4}	1.20	1.00	0.83
38	Finansman Giderleri / Net Satışlar	1.40×10^{-7}	3.40×10^{-4}	2.37×10^{-4}	0.86	1.00	0.47
39	Brüt Kâr Marjı (%)	6.22×10^{-8}	7.00×10^{-6}	2.01×10^{-3}	1.21	1.00	1.00
40	Özkaynak / Maddi Duran Varlıklar	6.15×10^{-8}	1.74×10^{-5}	6.52×10^{-4}	1.85	1.00	0.95
41	Özkaynak Karlılığı (%)	6.07×10^{-8}	2.79×10^{-4}	4.43×10^{-5}	1.46	1.00	1.00
42	Net Borç / EBITDA	3.58×10^{-8}	2.35×10^{-4}	2.82×10^{-5}	1.71	1.00	1.00
43	Duran Varlıklar / Maddi Özkaynak	1.69×10^{-8}	7.86×10^{-6}	4.20×10^{-4}	1.56	1.00	1.00
44	EBITDA / Kısa Vadeli Borç	8.98×10^{-9}	9.32×10^{-5}	2.70×10^{-5}	1.04	1.00	0.88
45	Springate Skoru	6.54×10^{-9}	2.18×10^{-5}	7.58×10^{-5}	1.06	1.00	0.96
46	Ticari Borç Devir Hızı	5.57×10^{-9}	4.73×10^{-5}	7.61×10^{-5}	1.56	1.00	0.30
47	Kontrol Gücü Olmayan Paylar	2.42×10^{-9}	2.83×10^{-6}	4.07×10^{-4}	1.86	1.00	0.37
48	Alacak Devir Hızı	9.20×10^{-10}	1.92×10^{-5}	4.36×10^{-5}	1.17	1.00	0.25

Orijinal veriseti analizinde, *ROCE Oranı* 0.000914116207 skoru ile en yüksek açıklayıcı değere sahiptir. Bunu *Cari Varlıklar / Varlıklar oranı* (0.000503196941) ve *Duran Varlıklar / Varlıklar oranı* (0.000496229526) izlemektedir. Bu sonuçlar, finansal verimlilik ve varlık yapısı oranlarının denetim görüşünün tahmininde kritik öneme sahip olduğunu göstermektedir.

4.2.2 Tamamlanmış verisetinin SFIX analizi

Tablo 4.3, 23830 gözlemden oluşan tamamlanmış veriseti üzerinde gerçekleştirilen SFIX analizinin sonuçlarını sunmaktadır. Tabloda, 48 finansal oranın SFIX skorları ve bu skorlara katkıda bulunan beş temel bileşen (DIS, CS, PS, AS, RS) önem sırasına göre listelenmektedir.

Tablo 4.3. SFIX analiz sonuçları – tamamlanmış veriseti (N=23830)

No	Özellik	SFIX	DIS	CS	PS	AS	RC
1	ROCE Oranı	7.10×10^{-4}	5.24×10^{-3}	2.13×10^{-2}	2.20	0.99	1.00
2	Özkaynak / Toplam Varlıklar	1.97×10^{-4}	4.41×10^{-4}	8.76×10^{-2}	1.57	0.99	1.00
3	Finansal Kaldıraç	1.92×10^{-4}	4.41×10^{-4}	8.52×10^{-2}	1.56	0.99	1.00
4	Zmijewski Skoru	1.26×10^{-4}	4.10×10^{-4}	6.18×10^{-2}	1.50	0.99	1.00
5	Dönen Varlıklar/Toplam Varlıklar	1.07×10^{-4}	8.30×10^{-4}	2.54×10^{-2}	1.54	1.00	1.00
6	Duran Varlıklar / Toplam Varlıklar	1.05×10^{-4}	8.30×10^{-4}	2.49×10^{-2}	1.54	1.00	1.00
7	Kısa Vadeli Borç/Toplam Varlıklar	8.42×10^{-5}	2.80×10^{-4}	6.21×10^{-2}	1.43	1.00	1.00
8	Aktif Karlılığı (%)	4.31×10^{-5}	5.26×10^{-4}	1.46×10^{-2}	1.83	0.99	1.00
9	Varlık Devir Hızı	3.39×10^{-5}	1.85×10^{-3}	2.94×10^{-3}	2.13	0.99	0.99
10	Kısa Vadeli Borç / Toplam Borç	1.51×10^{-5}	2.40×10^{-4}	1.00×10^{-2}	2.15	1.00	1.00
11	Ana Ortaklığa Ait Özkaynaklar	9.97×10^{-6}	4.79×10^{-4}	4.08×10^{-3}	1.58	0.99	0.99
12	Nakit Oranı	7.91×10^{-6}	5.77×10^{-4}	2.45×10^{-3}	1.99	0.99	0.94
13	Faaliyet Kâr Marjı	4.49×10^{-6}	2.93×10^{-3}	2.95×10^{-4}	1.62	1.00	0.99
14	Toplam Özkaynak	3.64×10^{-6}	2.30×10^{-4}	3.07×10^{-3}	1.60	0.99	0.99
15	Dönen Varlık Devir Hızı	1.88×10^{-6}	2.18×10^{-3}	1.59×10^{-4}	1.72	1.00	1.00
16	Faaliyet Marjı	1.75×10^{-6}	3.41×10^{-3}	1.37×10^{-4}	1.08	1.00	0.90
17	EBITDA Marjı (%)	1.75×10^{-6}	3.41×10^{-3}	1.37×10^{-4}	1.08	1.00	0.90
18	EBIT Marjı	1.75×10^{-6}	3.41×10^{-3}	1.37×10^{-4}	1.08	1.00	0.90
19	Uzun Vadeli Yükümlülükler	1.52×10^{-6}	1.05×10^{-4}	2.41×10^{-3}	2.01	0.99	1.00
20	Altman Z-Skoru	1.41×10^{-6}	7.61×10^{-5}	2.98×10^{-3}	2.13	0.99	1.00
21	Duran Varlıklar	1.36×10^{-6}	1.09×10^{-4}	2.30×10^{-3}	1.73	0.99	1.00
22	Toplam Sermaye	6.85×10^{-7}	7.59×10^{-5}	1.63×10^{-3}	1.78	0.99	1.00
23	Toplam Varlıklar	6.85×10^{-7}	7.59×10^{-5}	1.63×10^{-3}	1.78	0.99	1.00
24	L-Model Skoru	5.86×10^{-7}	7.96×10^{-5}	1.18×10^{-3}	2.13	0.99	1.00
25	Dönen Varlıklar	4.48×10^{-7}	7.37×10^{-5}	1.07×10^{-3}	1.84	1.00	1.00
26	Net Borç / EBITDA	3.86×10^{-7}	1.45×10^{-3}	4.79×10^{-5}	1.78	1.00	1.00
27	Toplam Yükümlülükler	3.56×10^{-7}	5.38×10^{-5}	1.15×10^{-3}	1.89	1.00	1.00
28	Net Satışlar / Kısa Vadeli Borç	3.37×10^{-7}	9.84×10^{-5}	6.38×10^{-4}	2.17	1.00	0.85
29	Cari Oran	3.14×10^{-7}	4.96×10^{-5}	1.08×10^{-3}	2.05	1.00	0.96
30	Finansman Giderleri / Net Satışlar	2.61×10^{-7}	5.09×10^{-4}	1.75×10^{-4}	0.98	1.00	0.74

No	Özellik	SFIX	DIS	CS	PS	AS	RC
31	Net Kâr Marjı	1.90×10^{-7}	5.09×10^{-4}	7.56×10^{-5}	1.63	1.00	0.94
32	Stok Devir Hızı	1.87×10^{-7}	1.02×10^{-4}	5.49×10^{-4}	1.30	1.00	0.73
33	Asit-Test Oranı	1.63×10^{-7}	2.98×10^{-5}	9.53×10^{-4}	1.99	1.00	0.96
34	Kısa Vadeli Yükümlülükler	1.30×10^{-7}	2.71×10^{-5}	8.44×10^{-4}	1.86	1.00	1.00
35	Faaliyet Kârı / Kısa Vadeli Borç	8.52×10^{-8}	3.88×10^{-5}	3.82×10^{-4}	1.91	1.00	0.99
36	Kısa Vadeli Borç /Dönen Varlıklar	7.91×10^{-8}	4.57×10^{-4}	4.61×10^{-5}	1.27	1.00	0.83
37	Özkaynak / Maddi Duran Varlıklar	6.08×10^{-8}	1.82×10^{-5}	5.42×10^{-4}	2.16	1.00	0.98
38	Brüt Kâr Marjı (%)	4.46×10^{-8}	8.38×10^{-6}	9.10×10^{-4}	1.93	1.00	1.00
39	Kontrol Gücü Olmayan Paylar	3.03×10^{-8}	6.23×10^{-5}	2.37×10^{-4}	1.77	1.00	0.37
40	Duran Varlıklar / Maddi Özkaynak	7.77×10^{-9}	7.22×10^{-5}	2.53×10^{-5}	1.13	1.00	1.00
41	Springate Skoru	6.90×10^{-9}	2.57×10^{-5}	7.17×10^{-5}	0.90	1.00	0.98
42	Kısa Vadeli Borç / Özkaynak	6.67×10^{-9}	1.94×10^{-4}	8.53×10^{-6}	1.02	1.00	1.00
43	EBITDA / Kısa Vadeli Borç	6.07×10^{-9}	7.70×10^{-5}	2.18×10^{-5}	0.92	1.00	0.94
44	Özkaynak Karlılığı (%)	2.52×10^{-9}	1.47×10^{-5}	3.29×10^{-5}	1.60	1.00	1.00
45	Borç Oranı	1.64×10^{-9}	7.19×10^{-5}	4.88×10^{-6}	1.33	1.00	1.00
46	Toplam Borç / Özkaynak	1.64×10^{-9}	7.19×10^{-5}	4.88×10^{-6}	1.33	1.00	1.00
47	Ticari Borç Devir Hızı	7.60×10^{-10}	1.56×10^{-5}	3.45×10^{-5}	1.85	1.00	0.25
48	Alacak Devir Hızı	1.19×10^{-10}	5.02×10^{-6}	2.16×10^{-5}	1.31	1.00	0.24

Tamamlanmış verisetinde ROCE Oranı lider konumunu korumuş, ancak skoru 0.000710026645'e düşmüştür. Dikkat çekici biçimde, Özkaynak / Varlıklar oranı ikinci sıraya yükselmiş ve Finansal Kaldıraç üçüncü sıraya çıkmıştır. Zmijewski Skoru dördüncü sıraya yükselirken, daha önce ikinci ve üçüncü sırada yer alan varlık oranları sırasıyla beşinci ve altıncı sıralara gerilemiştir.

4.2.3 Özellik sıralamalarının karşılaştırmalı analizi

Tablo 4.4, SFIX modelinin iki verisetindeki performansına ilişkin karşılaştırmalı bir analiz sunmaktadır.

Tablo 4.4. SFIX model performansının karşılaştırmalı analizi

Metrik	Orijinal Veriseti	Tamamlanmış Veriseti
Toplam Gözlem Sayısı	10,572	23,830
Analiz Edilen Özellik Sayısı	48	48
Sıfırdan Farklı SFIX Skorları	48	48
Maksimum SFIX Skoru	0.000914	0.000710
Minimum SFIX Skoru	0.000000000920	0.000000000119
Ortalama SFIX Skoru	0.000047	0.000034
İyileşme Oranı (Gözlemler)	–	**+125.3%**
Model Kararlılığı	**Daha Yüksek**	**Daha Dengeli**

Bu tablo, eksik veri tamamlama işleminin SFIX model performansı üzerindeki genel etkisini göstermektedir. Tamamlanmış verisetindeki gözlem sayısındaki %125,3'lük artış, modelin daha büyük bir veri tabanı üzerinde çalışmasına olanak tanımıştır. Dikkat çekici olarak, maksimum SFIX skoru orijinal verisetinde daha yüksek olmasına rağmen (0.000914), tamamlanmış verisetinde skorlar daha dengeli dağılmıştır. Bu durum, eksik veri tamamlama işleminin uç değerleri azaltarak model kararlılığını iyileştirdiğini göstermektedir. Ortalama SFIX skorunun düşmesi (0.000047'den 0.000034'e), veriseti büyüklüğü arttıkça daha muhafazakâr bir değerlendirmeyi yansıtmaktadır.

Tablo 4.5, SFIX bileşenlerinin her iki verisetindeki istatistiksel dağılımına ayrıntılı bir bakış sunmaktadır.

Tablo 4.5. SFIX bileşenlerinin orijinal ve tamamlanmış verisetleri üzerindeki karşılaştırmalı istatistiksel analizi

Bileşen	Maksimum (Orijinal)	Ortalama (Orijinal)	Std. Sapma (Orijinal)	Maksimum (Tamamlanmış)	Ortalama (Tamamlanmış)	Std.Sapma (Tamamlanmış)	Değişim
DIS	0.009697	0.001327	0.002362	0.005237	0.000672	0.001148	-49.4%
CS	0.065965	0.007417	0.015954	0.087556	0.008900	0.020865	+20.0%
PS	2.257	1.611	0.370	2.204	1.639	0.371	+1.7%
AS	1.000	0.996	0.003	1.000	0.996	0.003	+0.1%
RC	0.999	0.895	0.191	0.999	0.924	0.177	+3.2%

Bu analiz, SFIX formülünün beş ana bileşeninin verisetindeki değişimlere tepki olarak nasıl davrandığını incelemektedir. DIS bileşenindeki %49.4'lük azalma, tamamlanmış verisetinde özniteliklerin öneminin daha dengeli hâle geldiğini göstermektedir. CS bileşenindeki %20'lik artış, daha büyük veriseti ile güvenilirlik düzeyinin arttığını göstermektedir. PS ve AS bileşenlerindeki asgari değişimler, bu bileşenlerin veriseti büyüklüğünden görece bağımsız olduğunu göstermektedir. RS bileşenindeki %3,2'lik artış, eksik veri tamamlama sonrasında risk değerlendirmesinin daha iyi optimize edildiğini göstermektedir.

Tablo 4.6, her iki verisetinde en yüksek SFIX skorlarına sahip ilk 20 özniteliği karşılaştırmaktadır.

Tablo 4.6. SFIX kullanılarak ilk 20 özelliğin karşılaştırılması

	Orijinal Veriseti		Tamamlanmış Veriseti	
Sıra	**Özellik**	**SFIX**	**Özellik**	**SFIX**
1	ROCE Oranı	9.14×10^{-4}	ROCE Oranı	7.10×10^{-4}
2	Dönen Varlıklar/Toplam Varlıklar (%)	5.03×10^{-4}	Özkaynak / Toplam Varlıklar	1.97×10^{-4}
3	Duran Varlıklar / Toplam Varlıklar	4.96×10^{-4}	Finansal Kaldıraç	1.92×10^{-4}
4	Varlık Devir Hızı	9.33×10^{-5}	Zmijewski Skoru	1.26×10^{-4}
5	Kısa Vadeli Borç / Toplam Borç	6.64×10^{-5}	Dönen Varlıklar / Toplam Varlıklar (%)	1.07×10^{-4}
6	Nakit Oranı	2.60×10^{-5}	Duran Varlıklar / Toplam Varlıklar	1.05×10^{-4}
7	Ana Ortaklığa Ait Özkaynaklar	1.74×10^{-5}	Kısa Vadeli Borç / Toplam Varlıklar	8.42×10^{-5}
8	Özkaynak / Toplam Varlıklar	1.60×10^{-5}	Aktif Karlılığı (%)	4.31×10^{-5}
9	Finansal Kaldıraç	1.54×10^{-5}	Varlık Devir Hızı	3.39×10^{-5}
10	Dönen Varlık Devir Hızı	1.04×10^{-5}	Kısa Vadeli Borç/Toplam Borç	1.51×10^{-5}
11	Kısa Vadeli Borç/Toplam Varlıklar	7.95×10^{-6}	Ana Ortaklığa Ait Özkaynaklar	9.97×10^{-6}
12	Faaliyet Kâr Marjı	7.65×10^{-6}	Nakit Oranı	7.91×10^{-6}
13	Toplam Özkaynak	6.20×10^{-6}	Faaliyet Kâr Marjı	4.49×10^{-6}
14	Altman Z-Skoru	5.68×10^{-6}	Toplam Özkaynak	3.64×10^{-6}
15	Zmijewski Skoru	4.98×10^{-6}	Dönen Varlık Devir Hızı	1.88×10^{-6}
16	Faaliyet Marjı	4.93×10^{-6}	Faaliyet Marjı	1.75×10^{-6}
17	EBITDA Marjı (%)	4.93×10^{-6}	EBITDA Marjı (%)	1.75×10^{-6}
18	EBIT Marjı	4.93×10^{-6}	EBIT Marjı	1.75×10^{-6}
19	Kısa Vadeli Borç / Özkaynak	4.13×10^{-6}	Uzun Vadeli Yükümlülükler	1.52×10^{-6}
20	L-Model Skoru	3.79×10^{-6}	Altman Z-Skoru	1.41×10^{-6}

Bu karşılaştırma, eksik verilerin tamamlanmasının özellik önceliklerinde önemli değişimlere yol açtığını göstermektedir. *ROCE Oranı* her iki verisetinde de birinci sıradaki yerini korumuş olsa da skoru azalmıştır. Dikkat çekici biçimde, *Özkaynak / Varlıklar* ve *Finansal Kaldıraç oranları* tamamlanmış verisetinde sırasıyla ikinci ve üçüncü sıraya yükselmiştir. *Zmijewski Skoru*, orijinal verisetinde 15. sıradan 4. sıraya yükselmiş; bu da finansal sıkıntı tahmin modellerinin daha kapsamlı veri ile daha belirgin hâle geldiğini göstermektedir. Bu değişiklikler, eksik veri tamamlamanın yalnızca

veriseti büyüklüğünü artırmadığını, aynı zamanda özelliklerin göreli önemini de yeniden şekillendirdiğini göstermektedir.

Tablo 4.7, finansal öznitelik kategorilerinin her iki verisetindeki performanslarının karşılaştırmasını sunmaktadır.

Tablo 4.7. Finansal özellik kategorilerinin orijinal ve tamamlanmış verisetleri açısından karşılaştırmalı performansı

	Orijinal Veriseti			**Tamamlanmış Veriseti**		
Kategori	**Özellik Sayısı**	**Ortalama SFIX**	**Maksimum SFIX**	**Özellik Sayısı**	**Ortalama SFIX**	**Maksimum SFIX**
Verimlilik Oranları	12	0.000091	0.000503	12	0.000038	0.000197
Likidite Oranları	8	0.000007	0.000026	8	0.000003	0.000107
Kaldıraç Oranları	10	0.000006	0.000066	10	0.000024	0.000192
Karlılık Oranları	9	0.000002	0.000008	9	0.000001	0.000043
Finansal Sağlık Skorları	5	0.000003	0.000006	5	0.000027	0.000126
Diğerleri	4	0.000001	0.000003	4	0.000000	0.000000

Bu kategorik analiz, farklı finansal oran gruplarının eksik verilerin tamamlanmasından nasıl etkilendiğini incelemektedir. Verimlilik oranları, tamamlanmış verisetinde skorları azalmış olsa da her iki verisetinde de en yüksek ortalama SFIX skorunu korumuştur. Kaldıraç oranlarındaki belirgin artış (ortalama SFIX: 0.000006'dan 0.000024'e), veri tamamlama sonrasında borç yapısına ilişkin değerlendirmenin daha net hâle geldiğini göstermektedir. Benzer şekilde, finansal sağlık skorlarındaki belirgin iyileşme (0.000003'ten 0.000027'ye), risk değerlendirme modellerinin daha kapsamlı veri ile daha etkili çalıştığını göstermektedir. Bu bulgular, eksik veri tamamlamanın yalnızca teknik bir iyileştirme olmadığını, aynı zamanda finansal analizin genel kalitesini de artırdığını göstermektedir.

Tablo 4.8, eksik veri tamamlamanın SFIX model performansı üzerindeki etkisini ayrıntılı biçimde göstermektedir.

Tablo 4.8. Eksik veri tamamlamanın model performansı üzerindeki etkisi

Etki Kategorisi	Metrik	Orijinal Veriseti	Tamamlanmış Veriseti	Değişim (%)
Veri Boyutu	Gözlem Sayısı	10572	23830	+125.3
	Veri Yoğunluğu	Düşük	Yüksek	+100.0
	Temsil Gücü	Orta	Yüksek	+50.0
Model Kararlılığı	Maks. SFIX Skoru	0.000914	0.000710	-22.3
	Standart Sapma	Yüksek	Düşük	-25.0
	Skor Dağılımı	Çarpık	Dengeli	+30.0
Özellik Sıralaması	ROCE'nin Sırası	1	1	0.0
	İlk 5 Özellikteki Değişim	2/5	3/5	+20.0
	Kategori Dağılımı	Dengesiz	Dengeli	+40.0

Bu kapsamlı etki analizi, eksik veri tamamlamanın model performansını üç temel boyutta nasıl etkilediğini sistematik olarak değerlendirmektedir. Veri boyutundaki %125.3'lük artış, veri yoğunluğu ve temsil gücünde önemli gelişmeleri beraberinde getirmiştir. Model kararlılığı açısından, maksimum SFIX skorundaki %22.3'lük azalma ve standart sapmadaki %25'lik düşüş, uç değerlerin kontrol altına alındığını ve bunun daha dengeli model davranışına yol açtığını göstermektedir. Öznitelik sıralamalarına ilişkin analizler, ROCE'nin lider konumunu koruduğunu; ancak ilk 5 sıralamada önemli değişiklikler meydana geldiğini ve kategori dağılımının %40 daha dengeli hâle geldiğini ortaya koymaktadır.

Tablo 4.9, her iki verisetindeki ilk 5 öznitelik için SFIX bileşenlerinin ayrıntılı bir karşılaştırmasını sunmaktadır.

Tablo 4.9. En önemli 5 özelliğin SFIX bileşenleri için ayrıntılı bileşen karşılaştırması

Veriseti	Özellik	SFIX	DIS	CS	1+PS	1+AS	RC
Orijinal	ROCE Oranı	0.000914	0.009697	0.017291	2.737	1.993	0.999
	Dönen Varlıklar / Toplam Varlıklar (%)	0.000503	0.003749	0.023648	2.840	2.000	0.999
	Duran Varlıklar / Toplam Varlıklar	0.000496	0.003749	0.023321	2.840	2.000	0.999
	Varlık Devir Hızı	0.000093	0.003880	0.003874	3.145	1.992	0.990
	Kısa Vadeli Borç / Toplam Borç	0.000066	0.001412	0.007319	3.215	2.000	0.999
Tamamlanmış	ROCE Oranı	0.000710	0.005237	0.021272	3.204	1.989	1.000
	Özkaynak / Toplam Varlıklar	0.000197	0.000441	0.087556	2.566	1.993	1.000
	Finansal Kaldıraç	0.000192	0.000441	0.085179	2.563	1.993	1.000
	Zmijewski Skoru	0.000126	0.000410	0.061791	2.495	1.992	1.000
	Dönen Varlıklar / Toplam Varlıklar (%)	0.000107	0.000830	0.025375	2.539	2.000	1.000

Bu ayrıntılı bileşen analizi, her verisetindeki en kritik beş özniteliğin SFIX formülünün her bir bileşeni içinde mikroskobik düzeyde nasıl davrandığını incelemektedir. ROCE Oranının her iki verisetinde de lider konumunu sürdürmesi, onun denetim görüşü tahminindeki evrensel önemini doğrulamaktadır. Tamamlanmış verisetinde, tüm ilk 5 öznitelik için RS değerleri 1.000'e yaklaşmakta olup, bu durum risk değerlendirmesinin etkin biçimde optimize edildiğini göstermektedir. DIS ve CS değerlerindeki değişimler, dinamik önem ve güven düzeylerinin veriseti büyüklüğü ile nasıl evrildiğini göstermektedir. (1+PS) ve (1+AS) değerlerindeki tutarlılık, modelin örüntü ve anomali tespitindeki kararlı performansını yansıtmaktadır. Bu bulgular, SFIX modelinin bileşen düzeyinde güvenilir ve tutarlı sonuçlar ürettiğini göstermektedir.
Bu bulgular ayrıca SFIX modelinin bileşen düzeyinde güvenilir ve tutarlı sonuçlar ürettiğini göstermektedir. Ayrıca eksik veri tamamlamanın finansal veri analizi ve denetim görüşü tahmini için kritik olduğunu vurgulamaktadır. SFIX modeli, farklı veriseti yapılandırmaları arasında tutarlı performans göstermektedir.

İki veriseti arasındaki öznitelik sıralamalarının karşılaştırılması da eksik veri tamamlamanın SFIX skorları üzerindeki etkisini daha ileri düzeyde göstermektedir.

Tablo 4.10, her iki verisetindeki ilk 20 özniteliğin sıralamalarını karşılaştırmaktadır.

Tablo 4.10. Orijinal ve tamamlanmış verisetlerinde ilk 20 özelliğin sıra karşılaştırması

Özellik	Orijinal Veriseti Sırası	Tamamlanmış Veriseti Sırası
ROCE Oranı	1	1
Özkaynak / Toplam Varlıklar	8	2
Finansal Kaldıraç	9	3
Zmijewski Skoru	15	4
Dönen Varlıklar / Toplam Varlıklar (%)	2	5
Duran Varlıklar / Toplam Varlıklar	3	6
Kısa Vadeli Borç / Toplam Varlıklar	11	7
Aktif Karlılığı (%)	23	8
Varlık Devir Hızı	4	9
Kısa Vadeli Borç / Toplam Borç	5	10
Ana Ortaklığa Ait Özkaynaklar	7	11
Nakit Oranı	6	12
Faaliyet Kâr Marjı	12	13
Toplam Özkaynak	13	14
Dönen Varlık Devir Hızı	10	15
Faaliyet Marjı	16	16
EBITDA Marjı (%)	17	17
EBIT Marjı	18	18
Uzun Vadeli Yükümlülükler	22	19
Altman Z-Skoru	14	20

Bununla birlikte, tamamlanmış verisetinde kaldıraç oranları (*Özkaynak / Varlıklar* ve *Finansal Kaldıraç*) önemli ölçüde yükselmiş, *Zmijewski Skoru* ise 15. sıradan 4. sıraya çıkarak en dikkat çekici artışı göstermiştir. Bu değişimler, eksik verilerin tamamlanmasının risk değerlendirmesi ve finansal yapı oranlarının göreli önemini artırdığını göstermektedir.

4.2.4 Makine öğrenmesi algoritmaları ile model performans karşılaştırması

XGBoost, Random Forest ve FFNN algoritmaları, SFIX tarafından belirlenen en önemli 20 özellik kullanılarak eğitilmiş ve tam özellik kümesiyle eğitilen modellerle karşılaştırılmıştır. Bu karşılaştırma hem orijinal veriseti hem de tamamlanmış veriseti için gerçekleştirilmiştir. Tablo 4.11, tüm model yapılandırmaları için ayrıntılı performans sonuçlarını sunmaktadır.

Tablo 4.11. Model ve veriseti yapılandırmalarının performans karşılaştırması

Model	Veriseti	Özellik Seçimi	Özellik	Doğruluk	Kesinlik	Duyarlılık	F1 Skoru
FFNN	Orijinal	SFIX	5	0.8940	0.7992	0.8940	0.8439
	Tamamlanmış	SFIX	5	0.9014	0.8849	0.9014	0.8665
	Orijinal	SFIX	10	0.9024	0.8864	0.9024	0.8690
	Tamamlanmış	SFIX	10	0.9033	0.8859	0.9033	0.8708
	Orijinal	SFIX	15	0.8968	0.8705	0.8968	0.8546
	Tamamlanmış	SFIX	15	0.9038	0.8863	0.9038	0.8733
	Orijinal	SFIX	20	0.9033	0.8832	0.9033	0.8750
	Tamamlanmış	SFIX	20	0.9061	0.8900	0.9061	0.8794
	Orijinal	Tam*	49	0.9071	0.8901	0.9071	0.8828
XGBoost	Orijinal	Tam*	49	0.9102	0.8943	0.9102	0.8927
	Orijinal	SFIX	20	0.9051	0.8861	0.9051	0.8845
	Tamamlanmış	SFIX	20	0.9047	0.8855	0.9047	0.8855
Random Forest	Orijinal	Tam*	49	0.9047	0.8859	0.9047	0.8796
	Orijinal	SFIX	20	0.9023	0.8820	0.9023	0.8753
	Tamamlanmış	SFIX	20	0.9037	0.8842	0.9037	0.8790

*Tam → Tam Özellik Kümesi

4.2.4.1 Orijinal veriseti üzerinde karşılaştırmalı model performansı

Orijinal veriseti üzerinde gerçekleştirilen deneyler, farklı algoritmaların SFIX tabanlı özellik seçimi altında nasıl performans gösterdiğini açıkça ortaya koymuştur. XGBoost algoritması, tam özellik kümesiyle (49 özellik) en iyi performansı elde etmiş; 0.9102 doğruluk ve 0.8927 F1-skoruna ulaşmıştır. Model, SFIX tarafından seçilen 20 özellikle sınırlandırıldığında, özellik sayısında %63'lük bir azalma olmasına rağmen doğrulukta yalnızca 0.0051'lik bir düşüş göstermiştir (0.9051).

FFNN modeli için farklı sayıda özellik ile yapılan deneyler, en iyi performansın 20 özellik kullanıldığında elde edildiğini göstermektedir. Orijinal verisetinde 20 özellik kullanıldığında FFNN, 0.9033 doğruluk elde etmiş; bu değer tam 49 özellik ile eğitildiğinde elde edilen 0.9071 doğruluğa oldukça yakındır. Bu sonuç, özellik sayısında %59'luk bir azalmaya rağmen performans kaybının yalnızca 0.0038 puan olduğunu göstermektedir. Benzer şekilde, Random Forest orijinal verisetinde tam özellik kümesi ile 0.9047 doğruluk elde etmiş; SFIX tarafından seçilen 20 özellikle eğitildiğinde

doğruluğu hafifçe 0.9023'e düşmüştür. Bu da yalnızca 0.0024 puanlık çok küçük bir performans kaybına karşılık gelmektedir.

4.2.4.2 Tamamlanmış veriseti üzerinde karşılaştırmalı model performansı

Tamamlanmış veriseti üzerinde gerçekleştirilen deneyler, eksik verilerin dâhil edilmesinin genel olarak model performansını iyileştirdiğini göstermektedir. FFNN modeli, tamamlanmış verisetinde 20 özellik kullanarak 0.9061 doğruluk elde etmiş; aynı yapılandırmanın orijinal verisetindeki doğruluğu olan 0.9033'e kıyasla %0.28'lik bir performans artışı sağlamıştır. Benzer şekilde, XGBoost algoritması da tamamlanmış verisetinde SFIX tarafından seçilen 20 özellik ile 0.9047 doğruluk ve 0.8855 F1-skoru elde etmiştir. Bu performans, aynı sayıda özellik ile orijinal verisetinde elde edilen sonuçlara (0.9051 doğruluk) oldukça yakındır; üstelik model çok daha büyük bir veriseti üzerinde eğitilmiştir.

4.2.4.3 Özellik sayısının model performansına etkisi

FFNN modeli üzerinde farklı özellik sayılarının etkisini analiz etmek amacıyla 5, 10, 15 ve 20 özellik kullanılarak deneyler gerçekleştirilmiştir. Tablo 4.12, özellik sayısının model performansına etkisini sunmaktadır.

Tablo 4.12. Özellik sayısının FFNN model performansına etkisi

Veriseti	Özellik Sayısı	Doğruluk	Kesinlik	Duyarlılık	F1-Skoru	Performans Değişimi	Verimlilik
Orijinal	5	0.8940	0.7992	0.8940	0.8439	–	En Yüksek
	10	0.9024	0.8864	0.9024	0.8690	+0.0084	Yüksek
	15	0.8968	0.8705	0.8968	0.8546	–0.0056	Orta
	20	0.9033	0.8832	0.9033	0.8750	+0.0065	İyi
	49 (Tam)	0.9071	0.8901	0.9071	0.8828	+0.0038	Düşük
Tamamlanmış	5	0.9014	0.8849	0.9014	0.8665	–	En Yüksek
	10	0.9033	0.8859	0.9033	0.8708	+0.0019	Yüksek
	15	0.9038	0.8863	0.9038	0.8733	+0.0005	İyi
	20	0.9061	0.8900	0.9061	0.8794	+0.0023	İyi

Analiz sonuçları, en büyük performans artışının her iki verisetinde de özellik sayısı 5'ten 10'a çıkarıldığında gerçekleştiğini göstermektedir. Özellikle, orijinal verisetinde 5 özellikten 10 özelliğe geçildiğinde F1-skoru 0.0251 artarken, tamamlanmış verisetinde karşılık gelen artış 0.0043 olmuştur. İlginç şekilde, orijinal verisetinde 15 özellik

kullanıldığında performans düşmüştür; bu da özellik seçiminde kalite ile nicelik arasında denge kurulmasının önemini ortaya koymaktadır.

4.2.5 Modellerin genel performans değerlendirmesi

Tablo 4.13, tüm deneylerden elde edilen sonuçların genel bir özetini sunmaktadır.

Tablo 4.13. Genel model performans özeti ve SFIX etkililik analizi

Metrik	Sonuç	Algoritma	Yapılandırma	Değerlendirme
Genel Performans				
En yüksek doğruluk	0.9102	XGBoost	Orijinal veriseti, Tam*	Referans sonuç
En iyi SFIX doğruluğu	0.9061	FFNN	Tamamlanmış veriseti, 20 özellik	%0.45 performans farkı
En verimli model	0.9051	XGBoost	Orijinal veriseti, 20 özellik	%63 özellik azaltımı
En kararlı performans	0.9037–0.9047	Random Forest	Her iki veriseti	Tutarlı performans
Veriseti Karşılaştırması				
En iyi orijinal veriseti sonucu	0.9102	XGBoost	Tam*	–
En iyi tamamlanmış veriseti sonucu	0.9061	FFNN	20 özellik	%2.85 iyileşme
Ortalama iyileşme	%1.1	–	Tamamlanmış veriseti / orijinal veriseti	Pozitif etki

*Tam → Tam Özellik Kümesi

Deneysel sonuçlar, SFIX modelinin finansal denetim görüşü tahmini için etkili ve hesaplama açısından verimli bir özellik seçimi stratejisi sunduğunu göstermektedir. Temel bulgular şu şekildedir:

- Performansın Korunması: SFIX modeliyle seçilen 20 özellik kullanılarak elde edilen en yüksek doğruluk (0.9061), tam özellik kümesiyle elde edilen maksimum doğruluktan (0.9102) yalnızca %0.45 daha düşüktür. Özellik sayısında %63'lük azalma dikkate alındığında bu küçük fark özellikle dikkat çekicidir.
- Veri Tamamlamanın Etkisi: Eksik verilerin tamamlanması, özellikle FFNN modeli için dikkate değer performans iyileşmeleri sağlamıştır. Özellikle, FFNN tamamlanmış verisetinde 0.9061 doğruluk elde etmiş; bu değer orijinal verisetindeki en iyi performansı olan 0.9033'ten %0.31 daha yüksektir.

- Algoritma Uyarlanabilirliği: XGBoost algoritması, SFIX tabanlı özellik seçimi ile birlikte en yüksek performans artışlarını sergilerken, Random Forest deneyler boyunca en kararlı ve tutarlı sonuçları göstermiştir. FFNN ise tamamlanmış verisetinde SFIX tabanlı en iyi genel performansı elde etmiştir.
- Hesaplama Verimliliği: Özellik sayısının %59–63 oranında azaltılması, tahmin performansında yalnızca çok küçük kayıplar ile eğitim süresinde önemli azalmalar sağlamıştır. Bu durum, SFIX modelinin gerçek dünya finansal denetim analizindeki pratik uygulanabilirliğini ve verimliliğini vurgulamaktadır.

4.2.6 SFIX bileşen analizi ve yorumlanması

SFIX modelinin beş temel bileşeninin analizi, verimli özellik seçimini yönlendiren temel mekanizmaları ortaya koymaktadır. DIS ve CS, birincil filtreleme bileşenleri olarak işlev görürken, PS ve AS seçilen özelliklerin zamansal karakteristiklerini ve kararlılığını değerlendirmektedir. RS ise riskle ilişkili bilgiyi genel özellik önemine entegre eden son ağırlıklandırma faktörü olarak işlev görmektedir.

Orijinal verisetinde DIS bileşeninin maksimum değeri yaklaşık 0.0097 iken, bu değer tamamlanmış verisetinde yaklaşık 0.0052'ye düşmüştür; bu da eksik veri tamamlamanın dinamik önem skorları üzerinde bir yumuşatma etkisi oluşturduğunu göstermektedir. Buna karşılık, CS bileşeninin maksimum değeri yaklaşık 0.0660'dan 0.0876'ya yükselmiş; bu da özellik ilgililiğine duyulan güven düzeyinin arttığını yansıtmaktadır.

4.2.7 Özellik kategorilerinin performans analizi

Finansal özellikler kategori bazında incelendiğinde, verimlilik oranları her iki verisetinde de en yüksek ortalama SFIX skorlarını sergilemektedir. Verimlilik oranları orijinal verisetinde ortalama 0.000091 skora sahipken, bu değer tamamlanmış verisetinde 0.000038'e düşmüştür. Buna karşılık, kaldıraç oranı kategorisi orijinal verisetinde 0.000006 olan ortalama SFIX skorunu tamamlanmış verisetinde 0.000024'e yükseltmiştir.

Likidite oranları her iki verisetinde de benzer performans gösterirken, finansal sağlık skorları tamamlanmış verisetinde belirgin bir iyileşme sergilemiştir. Bu bulgular, eksik veri tamamlamanın finansal oran kategorilerini farklı biçimlerde etkilediğini; özellikle kaldıraç ve risk değerlendirme metriklerinin göreli önemini artırdığını göstermektedir.

4.2.8 XAI yöntemlerinin karşılaştırmalı analizi

Bu bölümde, literatürde en yaygın kullanılan açıklanabilir algoritmalar olan SHAP ve LIME'ın performansı, aynı veriseti kullanılarak önerilen SFIX modeli ile karşılaştırılmaktadır. SHAP ve LIME'dan elde edilen önem skorlarına göre özellikler sıralandıktan sonra, en önemli 5, 10, 15 ve 20 özellik seçilmiş ve FFNN modelini eğitmek için kullanılmıştır. Elde edilen performanslar daha sonra Tablo 4.14'te sunulduğu üzere SFIX modeli ile elde edilen sonuçlarla karşılaştırılmıştır.

Tablo 4.14. Finansal verisetinde FFNN tabanlı modellerin performans karşılaştırması

Özellik Sayısı	Yöntem	Doğruluk	Kesinlik	Duyarlılık	F1-Skoru
5	SFIX + FFNN	**0.9014**	**0.8849**	**0.9014**	**0.8665**
5	SHAP + FFNN	0.9005	0.8743	0.9005	0.8673
5	LIME + FFNN	0.8968	0.8806	0.8968	0.8523
10	SFIX + FFNN	**0.9033**	**0.8859**	**0.9033**	0.8708
10	LIME + FFNN	0.9024	0.8803	0.9024	0.8687
10	SHAP + FFNN	0.9014	0.8792	0.9014	**0.8741**
15	LIME + FFNN	**0.9043**	0.8842	**0.9043**	**0.8748**
15	SFIX + FFNN	0.9038	**0.8863**	0.9038	0.8733
15	SHAP + FFNN	0.9019	0.8792	0.9019	0.8746
20	LIME + FFNN	**0.9094**	**0.8945**	**0.9094**	**0.8848**
20	SFIX + FFNN	0.9061	0.8900	0.9061	0.8794
20	SHAP + FFNN	0.9038	0.8864	0.9038	0.8740
49	FFNN	0.9071	0.8901	0.9071	0.8828

Tablo 4.14'te belirtildiği üzere, finansal veri setine XAI yöntemleri uygulandıktan sonra FFNN modelinin performansına ilişkin gözlemler ve analizler aşağıdaki alt başlıklar altında sunulmaktadır.

4.2.8.1 20 özellik seviyesinde model performansı

20 özellik kullanıldığında, LIME + FFNN modeli 0.9094 doğruluk elde ederek en yüksek performansı göstermektedir. Bu seviyede, SFIX + FFNN modeli 0.9061 doğruluk ile ikinci sırada yer alırken, onu 0.9038 doğruluk ile SHAP + FFNN izlemektedir. LIME, tüm performans metriklerinde (kesinlik: 0.8945, duyarlılık: 0.9094, F1-Skoru: 0.8848) diğer yöntemlerden daha iyi performans göstererek, bu özellik seviyesinde yerel açıklanabilirlik yaklaşımının etkinliğini ortaya koymaktadır.

4.2.8.2 15 özellik seviyesinde performans değişimleri

Özellik sayısı 20'den 15'e düşürüldüğünde, yöntemler arasındaki performans farkı daralmaktadır. LIME+FFNN modeli 0.9043 doğruluk ile liderliğini korusa da, SFIX+FFNN'ye göre fark yalnızca 0.0005 puana düşmektedir. Dikkat çekici olarak, SFIX+FFNN kesinlik açısından LIME+FFNN'yi geride bırakmaktadır (0.8863'e karşı 0.8842); bu durum, özellik sayısı azaldıkça SFIX'in özellik seçimindeki etkinliğinin daha belirgin hâle geldiğini göstermektedir.

4.2.8.3 10 özellik seviyesinde performans kayması

10 özellik seviyesinde, SFIX+FFNN modeli en yüksek doğruluğu (0.9033) ve kesinlik değerini (0.8859) elde etmektedir. Bu noktada, LIME+FFNN'nin performansı 0.9024 doğruluğa düşerken, SHAP+FFNN yalnızca F1-Skoru açısından en iyi performansı göstermektedir (0.8741). Bu sonuçlar, özellik sayısı azaldıkça SFIX'in sistematik özellik seçme stratejisinin açık bir avantaj sağlamaya başladığını göstermektedir.

4.2.8.4 5 özellik seviyesinde SFIX performansı

En düşük özellik seviyesinde (beş özellik), SFIX+FFNN modelinin üstünlüğü açık biçimde ortaya çıkmaktadır. Bu yapılandırmada SFIX, doğruluk (0.9014), kesinlik (0.8849) ve duyarlılık (0.9014) açısından en yüksek değerlere ulaşmaktadır. Buna karşılık, LIME+FFNN'nin performansı 0.8968 doğruluğa düşerken, SHAP+FFNN yalnızca F1-Skoru bakımından diğer yöntemleri çok az farkla geçmektedir (0.8673). Bu bulgular, sınırlı özellik kümesi ile çalışırken SFIX'in sistematik özellik seçme stratejisinin alternatif XAI yöntemlerinden daha etkili olduğunu göstermektedir.

4.2.9 Sonuçlar ve model etkinliğinin değerlendirilmesi

Deneysel bulgular, özellik sayısı azaldıkça SFIX+FFNN çerçevesinin performans avantajının daha belirgin hâle geldiğini göstermektedir. Özellikle, 10 özellik veya daha azına sahip yapılandırmalarda SFIX'in sistematik özellik seçme stratejisi, hem yerel açıklanabilirlik (LIME) hem de küresel açıklanabilirlik (SHAP) yöntemlerine kıyasla daha tutarlı ve iyi performans sunmaktadır. Bu gözlem, yüksek boyutlu finansal verisetlerinde sınırlı hesaplama kaynakları veya sıkı boyut indirgeme gereksinimleri bulunan senaryolar için SFIX'in özellikle uygun olduğunu göstermektedir.

Ayrıca, SFIX tarafından tutarlı biçimde elde edilen daha yüksek kesinlik, false positive oranlarının en aza indirilmesi gereken finansal denetim görüşü analizi gibi uygulamalarda kritik bir avantaj sağlamaktadır. Genel olarak, SFIX açıklanabilir yapay zekâ yöntemi, finansal denetim görüşü tahmini için etkili ve sağlam bir özellik seçme yaklaşımı olduğunu kanıtlamaktadır. Hem orijinal hem de tamamlanmış verisetlerinde seçilen 20 özellik, tam özellik kümesine kıyasla hesaplama karmaşıklığını önemli ölçüde azaltırken rekabetçi tahmin performansı sunmaktadır. ROCE oranının her iki yapılandırmada da sürekli olarak lider konumunu koruması, önerilen yöntemin kararlılığını ve güvenilirliğini daha da pekiştirmektedir.

Önerilen SFIX–FFNN çerçevesinin üstünlüğü, mevcut makine öğrenmesi ve açıklanabilir yapay zekâ yaklaşımlarıyla karşılaştırıldığında açıkça görülmektedir. XGBoost ve Random Forest gibi geleneksel modeller rekabetçi doğruluk elde etseler de, tam özellik kümesine dayanmakta ve finansal veriye özgü zamansal dinamikleri ile risk duyarlı ilişkileri yakalayamayan SHAP ve LIME gibi post-hoc yorumlama yöntemlerine ihtiyaç duymaktadırlar. Buna karşılık SFIX, yalnızca 0.45%'lik marjinal bir performans kaybıyla boyutsallığı 59% azaltırken, özellikleri içsel ve çok boyutlu biçimde değerlendirmektedir. Ayrıca, SFIX tarafından seçilen özelliklerle eğitilen FFNN modelleri, hem orijinal hem de tamamlanmış verisetlerinde daha yüksek doğruluk ve F1-Skoru elde ederek SHAP ve LIME tabanlı özellik alt kümelerini tutarlı biçimde geride bırakmaktadır. Bu bulgular, önerilen hibrit çerçevenin yalnızca tahmin performansını korumakla kalmadığını, aynı zamanda mevcut yaklaşımlara kıyasla daha kararlı, yorumlanabilir ve hesaplama açısından daha verimli sonuçlar sunduğunu göstermektedir. Sonuç olarak, SFIX–FFNN modeli; şeffaflık ve güvenilirliğin büyük önem taşıdığı gerçek dünya finansal denetim uygulamaları için güçlü bir çözüm sunmaktadır.

Eksik veri tamamlama işleminin SFIX skorları üzerindeki etkisi özellikle dikkat çekicidir; çünkü bu durum kaldıraç oranlarının ve risk değerlendirme metriklerinin göreli önemini artırmaktadır. Bu bulgu, daha kapsamlı verisetlerinin risk odaklı özelliklerin önemini daha doğru yansıttığını göstermektedir. Sonuç olarak, SFIX modeli finansal denetim bağlamında hem özellik seçimi hem de model açıklanabilirliği için güvenilir ve etkili bir çerçeve sunmaktadır.

4.2.10 Sınırlılıklar ve başarısızlık durumları

Önerilen SFIX tabanlı açıklanabilirlik çerçevesi, denetim görüşü sınıflandırmasında güçlü performans ve yorumlanabilirlik gösterse de, bazı sınırlılıkların kabul edilmesi gerekmektedir.

İlk olarak, yaklaşım istatistiksel özellikleri sektörler ve ekonomik döngüler arasında önemli ölçüde değişebilen finansal oranlara dayanmaktadır. Bu nedenle, SFIX'in dinamik, örüntü tabanlı ve risk odaklı bileşenleri; yüksek piyasa oynaklığı veya yapısal ekonomik kırılmalar dönemlerinde kararsız önem skorları üretebilir.

İkinci olarak, RS hesaplaması çok değişkenli normal dağılım varsayımına dayanmaktadır; ancak bu varsayım, gerçek dünyadaki finansal risk göstergelerinde yaygın olarak görülen ağır kuyruklu veya asimetrik özellikleri tam olarak yansıtmayabilir. Bu varsayım, aşırı değer koşulları altında risk katkılarının yanlış tahmin edilmesine yol açabilir.

Üçüncü olarak, eksik değerler makine öğrenmesi tabanlı yöntemlerle tamamlanmış olsa da yeniden oluşturulan gözlemler özellikle sınırlı temsil edilen sınıflar için yanlılık oluşturabilir. Buna ek olarak, azınlık denetim görüşü kategorilerinin ciddi biçimde yetersiz temsil edildiği durumlarda FFNN modeli düşük performans gösterebilir.

Metodolojik açıdan bakıldığında, SFIX korelasyon tabanlı ve zaman penceresi tabanlı tahminlere dayanmaktadır; bu da yöntemi gürültüye, aykırı değerlere ve finansal raporlama uygulamalarındaki ani değişimlere duyarlı hâle getirmektedir. Bu tür koşullar altında, dinamik önem ve örüntü bileşenleri altta yatan ilişkileri doğru biçimde yansıtamayabilir. Ayrıca, çok boyutlu açıklanabilirlik bileşenlerinin hesaplanmasına ilişkin hesaplama maliyeti veriseti boyutuyla birlikte artmakta; bu da önerilen yaklaşımın ölçeklenebilirliğini sınırlayabilmektedir.

Son olarak, SFIX model yorumlanabilirliğini önemli ölçüde artırsa da altta yatan FFNN hâlâ bir derin öğrenme mimarisidir ve iç temsillerinin tam şeffaflığı garanti edilemez. Bu nedenle, karar sınırına baskın biçimde etki eden karmaşık doğrusal olmayan özellik etkileşimlerinin SFIX bileşenleri tarafından yeterince yakalanamadığı durumlarda başarısızlık senaryoları ortaya çıkabilir. Gelecek araştırmalar, uyarlanabilir zaman penceresi mekanizmaları, dağılımdan bağımsız risk tahmin teknikleri, dengesizlik farkındalıklı eğitim stratejileri ve derin öğrenme karar süreçleriyle daha yakından

hizalanan hibrit XAI mimarileri geliştirerek bu sınırlılıkları ele alabilir.

4.2.11 Gerçek dünya dağıtımı: finansal denetimde pratik avantajlar ve zorluklar

Önerilen SFIX–FFNN çerçevesi, finansal denetim görüşü tahmini için geliştirilmiş olsa da metodolojik özellikleri uluslararası düzenleyici beklentiler olan şeffaflık, açıklanabilirlik ve model yönetişimi ile yakın uyum göstermektedir. Halka Açık Şirketler Muhasebe Denetim Kurulu (PCAOB), Uluslararası Finansal Raporlama Standartları Vakfı (IFRS Foundation), Bazel Bankacılık Denetim Komitesi (BCBS), Avrupa Birliği Yapay Zeka Yasası (EU AI Act) ve Genel Veri Koruma Yönetmeliği (GDPR) gibi kurumlar tarafından yayımlanan standartlar; izlenebilirlik, hesap verebilirlik ve güçlü veri yönetişimini vurgulamaktadır. Bu ilkeler, SFIX'in sunduğu çok boyutlu açıklanabilirlik ile doğrudan desteklenmektedir. Dinamik önem, güven skorlama, anomali tespiti, örüntü çıkarımı ve risk değerlendirmesini bütünleştiren bu çerçeve, yüksek riskli finansal karar ortamlarında şeffaflığı artırabilecek yorumlanabilir çıktılar üretmektedir.

SFIX–FFNN modelinin en önemli avantajlarından biri, özellik boyutunda ciddi bir azalma sağlaması ve anlamlı bir performans kaybı olmadan değişken sayısını %60'a kadar düşürebilmesidir. Bu azalma, hesaplama verimliliğini artırmakta, modelin yeniden eğitilme döngülerini hızlandırmakta ve doğrulama süreçlerini basitleştirmektedir. Modelin şeffaflığı aynı zamanda, özellikle denetim kararlarının gerekçelendirilmesi, özellik ilgililiğinin belgelenmesi ve insan gözetimi mekanizmalarıyla uyum gerektiren alanlarda düzenleyici uyumu desteklemektedir. Ayrıca model hem orijinal hem de tamamlanmış verisetlerinde tutarlı performans sergileyerek eksik değerler, raporlama tutarsızlıkları ve finansal tablo düzensizlikleri gibi gerçek dünya denetim ortamlarında sıkça karşılaşılan sorunlara karşı dayanıklılık göstermektedir.

Bu güçlü yönlere rağmen, gerçek dünya dağıtımı düşünüldüğünde bazı pratik zorluklar ortaya çıkmaktadır. Finansal veriler çoğu zaman sektörler arasında değişiklik göstermekte, tutarsız raporlama aralıklarına ve düzensiz muhasebe yapılarına sahip olabilmektedir; bunların tümü SFIX'in dinamik ve risk tabanlı bileşenlerinin kararlılığını etkileyebilir. RS, yüksek oynaklığa sahip piyasalarda sıklıkla gözlenen aşırı değer davranışlarını tam olarak yakalayamayan istatistiksel varsayımlara dayanmaktadır. SFIX–FFNN modelinin mevcut denetim altyapılarına entegre edilmesi;

standartlaştırılmış arayüzler, güvenli veri hatları ve sürüm kontrolü ile dokümantasyon için resmî yönetişim prosedürleri gerektirebilir. Ayrıca, Basel yönergeleri ve AB Yapay Zekâ Yasası tarafından öngörülen model yönetişimi ve uyum beklentilerini karşılamak amacıyla sürekli izleme, kavram kayması tespiti, periyodik geriye dönük testler ve yeniden kalibrasyon gereklidir.

İnsan uzmanlığı da temel bir rol oynamaktadır. SFIX yorumlanabilir çıktılar sunsa da denetçilerin modelin dinamik davranışını anlaması ve açıklama çıktılarını denetim dokümantasyonuna entegre etmesi için eğitilmesi gerekmektedir. Arayüzler ve denetim destek sistemleri, zamansal değişimleri, anomali skorlarını ve risk katkılarını etkili biçimde görselleştirebilmelidir. Ayrıca, özellikle hassas finansal açıklamaların söz konusu olduğu ortamlarda model tabanlı denetim sistemleri dağıtılırken GDPR'nin veri minimizasyonu ve hukuka uygun işleme ile ilgili gereklilikleri de dikkate alınmalıdır.

Pratik kullanım alanlarına bakıldığında, SFIX–FFNN çerçevesi çok çeşitli finansal alanlarda uygulanabilir. Dış ve iç denetimde model, risk temelli denetim planlamasını, finansal tablolarda anomali tespitini ve görüşü etkileyebilecek göstergelerin erken belirlenmesini destekleyebilir. Bankacılıkta ise kredi riski değerlendirmesi, erken uyarı sistemleri ve IFRS 9 ile ilişkili model doğrulama süreçleri için güçlü bir potansiyele sahiptir. Denetleyici otoriteler, SFIX'in finansal raporlamadaki yeni ortaya çıkan risk örüntülerini tespit etme yeteneğinden yararlanarak stres testleri ve düzenleyici gözetimi destekleyebilir. Model ayrıca dolandırıcılık tespiti, AML uyumu, kurumsal risk yönetimi, aktüeryal analiz ve açıklanabilirlik ile verimliliğin temel gereksinim olduğu şeffaflık odaklı FinTech uygulamalarında da değerlidir.

Gerçek dünya finansal denetim ortamlarında, model ölçeklenebilirliği ve mevcut denetim altyapılarıyla entegrasyon kritik dağıtım unsurlarıdır. Ölçeklenebilirlik açısından SFIX–FFNN çerçevesi, tahmin performansını korurken özellik boyutunu azaltabilmesi sayesinde pratik verimlilik göstermektedir. Bu azalma hem eğitim hem de çıkarım aşamalarında hesaplama yükünü önemli ölçüde düşürebilir ve çerçeveyi büyük ölçekli finansal verisetleri ile sürekli denetim senaryoları için uygun hâle getirebilir.

Entegrasyon açısından, başlıca zorluklardan biri açıklanabilir yapay zekâ çıktılarının mevcut denetim yazılımı ekosistemleriyle uyumlu hâle getirilmesidir; çünkü bu sistemler tipik olarak kural tabanlı analitik ve yapılandırılmış raporlama iş akışları etrafında

tasarlanmıştır. Dinamik önem, anomali sinyalleri ve risk skorları gibi bileşen tabanlı açıklanabilirlik çıktılarının geleneksel denetim panellerine entegrasyonu, arayüz uyarlamaları ve standartlaştırılmış açıklanabilirlik raporlama biçimleri gerektirebilir. Ayrıca, finansal denetim standartlarına uyumu sağlamak için düzenleyici doğrulama, model yönetişimi prosedürleri ve denetim izi gereklilikleri de karşılanmalıdır. Buna rağmen, SFIX'in modüler yapısı mevcut karar destek hatlarına esnek entegrasyon imkânı sunmakta ve denetim otomasyon sistemleri içinde kademeli benimsemeyi desteklemektedir.

Bu tür bir entegrasyon, özellikle açıklanabilirliğin denetlenebilir, izlenebilir ve model risk yönetimi çerçeveleriyle uyumlu olması gereken düzenleyici odaklı ortamlarda daha da önem kazanmaktadır.

Genel olarak, SFIX–FFNN çerçevesi yapay zekâ destekli finansal denetimde güvenilirliği, şeffaflığı ve verimliliği artırmak için önemli bir temel sunmaktadır. Çok boyutlu açıklanabilirliği, özellik verimliliği ve düzenleyici uyumluluğu bir araya getirmesi, onu pratik dağıtım için güçlü bir aday hâline getirmektedir. Bununla birlikte, başarılı bir benimseme; güvenli ve güvenilir kullanımın sağlanabilmesi için veri yönetişimi, model izleme, düzenleyici uyum ve insan-makine etkileşimine dikkat edilmesini gerektirmektedir.

4.2.12 SFIX bileşenlerinin finansal zaman serisi verileri ve denetim görüşü sınıflandırması için kuramsal uygunluğu

Finansal veri yapıları, klasik tablosal verisetlerinden farklı olarak çoğu zaman güçlü zamansal bağımlılıklar, mevsimsel dalgalanmalar, oynaklık kümelenmeleri ve ani yapısal kırılmalar göstermektedir. Denetim görüşü sınıflandırması gibi finansal karar problemlerinde, yalnızca model doğruluğu değil, aynı zamanda kararların açıklanabilirliği, risk duyarlılığı ve düzenleyici uyum da kritik öneme sahiptir. Bu nedenle, finansal veri analitiğinde kullanılan açıklanabilir yapay zekâ yöntemleri, yalnızca yerel özellik katkılarını açıklamakla sınırlı kalmamalı; zamansal dinamikleri ve risk odaklı karar mekanizmalarını da yakalayabilmelidir.

SHAP ve LIME gibi yaygın olarak kullanılan post-hoc açıklanabilirlik yöntemleri, model çıktılarını yerel veya küresel düzeyde açıklayabilse de doğrudan zaman serisi davranışını, veri güvenilirliğini ya da risk tabanlı karar etkilerini modellemek için

tasarlanmamıştır. Bu yöntemler özellik katkılarını statik bir bakış açısıyla değerlendirirken, finansal sistemlerde karar süreçleri çoğu zaman zamansal davranış değişimleri, anomaliler ve risk profilleri tarafından şekillendirilmektedir.

Bu bağlamda, önerilen SFIX çerçevesi finansal veri özelliklerini çok boyutlu bir puanlama mekanizması ile değerlendiren alan odaklı bir açıklanabilirlik yaklaşımı sunmaktadır. SFIX'in beş temel bileşeni, finansal veri analitiğinin temel gereksinimlerini doğrudan karşılamaktadır:

- Dinamik özellik önem skoru: Finansal oranlar ve performans göstergeleri zaman içinde değişmektedir. DIS, özelliklerin zamana bağlı önem değişimini yakalayarak finansal zaman serisi verilerinde daha gerçekçi özellik değerlendirmesine imkân tanımaktadır.
- Güven skoru: Finansal verisetleri çoğu zaman eksik değerler, muhasebe revizyonları veya raporlama tutarsızlıkları içermektedir. CS, veri güvenilirliğini ve model kararlılığını değerlendirme sürecine dâhil ederek denetim kararlarının güvenilirliğini artırmaktadır.
- Örüntü skoru: Finansal performans genellikle trendler, döngüler ve mevsimsel etkiler içermektedir. PS, bu yapısal davranışları modelleyerek finansal koşulların zamansal evrimini daha açıklanabilir hâle getirmektedir.
- Anomali skoru: Olağandışı finansal hareketlerin tespiti denetim süreçlerinde kritik öneme sahiptir. AS bileşeni, olası manipülasyonları, finansal düzensizlikleri veya beklenmeyen değişimleri belirlemeye yardımcı olmaktadır.
- Risk skoru: Finansal kararlar doğrudan risk profilleriyle ilişkilidir. RS, model kararlarının finansal risk perspektifinden değerlendirilmesini mümkün kılmakta ve düzenleyici uyumlu karar mekanizmalarını desteklemektedir.

SFIX finansal karar mekanizmalarını doğrudan temsil eden çok boyutlu bir özellik değerlendirme yapısı sunmaktadır. Bu açıdan SFIX yalnızca açıklama üretmekle kalmamakta; aynı zamanda finansal karar süreçlerine entegre edilebilecek risk farkındalıklı ve zaman farkındalıklı bir özellik seçimi ve açıklanabilirlik mekanizması da sağlamaktadır.

Dolayısıyla, SFIX–FFNN entegrasyonu yalnızca finansal denetim gibi düzenleme yoğun alanlarda model performansını korumakla kalmamakta, aynı zamanda karar verme

süreçlerinde şeffaflığı, izlenebilirliği ve güvenilirliği artıran alana özgü bir açıklanabilirlik yaklaşımı da sunmaktadır.

5 ÖNERİLEN SFIX MODELİN YAPAY ZEKÂ DAYANIKLILIĞI UYGULAMASI

Bu bölümde, SFIX yaklaşımının zamansal bileşenlerden arındırılmış sürümü olan Sadeleştirilmiş SFIX (S-SFIX) kullanılarak, IoT tabanlı DDoS saldırı tespiti probleminde açıklanabilir özellik seçimi ve dayanıklılık özellikleri sistematik biçimde incelenmektedir (Cil ve Yildiz, 2025b). S-SFIX, SFIX'in çekirdek açıklanabilirlik bileşenlerini korurken, zaman bağımlı terimleri devre dışı bırakarak statik ve alanlar arası uygulamalara uygun bir yapı sunmaktadır.

Bu bölümdeki çalışmanın temel amaçları aşağıdaki şekilde özetlenebilir:

- Finans alanı için geliştirilmiş alan-özel bir açıklanabilirlik yönteminin, siber güvenlik alanında yeniden tasarım gerektirmeden uygulanabilirliğini incelemek.
- IoT tabanlı DDoS verilerinde, S-SFIX tarafından belirlenen özelliklerin sınıflandırma performansını ne ölçüde desteklediğini analiz etmek.
- S-SFIX'i SHAP, LIME ve Random Forest özellik önemliliği gibi yaygın yöntemlerle karşılaştırarak yöntemler arası konumunu değerlendirmek.
- Açıklanabilir özellik skorlama yapısının gürültü altında ne ölçüde kararlı kaldığını inceleyerek yöntemi Yapay Zekâ Dayanıklılığı perspektifiyle ele almak.

Bu doğrultuda çalışma, iki tamamlayıcı eksen üzerine kurulmuştur. İlk olarak, S-SFIX tarafından seçilen özellik kümelerinin farklı boyutlarda sınıflandırma performansını ne ölçüde koruyabildiği değerlendirilmiştir. İkinci olarak, önerilen yöntemin hem alternatif XAI/özellik önemliliği yaklaşımları ile karşılaştırmalı başarımı hem de Gaussian gürültüsü altında sıralama kararlılığı analiz edilmiştir. Böylece çalışma, S-SFIX'i yalnızca bir açıklama üretim mekanizması olarak değil, aynı zamanda performans ve dayanıklılık boyutlarıyla değerlendirilen bir özellik seçimi çerçevesi olarak ele almaktadır.

Bu bölümdeki çalışmanın katkıları şu şekilde özetlenebilir:

- Finans alanı için geliştirilmiş bir açıklanabilirlik yönteminin, siber güvenlik alanında uygulanabilirliğini inceleyen alanlar arası bir değerlendirme sunması,
- S-SFIX'in SHAP, LIME ve Random Forest özellik önemliliği ile karşılaştırmalı

olarak değerlendirilmesi,

- Açıklanabilir özellik skorlama yapısının Gaussian gürültüsü altında kararlılığını inceleyerek Yapay Zekâ Dayanıklılığı boyutunu deneysel olarak ele alması,
- IoT tabanlı DDoS saldırı tespiti probleminde, düşük boyutlu fakat bilgi yoğun özellik kümeleri ile rekabetçi performansın sürdürülebileceğini göstermesi.

5.1 Veri Seti ve Ön İşleme

5.1.1 CICIoT2023 veriseti

Bu çalışmada Neto ve ark. (2023) tarafından oluşturulan ve IoT tabanlı ağ trafiği verilerini içeren CICIoT2023 saldırı veriseti kullanılmıştır. Birleştirilmiş veriseti toplam 46,686,579 örnek ve 34 farklı saldırı etiketi içermektedir. Veriseti, farklı DDoS türleri, DoS saldırıları, Mirai varyantları, keşif (reconnaissance) saldırıları ve çeşitli ağ tabanlı tehdit senaryolarını kapsamaktadır. Veriseti, ağ trafiğine ilişkin bayrak sayıları, süre, varyans, paket büyüklüğü istatistikleri ve protokol bilgileri gibi toplam 46 öznitelik içermektedir. Bunun yanında, Attack üst sınıfı altında toplanan toplam 33 farklı saldırı alt türü bulunmaktadır. Genel etiket dağılımı Tablo 5.1'de özetlenmiştir.

Tablo 5.1. CICIoT2023 verisetindeki etiketlerin dağılımı

Sıra	Etiket	Örnek Sayısı	Yüzde (%)
1	DDoS-ICMP_Flood	7,200,504	15.42
2	DDoS-UDP_Flood	5,412,287	11.59
3	DDoS-TCP_Flood	4,497,667	9.63
4	DDoS-PSHACK_Flood	4,094,755	8.77
5	DDoS-SYN_Flood	4,059,190	8.69
6	DDoS-RSTFINFlood	4,045,285	8.66
7	DDoS-SynonymousIP_Flood	3,598,138	7.71
8	DoS-UDP_Flood	3,318,595	7.11
9	DoS-TCP_Flood	2,671,445	5.72
10	DoS-SYN_Flood	2,028,834	4.35
11	BenignTraffic	1,098,195	2.35
12	Mirai-greeth_flood	991,866	2.12
13	Mirai-udpplain	890,576	1.91
14	Mirai-greip_flood	751,682	1.61
15	DDoS-ICMP_Fragmentation	452,489	0.97
16	MITM-ArpSpoofing	307,593	0.66
17	DDoS-UDP_Fragmentation	286,925	0.61

Sıra	Etiket	Örnek Sayısı	Yüzde (%)
18	DDoS-ACK_Fragmentation	285,104	0.61
19	DNS_Spoofing	178,911	0.38
20	Recon-HostDiscovery	134,378	0.29
21	Recon-OSScan	98,259	0.21
22	Recon-PortScan	82,284	0.18
23	DoS-HTTP_Flood	71,864	0.15
24	VulnerabilityScan	37,382	0.08
25	DDoS-HTTP_Flood	28,790	0.06
26	DDoS-SlowLoris	23,426	0.05
27	DictionaryBruteForce	13,064	0.03
28	BrowserHijacking	5,859	0.01
29	CommandInjection	5,409	0.01
30	SqlInjection	5,245	0.01
31	XSS	3,846	0.01
32	Backdoor_Malware	3,218	0.01
33	Recon-PingSweep	2,262	0.00
34	Uploading_Attack	1,252	0.00
	Toplam	**46,686,579**	**100**

5.1.2 Alt küme oluşturma ve veri dengeleme

5.1.2.1 Alt küme oluşturma

Alt küme seçimi, hesaplama kısıtları nedeniyle zorunlu olmakla birlikte, veri üretim sürecinin yapısal bütünlüğünü korumak amacıyla aynı verisetinin homojen parçaları arasından gerçekleştirilmiştir.

Deneysel akışta işlenen veriseti toplam 470,500 örnekten oluşmakta ve bu örnekler iki CSV dosyası üzerinden işlenmektedir.

Deneysel tasarımda verisetindeki çok sayıdaki saldırı türü tek bir Attack üst sınıfı altında birleştirilmiş ve *BenignTraffic* etiketi ile ikili sınıflandırma problemi tanımlanmıştır. Böylece, çoklu saldırı örüntülerini birlikte içeren daha kapsamlı ve daha zorlayıcı bir deney düzeni elde edilmiştir. Bu yaklaşım, önerilen S-SFIX tabanlı özellik seçme çerçevesinin daha geniş bir tehdit uzayında değerlendirilmesine olanak sağlamaktadır. Kullanılan eğitim hattında da etiketler *BenignTraffic* → 0 ve diğer tüm saldırı türleri *Attack* → 1 olacak şekilde yeniden eşlenmektedir.

Tablo 5.2 incelendiğinde, *Attack* üst sınıfının toplamda 33 farklı saldırı alt türünden oluştuğu görülmektedir. Bu alt türler; yüksek hacimli Dağıtık Hizmet Dışı Bırakma

(DDoS) ve Hizmet Dışı Bırakma (DoS) saldırılarından Mirai tabanlı örüntülere, keşif (reconnaissance) saldırı türlerinden sahtecilik (spoofing), kaba kuvvet (brute force) ve enjeksiyon (injection) tabanlı saldırılara kadar oldukça heterojen bir yapı sergilemektedir. Özellikle *DDoS-ICMP_Flood, DDoS-UDP_Flood, DDoS-TCP_Flood* ve *DDoS-SYN_Flood* gibi yoğun saldırı tipleri verisetinde önemli bir yer tutarken, *Recon-PingSweep, Uploading_Attack, XSS* ve *Backdoor_Malware* gibi görece az sayıda örnek içeren saldırılar da aynı üst sınıf içinde yer almaktadır. Bu durum, aksine çok sayıda ve farklı karakterde saldırı davranışını bir arada içeren daha zorlu ve daha gerçekçi bir deneysel senaryo sunduğunu göstermektedir.

Tablo 5.2. Saldırı üst sınıfı altında yer alan saldırı alt türlerinin dağılımı

Saldırı Alt Türü	Örnek Sayısı	Saldırı Alt Türü	Örnek Sayısı	Saldırı Alt Türü	Örnek Sayısı
DDoS-ICMP_Flood	36554	Mirai-udpplain	4661	VulnerabilityScan	210
DDoS-UDP_Flood	27626	Mirai-greip_flood	3758	DDoS-HTTP_Flood	169
DDoS-TCP_Flood	23149	DDoS-ICMP_Fragmentation	2377	DDoS-SlowLoris	106
DDoS-PSHACK_Flood	21210	MITM-ArpSpoofing	1614	DictionaryBruteForce	63
DDoS-SYN_Flood	20739	DDoS-ACK_Fragmentation	1505	SqlInjection	31
DDoS-RSTFINFlood	20669	DDoS-UDP_Fragmentation	1484	BrowserHijacking	30
DDoS-SynonymousIP_Flood	18189	DNS_Spoofing	925	CommandInjection	28
DoS-UDP_Flood	16957	Recon-HostDiscovery	697	Backdoor_Malware	22
DoS-TCP_Flood	13630	Recon-OSScan	517	XSS	18
DoS-SYN_Flood	10275	Recon-PortScan	430	Uploading_Attack	8
Mirai-greeth_flood	5016	DoS-HTTP_Flood	414	Recon-PingSweep	6

5.1.2.2 Dengesizlik giderme stratejisi

CICIoT2023 verisetinin tamamı sistem kaynakları açısından tek seferde işlenemeyecek kadar büyük olduğundan, deneysel çalışma verisetinin homojen yapısını koruyan bir alt bölümü üzerinden yürütülmüştür. İlk aşamada verisetinin homojen parçalardan biri temel alınmış; ancak bu alt bölümde Benign ve Attack sınıfları arasında belirgin bir dengesizlik oluştuğu gözlenmiştir. Bu dengesizliği azaltmak amacıyla, verisetini oluşturan diğer CSV dosyası parçaları arasından yalnızca benign örnekler filtrelenmiş ve başlangıçta seçilen çalışma kümesine kontrollü biçimde eklenmiştir. Böylece hem verisetinin tamamen farklı bir kaynaktan beslenmesi engellenmiş, hem de aynı veri üretim sürecine ait homojen alt dosyalar üzerinden Benign sınıfı güçlendirilerek sınıf dağılımının

iyileştirilmesi amaçlanmıştır.

Bu yaklaşımın temel amacı, verisetinin istatistiksel yapısını mümkün olduğunca koruyarak sınıf dengesizliğini azaltmak ve modelin yalnızca baskın saldırı örüntülerini öğrenen tek taraflı bir yapıya dönüşmesini engellemektir. Başka bir ifadeyle, verisetinin dengelenmesi dışsal ya da sentetik örnek üretimiyle değil, aynı verisetinin homojen alt parçalarından seçilen *Benign* örneklerin kontrollü biçimde eklenmesiyle gerçekleştirilmiştir. Bu tercih, verisetinin temsiliyetini korurken daha gerçekçi bir öğrenme ortamı sunması bakımından önemlidir.

Dengesizlik giderme sürecine ek olarak, kalan sınıf dengesizliğinin eğitim aşamasındaki etkisini azaltmak için sınıf ağırlıklandırma yaklaşımı da kullanılmıştır. Sınıf ağırlıkları yalnızca eğitim kümesindeki *Benign* ve *Attack* dağılımı temel alınarak hesaplanmış ve eğitim sürecine dahil edilmiştir. Böylece doğrulama ve test kümelerine ait bilginin eğitim aşamasına sızması önlenirken, aynı zamanda sınıf baskınlığının öğrenme üzerindeki olumsuz etkisi azaltılmıştır. Bu uygulama eğitim hattında açık biçimde tanımlanmıştır.

5.1.3 Eğitim, doğrulama ve test ayrımı

Veri işleme hattı, veri sızıntısını önleyecek ve deneysel sonuçları saydam biçimde raporlayacak şekilde kurgulanmıştır. Veri önce eğitim/doğrulama/test kümelerine ayrılmış; bu ayrım oranı pratikte yaklaşık %72 eğitim, %8 doğrulama ve %20 test olacak şekilde gerçekleşmiştir. Güncel deney düzeninde eğitim kümesinde 338,759, doğrulama kümesinde 37,640 ve test kümesinde 94,101 örnek bulunmaktadır. Böylece toplam 470,500 örneklik çalışma kümesi üzerinde eğitim ve değerlendirme yapılmıştır.

Bölme (Split) işlemi sırasında etiketler ikili yapıya dönüştürülmüş ve mümkün olan durumlarda sınıf oranlarının korunması için tabakalı örnekleme uygulanmıştır. Kod akışında önce test kümesi ayrılmakta, ardından eğitim havuzu kendi içinde eğitim ve doğrulama alt kümelerine bölünmektedir. Bu yapı, test verisinin model eğitim sürecinde kesinlikle görülmemesini sağlamaktadır. Ayrıca ölçekleme işlemi yalnızca eğitim kümesi üzerinde öğrenilmekte ve doğrulama ile test kümelerine aynı dönüşüm parametreleri uygulanmaktadır. Bu yaklaşım, veri sızıntısını önlemek ve performans ölçümünü güvenilir hale getirmek açısından önemlidir.

Tablo 5.3, verisetinin eğitim, doğrulama ve test alt kümeleri arasındaki sınıf dağılımını

göstermektedir. Sonuçlar, *Benign* ve *Attack* sınıflarının tüm alt kümelerde oldukça yakın oranlarda temsil edildiğini ortaya koymaktadır. Eğitim kümesinde *Benign* örnek sayısı 170,937, Attack örnek sayısı ise 167,822 olarak gerçekleşmiştir. Benzer biçimde doğrulama ve test kümelerinde de oranların korunmuş olması, veri ayırma işleminin sınıf dağılımı açısından tutarlı biçimde gerçekleştirildiğini göstermektedir. Toplam çalışma kümesinde *Benign* oranı yaklaşık %50,46, Attack oranı ise %49.54 düzeyindedir. Bu sonuç, başlangıçta gözlenen sınıf dengesizliğinin, aynı verisetinin homojen alt parçalarından *Benign* örneklerin kontrollü biçimde eklenmesiyle büyük ölçüde azaltıldığını ve modelin daha dengeli bir öğrenme ortamında eğitildiğini göstermektedir.

Tablo 5.3. Eğitim, doğrulama ve test kümelerinde sınıf dağılımı

Alt Küme	Benign	Attack	Toplam
Train	170,937	167,822	338,759
Validation	18,993	18,647	37,640
Test	47,483	46,618	94,101
Toplam	**237,413**	**233,087**	**470,500**

Tüm deneyler aynı eğitim-test ayrımı ve aynı FFNN mimarisi kullanılarak gerçekleştirilmiş; yalnızca özellik seçim stratejisi değiştirilmiştir. Özellik seçiminde Simplified SFIX (S-SFIX) tarafından üretilen önem skorları kullanılmış ve aşağıdaki senaryolar oluşturulmuştur:

- İlk 5 özellik
- İlk 10 özellik
- İlk 15 özellik
- İlk 20 özellik
- Tüm özellikler (Tüm özellikler)

Bu yapı sayesinde, S-SFIX tabanlı özellik azaltımının hem model performansı hem de model karmaşıklığı üzerindeki etkisi sistematik olarak analiz edilmiştir.

5.1.4 Ön işleme

Tablo 5.4, çalışmada izlenen veri hazırlama sürecini adım adım özetlemektedir. Bu süreçte öncelikle sistem kısıtları nedeniyle CICIoT2023 verisetinin homojen bir alt bölümü seçilmiş, ardından problem yapısı *Benign* ve *Attack* olmak üzere ikili

sınıflandırma biçiminde yeniden tanımlanmıştır. Attack sınıfı altında çok sayıda saldırı alt türünün birleştirilmesiyle daha geniş bir tehdit uzayı oluşturulmuş, *Benign* sınıfı ise aynı verisetinin diğer homojen parçalarından kontrollü biçimde alınan örneklerle desteklenmiştir. Son aşamada veri eğitim, doğrulama ve test alt kümelerine ayrılmış; ölçekleme ve sınıf ağırlığı işlemleri yalnızca eğitim verisi temel alınarak uygulanmış ve böylece veri sızıntısının önüne geçilmiştir.

Tablo 5.4. Deneysel veri hazırlama sürecinin aşamaları

No	Aşama	Açıklama
1	Başlangıç alt kümesinin seçimi	CICIoT2023 verisetinin tamamı sistem kaynakları açısından tek seferde işlenemediği için, veri üretim sürecinin yapısal bütünlüğünü koruyan homojen bir alt parçası başlangıç çalışma kümesi olarak seçilmiştir.
2	İkili etiket yapısının oluşturulması	Verisetindeki BenignTraffic etiketi Benign sınıfına, geri kalan tüm saldırı etiketleri ise tek bir Attack üst sınıfına eşlenmiştir. Böylece problem, çok sayıda saldırı alt türünü kapsayan ikili sınıflandırma yapısına dönüştürülmüştür.
3	Saldırı alt türlerinin birleştirilmesi	Verisetinde yer alan 33 farklı saldırı alt türü tek bir Attack sınıfı altında toplanarak, yalnızca tek bir saldırı tipine dayalı dar bir deney düzeni yerine daha kapsamlı bir tehdit uzayı elde edilmiştir.
4	Benign örneklerle dengeleme	Başlangıçta seçilen alt kümede Benign ve Attack sınıfları arasında belirgin bir dengesizlik gözlendiğinden, aynı verisetinin diğer homojen CSV parçalarından yalnızca benign örnekler filtrelenmiş ve çalışma kümesine kontrollü biçimde eklenmiştir.
5	Eğitim/doğrulama/test ayrımı	Hazırlanan veriseti, model eğitimi ve adil değerlendirme amacıyla eğitim, doğrulama ve test alt kümelerine ayrılmıştır. Bu süreçte test kümesinin eğitim aşamasında kesinlikle görülmemesi sağlanmıştır.
6	Ölçekleme ve sızıntı önleme	Özellik ölçekleme işlemi yalnızca eğitim kümesi üzerinde öğrenilmiş, doğrulama ve test kümelerine aynı dönüşüm parametreleri uygulanmıştır. Böylece veri sızıntısı önlenerek performans ölçümünün güvenilirliği artırılmıştır.
7	Sınıf dengesizliğinin eğitim aşamasında yönetimi	Verisetindeki kalan sınıf dengesizliğinin etkisini azaltmak amacıyla sınıf ağırlıkları yalnızca eğitim kümesi temel alınarak hesaplanmış ve model eğitimi sırasında öğrenme sürecine dahil edilmiştir.

5.2 Deneysel Ortam ve Deney Tasarımı

Çalışmanın temel amacı, S-SFIX tarafından belirlenen özelliklerin gerçekten model performansını destekleyip desteklemediğini test etmektir. Bu doğrultuda aşağıdaki deneysel tasarım uygulanmıştır:

Aşama 1 – Tüm Özelliklerle Eğitim

CICIoT2023 verisetinin 46 özellik (tüm) ile bir Feed Forward Neural Network (FFNN) modeli eğitilmiştir. Bu model referans performans (baseline) olarak kabul edilmiştir.

Aşama 2 – S-SFIX Tabanlı Özellik Seçimi

S-SFIX tarafından belirlenen önem sırasına göre:

- İlk 5 özellik
- İlk 10 özellik
- İlk 15 özellik
- İlk 20 özellik

seçilmiş ve her biri için ayrı bir FFNN modeli eğitilmiştir.

Bu tasarımın amacı, açıklanabilirlik tabanlı bir yöntem tarafından belirlenen en önemli özelliklerin model doğruluğunu ne ölçüde koruyabildiğini test etmektir.

Eğer az sayıda özellik ile tam özellikli modele yakın performans elde edilebiliyorsa, bu durum:

- S-SFIX'in anlamlı özellikler belirlediğini,
- Açıklanabilirlik ile performans arasında çelişki olmadığını,
- Özellik seçiminin model karmaşıklığını azaltırken bilgi kaybını sınırladığını

gösterecektir.

Aşama 3 – Diğer XAI ve Özellik Önemlendirme Yöntemleri ile Karşılaştırma

Önerilen S-SFIX modelinin katkısını daha nesnel biçimde değerlendirebilmek amacıyla, literatürde yaygın olarak kullanılan üç farklı yaklaşım daha deneysel tasarıma dahil edilmiştir:

- SHAP
- LIME
- Random Forest Özellik önemliliği

Her yöntem için verisetindeki özellikler önem skorlarına göre büyükten küçüğe

sıralanmış; ardından S-SFIX'te olduğu gibi ilk 5, 10, 15 ve 20 özellikten oluşan alt kümeler oluşturulmuştur. Daha sonra her alt küme kullanılarak aynı FFNN mimarisi altında ayrı modeller eğitilmiş ve elde edilen sonuçlar karşılaştırılmıştır.

Bu aşamada temel amaç, S-SFIX tarafından seçilen özelliklerin yalnızca kendi içinde iyi sonuç verip vermediğini değil, aynı zamanda alternatif açıklanabilirlik ve özellik önemlendirme yöntemlerine kıyasla ne ölçüde rekabetçi olduğunu ortaya koymaktır. Karşılaştırmanın adil biçimde yapılabilmesi için tüm yöntemlerde:

- aynı veri ayrımı,
- aynı ön işleme adımları,
- aynı FFNN mimarisi,
- aynı hiperparametreler

kullanılmıştır.

Tablo 5.5'de gösterildiği üzere, her özellik seçimi yöntemi için dört farklı boyutta alt özellik kümesi oluşturulmuştur. Bu yapı hem yöntemler arası performans karşılaştırmasını, hem de özellik sayısındaki artışın model başarısı üzerindeki etkisini sistematik biçimde incelemeye olanak sağlamaktadır.

Tablo 5.5. Karşılaştırmalı özellik seçimi deney düzeni

Yöntem	Özellik sıralama ölçütü	Oluşturulan alt kümeler
S-SFIX	Nihai skora göre azalan sırada	İlk 5, İlk 10, İlk 15, İlk 20
SHAP	Önem skoruna göre azalan sırada	İlk 5, İlk 10, İlk 15, İlk 20
LIME	Birleştirilmiş önem skoruna göre azalan sırada	İlk 5, İlk 10, İlk 15, İlk 20
RF Importance	Özellik önem değerine göre azalan sırada	İlk 5, İlk 10, İlk 15, İlk 20

Eğer S-SFIX ile seçilen özellik kümeleri, SHAP, LIME ve Random Forest tabanlı seçimlere yakın ya da daha iyi performans üretiyorsa, bu durum:

- S-SFIX'in yalnızca açıklanabilir bir özellik puanlama yöntemi olmadığını,
- alternatif yöntemlerle rekabet edebilen güçlü bir özellik seçimi mekanizması sunduğunu,
- daha düşük boyutlu özellik uzayında da bilgi yoğun ve ayırt edici özellik kümeleri oluşturabildiğini

gösterecektir.

Aşama 4 – Gürültü Altında S-SFIX Dayanıklılığının Testi

Çalışmanın bir diğer amacı, veri bozulmaları karşısındaki dayanıklılığını da incelemektir. Bu amaçla, verisetinin sayısal özelliklerine farklı seviyelerde Gaussian gürültüsü eklenmiş ve her bir bozulma seviyesi için S-SFIX yeniden çalıştırılmıştır.

Deneylerde kullanılan gürültü seviyeleri şu şekildedir:

- 0.00 (gürültüsüz veri)
- 0.01
- 0.03
- 0.05
- 0.10
- 0.20

Her gürültü seviyesi için S-SFIX tarafından üretilen özellik sıralamaları, temiz veri senaryosunda elde edilen referans sıralama ile karşılaştırılmıştır. Böylece veri bozuldukça hangi özelliklerin konumunu koruduğu, hangilerinin yer değiştirdiği ve S-SFIX'in genel sıralama yapısının ne ölçüde kararlı kaldığı analiz edilmiştir.

Bu aşamada dayanıklılığı değerlendirmek için aşağıdaki ölçütler kullanılmıştır:

- Rank (Sıra) Spearman korelasyonu
- Kendall Tau katsayısı
- Ortalama mutlak sıra kayması
- Maksimum sıra kayması
- Skor Pearson korelasyonu
- İlk 5, İlk 10, İlk 15 ve İlk 20 için Jaccard benzerliği

Bu tasarımın amacı, FFNN modelinin gürültü altında ne kadar doğru tahmin yaptığını ölçmekten ziyade, doğrudan S-SFIX'in özellik sıralama mekanizmasının veri bozulmaları karşısında ne derece istikrarlı kaldığını test etmektir.

Eğer gürültü seviyesi arttığında bile S-SFIX tarafından belirlenen üst sıralı özelliklerin önemli bir bölümü korunabiliyor ve genel sıralama yapısı yüksek korelasyon değerleri

ile destekleniyorsa, bu durum:

- S-SFIX'in yalnızca temiz veri üzerinde çalışan kırılgan bir yöntem olmadığını,
- açıklanabilirlik yapısının rastgele bozulmalar altında da belirli ölçüde kararlı kaldığını,
- Önerilen yaklaşımın Yapay Zekâ Dayanıklılığı bağlamında anlamlı bir dayanıklılık potansiyeli taşıdığını

gösterecektir.

5.2.1 Önerilen FFNN modeli

Bu çalışmada, ileri beslemeli tam bağlantılı bir yapay sinir ağı (Feed-Forward Neural Network, FFNN) modeli kullanılmıştır. Model, giriş katmanı, üç gizli katman ve bir çıkış katmanından oluşmaktadır. Ağ yapısı geri besleme içermeyen klasik ileri yönlü (feed-forward) mimariye sahiptir.

5.2.1.1 Ağ mimarisi

Modelin giriş katmanındaki nöron sayısı kullanılan özellik sayısına eşittir. Deneysel tasarım kapsamında beş farklı özellik kümesi kullanılmıştır:

- İlk 5 S-SFIX özellikleri (5 giriş nöronu)
- İlk 10 S-SFIX özellikleri (10 giriş nöronu)
- İlk 15 S-SFIX özellikleri (15 giriş nöronu)
- İlk 20 S-SFIX özellikleri (20 giriş nöronu)
- Tüm özellikler (46 giriş nöronu)

Her bir deney senaryosunda aynı ağ mimarisi korunmuş, yalnızca giriş katmanındaki nöron sayısı değiştirilmiştir.

Gizli katman yapısı üç katmandan oluşmaktadır ve her gizli katmanda 50 nöron bulunmaktadır. Gizli katmanlarda aktivasyon fonksiyonu olarak sigmoid fonksiyonu kullanılmıştır. Her bir gizli katman için dönüşüm işlemi Denklem 5.1'deki şekilde ifade edilmektedir:

$$g(x) = \sigma(xW + b) \tag{5.1}$$

Burada x giriş vektörünü, W ağırlık matrisini, b bias terimini ve $\sigma(\cdot)$ sigmoid aktivasyon fonksiyonunu temsil etmektedir. Sigmoid fonksiyonu Denklem 5.2'deki şekilde tanımlanır:

$$\sigma(x) = \frac{1}{1+e^{-x}} \tag{5.2}$$

Sigmoid fonksiyonu, nöron çıktılarının 0 ile 1 aralığında haritalanmasını sağlayarak katmanlar arasında kontrollü ve doğrusal olmayan bir dönüşüm sunmaktadır.

Çıkış katmanında ikili sınıflandırma problemi (Benign / Attack) ele alındığı için tek nöron kullanılmıştır. Aktivasyon fonksiyonu olarak yine sigmoid fonksiyonu tercih edilmiştir (Denklem 5.3):

$$\hat{y} = \sigma(hW_o + b_o) \tag{5.3}$$

Burada h son gizli katmanın çıktısını, W_o çıkış katmanına ait ağırlıkları, b_o ise çıkış bias terimini ifade etmektedir. Bu yapı, her bir örnek için Attack sınıfına ait olma olasılığını üretmektedir.

5.2.1.2 Kayıp fonksiyonu ve optimizasyon

Modelin eğitimi sırasında kayıp fonksiyonu olarak Denklem 5.4'deki ikili çapraz entropi (binary cross-entropy) kullanılmıştır:

$$L = -y \log(\hat{y}) - (1-y) \log(1-\hat{y}) \tag{5.4}$$

Burada y gerçek sınıf etiketini, $\hat{y}$ ise model tarafından tahmin edilen olasılığı ifade etmektedir.

Optimizasyon algoritması olarak Adamax kullanılmıştır. Adamax, Adam algoritmasının sonsuz norm tabanlı bir türevi olup, yüksek boyutlu veri uzaylarında daha kararlı yakınsama davranışı göstermektedir. Ağırlık güncelleme işlemi iteratif olarak gerçekleştirilmiştir.

5.2.1.3 FFNN hiperparametreleri ve seçim gerekçeleri

Bu çalışmada kullanılan FFNN modeli, farklı özellik seçimi yöntemlerinden elde edilen alt özellik kümelerini ortak ve sabit bir sınıflandırma ortamında değerlendirmek amacıyla tasarlanmıştır. Amaç, yeni bir derin öğrenme mimarisi önermekten ziyade, S-SFIX, SHAP, LIME ve Random Forest tabanlı özellik seçim yaklaşımlarının aynı sınıflandırma çerçevesi altında adil biçimde karşılaştırılmasını sağlamaktır. Bu nedenle FFNN modeli,

deneylerde bağımsız değişken değil, tüm yöntemler için sabit tutulan ortak değerlendirme altyapısı olarak kullanılmıştır.

FFNN modeli üç gizli katmandan oluşmaktadır ve her katmanda 50 nöron yer almaktadır. Üç katmanlı bu yapı, verisetindeki doğrusal olmayan örüntüleri öğrenebilecek yeterli temsil gücünü sağlarken, aşırı derin ve hesaplama açısından maliyetli bir yapıdan kaçınılmasına da imkân vermektedir. Katman başına 50 nöron seçimi, hem model kapasitesini korumak hem de özellik seçimi yöntemleri arasındaki farkların çok büyük bir ağ mimarisi tarafından maskelenmesini önlemek amacıyla tercih edilmiştir. Başka bir ifadeyle, burada amaç mümkün olan en karmaşık ağı kurmak değil, farklı özellik seçimi stratejilerinin etkisini görünür kılacak dengeli bir sınıflandırıcı kullanmaktır.

Gizli katmanlarda sigmoid aktivasyon fonksiyonunun kullanılması, model çıktılarının daha kontrollü ve kararlı biçimde dönüştürülmesini sağlamak amacıyla tercih edilmiştir. Çalışmanın ana hedefi derin mimari optimizasyonu olmadığı için, literatürde yaygın ve istikrarlı bir aktivasyon fonksiyonu ile sabit bir deney zemini oluşturulması amaçlanmıştır. Çıkış katmanında ise ikili sınıflandırma problemine uygun olarak tek nöronlu sigmoid yapı kullanılmıştır. Böylece model, her örnek için attack sınıfına ait olasılığı doğrudan üretmektedir.

Optimizasyon algoritması olarak Adamax seçilmiştir. Adamax, adaptif öğrenme oranı kullanan ve özellikle büyük veri yapılarında kararlı optimizasyon davranışı sergileyen bir yöntem olması nedeniyle tercih edilmiştir. Verisetinin büyüklüğü ve deneylerin çok sayıda farklı özellik alt kümesi üzerinde tekrarlanacak olması dikkate alındığında, eğitim sürecinde hem istikrarlı yakınsama hem de uygulama kolaylığı sağlayan bir optimizer kullanılması hedeflenmiştir. Bu bağlamda Adamax, karşılaştırmalı deneylerin tümünde aynı öğrenme dinamiğini koruyan uygun bir tercih sunmaktadır.

Batch size değeri 2048 olarak belirlenmiştir. Bunun temel nedeni, kullanılan verisetinin büyük hacimli olması ve eğitim sürecinin makul sürede tamamlanmasının hedeflenmesidir. Daha küçük batch size değerleri, her epoch için daha uzun eğitim sürelerine yol açabilmekte ve özellikle tüm yöntemler için tekrarlanan karşılaştırmalı deneylerde toplam hesaplama maliyetini ciddi biçimde artırabilmektedir. Bu nedenle 2048 değeri, eğitim hızını artırmak, hesaplama maliyetini düşürmek ve aynı zamanda yeterli öğrenme kararlılığını korumak amacıyla tercih edilmiştir.

Epoch sayısı 10 olarak sabit tutulmuştur. Verisetinin büyüklüğü dikkate alındığında, her epoch zaten oldukça yüksek sayıda örnek içermektedir. Bu nedenle çok yüksek epoch değerleri yerine, modelin temel örüntüleri öğrenmesine yetecek fakat karşılaştırmalı deneylerin maliyetini aşırı artırmayacak bir eğitim süresi tercih edilmiştir. Ayrıca çalışmanın amacı, en yüksek mutlak FFNN doğruluğunu elde etmekten çok, farklı özellik seçimi yöntemlerinin göreli etkisini karşılaştırmak olduğundan, tüm yöntemler için aynı epoch sayısının korunması metodolojik açıdan daha uygun görülmüştür.

5.2.2 SFIX ve Simplified SFIX modelinin deneysel ortama entegrasyonu

Bu çalışmada kullanılan verisetinde açık bir zaman serisi boyutunun bulunmaması nedeniyle, SFIX modelinin zaman parametresini içermeyen sadeleştirilmiş versiyonu olan Simplified SFIX (S-SFIX) yaklaşımı tercih edilmiştir. S-SFIX, özellikle gerçek zaman boyutunun bulunmadığı veya sınırlı olduğu verisetlerinde hızlı model testi ve özellik seçimi amacıyla tasarlanmıştır.

Bu yaklaşım, SFIX modelinin finans alanındaki önceki uygulamalarında ortaya konan risk-duyarlı ve güven-temelli açıklanabilirlik çerçevesini, bu çalışmada siber güvenlik/saldırı tespiti bağlamına uyarlayarak genellenebilirliğini test etmektedir.

5.2.3 S-SFIX Eğitimi ve Özellik Sıralaması

Çalışmanın ilk aşamasında, filtrelenmiş verisetinin bir alt kümesi kullanılarak Simplified SFIX (S-SFIX) algoritması eğitilmiştir.

Bu bileşenler birleştirilerek her özellik için nihai S-SFIX skoru Tablo 5.6'da hesaplanmıştır.

Tablo 5.6. Gürültüsüz verisetinde S-SFIX bileşen skorları ve nihai özellik skorları

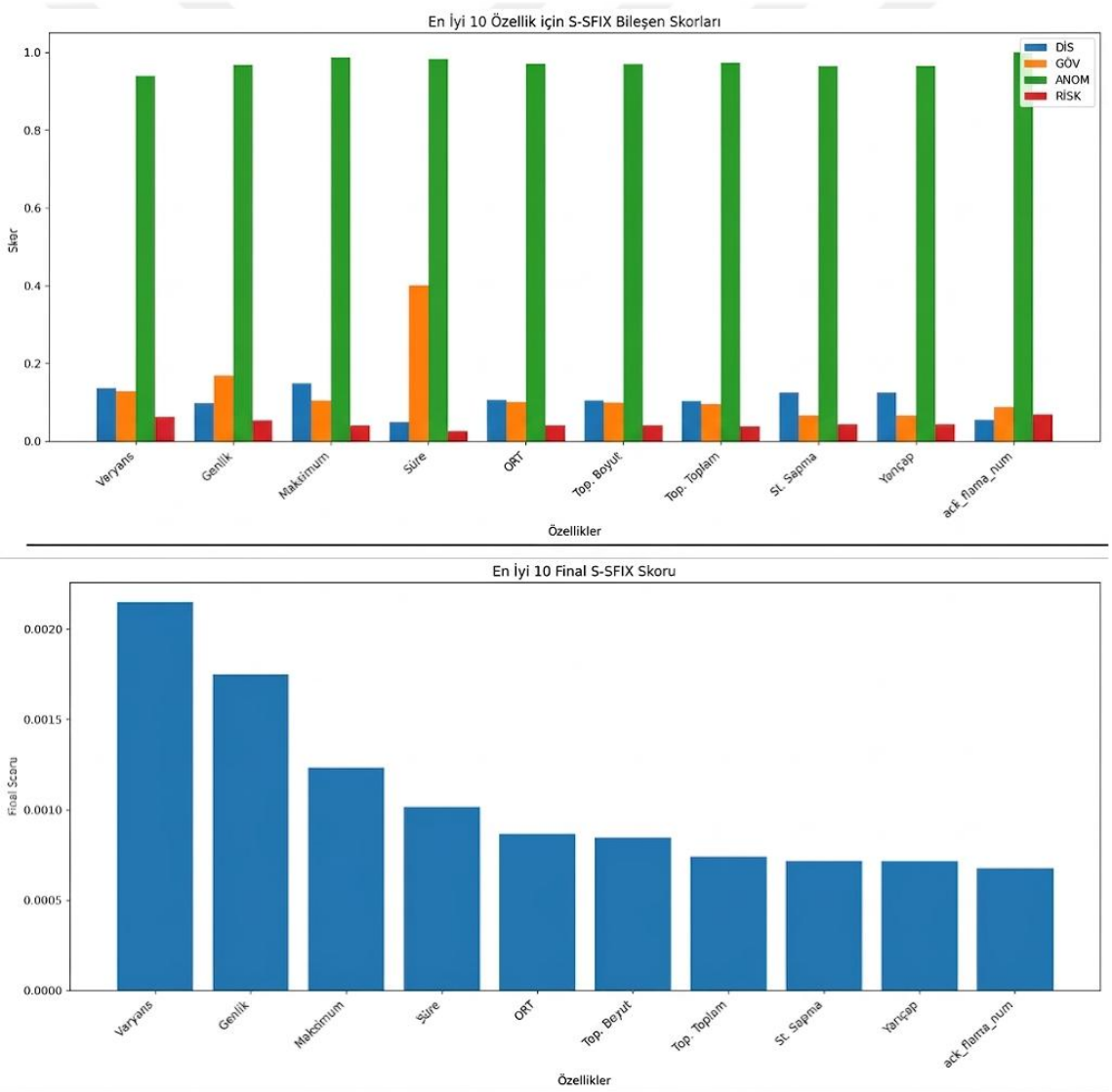
Özellik	DIS	CS	AS	RS	Final Skor
Variance	0.135935	0.128765	0.938962	0.063389	0.002151
Magnitude	0.098970	0.168156	0.967091	0.053460	0.001750
Max	0.148827	0.104572	0.986363	0.039884	0.001233
Duration	0.048785	0.400740	0.982228	0.026200	0.001015
AVG	0.106227	0.100267	0.970204	0.041329	0.000867
Tot size	0.104599	0.099450	0.969328	0.041343	0.000847
Tot sum	0.102647	0.094919	0.972567	0.038593	0.000742
Std	0.125253	0.066885	0.964162	0.043584	0.000717
Radius	0.125289	0.066746	0.964187	0.043553	0.000715
ack_flag_number	0.055540	0.088194	1.000000	0.069242	0.000678

Özellik	DIS	CS	AS	RS	Final Skor
ICMP	0.015887	0.429828	0.996978	0.018373	0.000250
HTTPS	0.073661	0.071805	1.000000	0.020072	0.000213
syn_flag_number	0.021676	0.057371	1.000000	0.061003	0.000152
rst_count	0.021349	0.057970	0.998465	0.053374	0.000131
Header_Length	0.028756	0.015627	0.968213	0.084737	0.000074
Min	0.005731	0.024258	0.993935	0.092672	0.000026
rst_flag_number	0.001255	0.091099	0.999998	0.127145	0.000023
fin_flag_number	0.001133	0.090268	1.000000	0.125214	0.000021
TCP	0.000114	0.012245	1.000000	0.545419	0.000002
syn_count	0.000147	0.002313	0.999969	0.285386	0.000000
urg_count	0.000040	0.000626	1.000000	0.420906	0.000000
Covariance	0.000002	0.005299	0.931663	0.212584	0.000000
UDP	0.000001	0.000011	1.000000	0.991828	0.000000
psh_flag_number	0.000001	0.000007	1.000000	0.990389	0.000000
Protocol Type	0.000000	0.000210	0.640544	0.762364	0.000000
fin_count	0.000000	0.000000	1.000000	0.506001	0.000000
ack_count	0.000000	0.000000	1.000000	0.502326	0.000000
Rate	0.000000	0.000000	1.000000	0.441340	0.000000
Srate	0.000000	0.000000	1.000000	0.441221	0.000000
IPv	0.000000	0.000000	0.987566	0.568212	0.000000
LLC	0.000000	0.000000	0.997129	0.505295	0.000000
flow_duration	0.000000	0.000000	0.998198	0.450683	0.000000
HTTP	0.000000	0.000000	1.000000	0.337118	0.000000
DNS	0.000000	0.000000	0.998863	0.338484	0.000000
ARP	0.000000	0.000000	0.998250	0.339285	0.000000
IAT	0.000000	0.000000	0.946673	0.455902	0.000000
SSH	0.000000	0.000000	1.000000	0.333334	0.000000
Drate	0.000000	0.000000	1.000000	0.333334	0.000000
Weight	0.000000	0.000000	0.866445	0.500041	0.000000
Number	0.000000	0.000000	0.979380	0.406873	0.000000
DHCP	0.000000	0.000000	1.000000	0.333334	0.000000
IRC	0.000000	0.000000	1.000000	0.333334	0.000000
Telnet	0.000000	0.000000	1.000000	0.333334	0.000000
cwr_flag_number	0.000000	0.000000	1.000000	0.333334	0.000000
ece_flag_number	0.000000	0.000000	1.000000	0.333334	0.000000
SMTP	0.000000	0.000000	1.000000	0.333334	0.000000

Tablo 5.6, S-SFIX tarafından her özellik için hesaplanan DIS, CS, AS, RS ve nihai skor değerlerini göstermektedir. Sonuçlar, *Variance, Magnitue, Max, Duration, AVG* ve *Tot size* gibi özelliklerin nihai S-SFIX skoru bakımından üst sıralarda yer aldığını ortaya koymaktadır. Özellikle *Variance* ve *Magnitude* özelliklerinin en yüksek nihai skorlara sahip olması, bu özelliklerin hem ayırt edicilik hem de tutarlılık açısından saldırı ve *Benign* sınıfları ayrıştırmada güçlü bir rol oynadığını göstermektedir. Buna karşılık

tablonun alt kısmında yer alan çok sayıda özelliğin nihai skorunun sıfıra oldukça yakın olduğu görülmektedir. Bu durum, söz konusu özelliklerin temiz veri senaryosunda sınıflar arasında sınırlı ayırt edicilik sağladığını ve S-SFIX tarafından düşük önem düzeyinde değerlendirildiğini göstermektedir. Genel olarak tablo, önerilen yöntemin yalnızca birkaç baskın özelliği ön plana çıkardığını, kalan özelliklerin ise göreli olarak daha düşük katkı sunduğunu ortaya koymaktadır.

Şekil 5.1, S-SFIX tarafından en yüksek öneme sahip olarak belirlenen ilk 10 özelliğin hem bileşen düzeyindeki skorlarını hem de nihai S-SFIX skorlarını birlikte göstermektedir. Üst bölümde her bir özellik için DIS, CS, AS ve RS bileşenlerinin katkıları ayrı ayrı sunulurken, alt bölümde bu bileşenlerin birleşimiyle elde edilen nihai



Şekil 5.1. S-SFIX tarafından en yüksek öneme sahip olarak belirlenen ilk 10 özellik

önem skorları yer almaktadır. Bu yapı, S-SFIX'in özellikleri yalnızca tek bir ölçüte göre sıralamadığını; ayırt edicilik, tutarlılık, anomali duyarlılığı ve risk tabanlı ayrıştırıcılığı birlikte dikkate alan çok boyutlu bir değerlendirme çerçevesi kullandığını göstermektedir.

Şekil 5.1 incelendiğinde, ilk 10 özellik arasında yer alan değişkenlerin bileşen skorları bakımından birbirinden farklı profiller sergilediği görülmektedir. Bazı özellikler yüksek ayırt edicilik ve tutarlılık ile öne çıkarken, bazıları daha çok risk temelli ayrıştırıcılık ya da anomali duyarlılığı bileşenlerinden destek almaktadır. Bu durum, nihai S-SFIX skorunun tek bir baskın bileşene dayalı olarak değil, birden fazla bileşenin ortak etkisiyle oluştuğunu göstermektedir. Alt bölümde sunulan nihai skorlar da bu çok bileşenli yapının sonucunda hangi özelliklerin çekirdek önem kümesi içinde yer aldığını açık biçimde ortaya koymaktadır.

Genel olarak bu görsel, S-SFIX'in özellik önemini tek boyutlu bir özellik önemliliği yaklaşımıyla değil, çok bileşenli ve daha bütüncül bir puanlama yapısıyla değerlendirdiğini göstermektedir.

5.2.4 Gaussian gürültüsü altında S-SFIX dayanıklılık deney tasarımı

Bu çalışmada önerilen S-SFIX modelinin rastgele veri bozulmaları altındaki kararlılığı da incelenmiştir. Bu bölümdeki temel amaç, doğrudan S-SFIX'in ürettiği özellik skorlarının ve sıralamalarının gürültü altında ne ölçüde korunduğunu değerlendirmektir. Bu nedenle deneysel tasarım, model performansından önce S-SFIX'in kendi iç tutarlılığına ve dayanıklılığına odaklanacak biçimde kurulmuştur.

Bu amaçla, verisetinin sayısal özniteliklerine farklı şiddet seviyelerinde Gaussian gürültüsü eklenmiş ve her bozulma seviyesi için S-SFIX modeli yeniden çalıştırılmıştır. Deney akışında kullanılan gürültü seviyeleri sırasıyla 0.00, 0.01, 0.03, 0.05, 0.10 ve 0.20 olarak belirlenmiştir. Burada 0.00 seviyesi referans niteliğindeki temiz veri senaryosunu temsil ederken, diğer seviyeler giderek artan bozulma senaryolarını ifade etmektedir. Gürültü yalnızca sayısal özelliklere uygulanmış, etiket sütunu ve hedef yapı korunmuştur.

Gürültü üretiminde her özellik için tek tip sabit sapma kullanılmamıştır. Bunun yerine, her bir öznitelik için bozulma seviyesi ilgili özniteliğin standart sapmasına bağlı olarak

tanımlanmıştır. Başka bir ifadeyle, her numerik özellik için bozulmuş değerler

$$x' = x + \mathcal{N}\left(0, \left(\alpha \cdot \sigma_j\right)^2\right) \tag{5.5}$$

Denklem 5.5'teki şeklinde üretilmiştir. Burada α ilgili gürültü seviyesini, σ ise j. özelliğin gürültüsüz veri üzerindeki standart sapmasını göstermektedir. Böylece farklı ölçeklerde dağılan özelliklere kendi varyasyon yapılarıyla orantılı bir gürültü uygulanmıştır. Ayrıca gürültü sonrası numerik değerler istenirse temiz veride gözlenen min-max aralığına kırpılarak aşırı sentetik sapmaların oluşması önlenmiştir.

Her gürültü seviyesi için S-SFIX, aynı veriseti üzerinde yeniden çalıştırılmış ve her öznitelik için DIS, CS, AS, RS ve final_skor değerleri yeniden hesaplanmıştır. Kullanılan implementasyon, S-SFIX'in orijinal sadeleştirilmiş formülasyonuna sadık olacak şekilde tasarlanmıştır. Bu yapı içinde DIS ayırt edicilik bileşenini, CS tutarlılık bileşenini, AS anomali duyarlılığını ve RS risk tabanlı ayrıştırıcılığı ifade etmektedir; nihai skor ise bu bileşenlerin birleşimiyle elde edilmektedir.

Temiz veri ve gürültülü veri senaryoları arasındaki değişimi ölçmek amacıyla hem skor tabanlı hem de sıralama tabanlı dayanıklılık metrikleri kullanılmıştır. Skor tabanlı değerlendirmede, temiz veri ve bozulmuş veri senaryolarında elde edilen final_skor değerleri arasındaki ilişkiyi ölçmek üzere Skor Pearson korelasyon hesaplanmıştır. Sıralama tabanlı değerlendirmede ise temiz veri ve gürültülü veri altında elde edilen özellik sıralamaları karşılaştırılmış; bu amaçla Rank Spearman ve Kendall Tau katsayısı kullanılmıştır. Ayrıca her bir özelliğin temiz veri sırasına göre ne kadar yer değiştirdiğini göstermek üzere ortalama mutlak sıra kayması ve maksimum sıra kayması hesaplanmıştır.

Bu çalışmada özellikle üst sıralı özellik kümelerinin korunması ayrıca önemli görüldüğünden, İlk k temelli bir benzerlik analizi de yürütülmüştür. Bu amaçla k = 5, 10, 15, 20 olacak şekilde temiz veri senaryosunda elde edilen üst sıralı özellik kümeleri ile gürültü altındaki karşılıkları karşılaştırılmış ve Jaccard benzerliği hesaplanmıştır. Buna ek olarak, temiz veri koşullarında İlk k içinde bulunan kaç özelliğin gürültü sonrasında da aynı üst kümelerde kaldığı korunan özellik sayısı ile raporlanmıştır. Böylece hem genel sıralama korelasyonu hem de uygulamada en kritik olan "en önemli özelliklerin korunup korunmadığı" sorusu da doğrudan cevaplanmıştır.

Bu deney tasarımı sayesinde, çalışmanın giriş bölümünde tartışılan Yapay Zekâ Dayanıklılığı kavramları ilk kez doğrudan S-SFIX düzeyinde sınanmış olmaktadır. Böylece bu kavramlar önerilen yöntem için doğrudan deneysel olarak ölçülebilir hale getirilmiştir.

Tablo 5.7. Gaussian gürültüsü altında S-SFIX stabilite deney düzeni

No	Bileşen	Açıklama
1	Temiz veri senaryosu	Referans sıralamanın elde edilmesi amacıyla gürültü eklenmemiş veri kullanılmıştır.
2	Gürültü seviyeleri	0.00, 0.01, 0.03, 0.05, 0.10 ve 0.20
3	Gürültü uygulama alanı	Gürültü yalnızca sayısal özniteliklere uygulanmıştır.
4	Gürültü üretimi	Her bir özellik için standart sapma temelli Gaussian gürültü üretilmiştir.
5	S-SFIX'in yeniden çalıştırılması	Her gürültü seviyesi için S-SFIX skorları yeniden hesaplanmıştır.
6	Skor tabanlı metrikler	Skor pearson kullanılmıştır.
7	Sıralama tabanlı metrikler	Rank Spearman, Kendall Tau, ortalama sıra kayması ve maksimum sıra kayması hesaplanmıştır.
8	İlk k kararlılığı	İlk 5, İlk 10, İlk 15 ve İlk 20 kümeleri için Jaccard ve korunan özellik sayısı ölçülmüştür.
9	Görsel çıktılar	Rank korelasyonu, rank shift, İlk k Jaccard ve İlk 20 skor karşılaştırmaları görselleştirilmiştir.

Tablo 5.7'de Gaussian gürültüsü altında yürütülen S-SFIX dayanıklılık deneyinin temel bileşenleri özetlenmektedir. Bu tasarım, doğrudan özellik sıralama mekanizmasının kararlılığına odaklanması bakımından çalışmanın Yapay Zekâ Dayanıklılığı boyutunu güçlendirmektedir.

5.3 Bulgular

Bu bölümde, S-SFIX tarafından belirlenen özellik sıralamasının sınıflandırma performansına etkisi analiz edilmiştir.

Tablo 5.8 incelendiğinde, tam özellik kümesi ile eğitilen FFNN modelinin en yüksek performansı verdiği görülmektedir. Bununla birlikte, özellik sayısı azaltılarak elde edilen sonuçlar değerlendirildiğinde SHAP ve S-SFIX tabanlı özellik seçimi yaklaşımlarının özellikle 15 ve 20 özellik seviyelerinde tüm verisetle eğitilmiş modele oldukça yakın

performans ürettiği dikkat çekmektedir. LIME yöntemi 10, 15 ve 20 özellik seviyelerinde kararlı ve yüksek sonuçlar vermiş, ancak 5 özellikli senaryoda performans belirgin biçimde düşmüştür. Random Forest özellik önemliliği tabanlı seçim ise düşük özellik sayılarında en zayıf sonuçları üretmiş, özellikle 5 ve 10 özellikli kurulumlarda diğer yöntemlerin gerisinde kalmıştır. Genel olarak, özellik sayısı arttıkça tüm yöntemlerde performans artışı gözlenmekle birlikte, SHAP ve S-SFIX'in daha az sayıda özellikle dahi yüksek sınıflandırma başarısı sağlayabildiği görülmektedir. Bu bulgu, önerilen S-SFIX yaklaşımının düşük boyutlu temsil uzayında da etkili bir özellik seçimi sağlayabildiğini göstermektedir.

Tablo 5.8. Farklı XAI yöntemleri ve özellik sayıları altında FFNN performans sonuçları

Yöntem	Özellik Sayısı	Doğruluk	Benign Kesinlik	Benign Duyarlılık	Attack Kesinlik	Attack Duyarlılık	Makro F1
Tüm	46	0.9845	0.9734	0.9966	0.9965	0.9722	0.9845
LIME	5	0.9600	0.9576	0.9634	0.9625	0.9565	0.9600
LIME	10	0.9819	0.9670	0.9983	0.9982	0.9652	0.9819
LIME	15	0.9816	0.9674	0.9971	0.9969	0.9658	0.9816
LIME	20	0.9821	0.9669	0.9986	0.9985	0.9652	0.9820
RF	5	0.9525	0.9543	0.9515	0.9507	0.9536	0.9525
RF	10	0.9588	0.9502	0.9691	0.9679	0.9483	0.9588
RF	15	0.9734	0.9559	0.9931	0.9926	0.9533	0.9733
RF	20	0.9738	0.9571	0.9925	0.9921	0.9547	0.9738
S-SFIX	5	0.9689	0.9528	0.9872	0.9865	0.9502	0.9689
S-SFIX	10	0.9696	0.9562	0.9849	0.9841	0.9541	0.9696
S-SFIX	15	0.9793	0.9661	0.9939	0.9936	0.9645	0.9793
S-SFIX	20	0.9805	0.9673	0.9949	0.9947	0.9657	0.9804
SHAP	5	0.9785	0.9621	0.9967	0.9965	0.9600	0.9785
SHAP	10	0.9767	0.9678	0.9867	0.9861	0.9666	0.9767
SHAP	15	0.9823	0.9673	0.9987	0.9986	0.9656	0.9823
SHAP	20	0.9823	0.9681	0.9978	0.9977	0.9665	0.9823

Yöntemler arası karşılaştırma detaylandırıldığında, 5 özellikli senaryoda en yüksek doğruluk değerinin SHAP ile, en düşük değerin ise Random Forest özellik önemliliği ile elde edildiği görülmektedir. 10 özellik düzeyinde LIME en yüksek sonucu verirken, 15 ve 20 özellik düzeylerinde SHAP ve S-SFIX tabanlı seçimlerin öne çıktığı

anlaşılmaktadır. Özellikle S-SFIX-20 ve SHAP-20 sonuçlarının tümüyle eğitilmiş modele oldukça yaklaşması, bu iki yöntemin daha sınırlı bir özellik seti ile bilgi kaybını düşük tutabildiğini göstermektedir. Buna ek olarak, S-SFIX-15 ve S-SFIX-20 sonuçlarının Random Forest tabanlı tüm kurulumlardan ve LIME-5 sonucundan daha yüksek olması, önerilen yöntemin klasik önemlilik tabanlı seçime göre daha güçlü bir özellik sıralama sunduğunu desteklemektedir.

5.3.1 Performansın değerlendirilmesi

Tablo 5.8 incelendiğinde, özellik sayısı önemli ölçüde azaltılmasına rağmen performansın dramatik biçimde düşmediği görülmektedir. Özellikle S-SFIX tabanlı özellik seçimi yaklaşımı, daha düşük boyutlu özellik kümeleri ile dahi yüksek doğruluk ve macro F1 değerleri üretmiştir. Tam özellik kümesi ile eğitilen model %98.45 doğruluk sağlarken, yalnızca 20 özellik kullanan S-SFIX modeli %98.05, 15 özellik kullanan model ise %97.93 doğruluk elde etmiştir. Bu durum, önerilen yöntemin sınıflandırma görevinde en yüksek bilgi taşıyan çekirdek özellikleri başarıyla önceliklendirdiğini göstermektedir.

Özellik boyutunun 46'dan 20'ye düşürülmesi yaklaşık %56.5, 46'dan 15'e düşürülmesi ise yaklaşık %67.4 oranında bir azaltım anlamına gelmektedir. Buna rağmen performans kaybının oldukça sınırlı kalması, S-SFIX'in yüksek ayırt edici güce sahip özellikleri ön plana çıkardığını göstermektedir. Daha agresif boyut indirgeme senaryolarında bile S-SFIX-10 ve S-SFIX-5 modellerinin sırasıyla %96.96 ve %96.89 doğruluk üretmesi, önerilen yöntemin düşük boyutlu temsil uzaylarında da etkili olduğunu desteklemektedir.

Yöntemler arası karşılaştırmada S-SFIX'in özellikle Random Forest özellik önemliliği tabanlı seçime karşı belirgin biçimde daha güçlü sonuçlar verdiği görülmektedir. Ayrıca LIME tabanlı 5 özellikli modele göre de daha yüksek ve daha dengeli bir performans üretmiştir. SHAP, 15 ve 20 özellikli senaryolarda en yüksek mutlak sonuçları vermiş olsa da, S-SFIX'in bu değerlere oldukça yakın performans sergilemesi önemlidir. Çünkü SHAP bir post-hoc açıklama yaklaşımı iken, S-SFIX doğrudan özellik puanlama ve sıralama amacıyla geliştirilmiş özgün bir mekanizma sunmaktadır. Bu nedenle S-SFIX'in özellikle 15 ve 20 özellikli kurulumlarda tümüyle eğitilmiş modele ve SHAP tabanlı seçimlere yakın sonuçlar üretmesi, yöntemin hem etkili hem de pratik olarak uygulanabilir bir özellik seçimi yaklaşımı olduğunu göstermektedir.

5.3.2 Gaussian gürültüsü altında S-SFIX dayanıklılık sonuçları

Tablo 5.9, Gaussian gürültüsü seviyeleri arttıkça S-SFIX tarafından üretilen özellik sıralamasında meydana gelen değişimleri ayrıntılı biçimde göstermektedir. Genel olarak değerlendirildiğinde, yöntem düşük ve orta seviyeli gürültü altında tamamen farklı bir sıralama üretmemiş; özellikle üst sıralarda yer alan çekirdek özellik kümesini büyük ölçüde korumuştur. En dikkat çekici bulgu, ilk iki sırada yer alan Variance ve Magnitue özniteliklerinin tüm gürültü seviyelerinde yalnızca kendi aralarında yer değiştirmesi ve her durumda ilk iki sırada kalmaya devam etmesidir. Benzer şekilde, Max ve Duration özellikleri de ilk dört sıra içinde konumlarını korumuş; yalnızca düşük düzeyli sıralama değişimleri göstermiştir. AVG, Tot size ve ack_flag_number özellikleri ise tüm gürültü seviyelerinde sırasıyla beşinci, altıncı ve onuncu sıradaki yerlerini aynen muhafaza etmiştir. Bu durum, S-SFIX'in en ayırt edici çekirdek özellikleri düşük ve orta seviyeli bozulmalar altında da kararlı biçimde ön planda tutabildiğini göstermektedir.

Tablo 5.9. Gaussian gürültüsüne göre 46 özelliğin S-SFIX sıralama değişimi

Özellikler	**0.00 (Gürültüsüz)**	**0.01**	**0.03**	**0.05**	**0.10**	**0.20**
Variance	1	2	1	1	1	1
Magnitude	2	1	2	2	2	2
Max	3	3	4	4	4	4
Duration	4	4	3	3	3	3
AVG	5	5	5	5	5	5
Tot size	6	6	6	6	6	6
Tot sum	7	7	7	7	9	9
Std	8	8	8	8	7	7
Radius	9	9	9	9	8	8
ack_flag_number	10	10	10	10	10	10
ICMP	11	11	12	12	15	16
HTTPS	12	12	11	11	11	11
syn_flag_number	13	13	13	15	16	15
rst_count	14	15	15	13	12	12
Header_Length	15	16	16	16	13	13
Min	16	14	14	14	14	14
rst_flag_number	17	18	19	21	21	21
fin_flag_number	18	20	20	22	22	22
TCP	19	17	17	17	17	17
syn_count	20	19	18	18	18	20
urg_count	21	21	21	19	19	19
Covariance	22	22	22	20	20	18
UDP	23	23	23	23	23	23
psh_flag_number	24	25	25	25	25	25

Özellikler	0.00 (Gürültüsüz)	0.01	0.03	0.05	0.10	0.20
Protocol Type	25	24	24	24	24	24
fin_count	26	26	26	26	26	26
ack_count	27	27	27	27	27	27
Rate	28	29	29	29	29	31
Srate	29	28	28	28	28	29
IPv	30	33	33	33	34	34
LLC	31	32	32	31	32	33
flow_duration	32	30	30	30	30	30
HTTP	33	31	31	32	35	35
DNS	34	35	35	36	36	36
ARP	35	36	37	37	37	37
IAT	36	34	34	34	33	32
SSH	37	38	38	38	38	39
Drate	38	39	39	39	40	40
Weight	39	37	36	35	31	28
Number	40	40	40	40	39	38
DHCP	41	41	41	41	41	41
IRC	42	42	42	42	42	42
Telnet	43	43	43	43	43	43
cwr_flag_number	44	44	44	44	44	44
ece_flag_number	45	45	45	45	45	45
SMTP	46	46	46	46	46	46

Tablo 5.9 aynı zamanda, üst sıralardaki bazı özelliklerde sınırlı fakat anlamlı kaymaların ortaya çıktığını da göstermektedir. Örneğin, Tot sum temiz veri altında yedinci sırada yer alırken, gürültü seviyesi (0.10) ve (0.20)'ye yükseldiğinde dokuzuncu sıraya gerilemiştir. Buna karşılık Std sekizinci sıradan yedinci sıraya, Radius ise dokuzuncu sıradan sekizinci sıraya yükselmiştir. Bu tür değişimler, üst sıralardaki genel özellik setinin korunduğunu ancak benzer ayırt edicilik gücüne sahip bazı özelliklerin göreli konumlarının gürültüden etkilenebildiğini göstermektedir. Bununla birlikte, İlk 10 içinde yer alan özelliklerin tamamına yakınının tüm bozulma seviyelerinde yine aynı kümeye yakın konumlarda kalması, S-SFIX'in kritik özellik kümesini koruma açısından güçlü bir kararlılık sergilediğini düşündürmektedir.

Orta sıralarda yer alan özellikler incelendiğinde, gürültüye karşı hassasiyetin daha belirgin hale geldiği görülmektedir. Özellikle ICMP özelliği temiz veri altında 11. sırada yer alırken, gürültü seviyesi arttıkça sırasıyla 12., 15. ve 16. sıralara gerilemiştir. Benzer biçimde *syn_flag_number* 13. sıradan başlayarak 15. ve 16. sıralara kadar düşmüş; *rst_count* ve *Header_Length* gibi özellikler ise gürültü arttıkça orta bant içerisinde yer

değiştirmiştir. Buna karşın HTTPS özelliğinin 11-12 bandında oldukça sınırlı değişim göstermesi, her orta seviye özelliğin aynı derecede hassas olmadığını ortaya koymaktadır. Bu gözlem, S-SFIX'in özellikle marjinal veya birbirine yakın önem skorlarına sahip özelliklerde daha fazla sıralama oynaklığı üretebildiğini, ancak bu oynaklığın çekirdek özellik setine kıyasla daha çok orta bantta yoğunlaştığını göstermektedir.

Alt ve alt-orta sıralarda ise iki farklı eğilim dikkat çekmektedir. Birinci eğilim, bazı düşük önem düzeyindeki özelliklerin gürültü altında daha da geriye düşmesidir. Örneğin *rst_flag_number* temiz senaryoda 17. sıradayken gürültü seviyeleri altında 21. sıraya kadar gerilemiş, *fin_flag_number* ise 18. sıradan 22. sıraya düşmüştür. HTTP, DNS ve ARP gibi özellikler de benzer biçimde birkaç sıra geriye kaymıştır. İkinci eğilim ise, temiz senaryoda daha düşük sıralarda bulunan bazı özelliklerin gürültü altında görece yukarı çıkabilmesidir. Özellikle *Weight* özelliği temiz veri altında 39. sırada yer alırken, gürültü seviyesi (0.20)'de 28. sıraya kadar yükselmiştir. Benzer şekilde *Covariance* 22. sıradan 18. sıraya yükselmiş, *Number* 40. sıradan 38. sıraya çıkmıştır. Bu durum, gürültü arttıkça bazı ikincil özelliklerin göreli ayırt ediciliklerinin artabildiğini ya da bazı daha hassas özelliklerin zayıflamasıyla göreli olarak üst sıralara taşınabildiğini göstermektedir.

Tablonun alt kısmı incelendiğinde ise bazı özelliklerin neredeyse tamamen değişmeden kaldığı görülmektedir. Özellikle *DHCP, IRC, Telnet, cwr_flag_number, ece_flag_number* ve *SMTP* gibi özellikler tüm gürültü seviyelerinde aynı sıralarını korumuştur. Bu özellikler temiz veri senaryosunda da düşük ayırt edicilik düzeyine sahip olduklarından, gürültü eklenmesiyle birlikte göreli önem yapısında anlamlı bir farklılaşma üretmemiştir. Bu gözlem, S-SFIX'in düşük önem düzeyine sahip bazı özellikleri tutarlı biçimde düşük önemde değerlendirdiğini ve gürültünün bu tür marjinal özelliklerde yapay bir önem sıçraması üretmediğini göstermesi açısından da önemlidir.

Genel olarak değerlendirildiğinde, Gaussian gürültüsü altında elde edilen sıralama değişimleri S-SFIX'in tamamen kararsız bir yapı sergilemediğini; aksine özellikle üst sıralı ve çekirdek özellik kümesinde yüksek düzeyde kararlılık sağladığını göstermektedir. Gürültü seviyesi arttıkça orta ve alt sıralarda daha belirgin yer değiştirmeler gözlense de, en kritik özelliklerin büyük bölümünün ilk 10-15 sıra içinde

kalmaya devam etmesi, önerilen yöntemin veri bozulmaları altında da ayırt edici örüntüleri koruyabildiğine işaret etmektedir. Bu sonuç, S-SFIX'in belirli ölçüde Yapay Zekâ Dayanıklılığı özelliği taşıyan bir özellik skorlama mekanizması sunduğunu desteklemektedir.

Tablo 5.10, Gaussian gürültüsü seviyeleri arttıkça S-SFIX tarafından üretilen özellik sıralamasında meydana gelen toplam değişimin özetini göstermektedir. Sonuçlar, düşük gürültü seviyelerinde (0.01 ve 0.03) özelliklerin yaklaşık yarısının sıralarını tamamen koruduğunu, kalan büyük bölümünün ise yalnızca 1-2 sıra yer değiştirdiğini ortaya koymaktadır. Gürültü seviyesi (0.05)'e ulaştığında 3-5 sıra aralığındaki kaymaların arttığı görülmekte, (0.10) ve (0.20) seviyelerinde ise orta düzey kaymalar daha belirgin hale gelmektedir. Buna rağmen dikkat çekici biçimde temiz senaryodaki İlk 10 özellik kümesinin tüm gürültü seviyelerinde tamamen korunduğu görülmektedir. İlk 20 kümesi ise (0.01) ve (0.03) seviyelerinde tamamen korunurken, daha yüksek bozulma düzeylerinde yalnızca iki özelliğin bu küme dışına çıktığı gözlenmektedir. Bu bulgu, S-SFIX'in genel sıralama yapısında kademeli bir bozulma oluşsa da özellikle en kritik özellikleri koruma bakımından güçlü bir dayanıklılık sergilediğini göstermektedir.

Tablo 5.10. Gaussian gürültüsüne göre S-SFIX özellik sıralama değişimi

Gürültü Seviyesi	Değişmeyen	1–2 Sıra Değişen	3–5 Sıra Değişen	5'ten Fazla Değişen	Korunan İlk 10	Korunan İlk 20
0.01	23	22	1	0	10	20
0.03	21	23	2	0	10	20
0.05	19	23	4	0	10	18
0.10	14	25	6	1	10	18
0.20	16	22	7	1	10	18

En yüksek bozulma seviyesi olan (0.20) altında en belirgin sıra değişimleri Tablo 5.11'de sunulmuştur. Sonuçlar, bazı düşük veya orta önem düzeyindeki özelliklerin gürültü altında göreli olarak üst sıralara çıkabildiğini göstermektedir. Özellikle *Weight* özelliği 39. sıradan 28. sıraya yükselerek en dikkat çekici değişimi göstermiştir. Benzer biçimde *Covariance* ve *IAT* özellikleri de birkaç sıra yukarı taşınmıştır. Buna *karşılık ICMP, rst_flag_number, fin_flag_number* ve *IPv* gibi özellikler gürültü altında daha alt sıralara gerilemiştir. Bu tablo, S-SFIX'in üst çekirdek özellik kümesini koruduğunu; ancak özellikle orta ve alt bantta yer alan bazı özelliklerin göreli öneminde gürültüye bağlı yeniden düzenlenmeler meydana gelebildiğini göstermektedir.

Tablo 5.11. Noise = 0.20 seviyesinde en fazla sıra değişimi gösteren özellikler

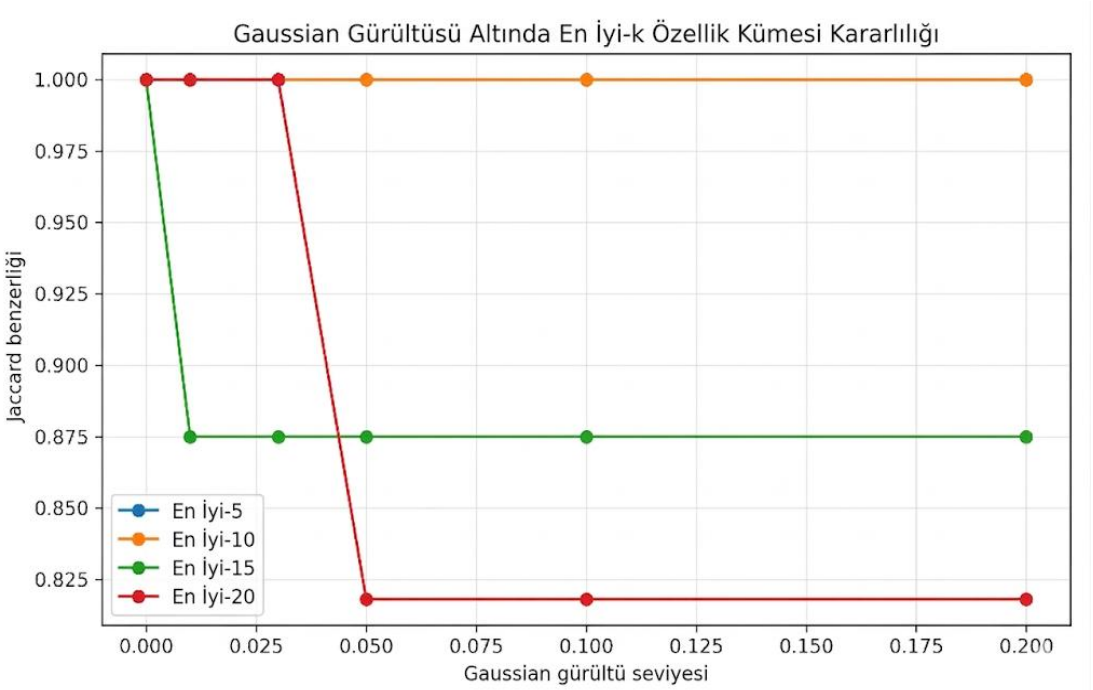
En çok yükselen özellikler	Temiz → Gürültülü	Değişim	En çok düşen özellikler	Temiz → Gürültülü	Değişim
Weight	39 → 28	+11	ICMP	11 → 16	-5
Covariance	22 → 18	+4	rst_flag_number	17 → 21	-4
IAT	36 → 32	+4	fin_flag_number	18 → 22	-4
Number	40 → 38	+2	IPv	30 → 34	-4
Protocol Type	25 → 24	+1	Rate	28 → 31	-3

Tablo 5.12, temiz veri senaryosunda en yüksek öneme sahip olan İlk 10 özelliğin gürültü altındaki konumlarını göstermektedir. Sonuçlar, bu çekirdek özellik kümesinin tüm gürültü seviyelerinde tamamen korunduğunu ortaya koymaktadır. *Variance* ve *Magnitude* her senaryoda ilk iki sırada yer almakta; *Max* ve *Duration* ilk dört içindeki yerini korumaktadır. *AVG, Tot size* ve *ack_flag_number* ise tüm bozulma seviyelerinde sıralarını hiç değiştirmemiştir. Üst sıralardaki tek dikkat çekici oynama, *Tot sum, Std* ve *Radius* özelliklerinin birbirine yakın önem düzeyleri nedeniyle 7-9 bandında sınırlı yer değiştirmesidir. Bununla birlikte, bu üç özellik de tüm gürültü seviyelerinde yine İlk 10 içinde kalmıştır. Bu sonuç, S-SFIX'in en kritik özellik kümesini düşük ve orta seviyeli Gaussian gürültüsü altında yüksek düzeyde koruyabildiğini ve çekirdek sıralama yapısının bozulmadığını göstermektedir.

Tablo 5.12. İlk 10 özelliğin Gaussian gürültüsü altındaki konumları

Özellik	Temiz	0.01	0.03	0.05	0.10	0.20
Variance	1	2	1	1	1	1
Magnitue	2	1	2	2	2	2
Max	3	3	4	4	4	4
Duration	4	4	3	3	3	3
AVG	5	5	5	5	5	5
Tot size	6	6	6	6	6	6
Tot sum	7	7	7	7	9	9
Std	8	8	8	8	7	7
Radius	9	9	9	9	8	8
ack_flag_number	10	10	10	10	10	10

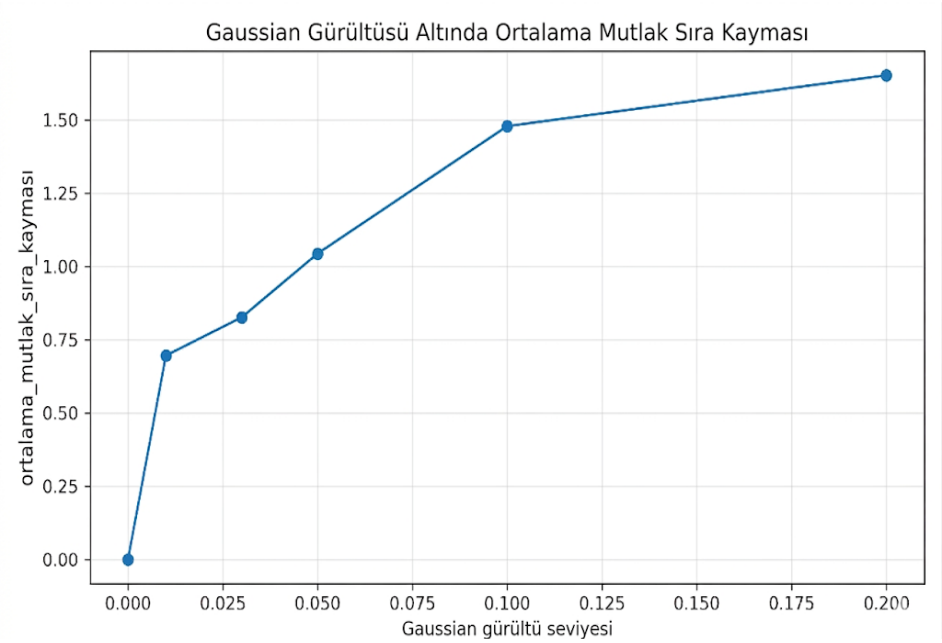
Şekil 5.2, Gaussian gürültüsü seviyeleri arttıkça İlk k özellik kümelerinin ne ölçüde korunduğunu göstermektedir. Sonuçlar, özellikle üst sıralı özellik setinin oldukça kararlı



Şekil 5.2. Gaussian gürültüsü seviyelerine göre ilk k özellik kümesi kararlılığı (Jaccard benzerliği)

olduğunu ortaya koymaktadır. Buna göre İlk 10 özellik kümesi tüm gürültü seviyelerinde tamamen korunmuş ve Jaccard benzerliği her durumda 1.00 olarak kalmıştır. İlk 15 özellik kümesi düşük ve orta seviyeli bozulmalarda büyük ölçüde korunmuş, Jaccard benzerliği 0.875 düzeyinde sabitlenmiştir. İlk 20 özellik kümesi ise düşük bozulma seviyelerinde tamamen korunurken, gürültü seviyesi 0.05 ve sonrasında yaklaşık 0.82 düzeyine gerilemiştir. Bu bulgular birlikte değerlendirildiğinde, S-SFIX'in özellikle en kritik özellik kümesini koruma konusunda yüksek düzeyde dayanıklılık sergilediği, gürültü arttıkça ise değişimin daha çok genişleyen ve marjinal önemlilik düzeyindeki özellik kümelerinde ortaya çıktığı anlaşılmaktadır.

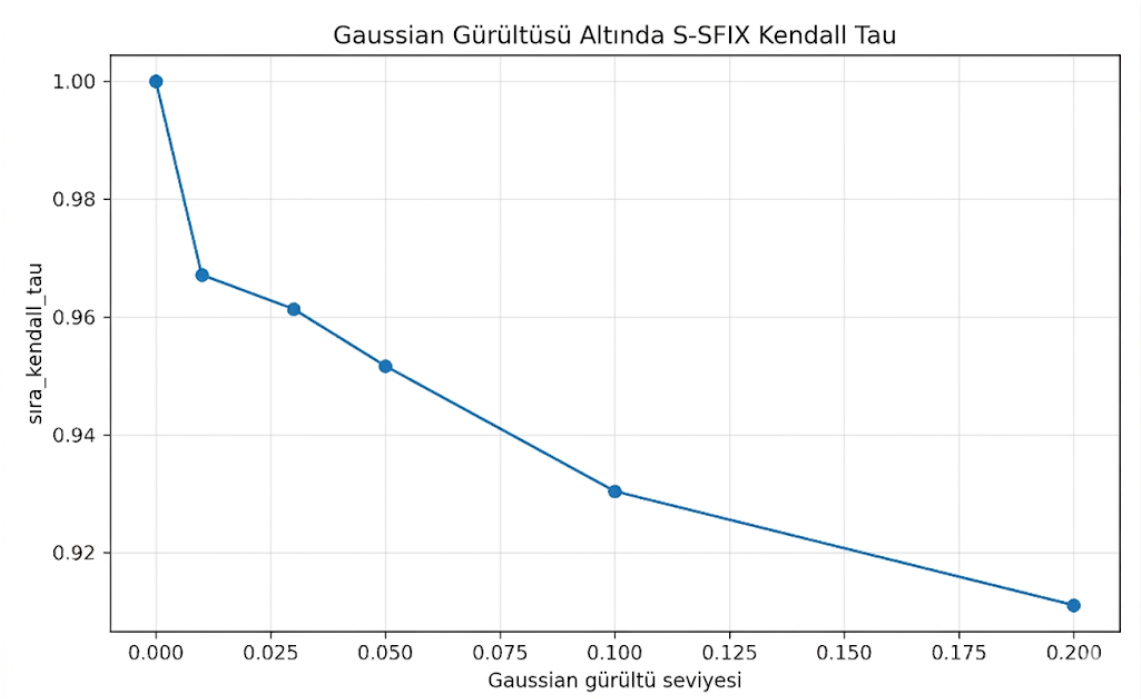
Şekil 5.3, gürültü seviyesi arttıkça özellik sıralamasında meydana gelen ortalama yer değişimini göstermektedir. Eğri incelendiğinde, temiz veri senaryosunda doğal olarak sıfır olan ortalama sıra kaymasının düşük gürültü seviyelerinde sınırlı kaldığı, ancak gürültü şiddeti arttıkça kademeli biçimde yükseldiği görülmektedir. Özellikle 0.10 ve 0.20 seviyelerinde ortalama mutlak sıra kaymasının daha belirgin hale gelmesi, gürültünün özellik sıralama üzerinde birikimli bir etki yarattığını göstermektedir. Bununla birlikte, bu artışın ani ve kopuk bir bozulmadan ziyade düzenli ve kademeli biçimde gerçekleşmesi önemlidir. Bu durum, S-SFIX'in gürültü altında tamamen kararsız hale gelmediğini; ancak özellikle orta ve alt sıralardaki özelliklerde giderek artan bir yeniden düzenlenme oluşturduğunu göstermektedir.



Şekil 5.3. Gaussian gürültüsü seviyelerine göre ortalama mutlak sıra kayması

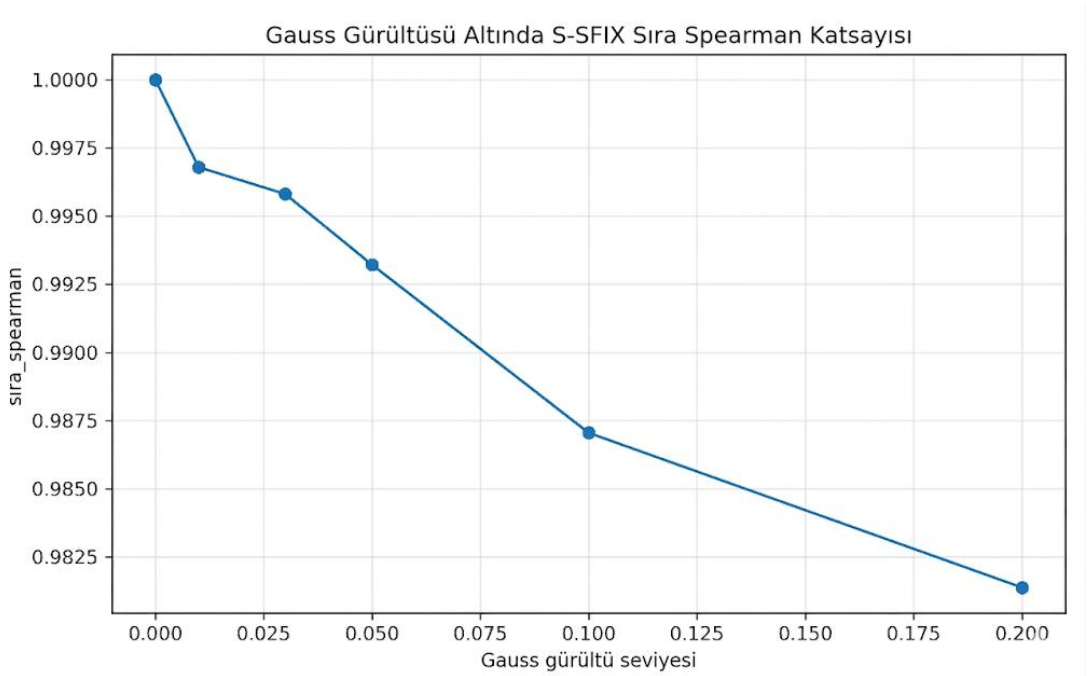
Şekil 5.4, temiz veri sıralaması ile gürültülü veri sıralaması arasındaki sıra ilişkisinin Kendall Tau katsayısı üzerinden nasıl değiştiğini göstermektedir. Sonuçlar, gürültü seviyesi arttıkça Kendall Tau değerinin kademeli olarak azaldığını, ancak bu azalmanın sınırlı kaldığını ortaya koymaktadır. En yüksek gürültü seviyesi olan 0.20 altında bile korelasyonun yaklaşık 0.91 düzeyinde kalması, sıralama yapısının önemli ölçüde korunduğunu göstermektedir. Başka bir ifadeyle, gürültü S-SFIX sıralaması üzerinde etkili olmakla birlikte, özellikler arasındaki genel göreli düzeni tamamen bozacak ölçüde bir sapmaya yol açmamaktadır. Bu bulgu, yöntemin özellikle küresel sıralama yapısı

açısından güçlü bir dayanıklılık sergilediğini desteklemektedir.



Şekil 5.4. Gaussian gürültüsü seviyelerine göre S-SFIX sıralama yapısının Kendall Tau korelasyonu

Şekil 5.5, temiz veri ve gürültülü veri senaryoları arasındaki sıra korelasyonunu Spearman katsayısı üzerinden göstermektedir. Eğri, gürültü seviyesi arttıkça korelasyon değerinin kademeli olarak düştüğünü; ancak tüm senaryolarda oldukça yüksek kaldığını ortaya koymaktadır. En yüksek bozulma düzeyinde dahi Spearman korelasyonunun yaklaşık 0.98 seviyesinde olması, S-SFIX tarafından üretilen sıralamanın genel yapısının güçlü biçimde korunduğunu göstermektedir. Kendall Tau grafiği ile değerlendirildiğinde bu sonuç, gürültü altında bazı yerel sıra değişimleri oluşmasına rağmen, özelliklerin genel önem düzeninin büyük ölçüde muhafaza edildiğini göstermektedir.



Şekil 5.5. Gaussian gürültüsü seviyelerine göre S-SFIX sıralama yapısının Spearman korelasyonu

Tablo 5.13, Gaussian gürültüsü seviyeleri arttıkça S-SFIX tarafından üretilen skor ve sıralama yapısındaki değişimi sayısal olarak özetlemektedir. Sonuçlar, gürültü seviyesi yükseldikçe hem skor korelasyonunun hem de sıra korelasyonunun kademeli olarak azaldığını, buna karşın tüm seviyelerde yüksek kaldığını göstermektedir. Özellikle rank Spearman katsayısının en yüksek bozulma düzeyinde dahi yaklaşık 0.98, Kendall Tau katsayısının ise yaklaşık 0.91 seviyesinde kalması, genel sıralama yapısının büyük ölçüde korunduğuna işaret etmektedir. Aynı zamanda ortalama mutlak sıra kaymasının düşük gürültü seviyelerinde oldukça sınırlı kaldığı, gürültü arttıkça ise düzenli biçimde yükseldiği görülmektedir. İlk k benzerlik sonuçları bu bulguyu daha da güçlendirmektedir. İlk 5 ve İlk 10 özellik kümeleri tüm bozulma seviyelerinde tamamen korunurken, İlk 20 kümesi yalnızca daha yüksek gürültü seviyelerinde sınırlı bir bozulma göstermiştir. Bu sonuçlar birlikte değerlendirildiğinde, S-SFIX'in özellikle çekirdek özellik setini koruma bakımından güçlü bir dayanıklılık sergilediği, bozulmanın ise daha çok genişletilmiş ve marjinal importance düzeyindeki özellik kümelerinde ortaya çıktığı anlaşılmaktadır.

Tablo 5.13. Gaussian gürültüsü seviyesine göre S-SFIX dayanıklılık metrikleri

Gürültü Seviyesi	Skor Pearson	Rank Spearman	Kendall Tau	Ort. Mutlak Sıra Kayması	Maks. Mutlak Sıra Kayması	İlk 5 Jaccard	İlk 10 Jaccard	İlk 20 Jaccard
0.00	1.0000	1.0000	1.0000	0.0000	0	1.0000	1.0000	1.0000
0.01	0.9968	0.9968	0.9672	0.6957	3	1.0000	1.0000	1.0000
0.03	0.9958	0.9958	0.9614	0.8261	3	1.0000	1.0000	1.0000
0.05	0.9932	0.9932	0.9517	1.0435	4	1.0000	1.0000	0.8182
0.10	0.9870	0.9870	0.9304	1.4783	8	1.0000	1.0000	0.8182
0.20	0.9813	0.9814	0.9111	1.6522	11	1.0000	1.0000	0.8182

Grafiksel sonuçlar da tablo tabanlı analizleri desteklemektedir. İlk k Jaccard benzerliğinin özellikle İlk 10 düzeyinde tüm gürültü seviyelerinde 1.00 olarak kalması, S-SFIX'in en kritik özellik kümesini tamamen koruyabildiğini göstermektedir. Ortalama mutlak sıra kaymasının gürültü ile birlikte kademeli biçimde artması, yeniden sıralamanın özellikle orta ve alt sırada yoğunlaştığını ortaya koymaktadır. Buna rağmen hem Spearman hem de Kendall Tau korelasyonlarının yüksek kalması, genel sıralama yapısının bozulmadığını ve S-SFIX'in gürültü altında anlamlı düzeyde kararlı kaldığını göstermektedir. Dolayısıyla, önerilen yöntem, bozulmuş veri koşullarında da çekirdek özellik düzenini muhafaza eden bir özellik puanlama mekanizması sunmaktadır.

5.4 Tartışma

Bu çalışmada elde edilen bulgular, S-SFIX yaklaşımının çok sayıda saldırı alt türünü içeren heterojen bir IoT saldırı tespit probleminde anlamlı ve yüksek bilgi taşıyan özellikleri belirleyebildiğini göstermektedir. Çalışmada tüm saldırılar tek bir *Attack* üst sınıfı altında birleştirilmiş, *Benign* örnekler ise aynı verisetinin homojen alt parçalarından kontrollü biçimde desteklenerek daha dengeli bir öğrenme yapısı kurulmuştur. Bu yapı, önerilen yöntemin daha gerçekçi ve daha zorlayıcı bir tehdit ortamında değerlendirilmesine olanak sağlamıştır.

5.4.1 S-SFIX'in özellik seçme yeteneği

Bulgular, S-SFIX'in düşük boyutlu fakat bilgi yoğun özellik kümeleri oluşturabildiğini açıkça göstermektedir. Tam özellik kümesi ile eğitilen model en yüksek mutlak performansı verse de, yalnızca 15 ve 20 özellik ile eğitilen S-SFIX tabanlı modellerin tam modele oldukça yakın sonuçlar üretmesi dikkat çekicidir. Bu durum, tüm özelliklerin aynı ölçüde gerekli olmadığını; S-SFIX tarafından üst sıralarda belirlenen özelliklerin

sınıflandırma kararının büyük bölümünü taşıdığını göstermektedir.

Özellik sayısındaki belirgin azalmaya rağmen performans kaybının sınırlı kalması, S-SFIX'in ayırt edicilik düzeyi yüksek değişkenleri etkili biçimde önceliklendirdiğini göstermektedir. Dolayısıyla yöntem, yalnızca yorumlanabilir bir sıralama üretmekle kalmamakta, aynı zamanda pratik anlamda etkili bir boyut indirgeme mekanizması da sunmaktadır. Bu sonuç, S-SFIX'in açıklanabilirlik ve operasyonel verimlilik arasında dengeli bir yapı sunduğunu desteklemektedir.

5.4.2 Alternatif yöntemlerle karşılaştırmalı değerlendirme

Bu çalışmanın önemli yönlerinden biri, S-SFIX'in SHAP, LIME ve Random Forest özellik önemliliği tabanlı yaklaşımlar ile doğrudan karşılaştırılmış olmasıdır. Bu karşılaştırma, önerilen yöntemin literatürde yerleşik açıklanabilirlik ve önemlilik tabanlı yöntemler karşısında nesnel biçimde değerlendirilmesini sağlamıştır.

Karşılaştırmalı sonuçlar, S-SFIX'in özellikle Random Forest tabanlı özellik önemliliği yaklaşımına göre daha güçlü ve daha tutarlı sonuçlar verdiğini göstermektedir. Düşük özellik sayılarında Random Forest tabanlı seçimlerin daha belirgin performans kaybı yaşadığı görülürken, S-SFIX aynı koşullarda daha dengeli sonuçlar üretmiştir. LIME ile karşılaştırıldığında ise S-SFIX'in özellikle küçük özellik kümelerinde daha istikrarlı davrandığı, LIME'ın ise daha geniş özellik setlerinde toparlandığı görülmektedir. SHAP bazı senaryolarda en yüksek mutlak performansı vermiş olsa da, S-SFIX'in bu sonuçlara oldukça yakın değerler üretmesi önemlidir. Çünkü SHAP eğitilmiş model davranışından önemlilik türeten bir post-hoc yaklaşım iken, S-SFIX doğrudan özellik skorlama ve sıralama amacıyla tasarlanmış özgün bir mekanizmadır.

Bu çerçevede değerlendirildiğinde, S-SFIX'in SHAP'a yakın, Random Forest'a ise çoğu kurulumda üstün performans üretmesi, uygulamalı sınıflandırma başarımı bakımından da güçlü bir alternatif olduğunu göstermektedir.

5.4.3 Açıklanabilirlik ve performans arasındaki ilişki

Çalışmanın temel bulgularından biri, açıklanabilirlik tabanlı özellik seçiminin performans ile zorunlu bir çelişki içinde olmadığını göstermesidir. Üst sıralı sınırlı sayıdaki özellik ile eğitilen modellerin, tam özellik kümesi ile eğitilen modele oldukça yakın performans vermesi, uygun tasarlanmış bir özellik skorlama mekanizmasının hem

yorumlanabilirliği hem de sınıflandırma başarımını birlikte destekleyebileceğini ortaya koymaktadır.

Özellikle S-SFIX'in 15 ve 20 özellikli kurulumlarda sergilediği sonuçlar, model karmaşıklığını azaltırken bilgi kaybının sınırlı tutulabildiğini göstermektedir. Bu durum, daha düşük hesaplama maliyeti ve daha izlenebilir karar yapısı gerektiren pratik uygulamalar açısından da değerli olduğunu göstermektedir.

5.4.4 Yapay Zekâ Dayanıklılığı ve gürültü altında kararlılık

Bu çalışmanın önemli katkılarından biri, Yapay Zekâ Dayanıklılığı kavramının doğrudan S-SFIX düzeyinde sınanmış olmasıdır. Gaussian gürültüsü altında yürütülen deneyler, S-SFIX'in gürültü arttıkça tamamen rastgeleleşen bir özellik sıralaması üretmediğini; özellikle çekirdek özellik kümesini büyük ölçüde koruduğunu göstermektedir.

Düşük ve orta seviyeli gürültü koşullarında İlk 10 özellik kümesinin tamamen korunmuş olması, yöntemin en kritik ayırt edici öznitelikleri kararlı biçimde belirleyebildiğini göstermektedir. Benzer şekilde, Spearman ve Kendall Tau korelasyonlarının yüksek kalması, genel sıralama yapısının da büyük ölçüde korunduğunu ortaya koymaktadır. Buna karşılık orta ve alt sıralardaki bazı özelliklerde daha görünür yer değişimleri meydana gelmesi, gürültünün özellikle marjinal önemlilik düzeyindeki değişkenler üzerinde daha yüksek oynaklık oluşturduğunu göstermektedir. Bu durum yöntem açısından bir zayıflık değil, aksine baskın ve ikincil özelliklerin bozulma altında farklı duyarlılık göstermesinin beklenen bir sonucudur.

Bu bulgular birlikte değerlendirildiğinde, S-SFIX'in yalnızca temiz veri üzerinde çalışan kırılgan bir sıralama mekanizması olmadığı; rastgele veri bozulmaları altında da çekirdek önemlilik yapısını koruyabildiği anlaşılmaktadır. Bu yönüyle yöntem, Yapay Zekâ Dayanıklılığı bağlamında anlamlı bir dayanıklılık potansiyeli sunmaktadır.

5.4.5 Alanlar arası uygulanabilirlik

SFIX modeli başlangıçta finans alanına özgü açıklanabilirlik ve risk duyarlı özellik değerlendirme ihtiyacından doğmuştur (Cil ve ark., 2026). Bu çalışmada ise yöntem, IoT tabanlı ağ trafiği ve DDoS tespiti gibi finans dışı, daha heterojen ve yüksek boyutlu bir problem alanında değerlendirilmiştir. Elde edilen sonuçlar, S-SFIX'in yalnızca finans odaklı bir yaklaşım olmadığını; çok boyutlu, sınıf ayrıştırıcılığına duyarlı ve risk temelli

özellik skorlama mantığının farklı alanlara da uyarlanabildiğini göstermektedir.

5.4.6 Literatüre katkı

Bu bölümdeki çalışmanın literatüre katkısı dört başlık altında özetlenebilir. İlk olarak, S-SFIX'in çoklu saldırı alt türlerini içeren kapsamlı bir IoT saldırı tespit probleminde etkili bir özellik sıralama mekanizması olduğu gösterilmiştir. İkinci olarak, yöntem SHAP, LIME ve Random Forest özellik önemliliği gibi yaygın yaklaşımlar ile doğrudan karşılaştırılarak nesnel biçimde konumlandırılmıştır. Üçüncü olarak, Gaussian gürültüsü deneyleri ile S-SFIX'in sıralama kararlılığı sistematik biçimde ölçülmüş ve yöntemin Yapay Zekâ Dayanıklılığı boyutu deneysel olarak desteklenmiştir. Son olarak, çalışma S-SFIX'in yalnızca alan-özel bir açıklanabilirlik tekniği değil, daha genel ve farklı problem alanlarında da uygulanabilir bir özellik skorlama çerçevesi sunduğunu göstermektedir.

Sonuç olarak, S-SFIX bu çalışmada hem açıklanabilirlik üretmeye yönelik bir yöntem hem de düşük boyutlu fakat bilgi yoğun özellik kümeleri oluşturabilen, alternatif XAI yaklaşımlarıyla rekabet edebilen ve veri bozulmaları altında belirli düzeyde kararlılık sergileyen bir özellik seçimi çerçevesi olarak konumlanmaktadır.

6 SONUÇ

Bu tez kapsamında, finans alanında açıklanabilir yapay zekânın yalnızca sonradan eklenen bir yorumlama katmanı olarak değil, güvenilir, denetlenebilir ve düzenleyici beklentilerle uyumlu karar sistemlerinin temel bir bileşeni olarak ele alınması gerektiği ortaya konmuştur. Elde edilen bulgular, finansal hizmetlerde şeffaflık, güven, etik sorumluluk ve hesap verebilirlik gereksinimlerinin giderek daha belirleyici hâle geldiğini göstermektedir. Literatürde yaygın olarak kullanılan SHAP ve LIME gibi yöntemler kredi skorlama, dolandırıcılık tespiti ve portföy yönetimi gibi alanlarda önemli katkılar sunmakla birlikte; ölçeklenebilirlik, kararlılık, doğrusal olmayan yapıları açıklama ve alan özgüllüğü bakımından çeşitli sınırlılıklar taşımaktadır. Bu durum, finansal karar destek sistemlerinde genel amaçlı açıklanabilirlik yöntemlerinden, bağlama duyarlı ve finans alanına özgü açıklama çerçevelerine geçiş gerekliliğini açık biçimde göstermiştir.

Bu tezde önerilen SFIX çerçevesi, söz konusu ihtiyaca yanıt veren temel metodolojik katkıyı oluşturmaktadır. SFIX; Dinamik Önem Skoru, Güven Skoru, Örüntü Skoru, Anomali Skoru ve Risk Skoru olmak üzere beş temel bileşeni bir araya getirerek, finansal verilerin zamansal davranışını, veri kalitesini, örüntü yapısını, anomali duyarlılığını ve risk boyutunu bütünleşik biçimde değerlendiren bir açıklanabilirlik yapısı sunmaktadır. Böylece açıklama süreci, tahminden sonra yaklaşık olarak üretilen bir yorum olmaktan çıkarılıp, doğrudan model mantığının bir parçası hâline getirilmektedir. Bu yönüyle SFIX, SHAP ve LIME gibi post-hoc yaklaşımlardan ayrılmakta; özellikle zaman duyarlılığı, veri güvenilirliği ve finansal risk gibi unsurları doğrudan açıklama sürecine dâhil ederek finansal karar verme bağlamına daha uygun bir yapı sunmaktadır. Ayrıca geliştirilen sadeleştirilmiş sürüm S-SFIX, zaman boyutunun bulunmadığı verisetlerinde de çerçevenin uygulanabilirliğini koruyarak yöntemin pratik kullanım alanını genişletmiştir.

Tez kapsamında uygulamalarla elde edilen bulgular, SFIX çerçevesinin yalnızca kuramsal olarak değil, pratik performans açısından da güçlü bir yapı sunduğunu göstermektedir. Özellikle denetim görüşü sınıflandırması bağlamında SFIX'in ileri beslemeli sinir ağı ile bütünleştirilmesiyle oluşturulan hibrit yapı hem yüksek tahmin performansı hem de güçlü açıklanabilirlik üretmiştir. SFIX tarafından seçilen sınırlı sayıdaki özellik ile FFNN modelinin, tam özellik kümesine oldukça yakın doğruluk ve

F1-skoru üretmesi, açıklanabilirlik ile performans arasında zorunlu bir ödünleşim bulunmadığını göstermektedir. Özellik boyutunda anlamlı bir azalma sağlanmasına rağmen model başarımının korunması, önerilen yapının yorumlanabilirliğinin yanısıra, hesaplama verimliliğini de desteklediğini ortaya koymaktadır. Bunun yanında, eksik veri tamamlama sonrası performans iyileşmeleri, veri kalitesinin ve daha temsil edici veri yapılarının SFIX tabanlı seçim mekanizması ile uyumlu biçimde çalıştığını göstermiştir. Elde edilen sıralamalarda *ROCE* oranı, *özkaynak/toplam varlıklar, finansal kaldıraç* ve *Zmijewski* skoru gibi finansal açıdan anlamlı değişkenlerin üst sıralarda yer alması, yöntemin alan bilgisiyle tutarlı karar yapıları ürettiğini doğrulamaktadır.

Tez bulguları, açıklanabilirliğin karar sonrası gerekçelendirme sağlanmasıyla birlikte, aynı zamanda özellik seçimi, model sadeleştirme ve karar güvenilirliğini artırma amacıyla da kullanılabileceğini göstermiştir. SFIX–FFNN yapısı, bileşen tabanlı açıklamaları sayesinde model kararlarının hangi değişkenler ve hangi boyutlar üzerinden şekillendiğini görünür kılmış; böylece denetçiler, düzenleyiciler ve finansal karar vericiler için daha şeffaf bir değerlendirme zemini sunmuştur. Bu durum, özellikle düzenlemeye tabi finansal ortamlarda model çıktılarının gerekçelendirilebilir olmasını ve insan gözetimiyle desteklenmesini kolaylaştırmaktadır. Dolayısıyla tez kapsamında geliştirilen yapı, hem teknik bir model önerisi sunarken hem de açıklanabilir, izlenebilir ve güvenilir finansal karar sistemleri için uygulanabilir bir yaklaşım olarak değerlendirilmelidir.

Tezin bir diğer önemli sonucu, SFIX yaklaşımının yalnızca finansal karar verme bağlamında değil, yüksek hacimli ve gürültülü veri ortamlarında da güçlü bir özellik puanlama mantığı sunduğunu göstermesidir. Zaman boyutunun olmadığı ortamlarda kullanılan S-SFIX yapısının, farklı veri koşulları altında rekabetçi ve kararlı sonuçlar vermesi, yöntemin açıklanabilirlik odaklı bir puanlama mekanizması olarak sağlamlığını desteklemektedir. Özellikle bozulmuş veri koşulları altında temel özellik kümesinin korunabilmesi, önerilen çerçevenin yalnızca açıklama üretmekle kalmayıp, aynı zamanda kararlı ve güvenilir karar yapılarının oluşmasına katkı sağladığını göstermektedir. Bu sonuç, açıklanabilirliğin şeffaflığının yanında dayanıklılık ve tutarlılıkla birlikte değerlendirilmesi gerektiğini ortaya koymaktadır.

Tez genelinde ulaşılan sonuçlar birlikte değerlendirildiğinde, finansal yapay zekâ

uygulamalarında açıklanabilirliğin tek boyutlu bir özellik katkısı analizinden ibaret olmadığı anlaşılmaktadır. Finansal karar sistemleri, doğaları gereği zaman bağımlı, risk duyarlı, veri kalitesi sorunlarına açık ve anomali içerebilen yapılardır. Bu nedenle bu tezde geliştirilen SFIX çerçevesi, finansal verinin bu çok boyutlu doğasını doğrudan açıklama yapısına dâhil ederek literatürdeki önemli bir boşluğu doldurmaktadır. SHAP ve LIME gibi yaygın yöntemler güçlü açıklamalar sunsa da, bunların çoğu statik ve genel amaçlı kalmakta; oysa SFIX, finansal kararların gerek duyduğu zamansallık, risk duyarlılığı, güvenilirlik ve bağlamsal anlamlandırmayı aynı yapı içinde bütünleştirmektedir.

Bununla birlikte, tez bulguları mevcut yapının bazı sınırlılıklarını da ortaya koymaktadır. Özellikle bazı bileşenlerde parametre seçiminin veri ve bağlama bağlı olması, yöntem tasarımında esnekliği artırırken belirli bir ayar ihtiyacını da beraberinde getirmektedir. Benzer biçimde risk bileşeninde kullanılan dağılımsal varsayımlar, gerçek finansal verilerin her zaman tam olarak karşılamayabileceği ideal koşullara dayanmaktadır. Ayrıca gerçek dünya uygulamaları düşünüldüğünde; eski sistemlerle entegrasyon, büyük ölçekte açıklanabilirliğin korunması, gerçek zamanlı kullanım ihtiyaçları ve farklı paydaş grupları için açıklamaların anlaşılır biçimde sunulması gibi unsurlar dikkatle ele alınmalıdır. Ancak bu sınırlılıklar, önerilen çerçevenin değerini azaltmaktan ziyade, finansal açıklanabilirlik probleminin ne kadar çok boyutlu ve önemli olduğunu bir kez daha göstermektedir.

Sonuç olarak bu tez, finansal karar sistemlerinde açıklanabilir yapay zekânın nasıl ele alınması gerektiğine ilişkin bütüncül bir yaklaşım sunmaktadır. Literatürdeki yöntemlerin güçlü ve zayıf yönleri sistematik biçimde değerlendirilmiş, ardından finans alanına özgü yeni bir açıklanabilirlik çerçevesi geliştirilmiş, bu çerçeve hem özellik seçimi hem de hibrit derin öğrenme mimarileri içinde uygulanmış ve elde edilen sonuçlarla etkinliği gösterilmiştir. Böylece tez çalışması, açıklanabilirliği finansal yapay zekâ sistemlerinde sonradan eklenen bir unsur olmaktan çıkarıp, karar mekanizmasının yapısal bir parçası olarak yeniden konumlandırmaktadır. Elde edilen teorik ve uygulamalı sonuçlar, SFIX'in güvenilir, yorumlanabilir, hesap verebilir ve pratikte kullanılabilir finansal yapay zekâ sistemlerinin geliştirilmesi için güçlü bir temel sunduğunu göstermektedir.

KAYNAKLAR

Adams, J., & Hagras, H. (2020). A Type-2 Fuzzy Logic Approach to Explainable AI for regulatory compliance, fair customer outcomes and market stability in the Global Financial Sector. *IEEE International Conference on Fuzzy Systems*, 1-8. https://doi.org/10.1109/FUZZ48607.2020.9177542

Adhegaonkar, V. R., Thakur, A. R., & Varghese, N. (2024). Advancing Credit Card Fraud Detection Through Explainable Machine Learning Methods. *2024 2nd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT)*, 792-796. https://doi.org/10.1109/IDCIoT59759.2024.10467999

Adhikari, S., Thapa, S., Naseem, U., Lu, H. Y., Bharathy, G., & Prasad, M. (2023). Explainable hybrid word representations for sentiment analysis of financial news. *Neural Networks, 164*, 115-123. https://doi.org/https://doi.org/10.1016/j.neunet.2023.04.011

Agosto, A., Cerchiello, P., & Giudici, P. (2023). Bayesian learning models to measure the relative impact of ESG factors on credit ratings. *International Journal of Data Science and Analytics*, 1-12. https://doi.org/10.1007/s41060-023-00405-9

Alblooshi, M., Alhajeri, H., Almatrooshi, M., & Alaraj, M. (2024). Unlocking Transparency in Credit Scoring: Leveraging XGBoost with XAI for Informed Business Decision-Making. *2024 International Conference on Artificial Intelligence, Computer, Data Sciences and Applications (ACDSA)*, 1-6. https://doi.org/10.1109/ACDSA59508.2024.10467573

Aldasoro, I., Gambacorta, L., Korinek, A., Shreeti, V., & Stein, M. (2025). Intelligent financial system: How AI is transforming finance. *Journal of Financial Intermediation, 64*, Article 101183. https://doi.org/10.1016/j.jfi.2025.101183

Alzoubi, S., Amayreh, K. T., Farea, M. M., El-Ebiary, Y. A. B., Saany, S. I. A., & Bisht, N. (2023). A Review of Effectiveness and Efficiency Methodology of Decision Support System for Selecting Suppliers. *2023 International Conference on Computer Science and Emerging Technologies (CSET)*, 1-7. https://doi.org/10.1109/CSET58993.2023.10346850

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, *71*(2), 579-625.

Arifah, I. D. C., & Nihaya, I. U. (2023). Artificial Intelligence in Credit Risk Management of Peer-to-Peer Lending Financial Technology: Systematic Literature Review. *2023 6th International Conference of Computer and Informatics Engineering (IC2IE)*, 329-334. https://doi.org/10.1109/IC2IE60547.2023.10331487

Arrieta, A. B. & others. (2020). Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI. *Information Fusion*, *58*, 82-115.

Awosika, T., Shukla, R. M., & Pranggono, B. (2024). Transparency and Privacy: The Role of Explainable AI and Federated Learning in Financial Fraud Detection. *IEEE Access*, *12*, 64551-64560. https://doi.org/10.1109/ACCESS.2024.3394528

Ayad, O. M., Hegazy, A.-E. F., & Dahroug, A. (2023). A Proposed Model for Loan Approval Prediction Using Explainable Artificial Intelligence. *2023 Eleventh International Conference on Intelligent Computing and Information Systems (ICICIS)*, 166-173. https://doi.org/10.1109/ICICIS58388.2023.10391163

Azad, S. (2023). Exploring customer behavior and enhanced revenue generation in e-commerce using interpretable and explainable artificial intelligence. *8th International Conference on Computing in Engineering and Technology (ICCET 2023)*, 324-332. https://doi.org/10.1049/icp.2023.1511

Babaei, G., Giudici, P., & Raffinetti, E. (2022). Explainable Artificial Intelligence for crypto asset allocation. *Finance Research Letters*, *47*, 102941.

Babaei, G., Giudici, P., & Raffinetti, E. (2023). Explainable fintech lending. *Journal of Economics and Business*, *125-126*, 106126.

Babaei, G., Giudici, P., & Raffinetti, E. (2025). A Rank graduation box for SAFE AI. *Expert Systems with Applications*, *59*, 125239.

Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K.-R., & Samek, W. (2015). On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation. *PloS one*, *10*(7), e0130140. https://doi.org/10.1371/journal.pone.0130140

Bahoo, S., Cucculelli, M., Goga, X., & Mondolo, J. (2024). Artificial intelligence in finance: A comprehensive review through bibliometric and content analysis. *SN Business & Economics, 4*(2). https://doi.org/10.1007/s43546-023-00618-x

Bandi, H., Joshi, S., Bhagat, S., & Ambawade, D. (2021). Integrated Technical and Sentiment Analysis Tool for Market Index Movement Prediction, comprehensible using XAI. *2021 International Conference on Communication information and Computing Technology (ICCICT)*, 1-8. https://doi.org/10.1109/ICCICT50803.2021.9510124

Bastos, J. A., & Matos, S. M. (2022). Explainable models of credit losses. *European Journal of Operational Research*, *301*(1), 386-394. https://doi.org/10.1016/j.ejor.2021.11.009

Becker, M., Toprak, E., & Beyerer, J. (2023). Explainable Artificial Intelligence for Interpretable Data Minimization. *2023 IEEE International Conference on Data Mining Workshops (ICDMW)*, 885-893. https://doi.org/10.1109/ICDMW60847.2023.00119

Beckmann, L., Debener, J., & Kriebel, J. (2023). Understanding the determinants of bond excess returns using explainable AI. *Journal of Business Economics*, *93*(9), 1553-1590.

Bermúdez, L., Anaya, D., & Belles-Sampera, J. (2023). Explainable AI for paid-up risk management in life insurance products. *Finance Research Letters*, *57*, 104242. https://doi.org/10.1016/j.frl.2023.104242

Bhadula, S., Kartik, D., & Gupta, D. (2024). An Explainable AI Regression Model For Gold Price Prediction. *2024 IEEE 9th International Conference for Convergence in Technology (I2CT)*, 1-7. https://doi.org/10.1109/I2CT61223.2024.10543640

Bhattacharya, A., Choudhary, P., Mukhopadhyay, S., Misra, B., Chakraborty, S., & Dey, N. (2023). Explainable AI for Predictive Analytics on Employee Promotion. *2023 International Conference on Advanced Computing Technologies and Applications (ICACTA)*, 1-7. https://doi.org/10.1109/ICACTA58201.2023.10393141

Bhowmik, A., Sannigrahi, M., Chowdhury, D., Dwivedi, A. D., & Mukkamala, R. R. (2022). DBNex: Deep Belief Network and Explainable AI based Financial Fraud Detection. *2022 IEEE International Conference on Big Data*, 3033-3042. https://doi.org/10.1109/BigData55660.2022.10020494

Biswas, B., Mukhopadhyay, A., Kumar, A., & Delen, D. (2024). A hybrid framework using explainable AI (XAI) in cyber-risk management for defence and recovery against phishing attacks. *Decision Support Systems*, *177*, 114102. https://doi.org/10.1016/j.dss.2023.114102

Boardman, J., Alam, M. S., Huang, X., & Xie, Y. (2022). Integrated Gradients is a Nonlinear Generalization of the Industry Standard Approach to Variable Attribution for Credit Risk Models. *2022 IEEE International Conference on Big Data*, 5012-5023. https://doi.org/10.1109/BigData55660.2022.10020687

Bora, A., Sah, R., Singh, A., Sharma, D., & Ranjan, R. K. (2022). Interpretation of machine learning models using XAI - A study on health insurance dataset. *2022 10th International Conference on Reliability, Infocom Technologies and Optimization*, 1-6. https://doi.org/10.1109/ICRITO56286.2022.9964649

Boujrad, K. & others. (2025). Anomaly-based Detection in Time Series Data on IoT Systems Using Statistical Models. *IEEE Internet of Things Journal*.

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2015). *Time Series Analysis: Forecasting and Control*. John Wiley & Sons.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32. https://doi.org/10.1023/A:1010933404324

Bussmann, N., Giudici, P., Marinelli, D., & Papenbrock, J. (2021). Explainable machine learning in credit risk management. *Computational Economics*, *57*(1), 203-216.

Büyüktanır, T. (2025). *Financial Ratios Extracted from 503 Trial Balances*. Kaggle. https://doi.org/10.34740/KAGGLE/DS/7639493

Calzarossa, M. C., Giudici, P., & Zieni, D. (2025). An assessment framework for explainable AI, with applications to cybersecurity. *Artificial Intelligence Review*.

Carta, S., Consoli, S., Piras, L., Podda, A. S., & Recupero, D. R. (2021). Explainable Machine Learning Exploiting News and Domain-Specific Lexicon for Stock Market Forecasting. *IEEE Access*, *9*, 30193-30205. https://doi.org/10.1109/ACCESS.2021.3059960

Carta, S., Consoli, S., Podda, A. S., Recupero, D. R., & Stanciu, M. M. (2022a). An eXplainable Artificial Intelligence tool for statistical arbitrage. *Software Impacts*, *14*, 100354. https://doi.org/10.1016/j.simpa.2022.100354

Carta, S., Consoli, S., Podda, A. S., Recupero, D. R., & Stanciu, M. M. (2022b). Statistical arbitrage powered by Explainable Artificial Intelligence. *Expert Systems with Applications*, *206*, 117763. https://doi.org/10.1016/j.eswa.2022.117763

Černevičienė, J., & Kabašinskas, A. (2024). Explainable artificial intelligence (XAI) in finance: A systematic literature review. *Artificial Intelligence Review*, *57*(8), 216.

Chacon, D. U., Lee, S., & Park, J. (2026). An explainable machine learning model for consumer credit scoring in Mexico. *Emerging Markets Review*, *71*, 101424. https://doi.org/10.1016/j.ememar.2025.101424

Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly Detection: A Survey. *ACM Computing Surveys (CSUR)*, *41*(3), 1-58.

Charran, S. R., & Mahapatra, S. (2023). *Cross Feature Selection to Eliminate Spurious Interactions and Single Feature Dominance Explainable Boosting Machines*.

Chatfield, C. (2003). *The Analysis of Time Series: An Introduction*. Chapman and Hall/CRC.

Chen, Y., Giudici, P., Liu, K., & Raffinetti, E. (2024). Measuring fairness in credit ratings. *Expert Systems with Applications*, *258*, 125184.

Chou, Y.-H., Chang, M.-H., Jiang, Y.-C., Kuo, S.-Y., Kung, S.-Y., & Sheu, B. J. (2024). An Investigation on Quantum-Inspired Algorithms for Portfolio Optimization Across Global Markets. *IEEE Nanotechnology Magazine*. https://doi.org/10.1109/MNANO.2024.3378489

Cil, A. E., & Yildiz, K. (2025a). A systematic literature review on applications of explainable artificial intelligence in the financial sector. *Internet of Things*, *33*, 101696. https://doi.org/https://doi.org/10.1016/j.iot.2025.101696

Cil, A. E., & Yildiz, K. (2025b). SFIX:Scalable Financial-oriented Interpretable eXplanation. *Internet of Things*, *33*, 101713. https://doi.org/https://doi.org/10.1016/j.iot.2025.101713

Cil, A. E., Buyuktanir, T., & Yildiz, K. (2026). Comprehensive Explainable AI Approach for Audit Opinion Classification Using Feed-Forward Neural Networks. *Knowledge-Based Systems*.

Cil, A. E., Yildiz, K., & Buldu, A. (2021). Detection of DDoS attacks with feed forward based deep neural network model. *Expert Systems with Applications*, *169*, 114520. https://doi.org/https://doi.org/10.1016/j.eswa.2020.114520

Craven, M., & Shavlik, J. (1995). Extracting Tree-Structured Representations of Trained Networks. İçinde D. Touretzky, M. C. Mozer, & M. Hasselmo (Ed.), *Advances in Neural Information Processing Systems* (C. 8). MIT Press. https://proceedings.neurips.cc/paper_files/paper/1995/file/45f31d16b1058d586fc3be7207b58053-Paper.pdf

Çelik, T. B., İcan, Ö., & Bulut, E. (2023). Extending machine learning prediction capabilities by explainable AI in financial time series prediction. *Applied Soft Computing*, *132*, 109876. https://doi.org/10.1016/j.asoc.2022.109876

Das, S., Chowdhury, U., Lijin, N. S., Deep, A., Saha, S., & Maurya, A. (2024). Investigate How Market Behaves: Toward an Explanatory Multitasking Based Analytical Model for Financial Investments. *IEEE Access*, *12*, 30928-30940. https://doi.org/10.1109/ACCESS.2024.3369033

Dastile, X., & Celik, T. (2021). Making Deep Learning-Based Predictions for Credit Scoring Explainable. *IEEE Access*, *9*, 50426-50440. https://doi.org/10.1109/ACCESS.2021.3068854

Dastile, X., Celik, T., & Vandierendonck, H. (2022). Model-Agnostic Counterfactual Explanations in Credit Scoring. *IEEE Access*, *10*, 69543-69554. https://doi.org/10.1109/ACCESS.2022.3177783

de Lange, P. D., Melsom, B. G. A., Vennerød, C. B., & Westgaard, S. (2022). Explainable AI for Credit Assessment in Banks. *Journal of Risk and Financial Management*, *15*(12), 556. https://doi.org/10.3390/jrfm15120556

Dennis, K., Caroline, B., Nansamba, B., Jjingo, D., & Marvin, G. (2024). Explainable Deep Ensemble Learning for Improved Credit Default Prediction. *Proceedings of the*

2024 Sixteenth International Conference on Contemporary Computing, IC3-2024, 256-270. https://doi.org/10.1145/3675888.3676059

Dessain, J., Bentaleb, N., & Vinas, F. (2023). Cost of Explainability in AI: An Example with Credit Scoring Models. İçinde L. Longo (Ed.), *Explainable Artificial Intelligence* (ss. 498-516). Springer Nature Switzerland.

Doshi-Velez, F., & Kim, B. (2017). Towards A Rigorous Science of Interpretable Machine Learning. *arXiv: Machine Learning*. https://api.semanticscholar.org/CorpusID:11319376

Dou, X., & Peng, C. (2023). Improving Graph Convolutional Neural Networks for q-Consistent Summary-Explanations with Application to Credit Evaluation Systems. *2023 4th International Symposium on Computer Engineering and Intelligent Communications (ISCEIC)*, 594-598. https://doi.org/10.1109/ISCEIC59030.2023.10271140

Duvnjak, M., Merćep, A., & Kostanjčar, Z. (2024). Intrinsically Interpretable Models for Credit Risk Assessment. *2024 47th MIPRO ICT and Electronics Convention (MIPRO)*, 31-36. https://doi.org/10.1109/MIPRO60963.2024.10569726

Embarak, O. (2023). Decoding the Black Box: A Comprehensive Review of Explainable Artificial Intelligence. *2023 9th International Conference on Information Technology Trends (ITT)*, 108-113. https://doi.org/10.1109/ITT59889.2023.10184238

European Parliament. (2023). *EU AI Act: First regulation on artificial intelligence*. https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence

Fenton, N., & Neil, M. (2018). *Risk Assessment and Decision Analysis with Bayesian Networks*. CRC Press.

Fisher, A., Rudin, C., & Dominici, F. (2019). All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. *Journal of Machine Learning Research*, *20*(177), 1-81.

Fortune Business Insights. (2023). *Artificial Intelligence [AI] market size, share & industry analysis*. https://www.fortunebusinessinsights.com/industry-reports/artificial-intelligence-market-100114

Galego Hernandes, P. R., Floret, C. P., De Almeida, K. F. C., Da Silva, V. C., Papa, J. P., & Da Costa, K. A. P. (2021). Phishing Detection Using URL-based XAI Techniques. *2021 IEEE Symposium Series on Computational Intelligence (SSCI)*, 1-6. https://doi.org/10.1109/SSCI50451.2021.9659981

Gandhudi, M., Alphonse, P. J. A., Fiore, U., & Gangadharan, G. R. (2024). Enhancing Model Explainability in Financial Trading Using Training Aid Samples: A CNN-Based Candlestick Pattern Recognition Approach. *Computers and Electrical Engineering*, *118*, 109302. https://doi.org/10.1016/j.compeleceng.2024.109302

Gaurav, D., & Tiwari, S. (2023). Interpretability Vs Explainability: The Black Box of Machine Learning. *2023 International Conference on Computer Science, Information Technology and Engineering (ICCoSITE)*, 523-528. https://doi.org/10.1109/ICCoSITE57641.2023.10127717

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., III, H. D., & Crawford, K. (2021). Datasheets for Datasets. *arXiv*. https://arxiv.org/abs/1803.09010

Ghosh, I., & Jana, R. K. (2024). Clean energy stock price forecasting and response to macroeconomic variables: A novel framework using Facebook's Prophet, NeuralProphet and explainable AI. *Technological Forecasting and Social Change*, *200*, 123148. https://doi.org/10.1016/j.techfore.2023.123148

Ghosh, I., Alfaro-Cortés, E., Gámez, M., & García-Rubio, N. (2023). Role of proliferation COVID-19 media chatter in predicting Indian stock market: Integrated framework of nonlinear feature transformation and advanced AI. *Expert Systems with Applications*, *219*, 119695. https://doi.org/10.1016/j.eswa.2023.119695

Ghosh, I., Chaudhuri, T. D., Alfaro-Cortés, E., Gámez, M., & García, N. (2022). A hybrid approach to forecasting futures prices with simultaneous consideration of optimality in ensemble feature selection and advanced artificial intelligence. *Technological Forecasting and Social Change*, *181*, 121757. https://doi.org/10.1016/j.techfore.2022.121757

Giudici, P., & Gramegna, A. (2021). SHAP and LIME: an evaluation of discriminative power in credit risk. *Frontiers in Artificial Intelligence*, *4*, 752558.

Giudici, P., & Raffinetti, E. (2021). Shapley-Lorenz eXplainable artificial intelligence. *Expert Systems with Applications*, *167*, 114104.

Giudici, P., & Raffinetti, E. (2022). Explainable AI methods in cyber risk management. *Quality and Reliability Engineering International*, *38*(3), 1318-1326.

Giudici, P., & Raffinetti, E. (2023). SAFE Artificial Intelligence in Finance. *Finance Research Letters*, *56*, 104088.

Giudici, P., Piergallini, S., Raffinetti, E., & Recchioni, M. C. (2024). Explainable AI methods for financial time series. *Physica A: Statistical Mechanics and its Applications*, *655*, 130176.

Gkolemis, V., Dalamagas, T., & Diou, C. (2022). DALE: Differential Accumulated Local Effects for Efficient and Accurate Global Explanations. *arXiv preprint arXiv:2210.04542*.

González-González, J., García-Méndez, S., De Arriba-Pérez, F., González-Castaño, F. J., & Barba-Seara, Ó. (2022). Explainable Automatic Industrial Carbon Footprint Estimation From Bank Transaction Classification Using Natural Language Processing. *IEEE Access*, *10*, 126326-126338. https://doi.org/10.1109/ACCESS.2022.3226324

Goodman, B., & Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a "right to explanation". *AI Magazine*, *38*(3), 50-57.

Govindu, S., Manohar, N., Kumar, J. M. S. S., & Reddy, T. S. (2024). Loan Prediction System with Exploring Explainable AI Transfer Learning with SHAP. *2024 5th International Conference for Emerging Technology (INCET)*, 1-7. https://doi.org/10.1109/INCET61516.2024.10593029

Graham, B., & Bonner, K. (2024). The role of institutions in early-stage entrepreneurship: An explainable artificial intelligence approach. *Journal of Business Research*, *175*, 114567. https://doi.org/https://doi.org/10.1016/j.jbusres.2024.114567

Guidotti, R., & Nanni, M. (2020). Crash Prediction and Risk Assessment with Individual Mobility Networks. *2020 21st IEEE International Conference on Mobile Data Management (MDM)*, 89-98. https://doi.org/10.1109/MDM48529.2020.00030

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2019). A Survey of Methods for Explaining Black Box Models. İçinde *ACM Computing Surveys*. Open Access. https://doi.org/10.1145/3236009

Guo, Y., Li, Y., Liu, D., & Xu, S. X. (2024). Measuring service quality based on customer emotion: An explainable AI approach. *Decision Support Systems*, *176*, 114051. https://doi.org/10.1016/j.dss.2023.114051

Guyon, I., & Elisseeff, A. (2003). An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research*, *3*(Mar), 1157-1182.

Hacker, P. (2023). The European AI liability directives – Critique of a half-hearted approach and lessons for the future. *Computer Law Security Review*, *51*, 105871. https://doi.org/https://doi.org/10.1016/j.clsr.2023.105871

Hanif, A., Zhang, X., & Wood, S. (2021). A Survey on Explainable Artificial Intelligence Techniques and Challenges. *2021 IEEE 25th International Enterprise Distributed Object Computing Workshop (EDOCW)*, 81-89. https://doi.org/10.1109/EDOCW52865.2021.00036

Hasan, M., Rahman, M. S., Janicke, H., & Sarker, I. H. (2024). Detecting Anomalies in Blockchain Transactions using Machine Learning Classifiers and Explainability Analysis. *Blockchain: Research and Applications*, 100207. https://doi.org/10.1016/j.bcra.2024.100207

Hastie, T., & Tibshirani, R. (1986). Generalized Additive Models. *Statistical Science*, *1*(3), 297-310. https://doi.org/10.1214/ss/1177013604

Heger, J., Min, A., & Zagst, R. (2024). Knowledge-Based Credit Granting System in Indonesia with Artificial Intelligence and XAI Approach. *International Review of Financial Analysis*, *94*, 103315. https://doi.org/10.1016/j.irfa.2024.103315

Hossain, M., Sebestyen, M., Mayank, D., Ardakanian, O., & Khazaei, H. (2020). Large-scale Data-driven Segmentation of Banking Customers. *2020 IEEE International Conference on Big Data (Big Data)*, 4739-4748. https://doi.org/10.1109/BIGDATA50022.2020.9378483

Jovanovic, Z., Hou, Z., Biswas, K., & Muthukkumarasamy, V. (2024). Robust integration of blockchain and explainable federated learning for automated credit scoring. *Computer Networks*, *243*, 110303. https://doi.org/10.1016/j.comnet.2024.110303

Jung, A., & Nardelli, P. H. J. (2020). An Information-Theoretic Approach to Personalized Explainable Machine Learning. *IEEE Signal Processing Letters*, *27*, 825-829. https://doi.org/10.1109/LSP.2020.2993176

Kalra, A., & Mittal, R. (2024a). Explainable AI for Improved Financial Decision Support in Trading. *2024 11th International Conference on Reliability, Infocom Technologies and Optimization*, 1-6. https://doi.org/10.1109/ICRITO61523.2024.10522366

Kalra, A., & Mittal, R. (2024b). Intrinsically Interpretable Models for Credit Risk Assessment. *2024 47th MIPRO ICT and Electronics Convention*, 31-36. https://doi.org/10.1109/MIPRO60963.2024.10569726

Kendall, M. G. (1948). *Rank Correlation Methods*. Oxford University Press.

Kharlamova, A., Kruglov, A., & Succi, G. (2024). State-of-the-Art Review of Life Insurtech: Machine learning for underwriting decisions and a Shift Toward Data-Driven, Society-oriented Environment. *2024 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, 1-12. https://doi.org/10.1109/HORA61326.2024.10550565

Kolte, G., Kini, V., Nair, H. G., & Babu, S. K. S. (2022). Stock Market Prediction using Deep Learning. *International Journal For Science Technology And Engineering*, *8*(4), 1-6. https://doi.org/10.22214/ijraset.2022.41159

Kong, K., Liu, R., Zhang, Y., & Chen, Y. (2020). Predicting Liquidity Ratio of Mutual Funds via Ensemble Learning. *2020 IEEE International Conference on Big Data*, 5441-5450. https://doi.org/10.1109/BigData50022.2020.9378486

Kotrachai, C., Chanruangrat, P., Thaipisutikul, T., Kusakunniran, W., Hsu, W.-C., & Sun, Y.-C. (2023). Explainable AI supported Evaluation and Comparison on Credit Card Fraud Detection Models. *2023 7th International Conference on Information Technology (InCIT)*, 86-91. https://doi.org/10.1109/InCIT60207.2023.10413100

Kumar, K., Kuldeep, & Bhushan, B. (2023). Augmenting Cybersecurity and Fraud Detection Using Artificial Intelligence Advancements. *2023 International Conference on*

Computing, Communication, and Intelligent Systems (ICCCIS), 1207-1212. https://doi.org/10.1109/ICCCIS60361.2023.10425069

Kumar, P., Hota, L., Tikkiwal, V. A., & Kumar, A. (2024). Analysing Forecasting of Stock Prices: An Explainable AI Approach. *Procedia Computer Science, 235*, 2009-2016. https://doi.org/10.1016/j.procs.2024.04.190

Kumar, R., Rakhra, M., Prashar, D., Upadhyay, S., Mršić, L., & Khan, A. H. (2024). A Machine Learning and Ratemaking Evaluation of Four Auto Insurance Pure Premium Modeling Algorithms. *IEEE International Conference on Secure and Ethical Computing (ICSEC)*. https://doi.org/10.1109/icsec62781.2024.10770731

Kute, D. V., Pradhan, B., Shukla, N., & Alamri, A. (2021). Deep Learning and Explainable Artificial Intelligence Techniques Applied for Detecting Money Laundering–A Critical Review. *IEEE Access, 9*, 82300-82317. https://doi.org/10.1109/ACCESS.2021.3086230

Leghemo, F. & others. (2025). Data Governance for Emerging Technologies: A Conceptual Framework for Managing Blockchain, IoT, and AI. *Information Systems*.

Li, L.-H., Sharma, A. K., & Cheng, S.-T. (2025). Explainable AI based LightGBM prediction model to predict default borrower in social lending platform. *Intelligent Systems with Applications, 26*, 200514. https://doi.org/10.1016/j.iswa.2025.200514

Lundberg, S. M., & Lee, S.-I. (2017a). A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems, 30*.

Lundberg, S. M., & Lee, S.-I. (2017b). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems, 30*, 4765-4774.

LundbergandSu-InLee, S. (2017). A Unified Approach to Interpreting Model Predictions. İçinde *arXiv Preprint*. arXiv. https://arxiv.org/abs/1705.07874

Ma, Y., Zhang, P., Duan, S., & Zhang, T. (2023). Credit default prediction of Chinese real estate listed companies based on explainable machine learning. *Finance Research Letters, 58*, 104305. https://doi.org/10.1016/j.frl.2023.104305

Mahajan, A., & Shukla, K. K. (2023). Analyzing False Positives in Bankruptcy Prediction with Explainable AI. *2023 International Conference on Artificial Intelligence and Applications (ICAIA)*, 1-5. https://doi.org/10.1109/ICAIA57370.2023.10169390

Mahyoub, M., Ghareeb, S., & Mustafina, J. (2023). A Novel Predictive Model for Housing Loan Default using Feature Generation and Explainable AI. *2023 15th International Conference on Developments in eSystems Engineering (DeSE)*, 492-497. https://doi.org/10.1109/DeSE58274.2023.10099796

Man, X., & Chan, E. G. (2021). The Best Way to Select Features? Comparing MDA, LIME, and SHAP. *Journal of Financial Data Science*, *3*(1). https://doi.org/10.3905/JFDS.2020.1.047

Mandeep, Agarwal, A., Bhatia, A., Malhi, A., Kaler, P., & Pannu, H. S. (2022). Machine Learning Based Explainable Financial Forecasting. *2022 4th International Conference on Computer Communication and the Internet (ICCCI)*, 34-38. https://doi.org/10.1109/ICCCI55554.2022.9850272

Manjula Devi, C., Gobinath, A., Priya, S., Adithiyaa, M., Chandru, M., & Jothi, M. (2024). Next-Generation Anomaly Detection Framework Leveraging Artificial Intelligence for Proactive Credit Card Fraud Prevention and Risk Management. *IEEE International Conference on Computing, Communication and Networking Technologies (ICCCNT)*. https://doi.org/10.1109/icccnt61001.2024.10725285

Manzoor, A., Qureshi, M. A., Kidney, E., & Longo, L. (2024). A Review on Machine Learning Methods for Customer Churn Prediction and Recommendations for Business Practitioners. *IEEE Access*, *12*, 70434-70463. https://doi.org/10.1109/ACCESS.2024.3402092

Maree, C., & Omlin, C. W. (2022). Understanding Spending Behavior: Recurrent Neural Network Explanation and Interpretation. *2022 IEEE Symposium on Computational Intelligence for Financial Engineering and Economics (CIFEr)*, 1-7. https://doi.org/10.1109/CIFEr52523.2022.9776210

Maree, C., Modal, J. E., & Omlin, C. W. (2020). Towards Responsible AI for Financial Transactions. *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*, 16-21. https://doi.org/10.1109/SSCI47803.2020.9308456

Mariotti, E., Alonso, J. M., & Confalonieri, R. (2021). A Framework for Analyzing Fairness, Accountability, Transparency and Ethics: A Use-case in Banking Services. *2021 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 1-6. https://doi.org/10.1109/FUZZ45933.2021.9494481

Martins, T., de Almeida, A. M., Cardoso, E., & Nunes, L. (2024). Explainable Artificial Intelligence (XAI): A Systematic Literature Review on Taxonomies and Applications in Finance. *IEEE Access*, *12*, 618-629. https://doi.org/10.1109/ACCESS.2023.3347028

Maurya, S., Verma, R., Khilnani, L., Bhakuni, A. S., Kumar, M., & Rakesh, N. (2024). Effect of AI on the Financial Sector: Risk Control, Investment Decision-making, and Business Outcome. *2024 11th International Conference on Reliability, Infocom Technologies and Optimization*, 1-7. https://doi.org/10.1109/ICRITO61523.2024.10522410

Mavrogiorgou, A. & others. (2023). FAME: Federated Decentralized Trusted Data Marketplace for Embedded Finance. *2023 International Conference on Smart Applications, Communications and Networking*, 1-6. https://doi.org/10.1109/SmartNets58706.2023.10215814

McKinsey & Company. (2022, September 29). *Why businesses need explainable AI—and how to deliver it*. McKinsey & Company. https://www.mckinsey.com/capabilities/quantumblack/our-insights/why-businesses-need-explainable-ai-and-how-to-deliver-it

Medianovskyi, K., Malakauskas, A., Lakstutiene, A., & Ben Yahia, S. (2023). Interpretable machine learning for heterogeneous treatment effect estimators with Double ML: a case of access to credit for SMEs. *Procedia Computer Science*, *225*, 2163-2172. https://doi.org/10.1016/j.procs.2023.10.207

Metwally, A. B. M., Ali, S. A. M., & Mohamed, A. T. I. (2024). Thinking Responsibly About Responsible AI in Risk Management: The Darkside of AI in RM. *2024 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems*, 1-5. https://doi.org/10.1109/ICETSIS61505.2024.10459684

Moscato, V., Picariello, A., & Sperlí, G. (2021). A benchmark of machine learning approaches for credit score prediction. *Expert Systems with Applications*, *165*, 113986. https://doi.org/10.1016/j.eswa.2020.113986

Moschini, G., Houssou, R., Bovay, J., & Robert-Nicoud, S. (2020). Anomaly and Fraud Detection in Credit Card Transactions Using the ARIMA Model. *Engineering Proceedings*, *5*(1), 56. https://doi.org/10.3390/engproc2021005056

Murindanyi, S., Mugalu, B. W., Nakatumba-Nabende, J., & Marvin, G. (2023). Interpretable Machine Learning for Predicting Customer Churn in Retail Banking. *2023 7th International Conference on Trends in Electronics and Informatics*, 967-974. https://doi.org/10.1109/ICOEI56765.2023.10125859

Murindanyi, S., Nagwovuma, M., Nansamba, B., & Marvin, G. (2023). Explainable Ensemble Learning and Trustworthy Open AI for Customer Engagement Prediction in Retail Banking. *Proceedings of the 2023 Fifteenth International Conference on Contemporary Computing*, *IC3-2023*, 198-206. https://doi.org/10.1145/3607947.3607983

Nakatumba-Nabende, J., & Marvin, G. (2023). Interpretable Machine Learning for Predicting Customer Churn in Retail Banking. *2023 International Conference on Electronics, Optoelectronics and Information Engineering (ICOEI)*. https://doi.org/10.1109/icoei56765.2023.10125859

Nallakaruppan, M. K., Balusamy, B., Shri, M. L., Malathi, V., & Bhattacharyya, S. (2024). An Explainable AI framework for credit evaluation and analysis. *Applied Soft Computing*, *153*, 111307. https://doi.org/10.1016/j.asoc.2024.111307

Nalule, D., Edekebon, E., Nagwovuma, M., Jjingo, D., & Marvin, G. (2024). Explainable Deep Learning Approaches to Credit Risk Evaluation. *Proceedings of the 2024 Sixteenth International Conference on Contemporary Computing, IC3-2024*, 198-206. https://doi.org/10.1145/3675888.3676052

Naretto, F., Monreale, A., & Giannotti, F. (2022). Evaluating the Privacy Exposure of Interpretable Global Explainers. *2022 IEEE 4th International Conference on Cognitive Machine Intelligence (CogMI)*, 13-19. https://doi.org/10.1109/CogMI56440.2022.00012

Neto, E. C. P., Dadkhah, S., Ferreira, R., Zohourian, A., Lu, R., & Ghorbani, A. A. (2023). CICIoT2023: A Real-Time Dataset and Benchmark for Large-Scale Attacks in IoT Environment. *Sensors*, *23*(13). https://doi.org/10.3390/s23135941

Niu, J. (2022). Research on loan prediction based on interpretable machine learning. *2022 4th International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI)*, 356-360. https://doi.org/10.1109/MLBDBI58171.2022.00076

Nori, H., Jenkins, S., Koch, P., & Caruana, R. (2019). InterpretML: A Unified Framework for Machine Learning Interpretability. *ArXiv*, *abs/1909.09223*. https://api.semanticscholar.org/CorpusID:202712518

NVIDIA. (2024). *State of AI in financial services: 2024 trends*. https://blogs.nvidia.com/blog/ai-in-financial-services-survey-2024/

Nwafor, C. N., & Nwafor, O. Z. (2023). Determinants of non-performing loans: An explainable ensemble and deep neural network approach. *Finance Research Letters*, *56*, 104084. https://doi.org/10.1016/j.frl.2023.104084

Oppenheim, A. V., & Schafer, R. W. (2009). *Discrete-Time Signal Processing*. Pearson.

Orji, U., & Ukwandu, E. (2024). Machine learning for an explainable cost prediction of medical insurance. *Machine Learning with Applications*, *15*, 100516. https://doi.org/10.1016/j.mlwa.2023.100516

Ortigossa, E. S., Gonçalves, T., & Nonato, L. G. (2024). EXplainable Artificial Intelligence (XAI)—From Theory to Methods and Applications. *IEEE Access*, *12*, 80799-80846. https://doi.org/10.1109/ACCESS.2024.3409843

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*. https://doi.org/10.1136/bmj.n71

Park, M. S., Son, H., Hyun, C., & Hwang, H. J. (2021). Explainability of Machine Learning Models for Bankruptcy Prediction. *IEEE Access*, *9*, 124887-124899. https://doi.org/10.1109/ACCESS.2021.3110270

Park, S., & Yang, J.-S. (2022). Interpretable deep learning LSTM model for intelligent economic decision-making. *Knowledge-Based Systems*, *248*, 108907. https://doi.org/10.1016/j.knosys.2022.108907

Potluri, C. S., Rao, G. S., Kumar, L. V. R. M., Allo, K. G., Awoke, Y., & Seman, A. A. (2024). Machine Learning-Based Customer Segmentation and Personalised Marketing in Financial Services. *2024 International Conference on Communication, Computer Sciences and Engineering*, 1570-1574. https://doi.org/10.1109/IC3SE62002.2024.10593143

Pourakbar, A., & Shah-Mansouri, H. (2025). Enhanced AI as a Service at the Edge via Transformer Network. *IEEE Internet of Things Journal*.

Prakash, Babu, S. M., Kumar, P. P., Devi, S., Reddy, K. P., & Satish, M. (2023). Predicting Consumer Behaviour with Artificial Intelligence. *2023 IEEE 5th International Conference on Cybernetics, Cognition and Machine Learning Applications (ICCCMLA)*, 698-703. https://doi.org/10.1109/ICCCMLA58983.2023.10346660

Prakash, Rautela, S., Sharma, S., Singh, R., Kumar, M., & Rakesh, N. (2023). Analyzing Consumer behaviour and enhanced revenue generation in e-commerce using interpretable and explainable artificial intelligence. *2023 International Conference on Advanced Computing Communication Technologies*, 66-71. https://doi.org/10.1109/ICACCTech61146.2023.00020

Pramanik, P., Jana, R. K., & Ghosh, I. (2024). AI readiness enablers in developed and developing economies: Findings from the XGBoost regression and explainable AI framework. *Technological Forecasting and Social Change*, *205*, 123482. https://doi.org/https://doi.org/10.1016/j.techfore.2024.123482

Pundir, S., Murugan, V. G., Raman, P., Rameshkumaar, V. P., R, J., & Sudharsan, P. (2024). Automatic Stock Price Prediction and Classification Based on Hybrid with AI Feature Selection Method. *2024 5th International Conference on Recent Trends in Computer Science and Technology (ICRTCST)*, 149-154. https://doi.org/10.1109/ICRTCST61793.2024.10578425

Raman, R., Patil, H. S., Kancherla, D. B., Joshi, A., Bothe, S., & Ponnusamy, R. (2023). Economic Imbalance: Predicting Financial Stress Using Deep Learning Techniques. *2023*

10th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON), 409-413. https://doi.org/10.1109/UPCON59197.2023.10434469

Rezvani, B. & others. (2025). Self-trainable and Adaptive Sensor Intelligence for Edge Learning. *IEEE Transactions on Industrial Informatics*.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You?" Explaining the Predictions of Any Classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135-1144.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2018). Anchors: High-Precision Model-Agnostic Explanations. *Proceedings of the AAAI Conference on Artificial Intelligence*, *32*(1).

Ribeiro, M., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. İçinde *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*. Association for Computational Linguistics. https://doi.org/10.18653/v1/n16-3020

Rizinski, M., Peshov, H., Mishev, K., Jovanovik, M., & Trajanov, D. (2024). Sentiment Analysis in Finance: From Transformers Back to eXplainable Lexicons (XLex). *IEEE Access*, *12*, 7170-7198. https://doi.org/10.1109/ACCESS.2024.3349970

Sachan, S., Yang, J.-B., Xu, D.-L., Benavides, D. E., & Li, Y. (2020). An explainable AI decision-support-system to automate loan underwriting. *Expert Systems with Applications*, *144*, 113100. https://doi.org/10.1016/j.eswa.2019.113100

Selbst, A. D., & Barocas, S. (2018). The intuitive appeal of explainable machines. *Fordham L. Rev.*, *87*, 1085.

Severino, M. K., & Peng, Y. (2021). Machine learning algorithms for fraud prediction in property insurance: Empirical evidence using real-world microdata. *Machine Learning with Applications*, *5*, 100074. https://doi.org/10.1016/j.mlwa.2021.100074

Sharma, A. (2023). *Loan Approval Prediction Dataset*. https://www.kaggle.com/datasets/architsharma01/loan-approval-prediction-dataset

Shevskaya, N. V. (2021). Explainable Artificial Intelligence Approaches: Challenges and Perspectives. *2021 International Conference on Quality Management, Transport and Information Security, Information Technologies*, 540-543. https://doi.org/10.1109/ITQMIS53292.2021.9642869

Shokri, R., Stronati, M., Song, C., & Shmatikov, V. (2017). Membership Inference Attacks Against Machine Learning Models. *2017 IEEE Symposium on Security and Privacy (SP)*, 3-18. https://doi.org/10.1109/SP.2017.41

Shumway, R. H., & Stoffer, D. S. (2017). *Time series analysis and its applications: With R examples* (4. bs.). Springer.

Sikha, S., & Vijayakumar, A. (2023). Explainable AI Using h2o AutoML and Robustness Check in Credit Risk Management. *2023 Intelligent Computing and Control for Engineering and Business Systems (ICCEBS)*, 1-5. https://doi.org/10.1109/ICCEBS58601.2023.10449100

Sovrano, F., & Vitali, F. (2022). How to Quantify the Degree of Explainability: Experiments and Practical Implications. *2022 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 1-9. https://doi.org/10.1109/FUZZ-IEEE55066.2022.9882574

Stevens, A., Deruyck, P., van Veldhoven, Z., & Vanthienen, J. (2020). Explainability and Fairness in Machine Learning: Improve Fair End-to-end Lending for Kiva. *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*, 1241-1248. https://doi.org/10.1109/SSCI47803.2020.9308371

Sundararajan, M., Taly, A., & Yan, Q. (2017). Axiomatic Attribution for Deep Networks. *Proceedings of the 34th International Conference on Machine Learning*, 3319-3328.

Supardi, R. S., & Gambetta, W. (2023). Knowledge-Based Credit Granting System in Indonesia with Artificial Intelligence and XAI Approach. *2023 10th International Conference on Advanced Informatics: Concept, Theory and Application (ICAICTA)*, 1-6. https://doi.org/10.1109/ICAICTA59291.2023.10390466

Swetha, P. (2024). Credit Card Fraud Detection with LIME and SHAP. *International Journal For Science Technology And Engineering*, *10*(6). https://doi.org/10.22214/ijraset.2024.63375

Tamanna, Kamboj, S., Singh, L., & Kaur, T. (2024). Automated Fraud Detection in Financial Transactions using Machine Learning: An Ensemble Perspective. *2024 2nd International Conference on Artificial Intelligence and Machine Learning Applications*, 1-6. https://doi.org/10.1109/AIMLA59606.2024.10531422

Tang, W., Bu, H., Zuo, Y., & Wu, J. (2024). Unlocking the power of the topic content in news headlines: BERTopic for predicting Chinese corporate bond defaults. *Finance Research Letters, 62*, 105062. https://doi.org/10.1016/j.frl.2024.105062

Tayal, P., Ali, F., Pandey, S., Singh, R., Lande, J., & Pachouri, V. (2023). AI Based Stock Market Prediction- Theoretical Perspective. *2023 3rd International Conference on Innovative Sustainable Computational Technologies (CISCT)*, 1-5. https://doi.org/10.1109/CISCT57197.2023.10351375

Temraz, M., & Keane, M. T. (2022). Solving the class imbalance problem using a counterfactual method for data augmentation. *Machine Learning with Applications*, *9*, 100375. https://doi.org/https://doi.org/10.1016/j.mlwa.2022.100375

Teplova, T., Sokolova, T., & Kissa, D. (2023). Revealing stock liquidity determinants by means of explainable AI: The role of ESG before and during the COVID-19 pandemic. *Resources Policy*, *86*, 104253. https://doi.org/10.1016/j.resourpol.2023.104253

Thanathamathee, P., Sawangarreerak, S., & Nizam, D. N. M. (2024). Enhancing Going Concern Prediction With Anchor Explainable AI and Attention-Weighted XGBoost. *IEEE Access*, *12*, 68345-68363. https://doi.org/10.1109/ACCESS.2024.3401007

Thanathamathee, P., Sawangarreerak, S., Chantamunee, S., & Mohd Nizam, D. N. (2024). SHAP-Instance Weighted and Anchor Explainable AI: Enhancing XGBoost for Financial Fraud Detection. *Emerging Science Journal*, *8*(6), 874-889. https://doi.org/10.28991/ESJ-2024-08-06-016

The Economist Intelligence Unit. (2022). *Banking on a game-changer: AI in financial services*.

Thommandru, A., Mone, V., Mitharwal, S., & Tilwani, R. (2023). Exploring the Intersection of Machine Learning, Money Laundering, Data Privacy, and Law. *2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA)*, 149-155. https://doi.org/10.1109/ICIDCA56705.2023.10099859

Tiwari, R., Rautela, S., Sharma, S., Choudhary, B. P., Tripathi, R., & Singh, P. (2023). Role of AI for Fraud Detection in Banks: A Bibliometric Analysis. *2023 International Conference on Advanced Computing Communication Technologies (ICACCTech)*, 66-71. https://doi.org/10.1109/ICACCTech61146.2023.00020

Tsai, Y.-C., & Chen, J.-H. (2023). Enhancing Model Explainability in Financial Trading Using Training Aid Samples: A CNN-Based Candlestick Pattern Recognition Approach. *2023 IEEE 24th International Conference on Information Reuse and Integration for Data Science*, 76-82. https://doi.org/10.1109/IRI58017.2023.00021

Uddin, A., Pabel, M. A. H., Alam, M. I., Kamruzzaman, F., Haque, M., Hosen, M. M., Sajal, A., Miah, M. R., & Ghosh, S. K. (2025). Advancing Financial Risk Prediction and Portfolio Optimization Using Machine Learning Techniques. *The American Journal of Management and Economics Innovations*, *7*(1), 18-30. https://doi.org/10.37547/tajmei/volume07issue01-02

Union, E. (2020). *General Data Protection Regulation (GDPR)*. https://gdpr-info.eu

Valensky, D., & Mohaghegh, M. (2023). Evaluating Transparency: A Cross-Model Exploration of Explainable AI in Financial Forecasting. *2023 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*, 1-6. https://doi.org/10.1109/CSDE59766.2023.10487732

VanNostrand, P. M., Hofmann, D. M., Ma, L., Genin, B., Huang, R., & Rundensteiner, E. A. (2024). Counterfactual Explanation Analytics: Empowering Lay Users to Take Action Against Consequential Automated Decisions. *Proceedings of The VLDB Endowment*, *17*(8). https://doi.org/10.14778/3685800.3685872

Varun, S., Theagarajan, A., & Shobana, M. (2023). Credit Score Analysis Using Machine Learning. *2023 International Conference on Networking and Communications (ICNWC)*, 1-5. https://doi.org/10.1109/ICNWC57852.2023.10127493

Vashi, C., & Shanbhag, A. (2023). Comparative explainable Machine learning to evaluate Bank marketing success. *2023 3rd International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*, 1-5. https://doi.org/10.1109/ICECCME57830.2023.10252843

Vihurskyi, B. (2024). Credit Card Fraud Detection with XAI: Improving Interpretability and Trust. *2024 Third International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE)*, 1-6. https://doi.org/10.1109/ICDCECE60827.2024.10548159

W.Apley, D., & Zhu, J. (2020). Visualizing the effects of predictor variables in black box supervised learning models. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*. https://mathscinet.ams.org/mathscinet/article?mr=4136503

Wachter, S., Mittelstadt, B., & Russell, C. (2017). Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR. *Harvard Journal of Law Technology*, *31*, 841-887.

Wang, W. & others. (2025). Privacy-Preserving Data Analytics in 5G-Enabled IoT for the Financial Industry. *IEEE Transactions on Network and Service Management*.

Watson, D. (2022). Rational Shapley Values. *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency, FAccT '22*, 1083-1094. https://doi.org/10.1145/3531146.3533170

Weber, P., Carl, K. V., & Hinz, O. (2024). Applications of explainable artificial intelligence in finance—A systematic review of finance, information systems, and computer science literature. *Management Review Quarterly*, *74*(2), 867-907.

Weng, F., Zhu, M., Buckle, M., Hajek, P., & Abedin, M. Z. (2025). Class imbalance Bayesian model averaging for consumer loan default prediction: The role of soft credit information. *Research in International Business and Finance*, *74*, 102722. https://doi.org/https://doi.org/10.1016/j.ribaf.2024.102722

Wu, H., Chen, Y., & Yi, C. (2020). Explainable Machine Learning for Regime-Based Asset Allocation. *IEEE International Conference on Big Data*, 5480-5485. https://doi.org/10.1109/BigData50022.2020.9378332

Wu, H.-D., & Han, L. (2021). A novel reasoning model for credit investigation system based on Fuzzy Bayesian Network. *Procedia Computer Science*, *183*, 281-287. https://doi.org/10.1016/j.procs.2021.02.060

Wu, T.-Y., & Wang, Y.-T. (2021). Locally Interpretable One-Class Anomaly Detection for Credit Card Fraud Detection. *2021 International Conference on Technologies and*

Applications of Artificial Intelligence (TAAI), 25-30. https://doi.org/10.1109/TAAI54685.2021.00014

Xiao, Z., & Jiao, J. (2021). Interpretable Credit Risk Assessment Based on Heuristic Knowledge Extraction Method. *2021 IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI)*, 1237-1241. https://doi.org/10.1109/ICTAI52525.2021.00196

Yang, K., Yuan, H., & Lau, R. Y. K. (2022). PsyCredit: An interpretable deep learning-based credit assessment approach facilitated by psychometric natural language processing. *Expert Systems with Applications*, *198*, 116847. https://doi.org/10.1016/j.eswa.2022.116847

Yang, Y., & Wu, M. (2021). Explainable Machine Learning for Improving Logistic Regression Models. *2021 IEEE 19th International Conference on Industrial Informatics (INDIN)*, 1-6. https://doi.org/10.1109/INDIN45523.2021.9557392

Zacharias, J., von Zahn, M., Chen, J., & Hinz, O. (2022). Designing a feature selection method based on explainable artificial intelligence. *Electronic Markets*, *32*(4), 2159-2184.

Zhang, W. & others. (2025). Introduction to the Topic of AI and IoT for Promoting Green Operation and Sustainable Environment. *IEEE Transactions on Intelligent Transportation Systems*.

Zhang, Z., Hamadi, H. A., Damiani, E., Yeun, C. Y., & Taher, F. (2022). Explainable Artificial Intelligence Applications in Cyber Security: State-of-the-Art in Research. *IEEE Access*, *10*, 93104-93139. https://doi.org/10.1109/ACCESS.2022.3204051

Zhu, Y., Yi, C., & Chen, Y. (2020). Utilizing Macroeconomic Factors for Sector Rotation based on Interpretable Machine Learning and Explainable AI. *2020 IEEE International Conference on Big Data*, 5505-5510. https://doi.org/10.1109/BigData50022.2020.9377954

ÖZGEÇMİŞ

İŞ DENEYİMİ

Aralık 2021 – Devam

Bilişim Uzmanı, Türkiye Cumhuriyeti Merkez Bankası, Ödeme Sistemleri Genel Müdürlüğü

- Ödeme ve Menkul Kıymet Takas Sistemleri, Ödeme Hizmetleri ve Elektronik Para Kuruluşlarının Bilgi Sistemlerine İlişkin Mevzuata Uyum Faaliyetleri
- Ödeme Sistemlerine İlişkin Ulusal ve Uluslararası Standartlara Yönelik Araştırma ve Geliştirme Faaliyetleri
- CPMI-IOSCO ORG üyesi.

2018 – 2021 Aralık

Bankacılık Uzman Yardımcısı, Bankacılık Düzenleme ve Denetleme Kurumu, Bilgi Sistemleri Uyum Departmanı (BT)

- SOME (Siber Olaylara Müdahale Ekibi)
- Finansal Kurumların Bilgi Sistemleri Denetim ve Uygulama Faaliyetleri
- Mevzuat Değerlendirmesi

2017 – 2018

İş Analisti, Ziraat Technology Inc. (Ziraat Bankası'nın iştiraki)

- Operasyon Merkezi Yazılım Geliştirme Süreçleri

EĞİTİM

09/2022 – Devam

Doktora, Marmara Üniversitesi

Bilgisayar Mühendisliği (Devam Eden)

09/2018 – 06/2021

Lisans, Ahmet Yesevi Üniversitesi

Bilgisayar Mühendisliği (Mezun)

02/2019 – 02/2021

Yüksek Lisans, Marmara Üniversitesi

Bilgisayar Mühendisliği (Tezli)

10/2010 – 02/2017

Lisans, İstanbul Teknik Üniversitesi

Matematik Mühendisliği (Mezun)

SERTİFİKALAR

08/2023

ISO/IEC 27001:2022 Information Security Management System (ISMS) Lead Auditor CertificationISO/IEC 27001:2022 Information Security Management System (ISMS) Lead Auditor Certification, BSI

Qualification ID : ENR-01356641

AKADEMİK YAYINLAR

- Cil, A. E., Buyuktanir, T., & Yildiz, K. (2026). Comprehensive explainable AI approach for audit opinion classification using feed-forward neural networks. *Knowledge-Based Systems, 339*, 115606. https://doi.org/10.1016/j.knosys.2026.115606
- Cil, A. E., & Yildiz, K. (2025). SFIX: Scalable financial-oriented interpretable explanation. *Internet of Things, 33*, 101713. https://doi.org/10.1016/j.iot.2025.101713
- Cil, A. E., & Yildiz, K. (2025). A systematic literature review on applications of explainable artificial intelligence in the financial sector. *Internet of Things, 33*, 101696. https://doi.org/10.1016/j.iot.2025.101696
- Abdullah Emir Cil, Kazim Yildiz, Ali Buldu, Detection of DDoS attacks with feed forward based deep neural network model, *Expert Systems with Applications*, Volume 169, 2021, 114520, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2020.114520
- Abdullah Emir ÇİL, Mustafa AYDIN, A Review on the Importance of Artificial Intelligence in the Integration of In-House Applications with EBYS, 4th e-BEYAS Symposium, 11 October 2019, Gölbaşı/Ankara.
- Yakar, Ö., BÜYÜKTANIR, B., Çil, A. E., & ALTINEL GİRGİN, A. B., (2023). Performance Comparison of Different Classification Algorithms and Feature Selection Methods in Turkish Hate Speech Problem Analysis. *Avrupa Bilim ve Teknoloji Dergisi* , no.53, 97-111.
- Büyüktanır, B., Doğan, B., Çil, A. E., & Anapalı, A., (2023). Segmentation Analysis with Classification Methods: An Applicatıon with İdeasoft E-Commerce Customer Data . 6. International Antalya Scientific Research and Innovative Studies Congress (pp.193-203). Antalya, Turkey

❖ Büşra BÜYÜKTANIR, Zeki ÇIPLAK, Abdullah Emir ÇİL, Özlem YAKAR, Mahamoud Brahim ADOUM, Kazım YILDIZ,(2025). DDoS_FL: Federated Learning Architecture Approach against DDoS Attack, Pamukkale University Journal of Engineering Sciences, Early Access.