MARMARA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

SOSYAL MEDYA ANALİTİĞİ İLE SİBER ZORBALIK TESPİTİ

FURKAN ZAHİT ATAY

YÜKSEK LİSANS TEZİ

Fen Bilimleri Enstitüsü

Bilgi Güvenliği Mühendisliği Programı

DANIŞMAN

Dr. Öğr. Üyesi Hüseyin YÜCE

EŞ-DANIŞMAN

Dr. Öğr. Üyesi İbrahim SABUNCU

İSTANBUL, 2021

MARMARA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

CYBERBULLYING DETECTION BY SOCIAL MEDİA ANALYTICS

FURKAN ZAHİT ATAY

(527319024)

YÜKSEK LİSANS TEZİ

Fen Bilimleri Enstitüsü

Bilgi Güvenliği Mühendisliği Programı

DANIŞMAN

Dr. Öğr. Üyesi Hüseyin YÜCE

EŞ-DANIŞMAN

Dr. Öğr. Üyesi İbrahim SABUNCU

İSTANBUL, 2021

ÖNSÖZ

Tez çalışmam boyunca verdiği desteklerden ötürü değerli danışman hocam sayın Dr. Öğr. Üyesi Hüseyin Yüce ‘ye teşekkür ediyorum.

Çalışmaya başlama konusunda öncülük eden ve değerli katkılarından dolayı sayın Dr. Öğr. Üyesi İbrahim SABUNCU’ya ayrıca teşekkür ediyorum.

Yüksek lisans eğitimim boyunca yardımlarını hep yanımda hissettiğim sevgili aileme teşekkür ediyorum.

Tüm yüksek lisans eğitimim boyunca ve tez çalışmam boyunca desteğini hiç esirgemeyen sevgili eşim Semra ATAY’a sonsuz teşekkür ediyorum.

TEMMUZ 2021 **Furkan Zahit ATAY**

İÇİNDEKİLER

ÖZET iv
ABSTRACT v
SEMBOLLER vi
KISALTMALAR vii
ŞEKİL LİSTESİ viii
TABLO LİSTESİ ix
1. GİRİŞ 1
2. LİTERATÜR ARAŞTIRMASI 4
3. GENEL KAVRAMLAR 8
3.1. Veri Madenciliği 9
3.1.1. Veri Madenciliği Teknikleri 11
3.1.1.1. Veri sınıflandırması 11
3.1.1.2. Kümeleme analizi 12
3.1.1.3. Birliktelik kuralları 12
3.1.1.4. Sıralı kalıpları bulma 12
3.1.1.5. Regresyon 12
3.1.1.6. Aykırı verileri belirleme 12
3.2. Sosyal Medya Analitiği 12
3.2.1. Sosyal Medya Kavramı 14
3.2.2. Sosyal Medya İstatistikleri 18
3.2.3. Sosyal Medyanın Hayattaki Yeri ve Sosyal Medya Analitiğinin Önemi 20
3.2.3.1. Pazarlama ve Finans 21
3.2.3.2. İnsan Kaynakları 22
3.2.3.3. Turizm 23
3.2.3.4. Spor Sektörü 24
3.2.3.5. Kamu Sektörü 28
3.2.3.6. Sağlık Sektörü 29
3.2.3.7. İstihbarat, Güvenlik ve Öngörüsel Polislik 30
3.3. Metin Madenciliği ve Duygu Analizi 32
3.3.1. Metin Madenciliği 32
3.3.2. Duygu Analizi 35

3.4. Siber Zorbalık 39
4. MATERYAL VE YÖNTEM 41
4.1. RapidMiner Studio 42
4.2. Veri Toplama 47
4.3. Veri Seti 50
4.4. Veri Etiketleme 50
4.5. Veri Önişleme 52
4.6. Yöntem 55
4.6.1. Sınıflandırma ve Tahmin 55
4.6.1.1. Naive Bayes 57
4.6.1.2. Genelleştirilmiş Doğrusal Model 57
4.6.1.3. Derin Öğrenme 58
4.6.1.4. Karar Ağacı 59
4.6.1.5. Random Forest 60
4.6.1.6. Gradyan Güçlendirilmiş Ağaçlar (Gradient Boosted Trees) 60
4.6.1.7. Destek Vektör Makinesi (Support Vector Machine – SVM) 60
5. BULGULAR VE TARTIŞMA 61
5.1. Auto Model Tahmin Modellerinin Sonuçları 63
5.1.1. Naive Bayes Model Sonuçları 66
5.1.2. Genelleştirilmiş Doğrusal Model Sonuçları 67
5.1.3. Derin Öğrenme Model Sonuçları 69
5.1.4. Karar Ağacı Model Sonuçları 73
5.1.5. Random Forest Model Sonuçları 74
5.1.6. Gradient Boosted Trees Model Sonuçları 75
5.1.7. Support Vector Machine Model Sonuçları 78
5.2. Modellerin Karşılaştırılması 80
6. SONUÇLAR 83
KAYNAKLAR 86
ÖZGEÇMİŞ 91
EK 92

ÖZET

SOSYAL MEDYA ANALİTİĞİ İLE SİBER ZORBALIK TESPİTİ

Veri madenciliği konusu günümüzde yapay zekâ ile makine öğrenmesi konularıyla da birleşerek gelişmeye devam eden bir konu haline gelmiştir. Bu alan için çeşitli yazılımlar da yapılmaktadır. Metin madenciliği (text mining) konusu da buna ek olarak ayrıca gelişen konulardandır.

Günümüzde sosyal medya kullanımının çok yaygın hâle gelmiş olması da metin madenciliği uygulamalarının bu alan üzerinde yapılmasını beraberinde getirmiştir. Özellikle de Twitter'da metin yoluyla insanların duygu, fikir ve görüşlerini paylaşması sebebiyle Duygu Analizi çalışmaları yine Twitter üzerinde yoğunlaşmıştır.

Her gün milyonlarca mesaj paylaşılan sosyal medyadaki mesajın okunup analiz edilmesi ve buralardaki siber zorbalıkların insan eliyle tespit edilmesi neredeyse mümkün değildir. Bu nedenle siber zorbalıkların yapay zekâ yardımıyla tespit edilmesi ve zararlı mesajların yaygınlaşmadan ilgililerce gerekli müdahalelerin yapılması önemlidir. Bu yüzden bu çalışmada siber zorbalık içeren tweetlerin yapay zeka destekli algoritmalarla tespit edilip bu mesajları paylaşan hesapların otomatik olarak belirlenmesine sağlayan bir sistemin geliştirilmesi amaçlanmıştır. Bu çalışmayı yapabilmek için öncelikle Twitter'da yüksek takipçi ve etkileşime sahip hesaplardan oluşan bir "Hedef Liste" oluşturulmuştur. Bu listedeki hesaplara atılan tweetler toplanmıştır. Siber zorbalık tespitlerini yapabilen makine öğrenmesi modelini oluşturabilmek için elde ettiğimiz bu veri seti kullanılmıştır.

Bölüm 1'de giriş yaptığımız bu çalışmada Bölüm 2'de literatür araştırması yapılmıştır. Bölüm 3'de kavramların genel tanımlarının yapılması ve sosyal medya analitiği çalışmalarının hangi alanlarda yapıldığından bahsedilmiştir. Bölüm 4'de ise kullandığımız çalışmamızdaki materyal ve yöntem ele alınmıştır. Çalışmayı yaparken bir veri madenciliği ve makine öğrenmesi uygulaması olan RapidMiner Studio uygulaması kullanılmıştır. Bölüm 5'de ise çalışmamızdan elde ettiğimiz bulgulardan bahsedilmiştir. Son bölümde sonuçlardan bahsedilerek çalışma tamamlanmıştır.

ABSTRACT

CYBERBULLYING DETECTION BY SOCIAL MEDİA ANALYTICS

The issue of data mining has become a subject that continues to develop, combining with artificial intelligence and machine learning. Various software are also made for this field. In addition, text mining is also one of the developing topics.

The fact that the use of social media has become widespread today has brought text mining applications to be made on this field. Sentiment Analysis studies have also concentrated on Twitter, especially because people share their feelings, ideas and opinions via text on Twitter.

It is not possible to read and analyze the messages on social media, where millions of messages are shared every day, and the bullying in these places can not be detected by human hands. Therefore, it is important that cyber bullying is artificial intelligence and necessary investigations are made by those concerned before harmful messages become widespread. Therefore, in this study, it is aimed to find a solution for detecting tweets containing bullying with artificial intelligence-supported algorithms and automatically determining the accounts that share these messages. In order to do this work, a "Target List" consisting of accounts with high followers and interaction on Twitter has been created. The tweets posted to the accounts in this list have been collected. This data set we obtained was used to create a machine learning model that can detect cyberbullying.

In this study, which we entered in Chapter 1, the results of the literature research are written in Chapter 2. In Chapter 3, general definitions of the concepts and the areas in which the study, which we call social media analytics, is carried out, is mentioned. In Chapter 4, the material and method in our study we used are discussed. RapidMiner Studio application, which is a data mining and machine learning application, was used while doing the study. In Chapter 5, the findings we obtained from our study were mentioned. In the last section, the study was completed by mentioning the results.

SEMBOLLER

N : Toplam tahmin sayısı

i : sıra sayısı

Yi : tahmin edilen i. sıradaki tahmin değeri

Xi : i. sıradaki gerçek değer

R : Duyarlılık

P : Kesinlik

F-Measure : F1 Ölçeği

r : Kolerasyon

KISALTMALAR

ARPANET : Gelişmiş Araştırma Projeleri Dairesi Ağı

KDK : Kamu Denetçiliği Kurumu

TÜİK : Türkiye İstatistik Kurumu

NLP **:** Doğal Dil İşleme

EMBERS : Early Model Based Event Recognition

MS : Microsoft

CRISP-DM : Cross Industry Standard Processing – Data Mining

SQL : Yapılandırılmış Sorgu Dili

URL : Tek Düzen Kaynak Bulucu

GLM : Genelleştirilmiş Doğrusal Model

RF : Random Forest

kNN : k En Yakın Komşu Algoritması

SVM **:** Support Vector Machine (Destek Vektör Makineleri)

RMSE **:** Root Mean Square Error - Kök Ortalama Kare Hata

MAE : Mean Absolute Error - Mutlak Hata

RE : Relative Error - Göreceli Hata

MSE : Mean Squared Error - Ortalama Kare Hata

TP : True Positive (Doğru Pozitif)

TN : True Negative (Doğru Negatif)

FP : False Positive (Yanlış Pozitif)

FN : False Negative (Yanlış Negatif)

ŞEKİL LİSTESİ

Şekil 1. Bilgi Keşif Sürecinde Veri Madenciliği 10
Şekil 2. Sosyal Medya Analitiği Süreci 14
Şekil 3. Metin Madenciliği ve İlişkili Olduğu Alanlar 33
Şekil 4. Metin Kümeleme 34
Şekil 5. Metin Sınıflandırma Adımları 34
Şekil 6. CRISP-DM Süreci 36
Şekil 7. Duygu Sınıflandırma Yöntemleri 38
Şekil 8. Önerilen Metodoloji 42
Şekil 9. RapidMiner Başlangıç Arayüzü 1 43
Şekil 10. RapidMiner Başlangıç Arayüzü 2 43
Şekil 11. RapidMiner Görsel İş Akışı Tasarımcısı 44
Şekil 12. RapidMiner Excel Veri Aktarma 44
Şekil 13. RapidMiner Veri Tabanı İşlemleri 45
Şekil 14. RapidMiner İstatistiklere Genel Bakış 45
Şekil 15. RapidMiner Turbo Prep Ekranı 46
Şekil 16. RapidMiner Auto Model 47
Şekil 17. RapidMiner Açıklanabilir Örnek Bir Model 47
Şekil 18. Twitter Hesapları için Döngü 49
Şekil 19. Search Twitter kullanımı 50
Şekil 20. Veri Setinden Örnek Bir Kesit 50
Şekil 21. Veri Önişleme Süreci 52
Şekil 22. Veri Önişleme Süreci 1 53
Şekil 23. Veri Önişleme Süreci 2 53
Şekil 24. Durak Kelimeleri Dosyası Ekleme 54
Şekil 25. Auto Model Başlangıç Ekranı 56
Şekil 26. RapidMiner Auto Model Tipleri 56
Şekil 27. Hata Matrisi [90] 63
Şekil 28. Deep Learning Simulatör (Basın-Yayın) 71
Şekil 29. Deep Learning Simulatör (Kamu ve Siyaset) 71
Şekil 30. Deep Learning Simulatör (Mobil) 72
Şekil 31. Deep Learning Simulatör (Web) 72
Şekil 32. Gradient Boosted Tree Model Sonucu Örnek Ağaç 77
Şekil 33. Gradient Boosted Tree Simulatör (Kamu ve Siyaset) 77
Şekil 34. Gradient Boosted Tree Simulatör (Kültür Sanat) 78
Şekil 35. Support Vector Machine Simülatör 79
Şekil 36. Modellerin Hata Türüne Göre Kıyaslanması 81
Şekil 37. Modellerin F1 Ölçeği ve Doğruluk Değerleri 82
Şekil 38. Modellerin Doğruluk Değerleri 83

TABLO LİSTESİ

Tablo 1. Sosyal Medya Kategorileri 16
Tablo 2. Dört Büyük Futbol Takımının Pazarlama Aracı Olarak Sosyal Medya Kullanımı 25
Tablo 3. Dört Büyük Spor Kulübü Sosyal Medya İstatistikleri 27
Tablo 4. Retrieve Twitter Hedef Liste Örneği 48
Tablo 5. Durak Kelimelerin Listesi 54
Tablo 6. Veri Seti Adetleri 61
Tablo 7. Kategorilerin Kişi Sayıları 61
Tablo 8. Kategorilerin Tweet Dağılımı ve Siber Zorbalık Tespit Oranları 62
Tablo 9. Sınıflandırma Performans Metriği 64
Tablo 10. Hata Türü Sembol ve Tanımları 65
Tablo 11. Naive Bayes Model Sonuçları 66
Tablo 12. Genelleştirilmiş Doğrusal Model Sonuçları 68
Tablo 13. Genelleştirilmiş Doğrusal Modelinde Niteliklerin Sonuca Etkisi 69
Tablo 14. Derin Öğrenme Model Sonuçları 70
Tablo 15. Karar Ağacı Model Sonuçları 73
Tablo 16. Random Forest Model Sonuçları 74
Tablo 17. Gradient Boosted Trees Model Sonuçları 76
Tablo 18. Support Vector Machine Model Sonuçları 79
Tablo 19. Modellerin Hata Türüne Göre Kıyaslanması 80
Tablo 20. Modellerin F-Ölçeği ve Doğruluk Değerleri 81

1. GİRİŞ

Sosyal medya, günümüzde artık neredeyse her kesim ve her yaştaki insan tarafından bilinen veya en azından duyulan bir kavramdır. Sosyal medyayı belki de bu kadar popüler yapan özelliklerin en başında kişilere serbest, özgür bir alan sunması ve bir tıkla birçok kaynağa erişim sağlayabilmesidir.

Sosyal medya; televizyon, radyo, dergi, gazete gibi geleneksel medyadan farklı olarak tek taraflı bir yayın değil çift taraflı bir yayındır. Farklı bir deyişle bireyler; tüketici değil, hem tüketici hem üretici konumundadır. İnsanlar mikroblog dediğimiz sosyal medya araçlarında fikirlerini, eleştirilerini, düşüncelerini ya da kendileriyle ilgili yazılarını yazıp başka insanlara içerik sunabilmektedir. Aynı şekilde fotoğraf paylaşımları yapılabiliyor, video atılabiliyor hatta canlı yayın yapılabiliyor. Durum böyle olunca sosyal medya dünyasından gerçek dünyadan neredeyse farksız hâle gelmeye başlamaktadır. Ayrıca Dünya'nın herhangi bir konumundaki bir kişi, diğer herhangi bir konumdaki bir kişiyi bulabilmekte ve takip edebilmektedir.

Tüm bu sosyal medya mecralarını düşününce üretilen verinin boyutunun çok büyük olduğu anlaşılmaktadır. Günümüzde bu verilerin işlenmesi ve anlamlı bilgiler çıkarılması konusunda çalışmalar gün geçtikçe daha da ilgi gören bir alan hâline gelmektedir. Sosyal medya madenciliği de denilen bu alana biz sosyal medya analitiği demeyi tercih etmekteyiz. İleriki bölümlerde de ele alacağımız sosyal medya analitiğinin kullanıldığı yerler arasında; pazarlama, insan kaynakları, halkla ilişkiler, turizm vb. alanları sayılabilir.

Veri madenciliği konusu günümüzde yapay zekâ ile makine öğrenmesi konularıyla da birleşerek gelişmeye devam eden bir konu haline gelmiştir. Bu alan için çeşitli yazılımlar da yapılmaktadır. Metin madenciliği (text mining) konusu da buna ek olarak ayrıca gelişen konulardandır.

Tezimizin ilk kısımlarında sosyal medya ile ilgili genel kavramlar ve sosyal medyanın kullanımına ilişkin istatistiki bilgiler vereceğiz. Ayrıca sosyal medyanın kullanıldığı alanları ele alarak aynı zamanda konumuzla olan ilgisi için de bilgiler vereceğiz. Son

kısımlarda sosyal medyadaki metinlerden elde ettiğimiz verileri analiz ederek bir mesajın siber zorbalık durumunu ölçeceğimiz bir sonuç elde etmeye çalışacağız.

Siber zorbalık, bilgi ve iletişim teknolojilerini kullanarak özel veya tüzel bir kişiye karşı teknik veya ilişkisel bir şekilde zararlı davranış olarak tanımlanabilir. Bilgi ve iletişim teknolojilerini kullanarak insanları sürekli taciz etmek (siber takip), insanlarla alay etmek, tehdit etmek, dedikodu yaymak, internet üzerinden hakaret etmek gibi ilişkisel saldırı davranışlarını içerir.[1]

Bu çalışmada siber zorbalığın Türkçe yazılmış metin içeriklerinden ve incelenen diğer özelliklerinden otomatik olarak tespit edilme başarısı araştırılmıştır. Bu bağlamda, ülkemizde en çok takipçiye ve etkileşime sahip olan Twitter hesapları araştırmamızda "Hedef Liste" olarak seçilmiştir ve bu kişilere atılan Twitter mesajları incelenmiştir. Hedef liste, 6 kategoriden oluşan 378 kişilik bir listedir. Bu kategoriler; Kamu ve Siyaset, Kültür Sanat, Spor, Fenomenler, Basın-Yayın ve İş Dünyası şeklindedir. 29198 adet tweetin okunması ve etiketlenmesi tamamen manuel olarak yapılmıştır. Tweetler, 1-7 Şubat 2021 tarihleri arasındaki en son ve popüler tweetler olarak çekilmiş ve MS Excel dosyasına aktarılmıştır.

Seçtiğimiz hedef listedeki hesaplardan metinler toplamak için RapidMiner Studio programı kullanılmıştır. Twitter'ın getirmiş olduğu kısıtlamalardan dolayı veri çekme işlemi tek seferde yapılamamıştır. 3 ayrı parçaya bölünerek veri çekme işlemi yapılmıştır. Bunun sonucunda otuzbin adete yakın örnek içeren bir veri çekilmiştir ve mükerrer olanlar silindiğinde 29198 adetlik veri seti hazırlanmaktadır. Bu veri seti, tespit edebildiğimiz kadarıyla Türkçe metin içeriklerinden siber zorbalığı tespit etmek için hazırlanmış en büyük veri seti olma özelliğini taşımaktadır.

Veri seti toplandıktan sonra, veriler tek tek okunmuş olup siber zorbalık içerikli metinlere sahip olanlar manuel olarak etiketlenmiştir. Veri çekme işlemi ve etiketleme işlemi haftalar süren bir çalışma olmuştur.

RapidMiner programını kullanarak binlerce tweet verisi elde ettiğimiz bu çalışmada metinlerin ön işlemesi ve analiz edilmesi gibi işlemleri de bu program üzerinden yapılmıştır. Makine öğrenmesi teknikleri kullanılmıştır. Makine öğrenmesi

tekniklerinden en verimli olan yöntem belirlenmiş ve olası siber zorbalık durumlarının tespiti bu yöntemle tespit edilecek bir sonuç elde edilmiştir.

Tezimizi diğerlerinden ayıran özelliklerini şu şekilde tanımlayabiliriz. Bu tez çalışması öncelikle literatürdeki diğer çalışmalara bakıldığında 30.000 adet Türkçe tweet çekilmesiyle özgün bir değer taşımaktadır. Türkçe metinlerde yapılan çalışmaların az olmasından kaynaklı bir durumun olmasının yanı sıra tweet sayısının fazla olması önemli bir değerdir. Diğer çalışmalara bakıldığında tweetin atıldığı kaynak ile siber zorbalık tespitinin ilişkisinin olup olmaması hiç araştırılmamıştır.

Çalışmamızın en önemli farklılıklardan bir diğeri ise, tweetin atıldığı kişiler kategorilendirilmiş ve bu kategorilerin siber zorbalık tespitiyle ilişkisi araştırılmıştır. Dolayısıyla çalışmamız, özellikle kategori özelliğinin siber zorbalık tespitinde kullanılması açısından uygulamaya katkı sağlamıştır. Çalışmamızın diğerlerinden ayıran bir diğer önemli özellik ise verileri etiketleme çalışmasının etiketleme işlemidir. Veriler detaylı bir şekile tek tek okunarak içerdikleri anlama göre karar verilmiş olup etiketlenmiştir. Dolayısıyla siber zorbalık tespiti açısında bütün kombinasyonlar gözle görülmüş ve buna göre değerlendirilmiştir. Bu da verimi arttırmıştır. Çalışmanın ilk aşaması olan veri setinin iyi hazırlanmış olması çalışmamızın sonuçlarını olumlu yönde etkilemiştir.

2. LİTERATÜR ARAŞTIRMASI

Bu bölümde konuyla ilgili daha önce yapılmış çalışmalara ilgili anahtar sözcüklerle ulaşılmaya çalışılmıştır. Konuyla ilgili Türkçe yapılmış çalışma sayısının az olduğu da fark edilmiştir. Çalışmaların birçoğunun İngilizce olduğu görülmüştür. YÖK Tez merkezi, DergiPark Akademik, IEEE Xplore, ResearhGate, Emerald Insight, Google Scholar, Web of Science gibi kaynalardan taramalar yapılmıştır. Anahtar sözcük olarak, "siber zorbalık, elektronik zorbalık, siber taciz, siber zorbalık tespiti vb." aramalar kullanılmıştır ve ilgili makale ve kaynaklar tespit edilip incelenmiştir. Bu makaleler kullandığı kaynaklar bakımından, kullandığı yöntemler bakımından, kullanılan ön işleme adımları bakımından ve kullandığı veri setlerin özellikleri bakımından incelenmiş ve sınıflandırmalar yapılmıştır.

Çalışmalar ağırlıklı olarak metinsel özellikler aracılığıyla siber zorbalığı tespit ettiği göz önüne alındığında, taramamız bu nedenle metinsel siber zorbalığa odaklanmıştır. Siber zorbalığın görüntü, video ve sözlü analizlerle tespit edilmesi gibi yeni ortaya çıkan alanlar araştırmamız kapsamında yer almamıştır.

İncelediğimiz çalışmalarda veri kaynağı olarak Twitter, Facebook, MySpace, Instagram, Youtube vb. sosyal mecralardaki yazılan mesajlar ve yorumlar ağırlıklı olarak kullanılmıştır. Tüm çalışmalarda veri kaynağı birbirinden farklı olsa da verinin elde edilişinin ardından yapılan işlemler metin madenciliği işlemlerinin alanına girdiği için çalışmalarda kullanılan yöntemler arasında yakınlık bulunmaktadır.

Mesajların tespitinde genellikle bu duygu ile ilgili kelime listeleri kullanılarak duygu analizi, denetimli öğrenme ve sözlük tabanlı sistemler ve analizler kullanılmıştır. En yaygın olarak da ikili sınıflandırma yapılmış ve siber zorbalık içerikli mesajları "zorbalık" sınıfının içine almıştır. Diğerleri ise "diğer" ya da "zorbalık içermeyen" sınıfına atanmıştır.

Yin ve ark.[2], 2009 yılında yapmış olduğu çalışma veri kaynağı olarak MySpace, Slashdot, Kongregate adlı web sitelerden faydalanmıştır. Elde edilen veri setinde manuel olarak etiketleme yapılmıştır. Zorbalığı tespit etmek için üç farklı temel yaklaşım izlenmiştir. İlk olarak kelime düzeyinde n-gramlar kullanılmış, ikinci olarak küfürlü kelimelerin derlendiği bir sözlük kullanılmıştır. Son olarak ve çalışmada performans

olarak en etkili olduğu söylenen TFIDF özellikleri kullanılmıştır. Denetimli öğrenme teknikleri kullanarak duygu analizi gerçekleştirmiştir.

Dinakar ve ark. [3], 2011 yılında 4500 Youtube yorumunu içeren bir veri seti üzerinde çalışmıştır. Youtube PHP API'si kullanılarak cinsellik, ırk ve kültür, zeka içerikli zorbalık mesajları toplanmıştır. Toplanan veri yine manuel olarak etiketlenmiştir. Küfürlü içerikleri de dikkate alan çalışma alaycı ifadelerin yanlış sınıflandırıldığı tespitinde de bulunmuştur. Denetimli öğrenme teknikleri kullanarak duygu analizi gerçekleştirmiştir.

Xu ve ark. [4], Twitter ve Wikipedia üzerinde çalışma yapmışlardır. Yaklaşım olarak n-gram yöntemini kullanmıştır. Naive Bayes, SVM ve Lojistik Regresyon modellerini kullanmıştır. Kullanıcıları suçlayıcı, zorbacı, haber veren ve mağdur olarak 4 farklı role ayırmıştır ve metinleri bu şekilde etiketlemiştir.

Rafiq ve ark. [5], sosyal medya olarak 6 saniyelik videolar paylaşılan Vine'ı seçmiştir. Yaklaşık 50 bin kişilik bir kullanıcı profili toplamış ve bu kullanıcıların konum bilgisi, profil açıklaması, video sayısı, yayın kimliği, takip ettiği sayısı ve takipçi sayısı gibi bilgilerini derlemiştir. Siber zorbalığın küfür içermeyen metinlerde de olabileceği kabulüne göre hareket etmiştir. Denetimli öğrenme teknikleri kullanarak duygu analizi gerçekleştirmiştir.

Burn-Thorton ve Burman [6], Kontostathis ve ark. [7], Nahar ve ark.[8], Munezero ve ark. [9], Nandhini ve Sheeba [10] ve Zhao ve ark.[11] yukarıda sözü edilen ikili sınıflandırmayı da uygularken mesaj sınıflandırmasını duygu analizi yöntemi yoluyla gerçekleştirmemiştir. Xu ve ark.[4] farklı olarak duygu analizini "zorbalık içeren" ve "zorbalık içermeyen" şeklinde ikili sınıflandırma amacıyla değil, "zorba" ve "zorbalık" kelimelerinin herhangi birini içeren tweetleri "zorba izleri" olarak ifade edilen duyguları anlamak için yapılan ender örneklerdendir.

Sanchez ve Kumar [12], Chen ve ark. [13], Dadvar ve ark. [14] rol tanımlama konusunda çalışma yapmıştır. Bu çalışmalarda 8 rol tespit edilmiş ve kullanılmıştır. Bunlar; zorba, mağdur, görgü tanığı, asistan, savunucu, muhabir/bildiren, suçlayıcı ve pekiştirendir. Yukarıda sözü edilen 8 roldeki görgü tanığı / seyirci, bir zorbalık olayına müdahale etmeyen tanıklardır. Asistanlar, suç işleyenlerdir ancak başlatan değildir. Pekiştirenler zorbalığa doğrudan dâhil olmamakla birlikte zorbaları teşvik eden ve devam

etmesi için katkı sağlayanlardır, zorbalığa gülen, sevinenlerdir. Suçlayıcı ile muhabir/bildiren arasındaki fark mağdurları ve zorbaları aktif olarak tanımlamaktır. Savunucular ise mağdurlara destek ve yardımda bulunurlar. Yukarıda sayılan çalışmalar bu rollerden bir veya birkaçını tespit etmeye yönelik yapılmış çalışmalardır.

Chen ve ark. [13], Dadvar ve ark. [15], Perez ve ark. [16], Del Bosque ve Garza [17], Potha ve Maragoudakis [18] zorbalık mesajlarının şiddetini belirleyen bir puan hesaplanarak siber zorbalığının etkisinin belirlenmesi ile ilgili çalışmalar yapmıştır. Böyle bir puanlama yaparak siber zorbalık metinlerinin şiddetini ve etkisini hesaplamaya çalışmıştır. Squicciarini ve ark.[19] çalışmasında ise siber zorbalık olayından sonra meydana gelen olayları tespit etmek ve sınıflandırmak gibi bir hedefte çalışmıştır.

Çalışmalardan siber zorbalık tespiti yapılırken 4 ana özellikten faydalanıldığı görülmüştür. İçerik tabanlı, duygu analizi tabanlı, kullanıcı tabanlı, ağ tabanlı şekilde özellikleri sayabiliriz. İçerik tabanlı özellikler; anahtar kelimeler, küfürler, zamirler gibi metinde yer alan öğeler olarak tanımlanabilir. Duygu tabanlı özellikler; metinde yer alan ifadelerden yola çıkarak duygu analizi yapılması olarak tanımlanabilir. Kullanıcı tabanlı özellikler; bir profilin yaş, cinsiyet, konum gibi özellikleridir. Ağ tabanlı özellikler ise, sosyal ağdan çıkartılabilen takipçi sayısı, arkadaş sayısı, gönderi sayısı gibi öğeleri içeren özelliklerdir.

İçerik tabanlı özellikleri kullanan çalışmalar; Dinakar ve ark. [3], Perez ve ark. [16], Kontostathis ve ark. [7], Nahar ve ark.[8] ile Bretschneider ve ark.[20] tarafından yapılmıştır. Çalışmalarında araştırmacılar tarafından derlenen veya www.noswearing.com ile urbandictionary.com gibi dış kaynaklardan alınan kelime listelerini kullanarak küfür sözlükleri oluşturulmuştur. Bunlar küfür etmenin siber zorbalıkla ilişkisi ve tek başına yeterli olup olmadığını araştıran çalışmalardır. Küfür içermese de ırk, fiziksel görünüm, zeka, cinsiyet, cinsellik, kültür temalı siber zorbalığı ifade eden ifadeler konusunda da çalışılmıştır. Benzer şekilde, belge uzunluğu, büyük-küçük harf kullanımı, imla ve özel karakterlerin kullanımı konusundaki araştırmalar [15] [14] [21] [8] da içerik özellikleri kullanan çalışmalara örnek olarak verilebilir.

Duygu temelli özellikleri kullanan çalışmalarda girdi olarak kullanmak için duygusal anahtar kelimelerine ihtiyaç bulunmaktadır. Böylelikle negatif, pozitif, nötr olarak duygu

analizinde sınıflandırma yapılabilecektir. Duygusal kelimeler genellikle WordNet vb kaynaklarda yer almaktadır. Xu ve ark. [4] yaptıkları çalışmada elde ettikleri tweetlerde 7 farklı duyguyu tespit etmişlerdir. Bunlar; öfkeli, utanç verici, empati yoksunu, korku dolu, gurur kırıcı, rahatlama ve üzüntü verici duygulardır. Sonraki çalışmalarında ise tek başına duyguları temel alan bir tespit sisteminin gerçek duygular ile alaycı olanları doğru şekilde ayırt edilemediği ifade edilmiştir. Munezero ve ark. [9] yalnızca içerik temelli özellikleri kullanan çalışmasına kıyasla duygu temelli özellikleri kullanan çalışmalarının düşük performanslı olduğu görülmüştür.

Kullanıcı temelli özellikleri kullanan çalışmalar; siber zorbalık tespit sistemlerine yaş, cinsiyet, cinsel yönelim ve ırk özelliklerini dâhil etmişlerdir. En yaygını ise yaş ve cinsiyet olduğu ortaya koyulmuştur. Dadvar ve De Jong [22] denetimli öğrenme yöntemini kullanırken cinsiyet özelliklerini kullanmış ve performansın büyük oranda arttığını ifade etmiştir. Ancak sosyal medyadaki cinsiyet, yaş gibi bilgilerin ne kadar doğruyu yansıttığı da bilinmemektedir. Serra ve Venter [23] kullanıcının yaşının yanında cep telefonu kullanım modeline göre kişilerin risk profilini tanımlamıştır.

Ağ tabanlı özellikleri kullanan çalışmalar; arkadaş sayısı, takipçi sayısı, etkileşim gibi verileri kullanmışlardır. Bu çalışmalarda egosantrik ağları kullanma yaygındır. Galan ve Garcia [24], gönderi zamanı, konum ve Twitter istemcisi gibi ağ tabanlı özellikleri kullanarak Twitter'da zorbalık yapan profilin gerçek sahibini bulmak için çemberi epey daraltmıştır.

Çalışmalar incelendiğinde siber zorbalık tespiti için kullanılan teknikler çoğunluğun denetimli öğrenme tekniği kullandığı ortaya çıkarılmıştır. Bu yöntemi kullanan ilk çalışma olarak Yin ve ark.[2] tarafından yapılan 2009 yılındaki çalışma tespit edilmiştir. kNN, Naive Bayes, SVM, Karar Ağacı, Nöral Ağlar, Lojistik Regresyon yöntemleri ağırlık kazanmıştır.

Literatür taramasında fark edildi ki; Twitter en yaygın veri kaynaklarındandır. Twitter'da metin ve duygu ifadelerinin ağırlıklı olması ve bireylere ulaşma ve mesaj gönderme açısından kolaylık sağladığı için sık kullanılıyor olmasından dolayı daha çok tercih edilen bir kaynaktır.

Araştırmalar bu konunun popüler bir hâle geldiğini ve önlem almaya yönelik çalışmaların arttığını göstermektedir. Bu çalışmalar şuanda metinsel açıdan ele alınmaktadır. İleriki dönemlerde resim ve video şeklinde görüntü verileri olarak ele alınması söz konusu olabilir.

3. GENEL KAVRAMLAR

Bu bölümde konuyla ilgili olduğu düşünülen kavramlarla ilgili açıklamalara yer verilmiştir.

3.1. Veri Madenciliği

Veri; gözlem, sayım, ölçüm, deney vb araştırma yoluyla elde edilen ham bir kayıttır. İşlenmeye ve yorumlanmaya imkân vermektedir. Bilgi ise, verinin analizi ve yorumu yapılmış anlamlı hâlidir.

Geçmişten günümüze veri madenciliği konusunda birçok tanım yapılagelmiştir. Bazıları şu şekildedir;

- Veri madenciliği ham datanın tek başına sunamadığı bilgiyi çıkaran, veri analizi sürecidir. [25]
- Hand (1998), veri madenciliğini istatistik, veritabanı teknolojisi, örüntü tanıma, makine öğrenme ile etkileşimli yeni bir disiplin ve geniş veritabanlarında önceden tahmin edilemeyen ilişkilerin ikincil analizi olarak tanımlamıştır.[26]

Tanımlamalardan özetle şöyle bir tanım yapmak mümkündür. Çeşitli yöntemlerle elde edilerek bir araya gelmiş çok büyük ve devasa verilerin bazı analizlerle amacımız doğrultusunda anlamlı ve yararlı bilgiler elde edilmesine veri madenciliği denilmektedir.

Veri madenciliği daha önce fark edilmemiş sonuçları elde edebilme yeteneğine sahip olması dolayısıyla kıymetli bir alandır. Günümüzde veri madenciliği uygulamaları sağlık, eğitim, bankacılık, pazarlama, endüstri, bilişim vb. birçok sektörde vazgeçilmez bir unsurdur.

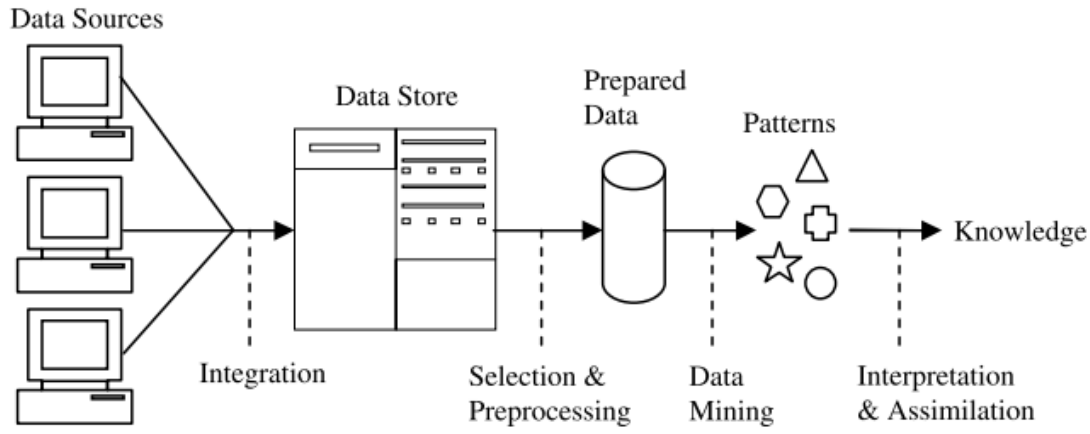
Veri madenciliği hedeflerini temelde ikiye ayırabiliriz; [27]

- Diğer değişkenlerden birini veya daha fazlasını tahmin etmek için değişkenlerin bir kısmı kullanılarak elde edilen *öngörücü hedefler* (örneğin, sınıflandırma, regresyon, anormallikler / aykırı değerler tespiti);
- Verileri tanımlayan ve kullanıcı tarafından kolayca anlaşılabilen kalıpların tanımlanmasıyla ulaşılan *tanımlayıcı hedefler* (örn. Kümeleme, ilişkilendirme kuralı keşfi).

Teknolojinin ilerlemesi ile birlikte hayatımızın her alanında veriler yer almaktadır. Örneğin bir alışveriş yapıldığında satıcının sattığı ürünün fiyatı, adedi, türü vb. her türlü

bilgisi birer veri olduğu gibi müşterinin yaşı, cinsiyeti, ödediği ücret vb. bilgisi de birer veridir. Satıcı şirketler için tüm veriler bir pazarlama stratejisi oluşturmak etkili faktörlerdir. Bu verileri anlamlı bir hale getirmesiyle beraber etkin pazarlama stratejileri yapabilmektedirler. Aynı örneklere bankacılık, sağlık gibi diğer sektörlerde de karşılaşılabilmektedir.

Aşağıda gösterilen Şekil 1'de [28] bilgi keşif süreci gösterilmiştir.



Şekil 1. Bilgi Keşif Sürecinde Veri Madenciliği

Adımlar şu şekilde sıralanabilir;

Veri Temizleme:

Elimizdeki verilerde; hatalı kaydedilmiş, eksik, tekrarlanmış, gürültülü veriler var olabilir. Bunların olması işlemlerimizin sonucuna etki edeceği için bu verilerin düzeltilmesi, temizlenmesi gerekmektedir. Eksik değerlerin; genelin ortalaması ile, en yakın komşuları dikkate alınarak yada mod-medyan gibi çeşitli yöntemlerle yeri doldurulabilir. Aykırı değerler bazı grafik çeşitleri ile görülebilir ve silinebilir. Hatalı girişler ise tespit edilip doğruları ile düzeltilmelidir.

Veri Entegrasyonu

Birden fazla veri kaynağının birleştirilebileceği yerlerde verilerin uyumlu ve entegre hale getirilmesi gerekmektedir.

Veri Seçimi

Veri madenciliği tekniklerini daha verimli kullanabilmek için tümünü temsilen daha küçük parça verinin seçilmesi işlemidir.

Veri dönüşümü

Verilerin özet veya toplama işlemleri gerçekleştirilerek madencilik için uygun biçimlere dönüştürülüp konsolide edildiği durumlarda kullanılır.

Veri madenciliği (veri modellerini çıkarmak için akıllı yöntemlerin uygulandığı önemli bir süreç)

Model değerlendirmesi (bilgiyi temsil eden doğru modelleri belirlemek)

Bilginin sunumu (çıkarılmış sonucu kişilere sunmak için görselleştirme ve bilgi sunma tekniklerinin kullanılması)

Kaynakların bazıları Bilgi Keşif Süreci'ni veri madenciliğe il eş anlamda görürken, bazı kaynaklarda da veri madenciliğinin bilgi keşif süreci içerisinde sadece bir adım olduğu söylenmektedir.

3.1.1. Veri Madenciliği Teknikleri

Veri madenciliği teknikleri genel anlamda tanımlayıcı/açıklayıcı ve öngörücü/tahmin edici olarak iki kategori olarak sınıflandırılır. Tanımlayıcı/açıklayıcı girdi verilerinin tüm özelliklerini anlatan tekniktir. Tahmin edici tekniklerde ise girdi verilerinden çıkarımlar yaparak gizli veya bilinmeyen veriler için tahminlerde bulunan bir tekniktir. Farklı teknikler şu şekildedir:

3.1.1.1. Veri sınıflandırması

Sınıflandırma, tanımlı ayırt edici sınıflar arasında veriyi dağıtma işlemidir. Bu yöntemle oluşturulan modeller ile yeni girdi verilerinin sınıfını tahmin etmek ve tanımlamak için kullanılır. Karar ağaçları, Naive Bayes, k-NN gibi yöntemler örnek olarak verilebilir.

3.1.1.2. Kümeleme analizi

Kümeleme, bir veri kümesini (kayıtlar / vektörler / örnekler / nesneler) önceden belirlenmiş belirli benzerliklere dayanarak birkaç gruba (kümeler) bölme yöntemidir. [27]

İki nesne arasındaki benzerliği araştırmak için, verilerin niteliğine ve amaçlanan amaca bağlı olarak seçilen benzerlik ölçüleri kullanılır.

3.1.1.3. Birliktelik kuralları

Olayların birlikte gerçekleşme durumlarını çözümleyen veri madenciliği yöntemlerine birliktelik kuralları denir.

3.1.1.4. Sıralı kalıpları bulma

Tıpta, bir hastalığı teşhis ederken, belirti kayıtları gerçek zamanlı olarak analiz edilir ve sıralı belirtiler için aslında bu yöntem kullanılıyor demektir.[27]

3.1.1.5. Regresyon

Regresyon analizi, iki ya da daha çok değişken arasındaki ilişkiyi ölçmek için kullanılan analiz metodudur. Veriyi gerçek değerli bir fonksiyona dönüştürme – eğri uydurma işlemidir.

3.1.1.6. Aykırı verileri belirleme

Adından da anlaşılacağı gibi, sapmaların / anormalliklerin / aykırı değerlerin tespit edilmesidir. Normal davranıştan farklı olarak önemli sapmaların keşfedilmesidir.

3.2. Sosyal Medya Analitiği

İnsanların birbirleriyle iletişim ve bağlantı kurma isteği teknolojinin ve internetin doğuşundan da önce gelen sosyal bir ihtiyaçtır. Geçmişte iletişim yüz yüze, mektup, telekomünikasyon teknolojileri ile gerçekleştiriliyordu. Web ve İnternet teknolojilerin gelişmesiyle bu iletişim farklı bir hâl aldı. Günümüzde hâlâ web tabanlı sosyal uygulamalar gelişmeye devam etmekte ve veri üretmeye sürekli devam etmektedir. Tüm

bu veriler kullanıcı görüşleri, bağlantıları ve diğerleri üzerindeki etkileri konusunda bilgiler sunabilmektedir. Bu nedenle bu değerli bilgileri elde edebilmek için bu verileri kullanmak doğaldır.[29]

Sosyal medyanın ortaya çıkışı iletişim ve bilgi aktarımını temelden değiştirdi. İnsanlar aktif olarak sosyal medyada bilgi oluşturur, paylaşır, bilgi alışverişi yapar ve ağ oluşturur. Günümüzde, insanlar tweet atarak, görüntü paylaşarak, sosyal sitelere yorum yaparak fikirlerini ve duygularını ifade ediyorlar. Kullanıcı tarafından oluşturulan çok miktarda içerik, sosyal medya verilerinden metin analizi yöntemleriyle bilgileri araştırmak ve çıkarmak için sosyal davranışı anlama ve sosyal olarak akıllı sistemler oluşturma fırsatları sunar. Metin madenciliği, metin verilerinden bilgi, anlamlı içerikler, konular, kelime ilişkileri ve örüntüler bulmayı amaçlamaktadır. [30]

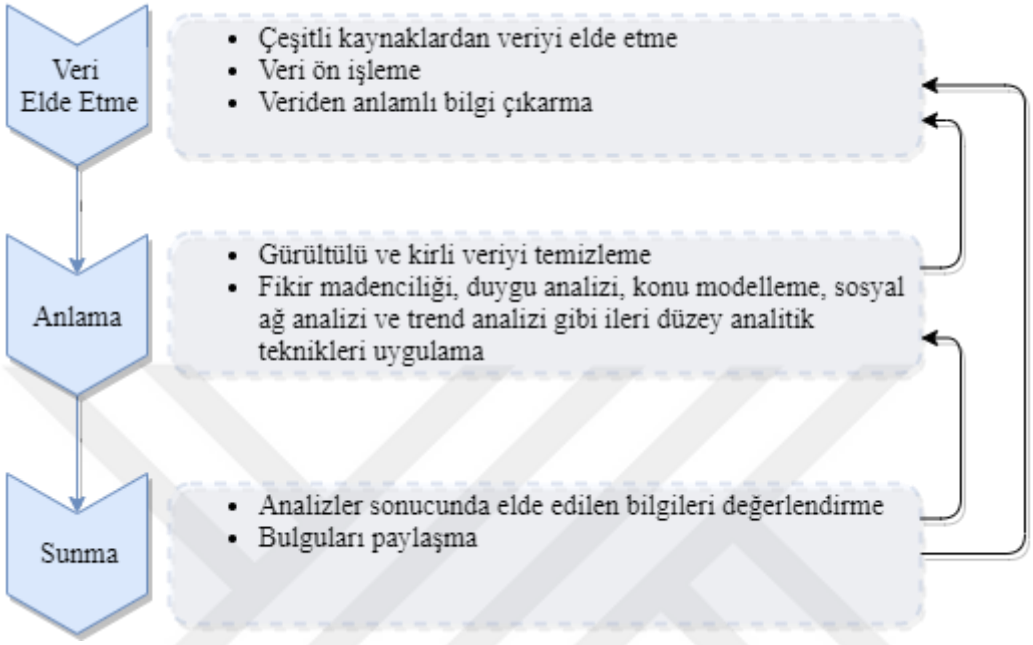
Sosyal medyanın dört genel nedenden dolayı kullanılabileceğini açıklanmaktadır: 1.si iletişim için, 2.si işbirliği için, 3.sü eğitim için ve 4.sü eğlence içindir.[31]

Geleneksel medyada hangi bilgilerin üretileceği bir takım otoriteler tarafından belirlenmektedir. Ancak sosyal medyada bu şekilde değildir. Kullanıcıların hem tüketici hem üretici rolündedir. Dolayısıyla bireylerin ürettiği her içerik, fikir ve veri tahmin edilemez büyüklükte olmaktadır.

Web 2.0'daki yeni yeniliklerden biri, Twitter gibi dünyanın dört bir yanından milyonlarca kullanıcıyla mikroblog uygulamalarıdır. Bu tür uygulamalar bir fikir paylaşımı ve bilgi yayma aracı ve aynı zamanda fikirleri kullanıcılar arasında yaymak için bir yol haline gelmiştir.[32]

Sosyal medya analitiği, etkili karar almayı desteklemek için sosyal medya verilerinin toplanmasını, analizini ve yorumlanmasını içerir. E-ticarette, müşteri ve pazar araştırmasında, tedarik zincirinde ve ürün geliştirmede, e-devlet ve politikada, satış sonrası ve müşteri ilişkilerinde, marka bilinirliği ve pazarlamada yaygın olarak kullanılmaktadır.[33]

Sosyal medyanın büyümesiyle birlikte toplumun fikirlerini anlamak daha önemli hale geldi. Elde edilen verilerin analizlerle anlamlı hale getirilmesi süreci aşağıda Şekil 2 [34]'de gösterilmiştir.



Şekil 2. Sosyal Medya Analitiği Süreci

3.2.1. Sosyal Medya Kavramı

İnternet teknolojisinin ve özellikle Web 2.0 kavramının ortaya çıkmasıyla yaygınlaşan sosyal medyanın tarihçesine bakıldığında ilk olarak 1979 yılında Jim Ellis ve Tom Truscott tarafından kurulan Usenet ortaya çıkmıştır. Sosyal medyanın günümüz kullanımına benzer şekilde meydana gelişi 1989 yılında Bruce ve Susan Alberson tarafından kurulan Open Diary Web sitesi ile olmuştur. [35] Bazı kaynaklara göre sosyal medyanın ilk doğuşu 1971 yılında ARPANET ile ilk ağ bağlantısının kurulması ve ilk e-posta gönderilmesi gösterilmektedir. 90'lı yıllarda ise kullanıcı sayılarında hızlı bir artış meydana gelmiştir.

Tarihten günümüze, sosyal medyada fazla bir çeşitlilik meydana gelmiştir. Bazılarını sıralayacak olursak; Linkedin 2003 yılında; MySpace, Facebook ve Flickr 2004 yılında;

Reddit ve Youtube 2005 yılında; Twitter 2006; Tumblr 2007; Foursquare ve Whatsapp 2009 yılında; Instagram ve Pinterest 2010 yılında; Google + ve Snapchat 2012 ylında; Vine ve Tinder 2012 yılında; Medium 2013; Periscope ve Scorp 2015 yılında; TikTok 2016 yılında doğarak bugünlerde sosyal medya platformları ailesine çeşitlilik kazandırmıştır. Her birinin uluslararası çapta milyonlarca ve hatta bazılarının milyarları bulan kullanıcısı bulunmaktadır.

Yapılan bir tanıma göre sosyal paylaşım ağları bireylere, genele tam ya da yarı açık profiller oluşturma zemini sunan, bilgileri paylaşacakları kişileri belirleyip başkalarının da aynı bilgileri görmelerine izin vermeyi sağlayan web tabanlı servislerdir. [36]

Sosyal medyayı en genel anlamda tanımlayacak olursak; bireylerin zaman ve mekân sınırlamalarına tabi olmaksızın paylaşımda ve tartışmada bulunabildikleri internet tabanlı iletişim şekli olarak tanımlamak mümkündür. [37]

Sosyal medya, katılımcıların sosyal medya platformları aracılığıyla paylaştığı video, fotoğraf, ses dosyaları, linkler, bloglar, vloglar gibi sosyal ağları oluşturan unsurları içinde bulunduran ortamlardır.[38]

Geleneksel medya olarak adlandırılan televizyon, radyo, dergi, broşür, film vb. medya araçları tek taraflı yayın ve bilgi akışına sahiptir. Bu tarz medya araçlarında kullanıcı konumundaki kişiler sadece izleyici ya da dinleyici durumunda olurken kişilere herhangi bir müdahale ya da paylaşım olanağı sunulmamaktadır. Geleneksel medya üreticileri aktif durumda iken tüketici durumunda olan kullanıcılar pasif durumda kalmaktadır. Sosyal medyaya baktığımızda bu durum geleneksel medyaya göre bambaşkadır. Yayın kaynağı artık medya kuruluşları değil aynı zamanda kullanıcıların da kendisidir. Dolayısıyla sosyal medya geleneksel medyaya göre daha güncel ve dinamiktir. Aynı zamanda sosyal medya maliyet açısından çok daha az maliyetlidir. Sosyal medya daha katılımcı bir mecradır. Bundan dolayıdır ki, sosyal medya büyük bir veri kaynağıdır.

Sosyal medya yukarıda sayılan özellikleri de düşünüldüğünde insanların hayatı için adeta vazgeçilmez bir unsur hâline gelmiştir. Kişilerin paylaşım yapma, aktif olma, beğeni kazanma, kendilerini ifade etme gibi birçok sosyolojik ve psikolojik boyutları da bulunmaktadır.

Sosyal medyada bir kategorilendirme yapacak olursak aşağıdaki gibi bir liste yapacak olursak aşağıdaki tabloyu düşünebiliriz. [39]

Tablo 1. Sosyal Medya Kategorileri

KATEGORİ	AÇIKLAMA
Sosyal Ağ	Sosyal ağ, sosyal medyanın kendisi değil aslında bir kategorisidir. Sosyal ağ çevrimdışı dünyada da var olan bir ilişki dünyasıdır. Kişilerin kendi veya ilgi alanlarıyla ilgili bilgileri arkadaş, profesyonel çevre veya diğerleriyle paylaşmasına olanak sağlar. Bu kategori, bugün sosyal medyada bağlantı kurmak, paylaşmak, eğitmek, etkileşim kurmak ve güven inşa etmek için kullandığımız birçok platformu tartışıyor. Örneğin; Linkedin, Facebook, MySpace vb.
Fotoğraf Paylaşımı	Bu kategoride, fotoğraf paylaşımını yönetmeye yardımcı olan araçlar listenebilir. Şirketlerin müşterilerine işlerin görüntülerini paylaşmaları için kullandığı araçlar da bu kategoridedir. İnsanların da gezdiği yerleri ve yaşadığı anları paylaşması bu kategori altındadır. Örneğin; Flickr, Picasa, Instagram vb.
Ses	Cebimize giren bir alet ve kulaklıklar sayesinde saatlerce ses dinlemeyi kolaylaştıran araçlardır. Akademik dersler, turistik geziler, şarkı yayını ve sesli kitaplar vb. bu kategori altındadır. Örneğin; iTunes, Spotify, SoundCloud vb.
Video	Bir resmin bin kelimeye bedel olduğu söylenir. Dolayısıyla videonun değeri kat kat fazla olabilir. Artık televizyonların

	yerini alan video yayınlarına cep telefonumuz sayesinde her an ulaşılabilmektedir ve bilgi aktarımı için de ilk tercih edilen seçenektir.
Mikroblog	Özellikle 140 karakterlik paylaşımları seviyoruz çünkü saniyeler içinde okuyup anlayabiliyoruz. Ayrıca son dakikaları da takip edip insanlarını yorumlarını okuyabildiğimiz sosyal araçlardır.
Canlı Yayın	Videonun canlı olarak yayınlanmasıdır. Ücretsiz olarak kendi televizyon şovunuzu yapabileceğiniz araçlardır.
Sanal Dünyalar	Sanal Gerçeklik, basit ifadeyle gerçekte var olmayan bir dünyayı varmış gibi yansıtmak olarak tanımlanabilir.
Oyun	Cep telefonu oyun uygulamaları, akıllı telefonlar için en hızlı büyüyen uygulama kategorilerinden biridir. Birçok kişinin aynı anda çevrimiçi olarak buluştuğu mecralardır. Ek olarak günümüzde e-spor adında yeni bir kavram oluşmuştur.
RSS ve Toplayıcılar	Web sitesindeki bir RSS özelliği istenildiğinde herhangi bir güncellemede kişilere otomatik olarak bildirim göndermektedir.
Arama	İnternet araması, internetin en önemli işlevlerindendir. Arama optimizasyonu da internet aramasında en üstlerde gözükmek, fark edilebilmek ve ilgi çekmek için oldukça önemlidir.

Mobil	İnsanların anlık olarak internete bağlanması, e-posta alıp göndermesini, fotoğraf-video-ses paylaşımlarını takip etmesini kısacası tüm sosyal medya kategorilerini sürekli takip edebilmesini sağlayan araçlardır.
Etkileşim	Bu kategorideki araçlar, kişiler arası iletişimi ve işbirliğini kolaylaştırır.

3.2.2. Sosyal Medya İstatistikleri

"We Are Social" ve "Hootsuite"in yaptıkları ve her yıl yayınladıkları araştırma sosyal medyadaki kullanım artışını gözler önüne sermektedir. [40] Araştırmanın Ocak 2020 verilerine göre Dünya geneli sonuçlar şu şekildedir;

- Nüfusun %67'si (5.19 milyar kişi) cep telefonuna sahiptir.
- %59'u (4.54 milyar kişi) internet kullanıcısıdır.
- %49'u (3.8 milyar kişi) sosyal medya kullanıcısıdır.
- Sosyal medya kullanımında bir önceki yıla göre değişim ise 321 milyon kişi ile %9,2 artış olarak gözlenmiştir.
- Paylaşımların %16,5'i Çince, %4,9'u da İngilizcedir. Türkçe paylaşımın oranı ise %1'dir.
- Sosyal medya kullanım oranı en yüksek olan ülke %99 ile Birleşik Arap Emirlikleri'dir.
- Dünya'daki ortalamaya göre insanlar internette günlük 6 saat 43 dakika vakit geçirmektedir.
- Sosyal medyada vakit geçirme süresi 2 saat 24 dakikadır.
- Her bir internet kullanıcısı başına ortalama 8,6 sosyal medya hesabı bulunmaktadır.
- İnternet kullanıcılarının %74'ü internetten bir ürün satın almışır.

- Yıllık ortalama 73,72 milyar dolar online alışveriş yapılmaktadır. Bu da akıllı telefonda barındıran ev başına yıllık 550 dolarlık bir tutara denk gelmektedir.

Dünya genelinde en çok kullanılan ilk 15 sosyal medya platformu sırasıyla;

1. Facebook
2. Youtube
3. Whatsapp
4. Facebook Messenger
5. Weixin / WeChat
6. Instagram
7. Douyin / TikTok
8. QQ
9. Qzone
10. Sina Webio
11. Reddit
12. Snapchat
13. Twitter
14. Pinterest
15. Kuaishou

Türkiye'deki durum ise [41] ;

- Nüfusun %92'si cep telefonuna sahiptir.
- %74'ü internet kullanıcısı
- %64'ü sosyal medya kullanıcısıdır. Bu oran dünya ortalamasının üstündedir ve dünyada 28. sıradadır. (nüfusa göre oran)
- Sosyal medya kullanım oranı önceki yıla göre %4,2 artış göstermiştir.
- Sosyal medya kullanıcılarının yaş ortalaması 31,5 olarak hesaplanmıştır.
- Kadınların oranı %51 ve erkeklerin oranı %49'dur.
- Türkiye'de internet kullanım süresi günlük ortalama 7 saat 29 dakika iken sosyal medyada bulunma süresi 2 saat 51 dakika olarak saptanmıştır.
- Televizyon izlemek için harcanan günlük ortalama süre 3 saat 4 dakikadır.
- Her bir internet kullanıcısı başına ortalama 9,1 sosyal medya hesabı bulunmaktadır.
- İnternet kullanıcılarının %63'ü internetten herhangi bir ürün alışverişi yapmıştır.

Türkiye genelinde en çok kullanılan sosyal medya platformları sırasıyla;

1. Youtube
2. Instagram
3. Whatsapp
4. Facebook
5. Twitter
6. Facebook Messenger
7. Pinterest
8. Linkedin
9. Snapchat
10. TikTok
11. Skype
12. Twitch
13. Tumlbr
14. Reddit
15. Wechat
16. Line

Türkiye'nin Twitter profilini incelediğimizde 11,8 milyon kişi bulunmaktadır ve %78,4'ü erkek %21,6'sı da kadınlardan oluşmaktadır. [41]

3.2.3. Sosyal Medyanın Hayattaki Yeri ve Sosyal Medya Analitiğinin Önemi

Teknolojinin ve internetin hızla hayatımıza girmesi, büyümesi ve yayılması insanların birbiriyle irtibatta kalma işlemini kolaylaştırmaktadır. Fikir, düşünce ve yorumlarını ve hatta neredeyse yaptığı her şeyi ve yaşadığı her anı paylaşan insanların sayısı da çoğalmaktadır. Bir düşüncenin ne kadar büyük veya önemli olduğu ancak başkaları tarafından doğru bir şekilde anlaşılınca önem kazanır.

Sosyal medya Dünya genelinde farklı insanlar tarafından farklı konular hakkındaki görüşlerini ifade etmek için de kullanılan bir yerdir. Bu nedenle insanların fikirlerini öğrenmek açısında çok değerli bir kaynaktır. Çok sayıda paylaşım içerir ve büyük bir veriye sahiptir. Kullanıcılar/İnsanlar din, dil, ırk, kültür, yaş, cinsiyet, eğitim vb. düzeylere göre değişmektedir. Bu sayede birçok ülkeden benzer konular için veri toplamak mümkündür.

Sosyal medyanın bu kadar çok kullanılması ve dolayısıyla sürekli dolaşan fikir akışının da anlamlı hâle gelebilmesi için sosyal medyada veri madenciliği bir diğer adıyla sosyal medya analitiği kavramı da önemli hâle gelmektedir.

Hızla gelişen ve büyüyen sosyal medya hayatımızın her alanında yer almaktadır. İş hayatında da önemli hale gelen sosyal medya için işletmeler de ilgi göstermiştir ve göstermelidir.

3.2.3.1. Pazarlama ve Finans

Sosyal medya pazarlaması, sosyal medya vasıtasıyla mal veya hizmetin satış ve tanıtım faaliyetlerini potansiyel ve mevcut tüketicilere ulaştırmasıdır. Bunun için işletmeler; farkındalık sağlama, tüketicilerin ihtiyaçlarını öngörebilme, beklentilerini belirleme, tüketicilerle bağ kurma ve güven oluşturma gibi tüm faaliyetleri en kolay ve hızlı şekilde sosyal medyayı iyi bir şekilde kullanarak yapabilmektedir. [42]

Günümüzde sosyal medya işletmeler için önemli ve gerekli konulardan biri hâline gelmiştir. Yöneticiler, Facebook ve Twitter gibi sosyal medya platformlarının işletmenin karlılığını nasıl arttırabileceği konusunda arayışlar içindedir. İşletmeler genellikle tüketicilerin aktif olduğu platformlara yönelmektedir. Kısacası artık işletmeler kolaylığı ve etkinliği dolayısıyla pazarlama faaliyetlerini yürütürken sosyal medya platformlarını kullanmaktadır.[35]

Sosyal medya pazarlamasının, geleneksel pazarlama yöntemlerine göre birçok faydası da mevcuttur. İlk olarak sosyal medya pazarlamasının daha az maliyetli olduğu söylenebilmektedir. Ayrıca özel filtreleme imkânları ile istenilen hedef kitlesine ulaşma kolaylığı açısında sosyal medya daha avantajlıdır. Sosyal medya zamandan tasarruf da sağlamaktadır. Esnek bir şekilde günün 24 saati kullanılabilmektedir. Sosyal medyada internet sitelerine erişim ve yönlendirme avantajı mevcuttur. Aktif sosyal medya hesaplarına sahip firmalar arama motorlarında üst sıralarda yer alacaktır. Sosyal medyanın getirdiği çift taraflı iletişim ve etkileşim sayesinde tüketicilerle daha iyi bir iletişim kurma, markalarını daha iyi tanıtma ve müşteri bağı kurma avantajları da mevcuttur.[43]

Facebook, Instagram, Twitter, Youtube gibi popüler sosyal medya araçlarında tüm bu pazarlama teknikleri sanal olarak çok kolay bir şekilde yapılabilmektedir. Dünya genelinde bu kavram “social media marketing” olarak literatüre girmiştir.

Konuyu bir örnekle biraz zenginleştirmek istersek; Mavi Jeans’in sosyal medya ile pazarlama başarısındaki anlatan çalışmayı gösterebiliriz. Firma sosyal medya pazarlama stratejilerini kullanarak, interaktif iletişim pazar nabzını tutma ve trendleri yönetme konularında beceriler kazandı. Ayrıca firmanın özgünlük ve yenilikçilik konularında da sosyal medyanın yeni nesil pazarlama için güçlü bir araç olduğu sonucuna varmıştır. Çalışmada firmanın bu aracı etkin kullanarak diğer firmalara da örnek olduğu söylenmiştir. Amerikan Stylophane şirketinin

araştırmasına göre dünyadaki tüm Jean markaları arasında 8. sırada müşterileriyle etkileşimde ise 1. sırada olduğu açıklanmıştır. Mavi Jeans, Instagram'ın yanı sıra Faceboook, Twitter, FriendFeed, Youtube, Flickr ve Foursquare'i aktif olarak kullanıyor. Firma televizyon reklamındaki oyuncunun hapşıran kız arkadaşına "çok yaşa" yerine "çok sev" demesinin çok sevilmesi üzerine bunu sosyal medyada bir trend olarak kullandı. İnsanların bu etiketle tweetler atmasını sağladı ve bunları bazı mağazalarında yayınlayarak en iyi tweeti atan kişilere yeni yıl hediyesi kazanmasına yönelik bir kampanya başlattı. Firma kampanya sürecinde takipçi sayısını %20 arttırmıştır. [44]

Finans dünyası günümüzde son dakika gelişmeleri, haberler ya da eğilimlere göre hızlı değişim yaşayan sektörlerdendir. Yapılan bir araştırmada [45] borsa gidişatını tahmin edebilmek için sosyal medyada atılan tweetlerden elde edilen duyguların analiz edilerek DJIA kapanış değerlerini etkileyip etkilemediği araştırılmıştır. Kamusal ruh halinin twitterdaki metinlerle tespit edilip bu ruh hallerinin de borsa değerlerini ne yönde etkileyeceği %86,7 gibi büyük bir doğruluk oranıyla tahmin edilebildiği tespit edilmiştir.

3.2.3.2. İnsan Kaynakları

İnsan kaynakları şirketleri veya personel alımı yapacak olan kurumlar sosyal medyayı artık daha fazla kullanmaktadır. Bullhorn şirketinin araştırmasına göre ABD'de işe alım süreçlerinde şirketlerin %98'i sosyal medyayı kontrol etmektedir. Linkedin %98, Facebook %51, Twitter %49 oranında kontrol edilmektedir. [46]

Adayların şirketlere verdikleri özgeçmişlere göre sosyal medyadan tespit edilen özgeçmiş verileri daha gerçekçi veriler olduğu saptanmıştır. Ayrıca personel adayının sosyal medyadaki tutumu ve davranışlarına bakılarak karakteri hakkında ipuçları elde edilebilir. Bu da işe alım sürecinde şirketin, aday hakkında vereceği kararda büyük etkiye sahip olacaktır.

Sosyal medya, şirketlerin ilgili pozisyon için en uygun adayı bulabilme kolaylığını arttırmaktadır. Sosyal medya aracığıyla daha geniş bir özgeçmiş/aday havuzuna kavuşmak mümkündür. Ayrıca ilanların paylaşılması ve yayılması konusunda en az maliyetli yöntemdir.

Şirketler açısından faydalar sağladığı gibi, iş arayanlar açısından da faydalar bulunmaktadır. İnsanlar daha hızlı ve aktif şekilde kendisi için uygun olduğunu düşündüğü ilanlara ulaşabilmektedir ve firma ile iletişime geçebilmektedir.

Yapılan araştırmaya göre sosyal medyanın işe alım süreçlerinde de etkisi aşağıda listelenmiştir; [47]

- Aday havuzu oluşturulması konusunda olumlu etkisi olduğu,
- İşe alım performansında olumlu etkisi olduğu,
- Aday havuzundaki personelin kalitesinde artış olduğu,
- Aday havuzundaki personelin çeşitliliğinde artış olduğu,
- İşe alım maliyetlerinde ciddi bir tasarruf sağlandığı,
- Hedef kitleden gelen başvuru sayısında artış olduğu,
- Referans araştırması ve kontrol süreçlerinde kolaylıklar sağladığı saptanmıştır.

3.2.3.3. Turizm

Turizm sektörü, Türkiye'de hızla büyüyen ve önemli gelir kaynağı oluşturan sektörlerin başında gelmekledir. Ülkelerin tarihi, kültürel ve manevi değerlerini gezip görmenin yanı sıra tatil beldelerini keşfetmek de turizmin önemli parçalarıdır. İnsanlar bu tarz yerleri keşfetmek için interneti aktif olarak kullanmaktadır. Hem keşfetmek hem de yapılan yorumları okumak ve edinilen izlenimleri görmek için internetin ve sosyal medyanın kullanılması kaçınılmazdır.

Turistik faaliyetlere katılan kişiler gezilerindeki izlenimlerini, çektikleri fotoğrafları ve videoları paylaşmak için yorumlarını da katarak sosyal medyayı kullanmaktadır. Yapılan araştırmaya göre katılımcıların %48'ine yakını turistik seyahati hakkındaki iyi ya da kötü görüşlerini sosyal medya üzerinden paylaşacaklarını ifade etmiştir. [48]

33 ülkede yapılan başka bir çalışmada insanların tatil ve seyahat yapmak üzere karar verecekken başvurduğu veya etkilendiği konular araştırılmıştır. İnsanlar %27'lik oranla en fazla arkadaş önerilerini dikkate alırken; %24'lük bir oranla internet ikinci sırada yer almıştır. [49]

Son yıllarda artış gösteren Kars Doğu Ekspresi ve Kars turizmine yönelik yapılan bir araştırmada katılımcıların %51 Doğu Ekspresi'nin ilk kez sosyal medya aracılığına ile keşfettiğini bildirmiştir. [50] Bu orandan anlaşılacağı üzere sosyal medya; keşfedilmeyi bekleyen yeni turizm bölgeleri için güçlü bir tanıtım aracı olduğuna dair güzel bir örnek olarak karşımıza çıkmaktadır.

Kastamonu İl Kültür ve Turizm Müdürlüğü için yapılan bir çalışmada da aşağıdaki sonuçlara varılmış ve öneriler sunulmuştur; [51]

1. Sosyal medya hesabının uzman kişiler tarafından yönetilmesi gerektiği,
2. Sosyal medya ortamlarında hesaplar açılarak aktif şekilde kullanılması gerektiği,
3. Hesapların sadece bilgilendirme amaçlı değil tanıtım amacıyla da kullanılması gerektiği ve vatandaşlarla soru-cevap ile istek-şikayet almak gibi etkileşim hâlinde olunması gerektiği,
4. İlin tarihi eserlerini, sanat eserlerini, müzeleri, mimari yapılarını, mutfağını, doğal yerleri vb. paylaşımların yapılması gerektiği,
5. Takipçi sayısının artması için vatandaşlara yönelik kampanyaların da yapılması gerektiği,
6. Paylaşımların Türkçe ve İngilizce olarak yapılması gerektiği,

çalışmada belirtilen maddelerden bazılarıdır.

Dünyadaki örneklerine bakacak olursak; yapılan bir çalışmaya göre turizmde ilk sıralarda yer alan ülkeler arasında sosyal medyayı en yoğun ve etkin kullananlar; İspanya, Fransa ve İngiltere olduğu tespit edilmiştir. Fransa'nın 6 sosyal medya aracı kullandığı ve mobil tanıtım uygulamalarına da sahip olduğu belirtilmiştir. [52]

3.2.3.4. Spor Sektörü

Tüm dünyada ve ülkemizde özellikle Futbol sektörü iyi takip edilen alanlardandır. Sosyal medyada yüksek takipçili hesaplara baktığımızda spor kulüpleri, teknik direktörleri sporcuları ilk sıralarda gelmektedir. Sosyal medya spor kulüpleri için taraftarla iletişim kurma yöntemlerinden biridir. Hatta son yıllarda ilginç transfer haberleri paylaşımları ile yüksek etkileşim alan kulüpler bu alanda da rekabete girmiş bulunmaktadır. Neredeyse bütün kulüplerin Facebook, Twitter, Instagram ve Youtube gibi popüler sosyal medya mecralarında hesapları bulunmaktadır.

Geçmişten günümüze bakıldığında spor sektörü gittikçe büyüyen ve milyarlarca doların döndüğü bir sektör hâline gelmiştir. Bunun en önemli sebepleri ise spor yapmaktan veya sağlıklı yaşamdan ziyade eğlence, hobi veya fanatizmdir. Dünyada televizyonlarda en çok izlenen programlara baktığımızda spor programları ilk sıralarda yer almaktadır. Örneğin 2003 yılında Bu kadar takip edilen ve büyük paraların harcandığı bu endüstri de spor pazarlaması olarak anılabilmektedir.

Dünya'da spor pazarlaması ilk olarak 1978 yılında Advertising Age Magazin konuşulmaya başlanmıştır. Bu kavram Türkiye'de ise 8-9 Nisan 1991 yılında yapılan 'Sporda Marketing Semineri'nde konuşulmuştur. Bu kavramlar spordaki ekonomik desteklerin, sponsorlukların artmasına ve spor sektörünün büyümesine yardımcı olmuştur. [53] Spor pazarlaması; pazarlama konseptlerinin spor ürün ve hizmetlerine uygulanması ve spor dışı ürünlerin sporla da ilişkilendirilerek pazarlanmasına denilmektedir. Bu da genel pazarlama uygulamalarının sporla ilgili ürün ve hizmetlere uygulayarak veya diğer tüketici ve endüstriyel ürün ve hizmetlerin spor yoluyla pazarlanmasıyla yapılmaktadır.[54]

2012 yılında yapılan bir araştırmada Türkiye'deki dört büyük kulübün sosyal medyada spor pazarlaması konusundaki tabloya aşağıda yer verilmiştir. [55]

Tablo 2. Dört Büyük Futbol Takımının Pazarlama Aracı Olarak Sosyal Medya Kullanımı

Ürün Tanıtımı	Beşiktaş	Fenerbahçe	Galatasaray	Trabzonspor
Yeni ürünleri tanıtma	Evet	Evet	Evet	Evet
Mevcut ürünleri canlandırmak veya yeniden konumlandırmak	Hayır	Evet	Hayır	Hayır
Eski ürünlerin yeni faydalarını anlatmak	Hayır	Evet	Hayır	Hayır
İnsanları ürünlere dâhil etmek	Evet	Evet	Evet	Evet
Çevrimiçi ürünlerle müşterilerle etkileşim kurmak	Evet	Evet	Evet	Evet
Bir ürün kategorisine ilgi oluşturup devam ettirmek	Evet	Evet	Evet	Evet
Reklam Desteği	Beşiktaş	Fenerbahçe	Galatasaray	Trabzonspor
Reklamın erişimini genişletmek	Evet	Evet	Evet	Evet
Reklamcılığa karşı tüketici	Hayır	Hayır	Hayır	Hayır

direncini yok etme				
Ticari karmaşadan kurtulmak	Evet	Evet	Evet	Evet
Reklam aralarından önce haber yapmak	Evet	Evet	Evet	Evet
Reklamları haber değeri haline getirmek	Evet	Evet	Evet	Evet
Mesajları güçlendirerek ve iddiaları meşrulaştırarak reklamı tamamlamak	Evet	Evet	Evet	Evet
Ek ürün avantajlarını bildirerek reklamları tamamlamak	Evet	Evet	Evet	Evet
Ziyaretçileri ticari web sitelerine çekmek	Evet	Evet	Evet	Evet
Ürün reklamının yapılmadığı medyada farkındalık kazanmak	Evet	Evet	Evet	Evet
Pazarlama Desteği	Beşiktaş	Fenerbahçe	Galatasaray	Trabzonspor
Pazarlama kavramlarını test etmek	Hayır	Hayır	Hayır	Hayır
Satış promosyon kampanyalarını güçlendirmek	Evet	Evet	Evet	Evet
Pazarlama programlarını yerel kitlelere uyarlamak	Hayır	Hayır	Hayır	Evet
Başlık sponsorlukları aracılığıyla marka bilinirliğini artırmak	Evet	Evet	Evet	Hayır
Tüketicilere ulaşmak için yeni	Evet	Evet	Evet	Evet

medya ve yeni yollar oluşturmak				
Satış Desteği	Beşiktaş	Fenerbahçe	Galatasaray	Trabzonspor
Dağıtım kazanmak	Evet	Evet	Evet	Evet
Mağaza trafiği oluşturmak	Evet	Evet	Evet	Evet
Satış sorguları oluşturmak	Hayır	Hayır	Hayır	Hayır
Satış ekibini motive etmek	Hayır	Hayır	Hayır	Hayır
Perakendeci desteği kazanmak	Hayır	Hayır	Hayır	Hayır

Tabloda görüldüğü üzere Türk futbol takımlarımızın sosyal medyada başarılı olduklarını ifade edebiliriz. Ayrıca bu sözü edilen çalışmanın 2012 yılında yapıldığı hatırlanacak olursa yıllar geçtikten sonra günümüzde daha da başarılı şekilde devam ettiklerini ifade edebiliriz.

Yine dört büyük spor kulübümüzün sosyal medyadaki istatistiklerini özetleyen tablo aşağıda gösterilmiştir.[56] *(12 Ekim 2020 tarihindeki durumu gösterilmiştir.)*

Tablo 3. Dört Büyük Spor Kulübü Sosyal Medya İstatistikleri

Facebook	**Galatasaray**	**Fenerbahçe**	**Beşiktaş**	**Trabzonspor**
Hayran Sayısı	12.187.402	9.044.952	5.654.109	941.514
Sayfa Hakkında Konuşanlar	81.297	282.667	50.405	19.588
Haftalık Artış	-3.851	-2.024	-1.840	-116
Haftalık Artış Oranı	%-0,03	%-0,02	%-0,03	%-0.01
Etkileşim Oranı	%0,06	%0,31	%0,13	%0,31
Twitter	**Galatasaray**	**Fenerbahçe**	**Beşiktaş**	**Trabzonspor**
Takipçi	9.528.986	7.661.300	4.325.935	1.779.472
Haftalık Artış	8.277	13.252	4.555	635
Haftalık Artış Oranı	%0,09	%0,17	%0,11	%0,04
Etkileşim Oranı	%0,08	&0,59	%0,25	%0,21
Instagram	**Galatasaray**	**Fenerbahçe**	**Beşiktaş**	**Trabzonspor**
Takipçi	9.164.892	5.787.204	3.508.801	786.995
Haftalık Artış	-1.724	12.382	147	742

Haftalık Artış Oranı	%-0,02	%0,21	%0,00	%0,09
Etkileşim Oranı	%1,93	%4,68	%3,69	%5,92

Bu sayılara bakıldığında ülkemizdeki spor sektörünün sosyal medya ayağının oldukça etkili ve yükselişte olduğunu göstermektedir. Sonuç olarak sosyal medyanın spor sektöründe ve spor pazarlamasında ne kadar gerekli ve ihtiyaç olduğu görülmektedir.

3.2.3.5. Kamu Sektörü

Kamu kurumları tanıtımlarını paylaşmak için Youtube'u, daha erişilebilir olmak için Facebook'u, kendilerine veya yaptıklarına dair bilgileri sunabilmek için Twitter gibi mikroblogları kullanmaya başlamışlardır. [57]

Pazarlama sektöründe kurum-müşteri arasındaki ilişkiye benzer şekilde devlet-vatandaş ilişkilerinde de sosyal medya yeni ufuklar açmıştır. Bilgi Edinme Hakkı Kanunu'nun yürürlüğe girmesiyle kamu yönetiminde halkla ilişkiler için internet kullanımı ilerlemiştir.

Şuanda sosyal medyada aktif olan kamu kurum ve kuruluşlarını sayacak olursak; Cumhurbaşkanlığı, Meclis Başkanlığı, Bakanlıklar, Bakanlıklara bağlı diğer kurumlar, Valilikler, Belediyeler gibi birçok kurumu sıralayabiliriz. Kamu yönetimi açısında ilk uygulama 1 Mart 2010 tarihinde Dışişleri Bakanlığı'nın Facebook ve Twitter sayfalarını açmasıyla başlamıştır. Bu kararın da Türkiye'nin uluslararası ortamda kendisini daha fazla kişiye ifade edebilme amacıyla faydalı olduğu ifade edilmiştir. [58]

Günümüzde ülkemizde Twitter'da en çok takipçiye sahip olan kişi Cumhurbaşkanımız Recep Tayyip Erdoğan iken Instagram'da ise Kamu ve Siyaset alanında baktığımız Sağlık Bakanı Dr. Fahrettin Koca ilk sıradadır. Bakan Koca Türkiye geneli Instagram'da tüm hesaplar arasında ise 9. Sıradadır. Bakan Fahrettin Koca'nın bu kadar yüksek takipçiye sahip olmasının sebebi ise hiç kuşkusuz Covid19 salgını için yaptığı düzenli bilgilendirmelerdir. Bu bilgilendirmeleri gerek yazılı, gerek görsel paylaşımı, gerek video paylaşımı gerek de canlı yayın olarak yapmıştır. Her akşam düzenli olarak yaptığı ve grafiklerle açıkladığı salgın durumu kendisini yüksek takipçili hesaplar arasına kısa süre dahil etmiştir.

Sosyal medya sayesinde anlık iletişim yoluyla haberleşme vatandaş-devlet arasındaki güven kavramını pekiştirebilmektedir. Vatandaş sosyal ağlardan başarılı bir hizmet aldığında

kuruma karşı olumlu düşünceler ve izlenimler ortaya çıkmaktadır. Web site ve sosyal ağların vatandaş-devlet ilişkisi açısında son derece önemli olduğu aşikârdır.[59] Ülkemizde e-devlet uygulaması başarılı bir şekilde devam etmektedir.

Kamu Denetçiliği Kurumu (KDK) için yapılan bir araştırmayı da bu bölümde ele alabiliriz. KDK 2012 yılında ulusal çapta kamu denetçisi (ombudsman) kurumudur. Kurumun resmi sosyal medya hesapları bulunmaktadır. Böyle bir kurumun başarılı olması için vatandaş açısından görünür ve erişilebilir olması gerekmektedir. KDK ne kadar saygın da olsa başvuru süreçleri ile ilgili bilgilendirmeler yapmalı, vatandaşlar açısında erişilebilir olmalı ve bunlar için tanıtım faaliyetleri yaparak farkındalık arttırılmalıdır. Bunun için sosyal medya vazgeçilmez bir araç olmalıdır. Kurum Facebook hesabını Haziran 2012, Twitter hesabını Temmuz 2014 tarihinde açmıştır. Kurumun gönderdiği tweetler incelendiğinde ortalama 3-4 retweet aldığı görülmüştür. [60]

3.2.3.6. Sağlık Sektörü

Sosyal medya hayatımızın her alanında önemli bir parça haline geldiği günümüzde hayati öneme sahip olan sağlık konusunda önemli bir sağlık enformasyon kaynağı haline gelmiştir.

Amerika Birleşik Devletleri Hastalık Kontrol ve Korunma Merkezi'nin yayınlamış olduğu el kitabında Youtube, Facebook, Twitter gibi sosyal medya mecralarının sağlık iletişimi alanında bir hayli kullanımının arttığı ve sağlık enformasyonu milyonlarca kişiye ulaşım noktasında büyük imkânlar sağlayan araçlar olduğunu ifade etmiştir. [61]

Sosyal medya sağlık alanında verdiği faydaları sıralayacak olursak;

- Sağlık ve korunmayla ilgili bilgilerin daha etkin ve hızlı şekilde yayılmasını sağlar.
- Bilginin daha geniş bir kitleye yayılmasını sağlar.
- İnsanları daha güvenli ve sağlıklı karar almalarına katkı sağlar.
- Bireylerin sağlık bilgisini istedikleri anda istedikleri yerde almasını sağlar.

Ancak sosyal medya tek taraflı yayın aracı olmadığı için ve herkesin bilgi paylaşma imkânına açık bir alan olduğu için ilgisiz kişilerin yaptığı yanlış ve yanıltıcı bilgiler de yer almaktadır. Bu bilgiler dolayısıyla zarar görme riski de epey fazladır. Yapılan araştırmalara sağlık konusunda sosyal medya çeşitli amaçlarla kullanılmaktadır. Bunlardan bazıları;

ilaçlarla ilgili bilgi almak, hastane ve doktor seçimi için araştırma yapmak, hastalıklarla ilgili bilgi almak, randevu almak içindir. [62]

Bu alanda yapılmış çalışmalardan biri de grip hastalığını tespit etmek için Twitter mesajlarını analiz ederek tahminlerde bulunmaya çalışan bir çalışmadır. Hastalıkla ilgili mesajları derleyerek analiz eden çalışma türlü algoritmalar geliştirerek çeşitli istatistiki verilere kaynak olmuştur. [63]

3.2.3.7. İstihbarat, Güvenlik ve Öngörüsel Polislik

Şirketlerin ihtiyaçları doğrultusunda istihbarat faaliyetleri yapmalarının yanında artık devletler de güvenlik amaçlı istihbarat faaliyetleri yapmaktadır. Bu yeni istihbarat kavramı "siber istihbarat", sosyal medyada yapılan istihbarat ise "sosyal medya istihbaratı" olarak adlandırılmaktadır.[64]

Sosyal ağların günlük yaşamdaki yansıması ile gerçek yaşamın kendisine dönüşen sanal dünyada bireylerin ve kurumların gerçek hayatta karşılaştıkları benzer tehditler vardır. Bireysel, kurumsal ve devlet güvenliği sorunları siber düzeyde ortaya çıkmıştır. Sosyal medya istihbaratı adı verilen yeni bir istihbarat türü tanımlanmış ve sosyal zekânın nasıl yapılacağı açıklanmıştır. Siber saldırıların yeni bir savaş türü olarak sınıflandırılabileceği de açıklanmaktadır, çünkü ulusal güvenliği tehdit etmektedirler ve uluslararası ilişkilerde tanımlanmış sınırları olmayan siber alanda meydana gelmektedirler. Dünyadaki Arap Baharı olaylarından sonra sosyal medya sitelerinin kitleler üzerindeki etkileri daha dikkatli bir şekilde incelenmeye başlandı ve araştırmacılar sosyal medyanın etkisini ve hükümetlerin ve devletlerin buna nasıl katıldığını incelediler.[65]

2013 TÜİK verilerinde Türkiye'deki ceza davalarına göre dokuz farklı anahtar kelime belirleyen çalışmada kelimeler şu şekilde seçilmiştir; "Bomba", "terör", "velayet", "saldırı", "gösteri", "kaçakçılık", "uyuşturucu", "fuhuş", "tecavüz". Çalışmada bir aylık dönemde Twitter'dan atılan 158.463 mesajın çekilmesiyle yapılmıştır. Metinlerde 2 milyondan fazla kelime ve 19 milyona yakın karakter vardır. Verilerin analizi, Türkçe için en çok kullanılan NLP kütüphanelerinden Zemberek-NLP kullanılmıştır. Kelimler köklere bölündüğünde 301.690 kök elde edilmiştir. Çalışmada aynı zamanda R Studio programı kullanılarak analiz yapılmıştır. "Bomba" kelimesi %24 ile en fazla kullanım oranına sahip kelime olmuştur. [65]

Sosyal medyanın hızlı bir şekilde kendisini kitle iletişimi için ucuz, kolay ve etkili bir araç olarak sundu. Bundan dolayı terör örgütleri için de kitlelerini genişletmek ve daha fazla insan çekmek için bir fırsat oluşmuş oldu. Bu da devletler veya güvenlik kuvvetleri için zorlukların oluşmasına sebep oldu. Terörist gruplar sosyal medyayı para toplamak için de kullanabilmektedirler. Son yıllarda IŞİD terör örgütü masum kişilere yaptıkları katliamları özellikle sosyal medyada paylaşmasıyla daha hızlı bir şekilde ün kazanmıştı. Twitter bu hesapları kapatmasına rağmen örgüt bir yenisi açtı ve devam etti. Twitter hızla kapatmaya devam ettikçe örgüt bu sefer Twitter kurucularından birini tehdit etmeye başlamıştı. Bu konuda metin madenciliği yöntemleriyle Dünya'da IŞİD'e karşı duyulan duyguların analiz edildiği bir çalışma mevcuttur. Hem Batı hem Doğu ülkelerindeki insanların terör örgütü hakkındaki görüşlerini belirten kelimeleri inceleyerek IŞİD'i terör örgütü olarak görüp görmediğini tespit etmeye çalışmıştır. Çeşitli yöntemlerle 4 Batı, 4 Doğu ülkesi olmak üzere 8 farklı ülkeden tweetler toplandı bunlarla bir veri tabanı oluşturuldu. Duygu analizi yöntemleriyle kelimeleri pozitif, negatif veya nötr olarak kategorilendirildi. Bu sayede her bir görüşün olumlu ya da olumsuz oldukları tespit edildi. Çalışma 8 ülkeden toplanan metinlerin neredeyse aynı kelimeleri içerdiğini gözler önüne sermektedir. En sık kullanılan negatif kelimeler ise; tehdit, saldırı, bomba ve korku olarak sıralanmıştır. Negatif kelime oranı toplam kelime sayısına göre %67-71 arasında olduğu belirlenmiştir. Toplanan kelimelere bakıldığında da ülkelerin IŞİD'i bir terör örgütü olarak gördüğü rahatlıkla söylenebilmektedir. Hatta haberlerin ana odağında Suriye, Türkiye, Rusya, Obama ve Fransa olduğu da çalışmadan anlaşılmaktadır. [66]

Öngörüsel ve tahmine dayalı güvenlik önlemleri konusunda ABD Gelişmiş İstihbarat Araştırma Projesi Ajansı'nın projesi olan EMBERS programı örnek gösterilebilir. Sivil huzursuzluk, hastalıklar, protestolar, salgınlar ve seçimler gibi kritik olayların tahmin edilmesinde kullanıldığı ifade edilmiştir. EMBERS Twitter verileri üzerinden de tahminlerde bulunmuştur. Güney Amerika coğrafyasına yönelik yapılan çalışmada günlük yaklaşık 19 GB boyutunca İspanyolca, Portekizce ve İngilizce Twitter mesajları üretmiştir. Daha sonra bu veriler coğrafi kodlama, duygu analizi, metin madenciliği gibi çalışmalarla analiz edildi. Yaklaşan sosyal etkinlikler öncesinde de sürekli benzer çalışmalar tekrar edildi. Çalışmada olumsuz kabul edilen 800 spesifik kelimeyi de araştırdı. Sistem genellikle olay türü başına 6-8 algoritma arasında birden çok makine öğrenmesi modeli geliştirdi. Modeller daha sonra en doğru tahmini oluşturacağı düşünülen bir şekilde modelleri birleştiren bir ana füzyon modülü tarafından optimize edildi. Ortalama olarak günde 50 uyarı üreten sistemin öngördüğü önemli olaylardan bazıları şu şekildedir; Paraguay Devlet Başkanı'nın görevden alınmasının ardından

protestolar, Brezilya Baharı (Brezilya'nın bazı şehirlerindeki gösteriler), Arjantin ve Şili'de hantavirüs salgınları ve Venezuelalı öğrencilerin yaygın protestolarıydı. Sistem beş yıl çalıştırıldı ve %65 oranında doğru tahminler verdiği tespit edilmiştir. [67]

Terörle ilgili haberlerin yayılması neticesinde sosyal medya üzerinden sivil yanıtın yaygın olması yadsınamaz bir gerçektir. 2008 Mumbai saldırıları ile ilgili haberlerin öncelikle vatandaşlar tarafından bulundukları yerde Twitter üzerinden paylaşması konuya örnek olarak verilebilir. Aynı şekilde 2009 Jakarta'daki bombalı terörist saldırısı örnek olarak verilebilir. Bir diğer örnek ise Hudson Nehri uçak kazasının son dakika olarak TwitPic üzerinden ilk görüntülerinin yayınlanması verilebilir. [68]

3.3. Metin Madenciliği ve Duygu Analizi

3.3.1. Metin Madenciliği

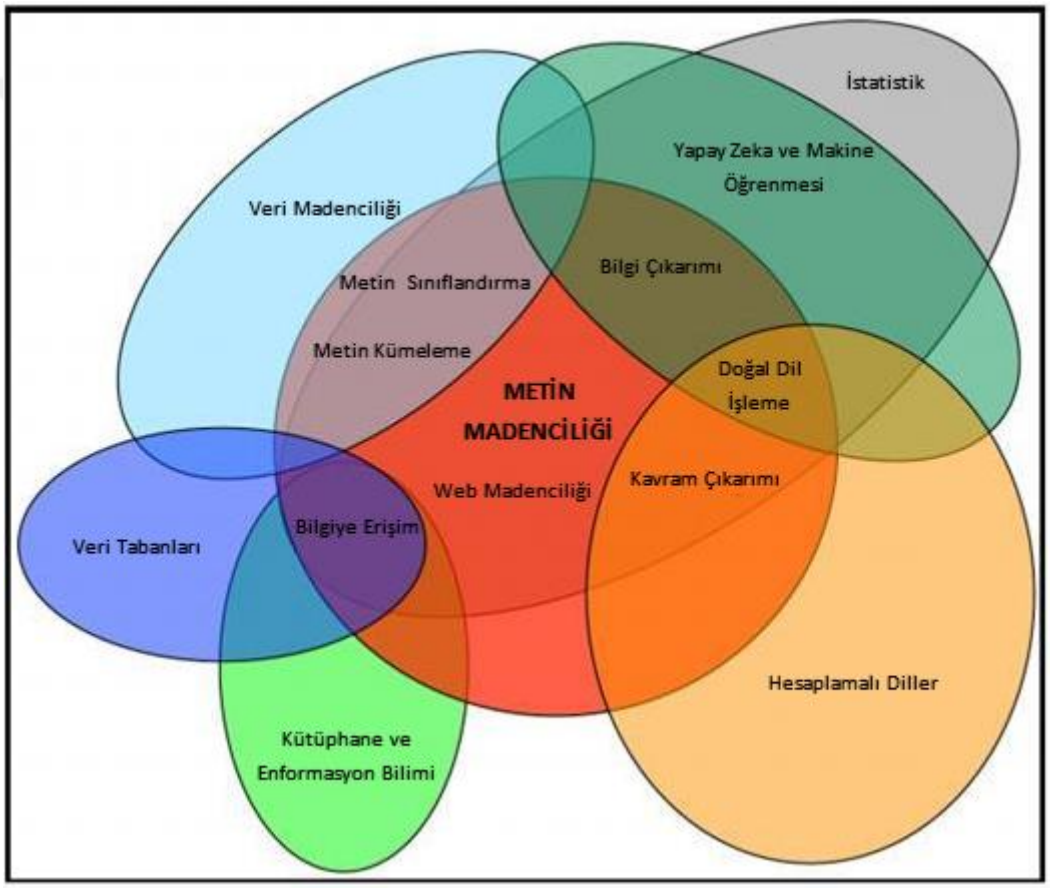
Basit ifadeyle metin madenciliği, veri madenciliği yönteminin bir alt dalı denilebilir ve günümüzde yaygın olan bir araştırma konusudur. Metin içerikli verileri kullanarak sonucunda anlamlı ve keşfedilmiş bilgiler oluşturan ve büyük bir veriye sahip olduğu için bilgisayar destekli yapılan bir çalışmadır.

Metin, kelimelerden meydana gelen yapılandırılmamış veriler olarak tanımlanır. Metni oluşturmak için tek tek dizelerin anlamlarını ve bunların dil bilgisi adı verilen kurallara göre kombinasyonunu gerektirir. Kaynak kodu veya matematiksel denklemler gibi yapay dillerle yazılmış sözcükler metin kapsamının dışındadır. Metin dediğimiz kavram; kelimeler, cümleler ve paragraflardan oluşturmaktadır. Kelime anlamlı en temel metin birimidir. İngilice'de durdurma sözcükleri olarak adlandırılan "a" veya "the" gibi dilbilgisi işlevleri için kullanılan etkisiz kelimelerin hiçbir anlamları yoktur. Bu nedenle metin ön işlemesinde bu tarz kelimeler hariç tutulmaktadır. (Türkçede bu kelimelere örnek olarak; "bu" , "şu", "ve", "ise", "ile" gibi kelimeler verilebilir.) Metin işlenmesinde çeşitli metin biçimleri mevcuttur. Popüler metin biçimleri uzantısı 'doc' olan MS Word dosyası, 'ppt' uzantısı olan MS PowerPoint dosyası ve uzantısı 'xls' olan MS Excel dosyası gibi MS Office yazılımı tarafından yapılanlardır. Metinler kullanıcılar arasında en popüler 'PDF' formatında karşıya yüklenerek ve indirilerek aktarılır. Bir e-posta mesajı da metin türü olarak değerlendirilebilir. İstenmeyen e-postaları filtreleme işlemi de metin madenciliği görevi olarak ele alınabilir. 'XML' de başka bir metin formatı

olarak düşünülebilir. Web tarayıcılarında biçim olarak HTML'ye dayalı olarak tasarlanan daha esnek web belgesi biçimidir. [69]

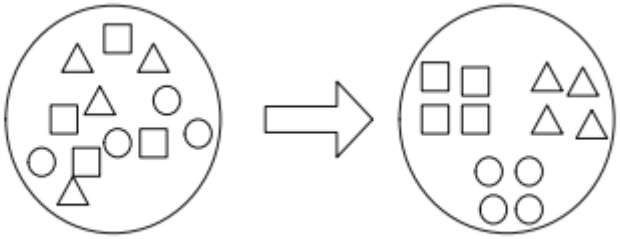
Metin madenciliği birçok bilim dalıyla kesişimi olan bütünleşik bir veri bilimidir. Şekil 3'teki Venn şemasında bu bilim dalları ve uygulama alanı metin madenciliği ile ilgili olarak gösterilmiştir. [70] Kavramların kısa açıklamaları da şu şekilde sıralanmıştır.

i. **Bilgiye Erişim (Information Retrieval / IR):** metin madenciliğinde ilk adımdır. Amacı ise kullanıcıların bilgi ihtiyaçlarını karşılayacak olan belgeleri bulmasına yardımcı olur.



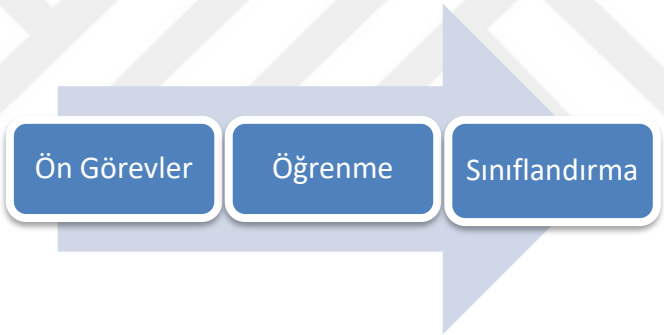
Şekil 3. Metin Madenciliği ve İlişkili Olduğu Alanlar

ii. **Metin Kümeleme:** Bir metin grubunu içerik tabanlı benzer alt gruplara ayırma sürecini ifade eder. Metin kümeleme büyük gruptaki metinler arasındaki benzerliklere bağlı olarak ilerler, benzerlik ölçütlerinin nasıl tanımlanacağı çok önemli bir görevdir. Metin kümele işlemini temsilen aşağıdaki Şekil 4'te özet gösterilmiştir.



Şekil 4. Metin Kümeleme

iii. **Metin Sınıflandırma:** Şekilde gösterilen üç adımın her metne önceden tanımlanmış kategorilerden birini veya birkaçını atama işlemi olarak tanımlanır. Kategori ağacını önceden tanımlanması gerekir. Hangi sınıflandırma algoritmasının benimseneceğine dair karar verme sürecine de öğrenme süreci denir. [69] Metin sınıflandırma adımları Şekil 5'te gösterilmiştir.



Şekil 5. Metin Sınıflandırma Adımları

iv. **Web Madenciliği:** Metin madenciliğindeki tüm işlem ve süreçlerin web üzerindeki bilgilerin işlenmesi ve analiz edilmesi için kullanılmasına denilmektedir. Günümüzde popüler olan alanlardandır.

v. **Bilgi Çıkarımı (Information Extraction / IE):** Geniş ölçekli bilgilerden ilgili gerçeklerin ve ilişkilerin ya da özetin çıkarılması olarak kısa bir tanım yapılabilir.

vi. **Doğal Dil İşleme:** Daha çok NLP (natural language processing) olarak bilinen bu uygulama bilgisayar tarafından doğal dilin çözümlenip anlaşılması ve yorumlanması sürecidir.

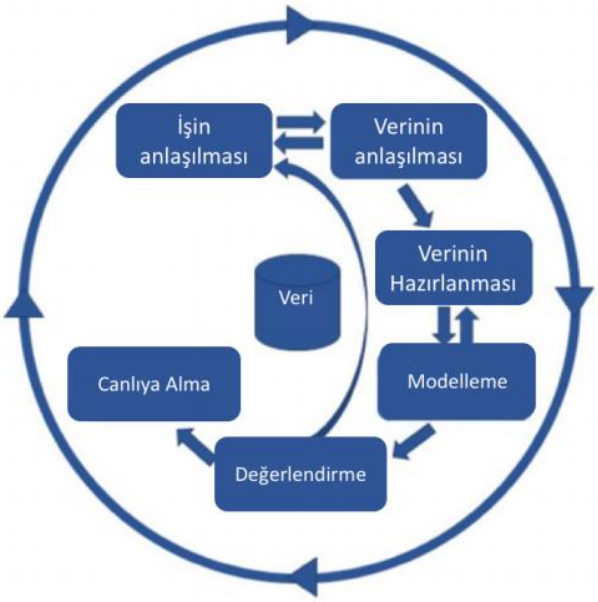
vii. **Kavram Çıkarma:** Kelimelerin ve cümlelerin anlamsal olarak benzer gruplara ayrılmasıdır.

3.3.2. Duygu Analizi

Görüş veya fikir madenciliği olarak da isimlendirilen duygu analizi, kişilerin varlıklara ve yazılı metinde ifade edilen niteliklerine yönelik fikirlerini, duygularını, değerlendirmelerini, tutumlarını ve duygularını analiz eden çalışma alanıdır. Bu varlıklar ürün, hizmet, kuruluş, birey, etkinlik, sorun veya konu olabilir. Duygu analizi, fikir madenciliği, fikir analizi, fikir çıkarma, duygu madenciliği, öznellik analizi, etki analizi, his analizi ve gözden geçirme madenciliği gibi pek çok ilgili isim ve biraz farklı görevler artık duygu analizi şemsiyesi altındadır. Duygu analizi terimi belki ilk olarak Nasukawa ve Yi tarafından 2003 yılında, fikir madenciliği terimi de ilk olarak Dave tarafından 2003 yılında ortaya atılmıştır. [71]

Duygu analizi, doğal dil işleme, veri madenciliği ve metin madenciliği gibi çeşitli araştırma alanlarını bir araya getirir ve hesaplamalı zeka yöntemlerini operasyonlarına entegre etmeye çabaladıkça ve daha fazla ışık tutmaya ve iyileştirmeye çalıştıkça kuruluşlar için hızla büyük önem kazanmaktadır. Duygu analizinde veya fikir madenciliğinde (DAFM) amaç, insanların yazılı dilde (metin) ifade edilen fikirlerini keşfetmektir. Terim olarak duygu analizi, "kişinin bir şey hakkında ne hissettiği", "kişisel deneyim, kendi duyguları", "bir şeye karşı tutum" veya "bir fikir" anlamına gelir. [72]

Duygu analizin işleyiş süreci için CRISP-DM (Cross Industry Standard Processing – Data Mining) standart süreci uygulanabilir. En temel tanımı itibariyle 6 adımlı bir süreçten ibarettir. Görsel olarak aşağıdaki Şekil 6’da gösterilmiştir.



Şekil 6. CRISP-DM Süreci

1. **İşin anlaşılması:** Duygu analizi için metin madenciliği sürecinin ilk adımıdır. Bu aşamada problem tanımı yapılır. Yapılacak olan analiz çalışması niçin yapılmaktadır, bu çalışmadan nasıl bir sonuç amaçlanmaktadır ve elde edilen sonuçlardan nasıl çıktılar çıkabileceği tanımlanır. Örneğin; Twitter'da yazılmış metinlerdeki ifadelerle kişilerin herhangi bir konuyla ilgili görüşlerinin ne olduğu çalışmada elde edilmek istenmiş olabilir. Bu konu Bilgi Güvenliği alanında olabilir.
2. **Verinin anlaşılması:** Birinci adımdaki işin anlaşılması adımıyla peş peşe ve tekrarlı yapılabilecek bir adımdır. Probleme uygun verilerin toplanması ve gözden geçirilmesidir. Elde edilen verilerdeki problemler varsa bunların tespiti yapılır. Gerekli görülürse verilerin düzeltilip zenginleştirilmesi veya ilave verilerin toplanması kararı verilir.
3. **Verinin Hazırlanması:** Verinin analiz edilebilmesi için hazır hale getirilmesidir. Veri üzerinde bazı işlemlerin yapılması gerekir ve bunların hangi yöntemlerle yapılması gerektiğinin kararı verilir. Veri ön işleme aşamasıdır. Metinlerin normalleştirilmesi, gereksiz/anlamsız/durak kelimelerin çıkarılması, kelimelerin köklerinin tespit edilmesi, tokenization işlemlerinin yapılması gibi işlemler bu aşamadadır.

   Bu aşamada bazı zorluklar çıkabilmektedir. Bu zorlukların bazılarından aşağıda bahsedilmektedir [72];

- Eş Anlamlı ve Eş Sesli Kelimeler: Aynı konuda ve aslında aynı anlama gelmesine rağmen farklı kullanıcılar tarafından farklı kelimeler kullanılması bazı zorlukları da beraberinde getirecektir. Örneğin; "araba" kelimesi bazı kullanıcılar tarafından "otomobil" ya da "taşıt" kelimesiyle ifade edilmiş olabilir. Bir başka örnek olarak; "kötü" kelimesi başka kullanıcılar tarafından "çirkin, berbat, pis" gibi kelimelerle ifade edilebilir.

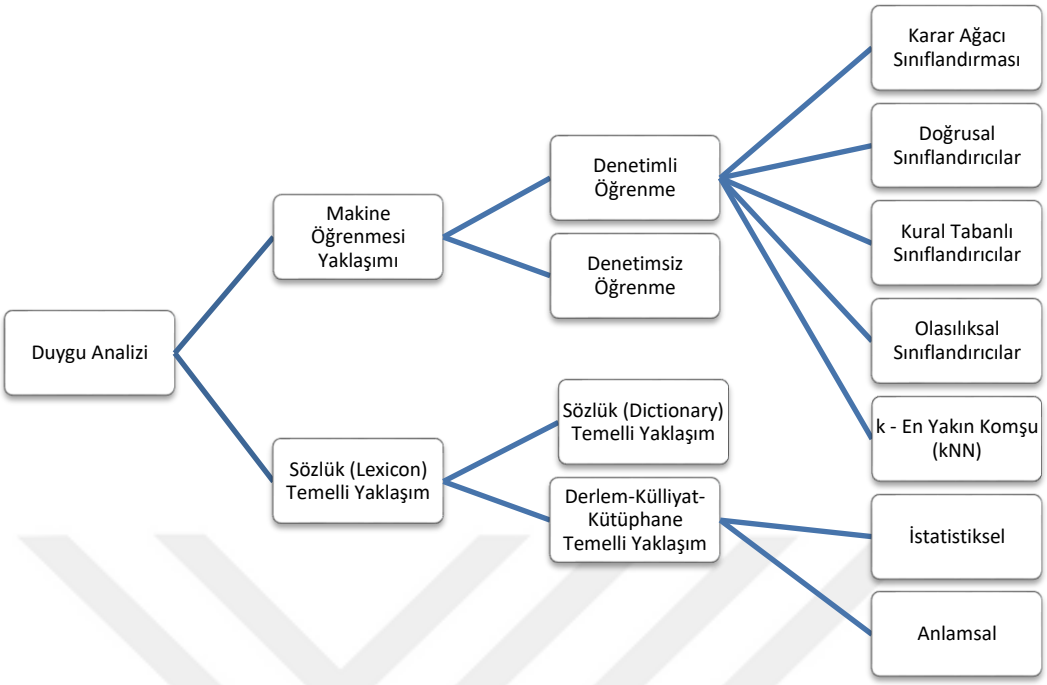
  Yine benzer şekilde, yazılışı aynı fakat manaları farklı olan eş sesli kelimeler de duygu analizi yöntemleri uygulanırken zorluklar çıkaracaktır. Örneğin; Yüz kelimesi ve Gül kelimesi için farklı anlamlar yüklenebilmektedir.

  - "Dün akşam geldiğinde yüzü gülüyordu."
  - "Gülün dikeni eline battığı için ağladı."
  - "Yarın sabah denize gidip yüzelim."
  - "Bahçede tam yüz adet gül vardı."

Aynı şekilde "tez" kelimesi bilimsel ve akademik eser manasında kullanıldığı gibi çabuk, süratli anlamlarında da kullanılmaktadır. Bu örneklere birçok kelime örnek olarak verilebilir.

- Alay: Bu ifadeler insanlar tarafından bile anlaması zor ifadeler olduğu gibi bilgisayarlar için de anlaşılması zor olacaktır. Alay ifadeleri genellikle ters anlam ifade etmektedir. Dilimizde buna iğneleme ve tariz sanatı da denmektedir. Örnek olarak "Ne kadar kültürlü olduğu yazdıklarından belli…" cümlesi verilebilir. Burada olumsuz bir kelime olmamasına karşın anlam olarak olumsuz bir duygu bulunmaktadır. Bu tarz ters anlam içeren alay kelimelerini anlayabilmek için bazı çalışmalar [73] mevcuttur ve gelişime açıktır.
- Birleşik Cümleler: Birden fazla cümlenin bulunduğu ve "ve", "veya" , "ama", "için" gibi bağlaçlarla ya da noktalı virgül ile birleştirilen birbirine bağlı cümlelerdir. "Çocuklar filmi çok beğendi ama biz sevmedik." cümlesinde olduğu gibi bu tarz birleşik cümleler duygu analizi için zorlayıcıdır.

4. **Modelleme:** Bu aşamada yapılacak çalışma ile ilgili bir makine öğrenmesi veya istatistiksel model oluşturulur. Burada birden çok model denenip, doğruluk oranı en yüksek model tercih edilir. Hata payının en aza indirilebilmesi için bu işlem tekrarlanmalıdır. Duygu analizi sınıflandırma yöntemleri için en yaygın kullanılan teknikler Şekil 7'de gösterilmiştir. [74]



Şekil 7. Duygu Sınıflandırma Yöntemleri

Makine öğrenimi teknikleri, denetimli ve denetimsiz öğrenme olarak ikiye ayrılır. Denetimli öğrenme yöntemleri, büyük miktarda önceden işlenmiş (etiketlenmiş) eğitim verisi gerektirirken, denetimsiz öğrenme yöntemleri böyle bir gerekliliği dayatmaz.

Sözlük tabanlı yaklaşımlar, metinlerdeki duyguları tanımak için duygu sözlükleri kullanır. Sözlüğe dayalı bir yaklaşım, duygu yönelimi ile ilişkili manuel olarak toplanan temel duygu sözcüklerinden başlar. Daha sonra sözlükte (WordNet veya başka bir çevrimiçi sözlük) yeterli eşanlamlılar ve zıt anlamlılar aranır. Derlem temelli yaklaşımlar, analizin konusu olan alana özgü duygu sözcükleri listesinden başlar ve diğer duygu sözcükleriyle birlikte araştırılır.

5. **Değerlendirme:** Bu aşamada şuanki mevcut durum ve ilk durumda hedeflenenin ne kadarı gerçekleştirilebildiği değerlendirilmektedir. Ayrıca uygulama aşamasına geçmeden önce eldeki durumla ilgili karar verilmesi gerekmektedir. Sonuçların kullanımına karar verilmesi hâlinde bir sonraki aşamaya geçilecektir.
6. **Uygulama:** Önceki aşamalardaki tüm süreçlerin tamamlanmasıyla artık eldeki verilerin ve modelin kullanıma hazır olması anlamına gelir. Diğer başka süreçlere girdi olarak

hazır olması için uygulama yapılması ya da çıktı olarak kullanılacaksa uygulamanın gerçekleştirilmesi adımıdır.

3.4. Siber Zorbalık

Siber zorbalık, bilgi ve iletişim teknolojilerini kullanarak özel veya tüzel bir kişiye karşı teknik veya ilişkisel bir şekilde zararlı davranış olarak tanımlanabilir. Bilgi ve iletişim teknolojilerini kullanarak insanları sürekli taciz etmek (siber takip), insanlarla alay etmek, tehdit etmek, dedikodu yaymak, internet üzerinden hakaret etmek gibi ilişkisel saldırı davranışlarını içerir.[1]

Siber zorbalık için yapılan tanımlar bazıları şu şekildedir;

Kanadalı eğitimci Bill Belsey; "Başkalarına zarar vermek maksadıyla bilgi ve iletişim teknolojilerini kullanarak kişi veya grup tarafından tekrarlı ve kasıtlı olan düşmanca hareketlerdir."

Cinsel kışkırtma dışındaki tehdit veya saldırganca tavırların başka kimselerin görmesi amacıyla çevrimiçi olarak yayınlanmasıdır.

Birine hakaret etmek ve tehdit etmek maksadıyla internet veya benzer dijital iletişim araçlarının kullanılmasıdır. [75]

Kişisel elektronik araç, anlık mesajlaşma, internet veya elektronik iletişim gereçleriyle yapılan zorbalıklardır.[76]

Elektronik yazışma ortamında bilerek tekrar edilen zarar verme, art niyet davranışıdır.[77]

Özellikle mobil cihaz ve/veya internet üzerinden yapılan saldırgan davranışlardır. Bireysel veya bir grup tarafından kendisini kolayca savunamayan bir mağdura tekrarlı ve kasıtlı olarak yapılan saldırganlıktır. [78]

İnternet veya dijital teknolojileri kullanarak zararlı materyal veya zararlı metinleri gönderip, yaymak veya diğer sosyal saldırganlık yöntemleriyle diğerlerine karşı saldırganca ve zalimce davranmaktır. [79]

2000'li yıllardan itibaren araştırmacıların ilgisini çekmeye başlayan bu kavram, araştırmacılar tarafından farklı özelliklerine odaklanılarak çeşitli şekillerde isimlendirilmiştir.

Buna göre siber zorbalık literatürde elektronik zorbalık, çevrimiçi zorbalık, internet zorbalığı, dijital zorbalık ve çevrimiçi zarar verme olarak çeşitli isimlerle anılmıştır. [80]

Siber zorbalık, bilgi ve iletişim teknolojisini kullanan bir kişi veya grup tarafından kendini koruyamayan bir mağdura karşı zaman zaman veya birkaç kez kasıtlı olarak yapılan olumsuz davranışlardır [81]. Siber zorbalık, zarar verme amacıyla tekrarlanan bir davranış bozukluğu olarak tanımlanmıştır.[82] Öte yandan, bir kişiyi veya bir grubu aşağılamak için anonim çağrılar, kimlik gizlenmiş hakaret / tehdit, gönderilen e-postalar, metinler, videolar veya görüntüler gibi siber zorbalığın tüm eylemlerini aynı başlık altında toplanmıştır. [83]

Bir başka tanımda ise siber zorbalık; bilgisayar, cep telefonları ve diğer teknolojik araçları kullanarak tekrarlayan kasıtlı zarar verme eylemi olarak tanımlanmıştır.[84]

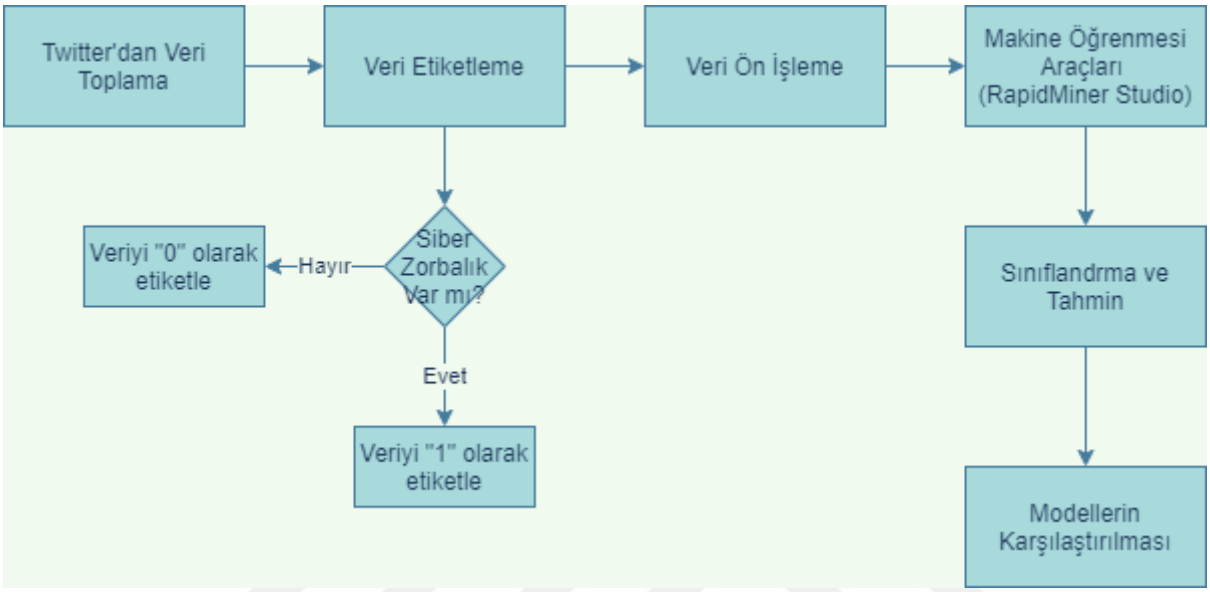
Siber zorbalık, saldırı ve tacizin elektronik bir şeklidir. Türkiye ve hatta birçok Avrupa ülkesi için yeni bir konu olan siber zorbalık konusu son 10 yıldır ABD ve Kanada'daki eğitim bilimcileri ve psikologlar başta olmak üzere diğer uzmanlar tarafından araştırılıyor. Siber zorbalık tehditlerinin önüne geçmek için Suicide Prevention Center ve Child Focus gibi bir dizi ulusal ve uluslararası çocuk koruma girişimi başlatıldı. Bu çabalara rağmen, birçok mesaj ve istenmeyen tehditler çevrimiçi olarak kalmaktadır. Ebeveynler, çocuklarını internet başta olmak üzere elektronik iletişim araçlarından korumak için bazı önlemler almaya çalışsalar bile çocuklarına tam kontrol sağlayamamakta ve bunun sonucunda siber zorbalık birçok kişisel soruna yol açmaktadır. Bu sorunlar kişinin psikososyal sorunlar, akademik sorunlar ve saldırgan davranışlar sergilemesine neden olur. Facebook, Twitter, Youtube, Instagram, Tumblr ve WhatsApp gibi son yıllarda gelişen sosyal paylaşım platformlarında sıklıkla karşılaşılan siber zorbalık mesajları, bireylerin psikolojisini derinden etkilemektedir. Bu noktada elektronik ortamda siber zorbalığa neden olan metinsel öğelerin tespiti ve filtrelenmesi ve bireye ulaşmadan önlem alınması, soruna ana çözüm bulmak için etkili bir yöntem olacaktır.[85]

4. MATERYAL VE YÖNTEM

Bu tez çalışması Twitter sosyal ağındaki Türkçe metinlerin derlenerek siber zorbalık içerenleri tespit edebilmek için bir yöntem uygulamak ve önermek için gerçekleştirilmiştir. Twitter, özellikle fikir beyan etme ve duygu paylaşımı yapmak üzere büyük oranda metin paylaşımı yapılan bir sosyal ağ olduğu için çalışma için en uygun alan olma özelliğini taşımaktadır. Bu çalışmaya benzer şekilde yapılan Türkçe dilinde pek çalışma olmadığı için veri setinin de hazırlanmasına ihtiyaç duyulmuştur.

Çalışmamızda Şekil 8'de gösterildiği gibi bir metodoloji önerilmektedir. Buna göre öncelikle seçilen kişilere atılan tweetleri toplayabilmek için hedef listesi oluşturulmuştur ve bu kişilere atılan Tweetler toplanmıştır. Toplanan veriler tek tek okunup siber zorbalık içerikli

olanları belirtebileceğimiz şekilde etiketlenmiştir. Etiketleme işleminden sonra veri ön işlemeden geçirilmiş ve temizlenmiştir. RapidMiner Studio programı kullanılarak makine öğrenmesi metotları ile verimizde sınıflandırma işlemi yapılmış ve tahmin modelleri oluşturulmuştur. Şekil 8'de gösterilen metodolojiye göre tüm süreçler bu bölümde ilerleyen başlıklar altında anlatılmıştır.

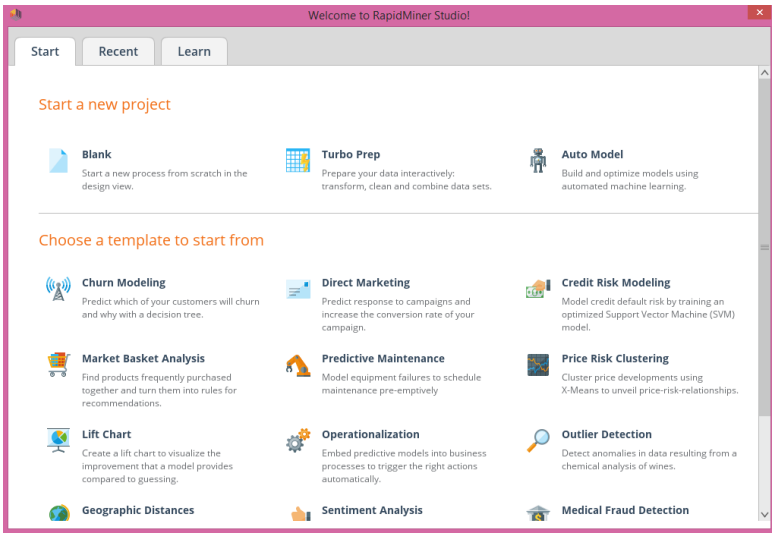


Şekil 8. Önerilen Metodoloji

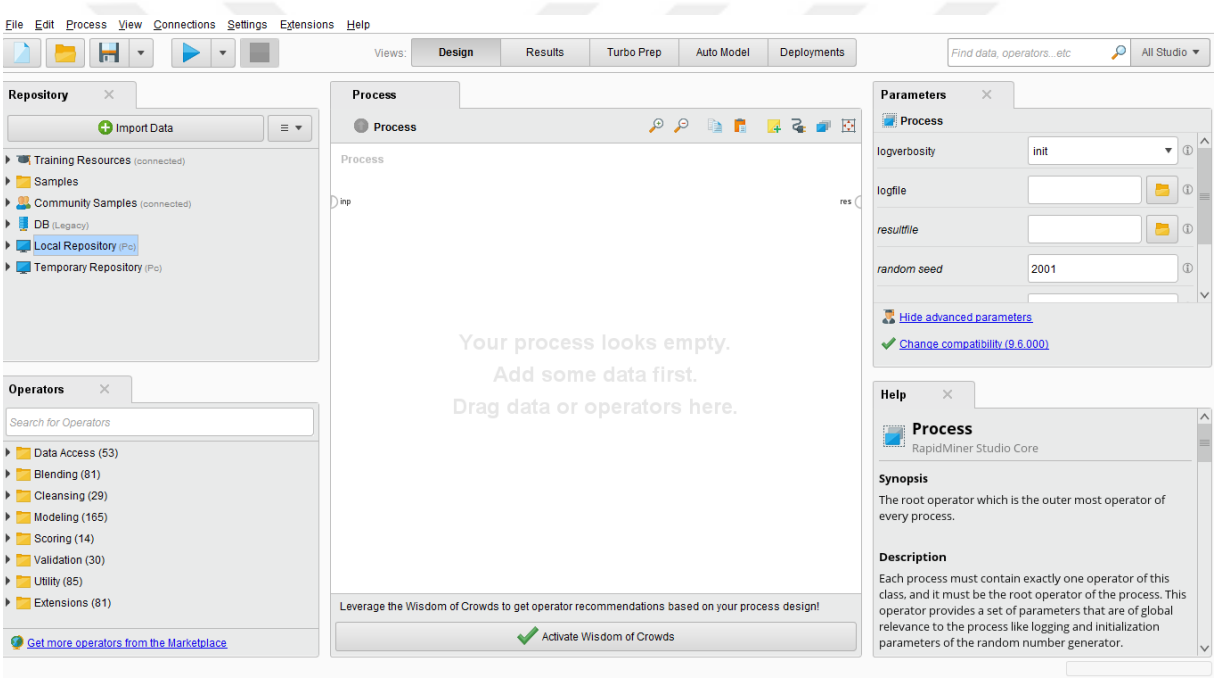
4.1. RapidMiner Studio

Çalışmamızda sonraki bölümlerde de anlatılan işlemleri yapmak için RapidMiner Studio [86] programı kullanılmıştır. Bu program makine öğrenmesi, veri madenciliği, metin madenciliği, tahmin edici analiz ve iş analizi amaçlarına yönelik olarak geliştirilmiş bir yazılım platformudur. Genel olarak iş ve ticaret dünyasında kullanıldığı gibi aynı zamanda araştırma, eğitim, hızlı prototipleme ve uygulama geliştirme gibi amaçlarla da kullanılabilir. Ayrıca veri madenciliği sürecinin tüm adımları RapidMiner tarafından desteklenmektedir, bu yüzden veri hazırlama, veri modelleme, doğrulama ve optimizasyon gibi amaçlarla da kullanılmaktadır.

Veri Hazırlama, makine öğrenmesi ve modelleme şeklinde 3 temel işlev için kullanılmaktadır. Açılış ekranı Şekil 9 ve Şekil 10'daki gibi olan RapidMiner, görsel açıdan da kullanıcı dostu bir program olma özelliğine sahiptir.

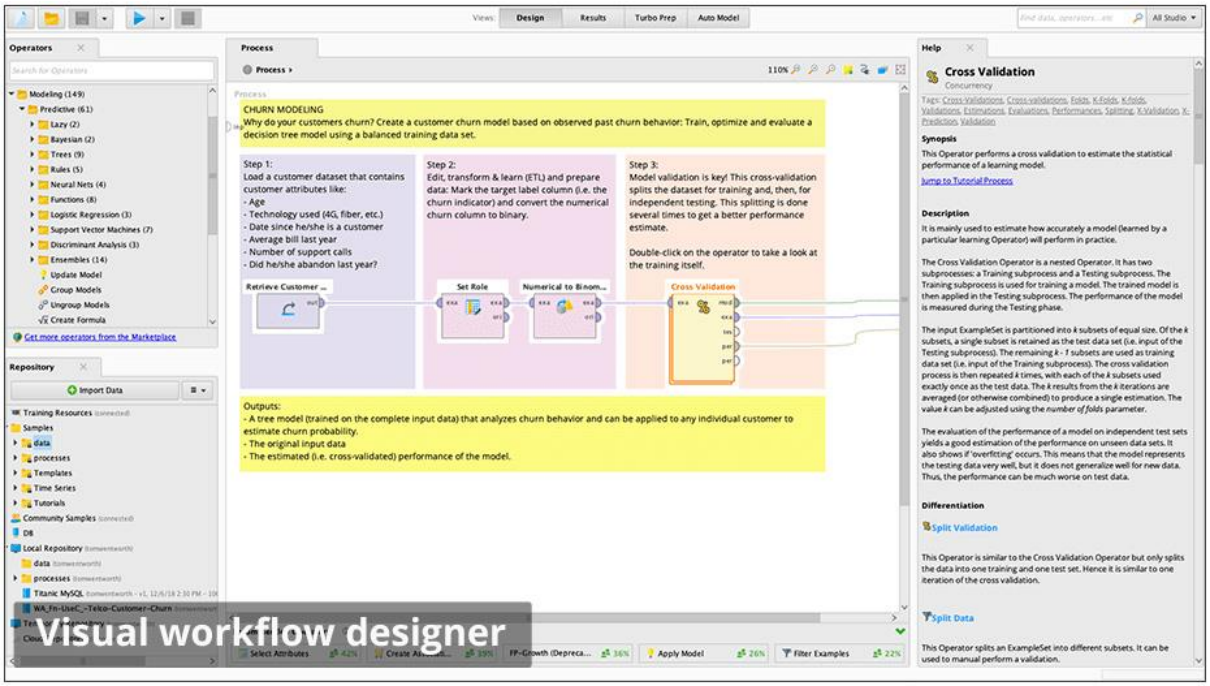


Şekil 9. RapidMiner Başlangıç Arayüzü 1



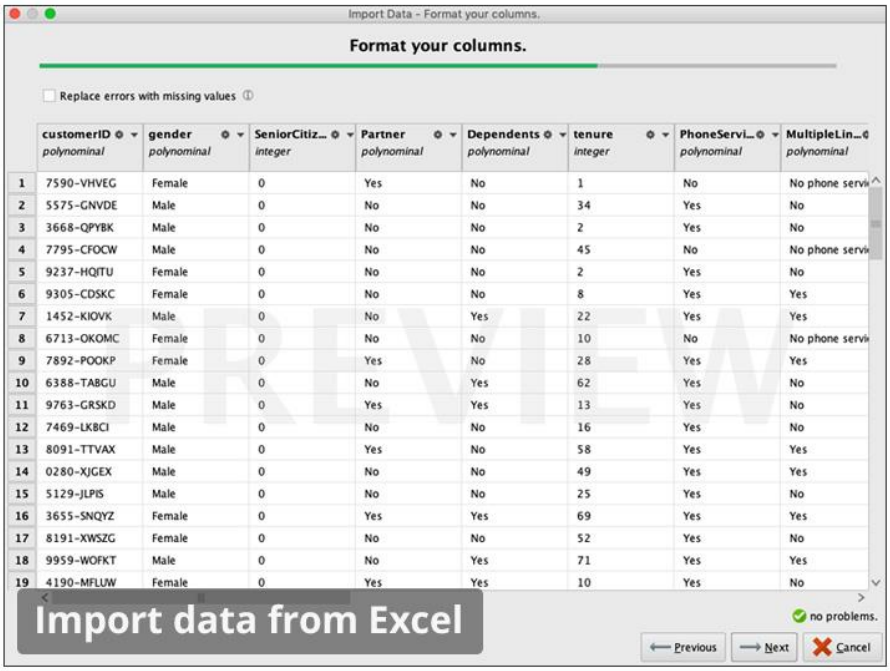
Şekil 10. RapidMiner Başlangıç Arayüzü 2

Sürükle ve bırak görsel arayüzüne sahip olması tahmini modellerin oluşturulmasını hızlandırır ve otomatikleştirir. 1500'den fazla algoritma ve fonksiyondan oluşan zengin kütüphane, her kullanım durumu için en iyi modeli sağlar. Yaygın kullanım örnekleri için oluşturulmuş hazır şablonlara sahiptir. Şekil 11'de RapidMiner'ın görsel tasarım ekranı gösterilmiştir.



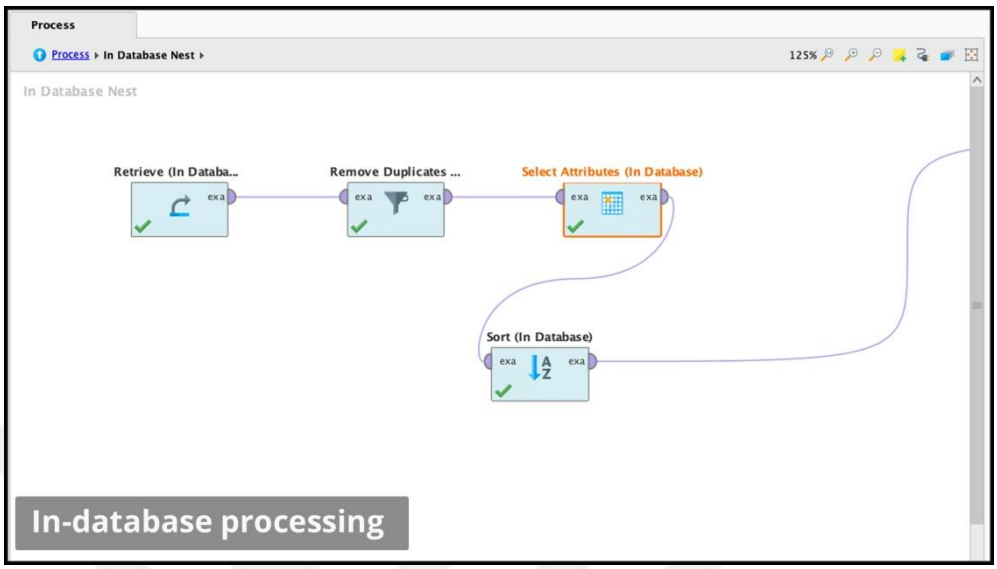
Şekil 11. RapidMiner Görsel İş Akışı Tasarımcısı

Herhangi bir veri kaynağına bağlanmayı da sağlar. Tüm veri kaynaklarına, kurumsal veri ambarlarına, veri göllerine, bulut uygulamalarına, iş uygulamalarına ve sosyal medyaya anında point + click bağlantıları oluşturabilmektedir. Şekil 12'de Excel'e veri aktarma ekranı gösterilmiştir.



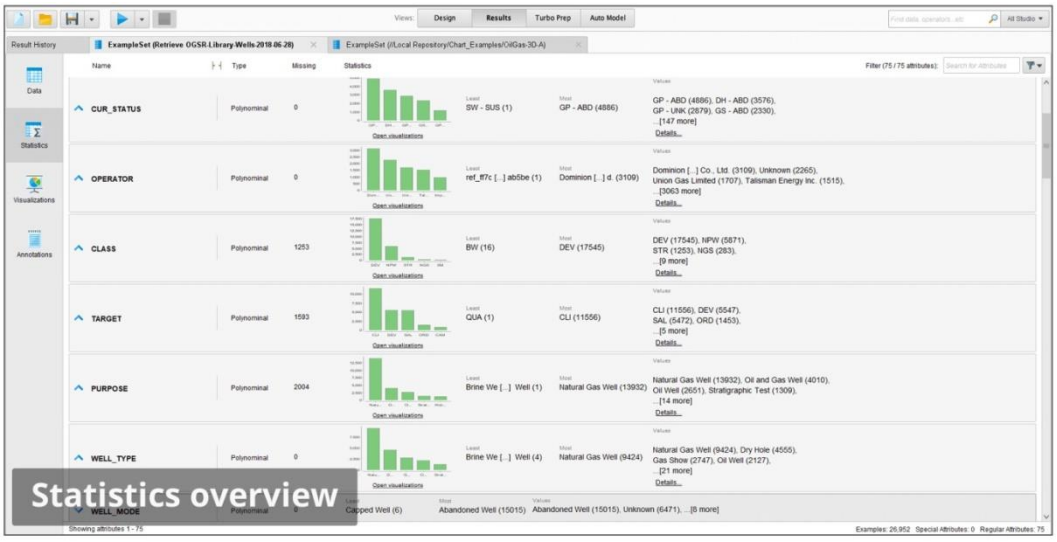
Şekil 12. RapidMiner Excel Veri Aktarma

Veri tabanı için otomatik işlemler yapmayı sağlar. Karmaşık SQL sorguları yazmadan verileri sorgulamayı ve almayı sağlar. MySQL, PostgreSQL ve Google BigQuery'yi destekler. Veri tabanı işlemlerine örnek görsel olarak Şekil 13'de gösterilmiştir.



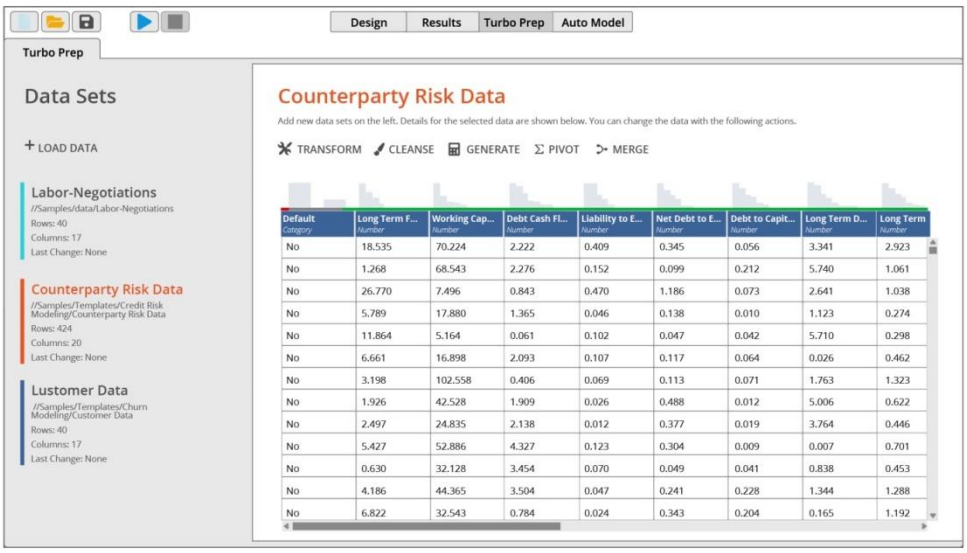
Şekil 13. RapidMiner Veri Tabanı İşlemleri

Veri keşfi ve görselleştirme özellikleri oldukça yeterli ve fazladır. Dağılım grafikleri, histogramlar, çizgi grafikleri, paralel koordinatlar, kutu grafikleri ve daha fazlasıyla desenleri, eğilimleri ve dağılımları anlama imkânı sağlar. Eksik değerler ve aykırı değerler de dahil olmak üzere yaygın veri kalitesi sorunlarını hızla bulup ve düzeltmeyi sağlar. Sağlam istatistiksel genel bakışlar ve 30'dan fazla etkileşimli görselleştirme kullanarak verileri sunar. İstatistiklere genel bakış alanı Şekil 14'te gösterilmiştir.



Şekil 14. RapidMiner İstatistiklere Genel Bakış

Tahmini modelleme için veri hazırlama zorluğunu ortadan kaldırır. Şekil 15'te örnek bir ekranı gösterilen RapidMiner Turbo Prep özelliği, tamamen etkileşimli bir point + click veri hazırlama deneyimi sunar. İstenilen sayıda kaynakta veri ayıklamayı, birleştirmeyi, filtrelemeyi ve gruplamayı sağlar.



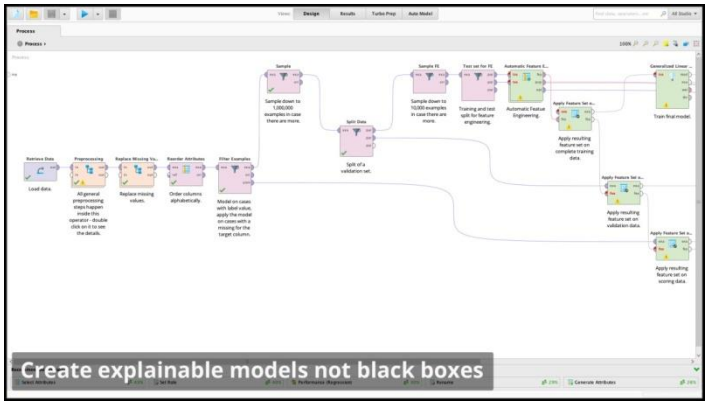
Şekil 15. RapidMiner Turbo Prep Ekranı

RapidMiner kod yazmadan hızlı bir şekilde etkili makine öğrenme modelleri oluşturmayı sağlar. Şekil 16'da örnek bir ekranı gösterilen RapidMiner Auto Model, otomatik makine öğrenimi kullanarak 5 tıklamayla modeller oluşturmayı sağlar. Yüzlerce denetimli ve denetimsiz makine öğrenme algoritması mevcuttur. Regresyon, kümeleme, zaman serileri, metin analitiği ve derin öğrenme dâhil temel ve gelişmiş makine öğrenmesi tekniklerini uygulayabilmeyi sağlar.



Şekil 16. RapidMiner Auto Model

Açıklanması ve anlaşılması kolay görsel veri bilimi iş akışları oluşturabilmektedir. Veri hazırlama, modelleme ve doğrulama sürecindeki her adım tam şeffaf bir şekilde belgelenebilmektedir. Görsellerin anlaşılması ve diğer kişilere açıklanması kolaydır. Şekil 17'de örnek bir model gösterilmiştir.



Şekil 17. RapidMiner Açıklanabilir Örnek Bir Model

4.2. Veri Toplama

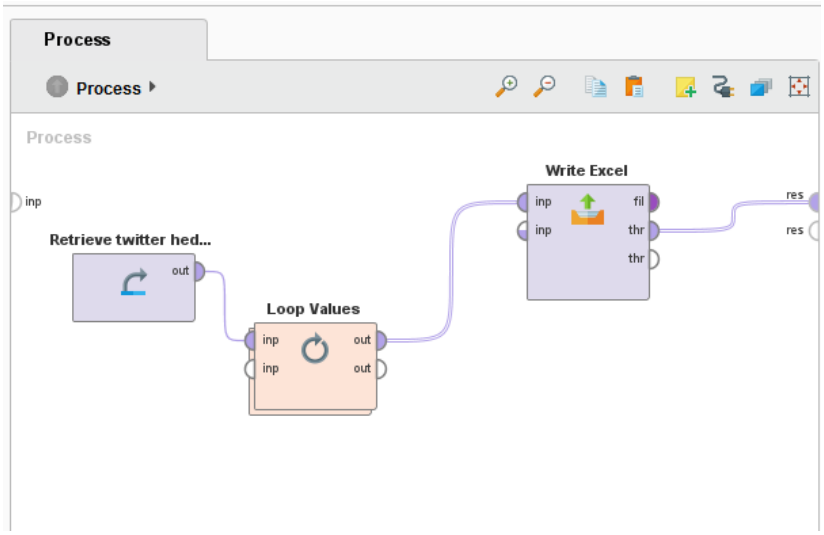
Çalışmamızda veri olarak Twitter'da yazılan Türkçe mesajlar kullanılmıştır. Bu yüzden çalışmanın ilk aşamasında Twitter mesajlarını çekmek için "Hedef Liste" adında bir liste hazırlanmıştır. Bu liste mesajların "kime" atıldığı, siber zorbalık var ise "kime" karşı yapıldığını

belirlemek için hazırlanmıştır. En çok siber zorbalığa maruz kaldığını düşündüğümüz ya da siber zorbalık örneklerine fazlaca rastlayabileceğimizi düşündüğümüz için bu listede kamu ve siyaset, kültür-sanat, basın-yayın, spor, iş dünyası, fenomenlerden oluşan kişiler bulunmaktadır. Ek 1'de paylaşılan "Hedef Liste" ve 378 kişi bulunan bu liste hazırlanırken yukarıda sayılan kategorilerdeki en çok takipçili ve/veya en aktif hesaplardan seçilmiştir. Bu liste, sosyal medya araştırmaları için kurulmuş olan web siteler yardımıyla listeler tespit edilerek hazırlanmıştır. Örnek olarak çalışmamızda da kullandığımız BoomSocial.com (sosyal medya ölçümleme& analiz) web sitesi verilebilir.

Veri toplama işlemi yapılırken RapidMiner Studio programından faydalanılmıştır. Twitter paylaşımlarını toplamak için, RapidMiner Studio programında oluşturduğumuz 378 kişilik twitter hesabı için aşağıdaki Şekil 18'de gösterildiği gibi bir döngü oluşturulmuştur ve her biri için atılan son ve popüler 100 tweet çekilmiştir. Normal şartlarda Twitter'ın kısıtlamalarından dolayı 15 dakika 450 sorgu yapılabilmektedir ve 10.000 adet tweet çekilebilmektedir. Bundan dolayı RapidMiner programına 3 twitter hesabıyla bağlanarak ve ayrıca 378 kişilik liste 3 ayrı parçaya bölünerek kullanılmıştır. Tweet çekmek için kullanılan Şekil 18'de gösterilen döngüde 15'er dakikalık beklemeler de ilave edilerek kısıtlamalara takılmadan toplam 29198 adet tweet saatlerce çalışan prosesle (süreç) çekilmiştir. Bu veri Write Excel operatörü sayesinde bir excelde listelenmiştir. Hedef liste 3 ayrı parçaya bölündüğü için oluşan 3 Excel dosyası daha sonra aynı Excel sayfasında birleştirilmiştir. Tweetleri çekmek için "Loop Values" döngüsü içerisinde iç süreç olarak Şekil 18'de gösterildiği gibi "SearchTwitter" operatörü kullanılmıştır. Şekil 18'de gösterilen "Retrieve twitter hedef liste 1" adlı dosya içerisinde Ek 1'de paylaşılan hedef listenin 1/3 'lük bir kısmı bulunmaktadır. Örnek olarak Tablo 4'te gösterilen şekilde bir liste bulunmaktadır.

Tablo 4. Retrieve Twitter Hedef Liste Örneği

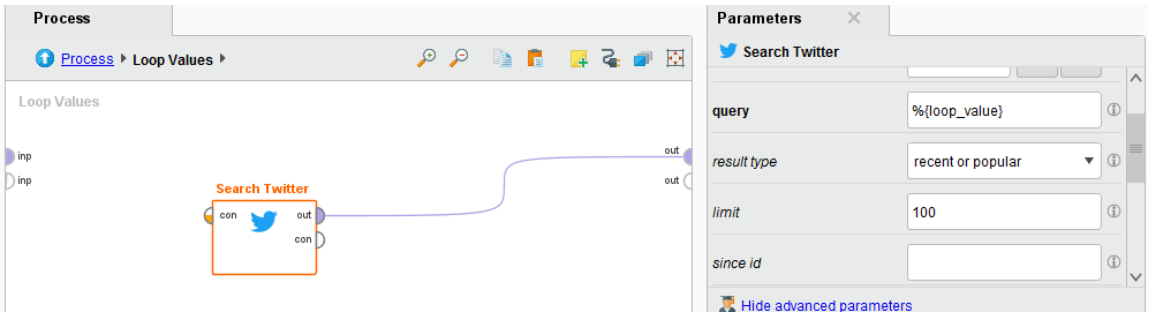
@username
@username1
@username2
@username3



Şekil 18. Twitter Hesapları için Döngü

Şekil 18'de yer alan "Loop Values" iç sürecinin detayında kullanılan "SearchTwitter" operatörü Şekil 19'da gösterildiği gibi sorgu (query) alanına %{loop_value} değeri yazılmıştır. Bu ifade aslında Tablo 4'te örnek olarak gösterilen kullanıcı adı bilgilerinin Twitter mesajları içerisinde aranmasıdır. Sonuç türü (result type) alanına "recent or popular" değeri kullanılarak çekilmesi istenilen Twitter mesajlarının son veya popüler tweetler arasından çekilmesi için kullanılan bir parametredir. Limit parametresi için girilen değer ile de her bir sorgu için kaç adet twitter mesajının çekileceği ayarlanmaktadır. Çalışmamızda her 378 kişi için de 100'er adet mesaj çekilmesi gerçekleştirilmiştir. Bu sorgular son 1 haftalık bir zaman dilimi içerisinden çekildiği için her bir kullanıcı için 100 mesaj olmayabilmektedir. Daha düşük mesaj sayısına sahip kullanıcılar da bulunmaktadır.

Şekil 18'deki Twitter mesajlarını toplama sürecimizi, hedef listemizi üç ayrı parçaya böldüğümüz için "Retrieve twitter hedef liste 2" ve "Retrieve twitter hedef liste 3" dosyaları ile de beraber olacak şekilde 3 eş zamanlı süreç olarak gerçekleştirilmiştir. "Write Excel" operatörü sayesinde oluşan 3 Excel dosyası birleştirilerek yaklaşık 30.000 adetlik Twitter mesajları toplanmıştır.



Şekil 19. Search Twitter kullanımı

4.3. Veri Seti

Kamu ve siyaset, kültür-sanat, basın-yayın, spor, iş dünyası, fenomenlerden belirlediğimiz toplam 378 kişiden oluşan listedeki isimlere atılan toplam yaklaşık otuzbin adetlik tweet verisi elde edilmiştir. Bu verideki tweet sayısı, yinelenenleri kaldırma işlemi yapılarak 29198 adete düşürülmüştür.

Veri setimizde; tweetin hangi kullanıcı tarafından atıldığı (From-User-Id), kime atıldığı (To-User), ne yazıldığı (Text), kaç retweet aldığı, hangi kaynaktan atıldığı (Twitter For Android, Twitter for iPhone, Twitter for Mac, Twitter For Web App vb.) bilgisi yer almaktadır. Kaynak bilgisi özetle Web veya Mobil olacak şekilde Excel formülü yardımıyla "Kaynak Cihaz" olarak bir sütun açılarak belirtilmiştir. Ayrıca bunlara ek olarak Excel'de düşey ara formülü yardımıyla, "To-User" alanı ile Ek 1'deki @username ve kategori alanları karşılaştırılarak tweetin atıldığı kişinin hangi kategoride yer aldığı bilgisi de veri setine eklenmiştir. Son olarak veride siber zorbalık bulunup bulunmadığını etiketleyebilmek için "Siber Zorbalık" sütunu adıyla bir bilgi daha eklenmiştir. Bu alan sonraki bölümde anlatılacağı şekilde manuel olarak tüm tweetler haftalar süren bir etiketleme çalışması neticesinde okunarak doldurulmuştur. Veri seti ile ilgili örnek ekran görüntüsü Şekil 20'de gösterilmiştir.

From-User-Id	To-User	Text	Retwee	Kaynak	Kaynak Ciha	Kategori	Siber Zorbalık
1127708375695990785	RTErdogan	@furkaansoslu @yedekpeyok @karab	,0	Twitter for iPhone	Mobil	Kamu ve Siyaset	1
1229102057195438080	RTErdogan	59.Ey Efendiler. Sorgulamayan İnsan C	,0	Twitter for Mac	Web	Kamu ve Siyaset	1
1078557217761214464	RTErdogan	9 yaşındaki Sena, @RTErdogan dedesi	,0	Twitter for iPhone	Mobil	Kamu ve Siyaset	0
950489492477173760	RTErdogan	@nihalerkoc06 @TC_Zonguldak @mtu	,0	Twitter for iPhone	Mobil	Kamu ve Siyaset	0
1255967988873863173	RTErdogan	@metinsahintkd @RTErdogan MaşaAl	,0	Twitter Web App	Web	Kamu ve Siyaset	0
1229102057195438080	RTErdogan	64.Şanlı Ordumuzun Kahraman Astsub	,0	Twitter for Mac	Web	Kamu ve Siyaset	0

Şekil 20. Veri Setinden Örnek Bir Kesit

4.4. Veri Etiketleme

Elde edilen 29198 tweet tek tek incelenmiştir. Her bir tweet manuel olarak siber zorbalık varsa "1", siber zorbalık yoksa "0" olarak etiketlenmiştir. Etiketleme işlemi yapılırken Yin ve

ark.[2] çalışmasında 3 olarak açıkladığı ancak bizim uygulamamızda 4 farklı örnek türüne rastlanılmıştır.

1) İlk olarak; siber zorbalık olduğu açıkça anlaşılan örneklerdir ve bunlar "1" (siber zorbalık içeriyor) olarak etiketlenmiştir. Örnek olarak aşağıdaki ifadeler verilebilir;

- "@username Bunu yazarken Amerika'dan izin aldın mı? Doğru söyle bak şerefsiz??????"
- "@ username 2 gram beynin var kabullen bunu laöepqçzxmvcödlap"

Siber zorbalık örnekleri etiketlenirken Xu ve ark. [4] yaptıkları çalışmada tespit ettikleri 7 farklı duygunun metin içerisinde varlığı kontrol edilerek etiketleme işlemi yapılmaya çalışılmıştır. Bu duygular; öfkeli, utanç verici, empati yoksunu, korku dolu, gurur kırıcı, rahatlama ve üzüntü verici duygulardır. Bu 7 duyguya sahip ve başkasına karşı atılan mesajlar siber zorbalık olarak tespit edilerek "1" (siber zorbalık içeriyor) ile etiketlenmiştir.

2) İkinci olarak; hiçbir siber zorbalık ifade içermediği açıkça anlaşılan örneklerdir ve bu metinler "0" (siber zorbalık içermiyor) olarak etiketlenmiştir. Örnek olarak aşağıdaki ifadeler verilebilir;

- "başkanm sizleri güzel izmir'e de bekliyoruz @username ."
- "derbide birinizden gol istiyoruz @username1 @username2 ."

Bu örnekler incelenirken kelimelerin anlamlarıyla beraber cümlenin bütünü de düşünülmüştür. Çünkü bazı örneklerde kelimeler düzgün seçilmesine rağmen açıkça olmasa da imalı ifadeler barındırdığı için siber zorbalık içeren örnekler de bulunmaktadır. Bunlar sonraki örneklerde açıklanmıştır.

3) Üçüncü olarak; kullanılan kelimelere bakıldığında siber zorbalık içermiyor gibi görünse de anlam bütünlüğü açısından yorumlandığında siber zorbalık içeren örneklerdir ve bu örnekler "1" (siber zorbalık içeriyor) olarak etiketlenmiştir. Bu örneklerde de, 1. maddedeki örneklerde olduğu gibi Xu ve ark. [4] yaptıkları çalışmada tespit ettikleri 7 farklı duygunun yer alıp almaması kontrol edilmiştir.

- "eyy @username cebin doldurdun tabi tatilinizi rahat yapabiliyor musunuz bukadar insanın hakkını yedikten sonra…."

Bu örnekte, kişiye hırsızlık, hak yeme iftirası atıldığı ve saldırı niyetiyle yazıldığı düşünülerek "1" (siber zorbalık içeriyor) olarak etiketlenmiştir.

4) Dördüncü ve son olarak; kullanılan kelimelere bakıldığında siber zorbalık içeriyor gibi görünse de başkasını hedef almayan örneklerdir ve bu örnekler "0" (siber zorbalık içermiyor)

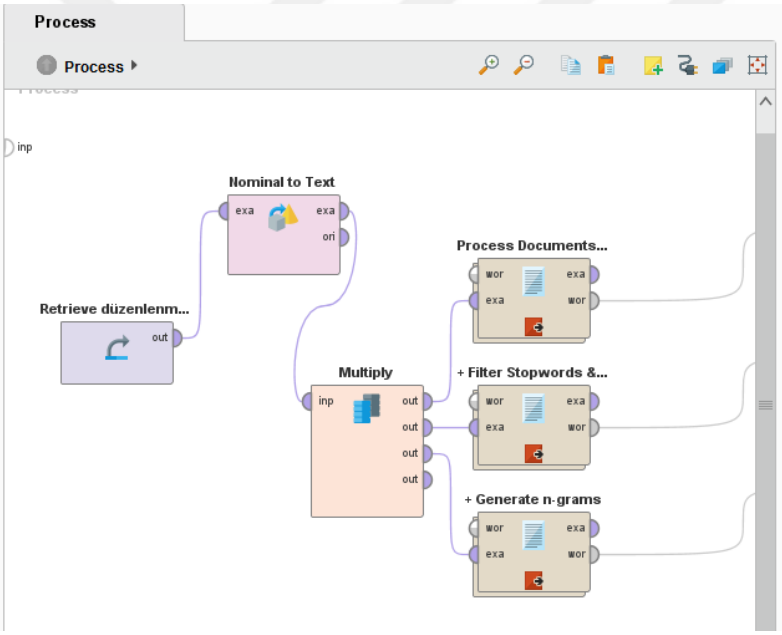
olarak etiketlenmiştir. Çünkü cümle içerisinde geçen küfürlü kelimeler mesajın siber zorbalık içerikli olarak etiketlememize yeterli değildir. Örnek olarak aşağıdaki cümleler verilebilir;

- "36 saattir elektrikler yok, k.çımız dondu."
- "Otobüs havalandırmayı açmadığı için ortam b.k gibi kokuyor."

Bu yüzden bu tarz metinler "0" (siber zorbalık içermiyor) olarak etiketlenmiştir.

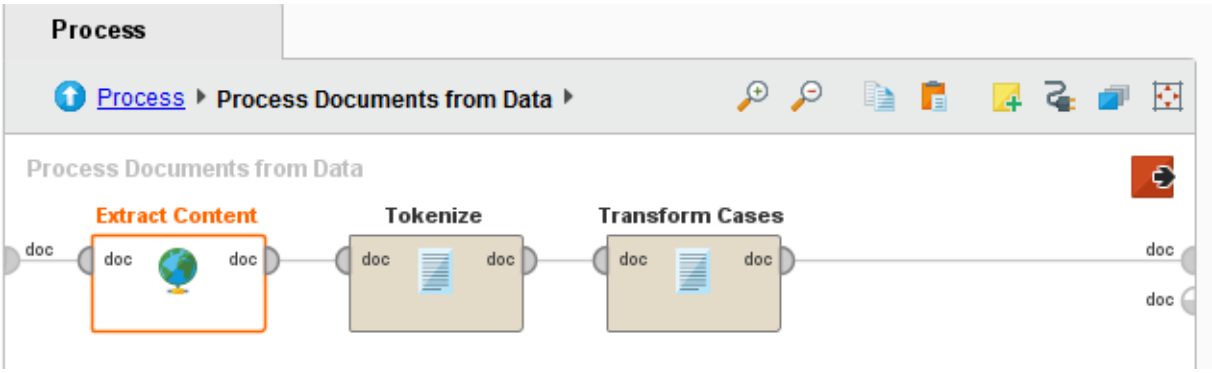
4.5. Veri Önişleme

Toplanan ve etiketlenen veri setinde Şekil 21'de gösterildiği gibi veri önişleme işlemleri yapılmıştır. Burada durak kelimeleri eleme, tokenize etme, küçük harfe çevirme, n-gramlar oluşturma, URL gibi anlamsız metinleri çıkarma işlemleri yapılmıştır. Önişleme işlemleri RapidMiner aracılığıyla yapılmıştır.



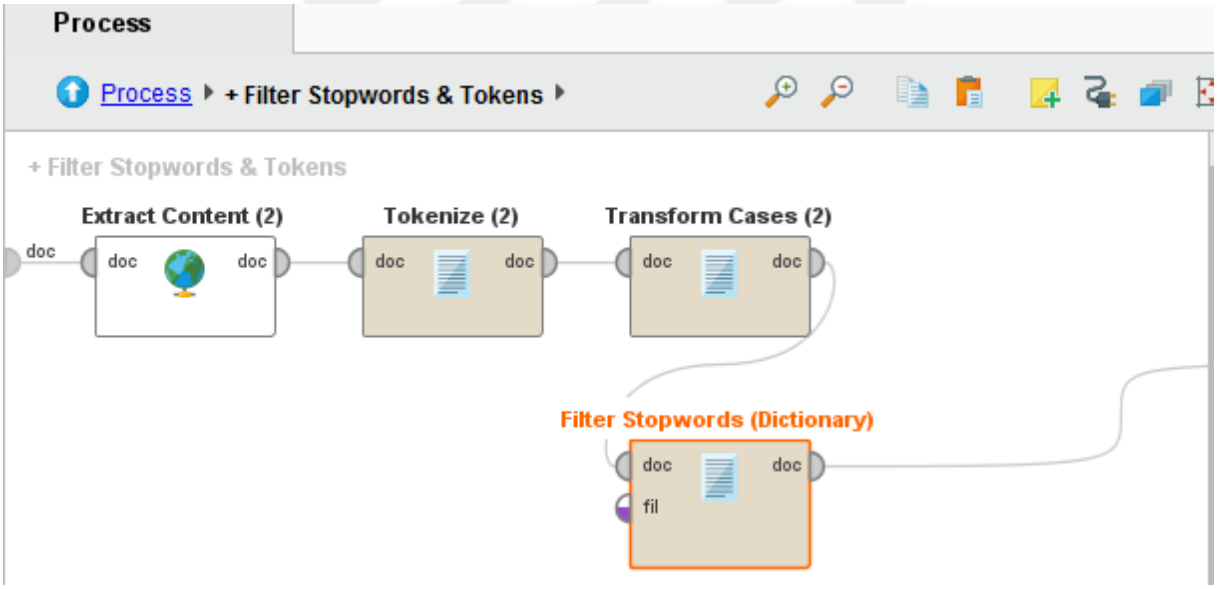
Şekil 21. Veri Önişleme Süreci

Burada girdi olarak oluşturduğumuz ve etiketleme işlemini tamamladığımız "Retrieve düzenlenmiş veri seti" dosyası verilmiştir. Verimiz "text" türünde olan tweetlerimizi alıp 3 alternatif iç sürece dâhil edilmiştir. Buradaki ilk iç sürecimizde Şekil 21'de gösterildiği gibi URL içeriklerini ayıklama, tokenize işlemleri ve küçük harfe çevirme işlemleri yapılmıştır.

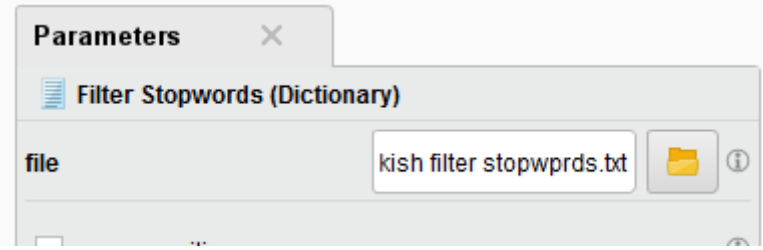


Şekil 22. Veri Önişleme Süreci 1

Şekil 23'teki diğer alternatife bakıldığında ise sürecimizde URL içeriklerini ayıklama, tokenize işlemleri ve küçük harfe çevirme işlemlerine ek olarak "acaba, ama, şey, bu, için vb." durak kelimelerini ayıklama işlemi de yapılmıştır. "Filter Stopwords (Dictionary)" operatörünün detayında ise kendi oluşturduğumuz durak kelimelerinin "txt" dosyası Şekil 24'te gösterildiği gibi parametre olarak verilmiştir.



Şekil 23. Veri Önişleme Süreci 2



Şekil 24. Durak Kelimeleri Dosyası Ekleme

Durak kelimelerin listesi de Tablo 5'te verilmiştir.[87]

Tablo 5. Durak Kelimelerin Listesi

1	**a**	**58**	da	**115**	hiç	**172**	olarak	**229**	sizden
2	acaba	**59**	daha	**116**	hiçbir	**173**	oldu	**230**	sizden
3	altı	**60**	dahi	**117**	hiçbiri	**174**	olduğu	**231**	size
4	altmış	**61**	dan	**118**	i	**175**	olduğunu	**232**	sizi
5	ama	**62**	de	**119**	ı	**176**	olduklarını	**233**	sizi
6	ancak	**63**	defa	**120**	için	**177**	olmadı	**234**	sizin
7	arada	**64**	değil	**121**	içinde	**178**	olmadığı	**235**	sizin
8	artık	**65**	diğer	**122**	iki	**179**	olmak	**236**	sonra
9	asla	**66**	diğeri	**123**	ile	**180**	olması	**237**	şöyle
10	aslında	**67**	diğerleri	**124**	ilgili	**181**	olmayan	**238**	şu
11	aslında	**68**	diye	**125**	ise	**182**	olmaz	**239**	şuna
12	ayrıca	**69**	doksan	**126**	işte	**183**	olsa	**240**	şunları
13	az	**70**	dokuz	**127**	itibaren	**184**	olsun	**241**	şunu
14	bana	**71**	dolayı	**128**	itibariyle	**185**	olup	**242**	ta
15	bazen	**72**	dolayısıyla	**129**	kaç	**186**	olur	**243**	tabii
16	bazı	**73**	dört	**130**	kadar	**187**	olur	**244**	tam
17	bazıları	**74**	e	**131**	karşın	**188**	olursa	**245**	tamam
18	belki	**75**	edecek	**132**	kendi	**189**	oluyor	**246**	tamamen
19	ben	**76**	eden	**133**	kendilerine	**190**	on	**247**	tarafından
20	benden	**77**	ederek	**134**	kendine	**191**	ön	**248**	trilyon
21	beni	**78**	edilecek	**135**	kendini	**192**	ona	**249**	tüm
22	benim	**79**	ediliyor	**136**	kendisi	**193**	önce	**250**	tümü
23	beri	**80**	edilmesi	**137**	kendisine	**194**	ondan	**251**	u
24	beş	**81**	ediyor	**138**	kendisini	**195**	onlar	**252**	ü
25	bile	**82**	eğer	**139**	kez	**196**	onlara	**253**	üç
26	bilhassa	**83**	elbette	**140**	ki	**197**	onlardan	**254**	un
27	bin	**84**	elli	**141**	kim	**198**	onları	**255**	ün
28	bir	**85**	en	**142**	kime	**199**	onların	**256**	üzere
29	biraz	**86**	etmesi	**143**	kimi	**200**	onu	**257**	var
30	birçoğu	**87**	etti	**144**	kimin	**201**	onun	**258**	vardı
31	birçok	**88**	ettiği	**145**	kimisi	**202**	orada	**259**	ve
32	biri	**89**	ettiğini	**146**	kimse	**203**	öte	**260**	veya

Tablo 3. (Devamı) Durak Kelimelerin Listesi									
33	birisi	**90**	fakat	**147**	kırk	**204**	ötürü	**261**	ya
34	birkaç	**91**	falan	**148**	madem	**205**	otuz	**262**	yani
35	birşey	**92**	filan	**149**	mi	**206**	öyle	**263**	yapacak
36	biz	**93**	gene	**150**	mı	**207**	oysa	**264**	yapılan
37	bizden	**94**	gereği	**151**	milyar	**208**	pek	**265**	yapılması
38	bize	**95**	gerek	**152**	milyon	**209**	rağmen	**266**	yapıyor
39	bizi	**96**	gibi	**153**	mu	**210**	sana	**267**	yapmak
40	bizim	**97**	göre	**154**	mü	**211**	sanki	**268**	yaptı
41	böyle	**98**	hala	**155**	nasıl	**212**	sanki	**269**	yaptığı
42	böylece	**99**	halde	**156**	ne	**213**	şayet	**270**	yaptığını
43	bu	**100**	halen	**157**	neden	**214**	şekilde	**271**	yaptıkları
44	buna	**101**	hangi	**158**	nedenle	**215**	sekiz	**272**	ye
45	bunda	**102**	hangisi	**159**	nerde	**216**	seksen	**273**	yedi
46	bundan	**103**	hani	**160**	nerede	**217**	sen	**274**	yerine
47	bunlar	**104**	hatta	**161**	nereye	**218**	senden	**275**	yetmiş
48	bunları	**105**	hem	**162**	neyse	**219**	seni	**276**	yi
49	bunların	**106**	henüz	**163**	niçin	**220**	senin	**277**	yı
50	bunu	**107**	hep	**164**	nin	**221**	şey	**278**	yine
51	bunun	**108**	hepsi	**165**	nın	**222**	şeyden	**279**	yirmi
52	burada	**109**	her	**166**	niye	**223**	şeye	**280**	yoksa
53	bütün	**110**	herhangi	**167**	nun	**224**	şeyi	**281**	yu
54	çoğu	**111**	herkes	**168**	nün	**225**	şeyler	**282**	yüz
55	çoğunu	**112**	herkese	**169**	o	**226**	şimdi	**283**	zaten
56	çok	**113**	herkesi	**170**	öbür	**227**	siz	**284**	zira
57	çünkü	**114**	herkesin	**171**	olan	**228**	siz		

4.6. Yöntem

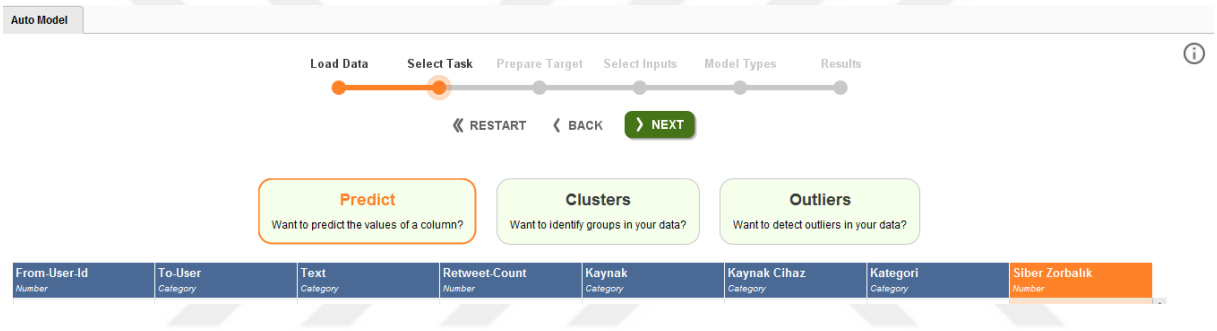
Bu bölümde elde ettiğimiz ve ön işlemeden geçirdiğimiz verileri sınıflandırma yöntemleri kullanıldığı gösterilmiştir. Veri setimizdeki tüm tweetler okunmuş ve manuel olarak etiketlenmiştir. Siber zorbalık içeren tweetler için “1” değeri içermeyenler için “0” değeri atanmıştır. Oluşturulan bu veri RapidMiner sistemine aktarılmıştır.

4.6.1. Sınıflandırma ve Tahmin

Bu bölümde oluşturduğum veriden yola çıkarak sınıflandırma yapılacak ve siber zorbalık tespitine etki eden faktörleri değerlendirerek tahmin yürütme işlemleri yapılmıştır. Bu işlemler için makine öğrenmesi teknikleri de kullanılmıştır. Bu teknikler, RapidMiner Auto Model özelliği kullanarak yapılmıştır.

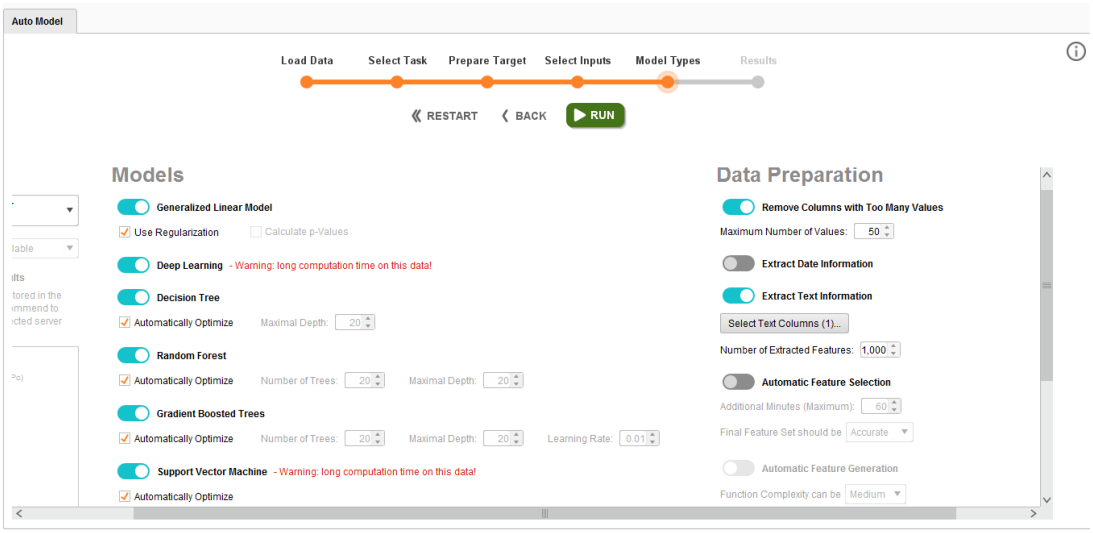
Auto Model özelliği; RapidMiner Studio'nun, model oluşturma ve modeli uygulama konusunda kolaylık sağlayan ve hız kazandıran bir özelliktir. Tahmin, Kümeleme ve Aykırı değerleri tespit etme olarak 3 temel işlevi bulunmaktadır. Çalışmamızda kullandığımız ise Tahmin işlevidir. Auto Model verileri otomatik olarak değerlendirir ve problem çözümü için modeller sunar ve hesaplamalar tamamlandıktan sonra bu modellerin sonuçlarını karşılaştırmada yardımcı olur.

Auto Model başlangıç ekranı Şekil 25'te gösterilmiştir. Verimiz eklendikten ve Auto Model özelliği seçildikten sonra "Predict" işlemi seçilir ve tahmin edilecek alan seçilir. Burada "Siber Zorbalık" alanı tahmin yapılacak alan olarak seçilmiştir. Verimiz, 2626 adet siber zorbalığa sahip veri ile eşit sayıda olacak şekilde 2626 siber zorbalık içermeyen veri ile birlikte 5252 tweet içermektedir.



Şekil 25. Auto Model Başlangıç Ekranı

Tahmin yapılırken hangi alanlarımızın girdi olarak kullanılacağı seçilirken veri setimizdeki To-User, Text, Kaynak Cihaz (mobil, web) ve Kategori girdileri seçilmiştir. Sınıflandırma için Şekil 26'da gösterilen model tiplerinden hepsi seçilmiştir.



Şekil 26. RapidMiner Auto Model Tipleri

4.6.1.1. Naive Bayes

Naive Bayes sınıflandırıcısının temeli Bayes teoremine dayanır. Naive Bayes sınıflandırıcı, örüntü tanıma problemine ilk bakışta oldukça kısıtlayıcı görülen bir önerme ile kullanılabilen olasılıksal bir yaklaşımdır. Bu önerme, örüntü tanımada kullanılacak her bir tanımlayıcı öznitelik ya da parametrenin istatistik açıdan bağımsız olması gerekliliğidir. Her ne kadar bu önerme Naive Bayes sınıflandırıcının kullanım alanını kısıtlasa da, genelde istatistik bağımsızlık koşulu esnetilerek kullanıldığında da daha karmaşık yapay sinir ağları gibi metotlarla karşılaştırabilir sonuçlar vermektedir. Bir Naive Bayes sınıflandırıcı, her özniteliğin birbirinden koşulsal bağımsız olduğu ve öğrenilmek istenen kavramın tüm bu özniteliklere koşulsal bağlı olduğu bir Bayes ağı olarak da düşünülebilir. Naive Bayes sınıflandırıcısı adını İngiliz matematikçi Thomas Bayes'ten almaktadır.

Naive Bayes Sınıflandırması Makine öğreniminde öğreticili öğrenme alt sınıfındadır. Daha açık bir ifadeyle sınıflandırılması gereken sınıflar(kümeler) ve örnek verilerin hangi sınıflara ait olduğu bellidir. E-posta kutusuna gelen e-postaların spam olarak ayrıştırılması işlemi buna örnek verilebilir. Bu örnekte spam e-posta ve spam olmayan e-posta ayrıştırılacak iki sınıfı temsil eder. Elimizdeki spam ve spam olmayan e-postalardan yaralanarak gelecekte elimize ulaşacak e-postaların spam olup olmadığına karar verecek bir Algoritma da öğreticili makine öğrenmesine örnektir.

4.6.1.2. Genelleştirilmiş Doğrusal Model

İstatistikte, genelleştirilmiş doğrusal model (GLM), normal dağılım dışında hata dağılım modellerine sahip yanıt değişkenlerine izin veren sıradan doğrusal regresyonun esnek bir genellemesidir. GLM, doğrusal modelin bir bağlantı işlevi aracılığıyla yanıt değişkeni ile ilişkilendirilmesine izin vererek ve her ölçümün varyans büyüklüğünün tahmin edilen değerinin bir işlevi olmasına izin vererek doğrusal regresyonu genelleştirir.

Genelleştirilmiş doğrusal modeller, John Nelder ve Robert Wedderburn tarafından doğrusal regresyon, lojistik regresyon ve Poisson regresyonu dahil olmak üzere çeşitli diğer istatistiksel modelleri birleştirmenin bir yolu olarak formüle edildi. [88] Model parametrelerinin maksimum olasılık tahmini için yinelemeli olarak yeniden ağırlıklandırılmış en küçük kareler yöntemi önerdiler. Maksimum olasılık tahmini popüler olmaya devam etmektedir ve birçok istatistiksel hesaplama paketinde varsayılan yöntemdir. Bayesci yaklaşımlar ve varyans stabilize yanıtlara uyan en küçük kareler dahil olmak üzere diğer yaklaşımlar geliştirilmiştir.

Sıradan doğrusal regresyon, belirli bir bilinmeyen miktarın (yanıt değişkeni, rastgele bir değişken) beklenen değerini, bir dizi gözlemlenen değerin (tahmin ediciler) doğrusal bir kombinasyonu olarak tahmin eder. Bu, bir yordayıcıdaki sabit bir değişikliğin, yanıt değişkeninde (yani bir doğrusal yanıt modeli) sürekli bir değişikliğe yol açtığı anlamına gelir. Bu, yanıt değişkeni her iki yönde süresiz olarak veya daha genel olarak tahmin değişkenlerindeki varyasyona kıyasla nispeten küçük bir miktarda değişen herhangi bir miktar için iyi bir yaklaşıma göre değişebildiğinde uygundur; örneğin; insan boyları.

4.6.1.3. Derin Öğrenme

Derin öğrenme, bir veya daha fazla gizli katman içeren sinir ağları ve benzeri makine öğrenme algoritmalarını kapsayan çalışma alanıdır. Yani en az bir adet yapay sinir ağınır kullanıldığı ve birçok algoritma ile bilgisayarın eldeki verilerden yeni veriler elde etmesidir. Derin öğrenme gözetimli, yarı gözetimli veya gözetimsiz olarak gerçekleştirilebilir. Derin yapay sinir ağları pekiştirmeli öğrenme yaklaşımıyla da başarılı sonuçlar vermiştir.

2006'dan beri, derin yapılandırılmış öğrenme veya daha yaygın olarak derin öğrenme veya hiyerarşik öğrenme olarak adlandırılan yeni bir makine öğrenimi araştırması alanı olarak ortaya çıktı. Geçtiğimiz birkaç yıl boyunca, derin öğrenme araştırmalarından geliştirilen teknikler, makine öğrenimi ve yapay zekanın temel yönleri dâhil olmak üzere geleneksel ve yeni, genişletilmiş kapsamlar dâhilinde çok çeşitli sinyal ve bilgi işleme çalışmalarını etkiledi. Son yıllarda bir dizi çalıştay, öğretici ve özel konu veya konferans özel oturumları, özellikle derin öğrenmeye ve çeşitli sinyal ve bilgi işleme alanlarına uygulanmasına ayrılmıştır. Derin öğrenmenin ayrıntılarını açıklamaya başlamadan önce, gerekli tanımları sağlayalım. Derin öğrenmenin birbiriyle yakından ilgili çeşitli tanımları şu şekildedir; [89]

Tanım 1: Denetimli veya denetimsiz özellik çıkarımı ve dönüşümü ile model analizi ve sınıflandırması için doğrusal olmayan bilgi işlemenin birçok katmanından yararlanan bir makine öğrenimi teknikleri sınıfıdır.

Tanım 2: Veriler arasındaki karmaşık ilişkileri modellemek için birden çok temsil düzeyini öğrenmek için algoritmalara dayanan makine öğrenimi içindeki bir alt alan. Daha yüksek seviyeli özellikler ve kavramlar bu nedenle daha düşük seviyeli özellikler açısından tanımlanır ve böyle bir özellik hiyerarşisine derin mimari denir. Bu modellerin çoğu, temsillerin denetimsiz olarak öğrenilmesine dayanmaktadır.

Tanım 3: Çeşitli temsil seviyelerini öğrenmeye dayanan, özelliklerin veya faktörlerin veya kavramların bir hiyerarşisine karşılık gelen, daha yüksek seviyeli kavramların daha düşük seviyeli olanlardan tanımlandığı ve aynı alt seviyeli kavramların yapabileceği bir makine öğrenimi alt alanı birçok üst düzey kavramın tanımlanmasına yardımcı olur. Derin öğrenme, öğrenme temsillerine dayalı daha geniş bir makine öğrenimi yöntemleri ailesinin bir parçasıdır. Bir gözlem (örneğin, bir görüntü) birçok şekilde temsil edilebilir (örneğin, bir piksel vektörü), ancak bazı temsiller, örneklerden ilgilenilen görevleri (örneğin, bu bir insan yüzünün görüntüsü mü?) ve bu alandaki araştırma, neyin daha iyi temsiller yaptığını ve bunların nasıl öğrenileceğini tanımlamaya çalışır.

4.6.1.4. Karar Ağacı

Karar ağacı, ağaç benzeri bir karar modelini ve bunların olası sonuçlarını, şans olaylarının sonuçları, kaynak maliyetleri ve fayda dahil olmak üzere kullanan bir karar destek aracıdır. Yalnızca koşullu kontrol ifadeleri içeren bir algoritmayı görüntülemenin bir yoludur. Karar ağaçları, bir hedefe ulaşma olasılığı en yüksek olan ancak aynı zamanda makine öğreniminde popüler bir araç olan bir stratejinin belirlenmesine yardımcı olmak için özellikle karar analizinde olmak üzere yöneylem araştırmasında yaygın olarak kullanılır.

Karar ağacı, her dâhili düğümün bir öznitelik üzerinde bir "testi" temsil ettiği, her dalın testin sonucunu temsil ettiği ve her yaprak düğümün bir sınıf etiketini temsil ettiği akış şemasına benzer bir yapıdır. Kökten yaprağa giden yollar, sınıflandırma kurallarını temsil eder. Karar analizinde, bir karar ağacı ve yakından ilişkili etki diyagramı, rakip alternatiflerin beklenen değerlerinin (veya beklenen faydasının) hesaplandığı görsel ve analitik bir karar destek aracı olarak kullanılır. Karar ağaçları genellikle yöneylem araştırması ve işlem yönetiminde kullanılır. Uygulamada, kararların eksik bilgi altında hatırlama olmadan çevrimiçi alınması gerekiyorsa, bir karar ağacı en iyi seçim modeli veya çevrimiçi seçim modeli algoritması olarak bir olasılık modeli ile paralelleştirilmelidir. Karar ağaçlarının başka bir kullanımı, koşullu olasılıkları hesaplamak için açıklayıcı bir araçtır. Karar ağaçları, etki diyagramları, fayda fonksiyonları ve diğer karar analizi araçları ve yöntemleri işletme, sağlık ekonomisi ve halk sağlığı okullarındaki lisans öğrencilerine öğretilir ve yöneylem araştırması veya yönetim bilimi yöntemlerinin örnekleridir.

En yüksek bilgi kazanımını sağlayan tahmin edici değişken belirlenir ve ağaç bu değişkenden dallanmaya başlar. Böylece veriler her şubenin altına eşit olarak dağıtılacaktır. İlk

tahminci değişkeni belirlendikten sonra aynı işlem bu sefer toplam entropide değil, belirlenen bu tahmin edici değişkenin bilgi değerinde tekrarlanır. Belirlenen bu değişkenin bölünmesinin daha fazla bilgi kazanımı sağlayacağı kalan tahmin değişkenlerinden hangisi ile hesaplanır. Bu işlem, tüm tahmin değişkenleri ağaca eklenene veya düğümdeki tüm veri örnekleri aynı sınıfa ait olana kadar devam eder.

4.6.1.5. Random Forest

Random Forest (RF) karar ağacı kümesinde, seçilen düğümleri en iyi özniteliklerden dallara ayırmak yerine, her düğümden en iyi rasgele alınan öznitelikleri seçerek tüm düğümleri dallara ayırır. Her veri kümesi, orijinal veri kümesinden yer değiştirebilir şekilde oluşturulur. Ağaçlar rastgele özellik seçimi kullanılarak ve budama yapılmadan geliştirilir. Diğer algoritmalardan daha hızlı ve daha doğrudur.

4.6.1.6. Gradyan Güçlendirilmiş Ağaçlar (Gradient Boosted Trees)

Gradyan artırma, tipik olarak karar ağaçları gibi zayıf tahmin modelleri topluluğu şeklinde bir tahmin modeli üreten regresyon ve sınıflandırma problemleri için bir makine öğrenme tekniğidir. Karar ağacı zayıf bir öğrenci olduğunda, ortaya çıkan algoritmaya, genellikle rastgele ormandan daha iyi performans gösteren gradyan destekli ağaçlar denir. Modeli, diğer artırma yöntemlerinin yaptığı gibi aşama bazında oluşturur ve keyfi bir türevlenebilir kayıp fonksiyonunun optimizasyonuna izin vererek genelleştirir.

4.6.1.7. Destek Vektör Makinesi (Support Vector Machine – SVM)

Destek Vektör Makineleri (Cortes, 1995) (SVM), regresyon analizi ve sınıflandırma için kullanılabilen denetimli bir öğrenme yöntemidir. Sınıflandırmada kullanılan etkili ve doğru yöntemlerden biridir. Sınıflandırma için, iki grup arasında bir düzlemde bir sınır çizilerek iki grubu ayırmak mümkündür. Bu sınırın çizildiği yer, her iki grubun üyelerinden en uzak olanı olmalıdır. Burada SVM, bu sınırın nasıl çizileceğini belirler. Bu işlemi gerçekleştirmek için birbirine yakın ve paralel iki sınır çizgisi çizilir ve bu sınır çizgileri birbirine yaklaştırılarak ortak bir sınır çizgisi oluşturulur.

5. BULGULAR VE TARTIŞMA

Bu bölümde kullanılan yöntemlerin deneysel sonuçları ve önerilen sınıflandırıcı sunulur ve karşılaştırılır. Bu çalışmada, ön işleme, özellik seçimi ve sınıflandırma için RapidMiner Studio programı tercih edilmiş; tüm işlemler RapidMiner Studio ortamında yapılmış ve sonuçlar alınmıştır.

Toplanan ve ardından manuel olarak etiketlenen veri setinde siber zorbalık içeren ve içermeyen verilerin sayısı Tablo 6'da gösterilmektedir.

Tablo 6. Veri Seti Adetleri

VERİ SETİ	TOPLAM İNCELENEN VERİ SAYISI	SİBER ZORBALIK İÇEREN VERİ SAYISI	SİBER ZORBALIK İÇERMEYEN VERİ SAYISI	SİBER ZORBALIK İÇEREN ORANI
TWEET SAYISI	29198	2626	26572	%9

Siber zorbalık içeren tweetlerin araştırılması için belirlenen kategorilerdeki en yüksek takipçiye sahip veya en aktif hesaplar listelenmiştir. Bu liste kamu ve siyaset, kültür-sanat, basın-yayın, spor, iş dünyası, fenomenler şeklinde kategorilerden oluşmaktadır. Bu kategorilerin seçilme sebebi kişilerin halk tarafından bilinen, çokça mesaja ve yoruma maruz kalan, gündemde ismi yer alan ve dolayısıyla siber zorbalığa maruz kalacaklarını düşündüğümüz içindir. Ek 1'de paylaşılan ve 378 kişi bulunan bu liste hazırlanırken yukarıda sayılan kategorilerdeki en çok takipçili ve/veya en aktif hesaplardan seçilmiştir. Bu liste, sosyal medya araştırmaları için kurulmuş olan web siteler yardımıyla listeler tespit edilerek hazırlanmıştır. Örnek olarak çalışmamızda da kullandığımız BoomSocial.com (sosyal medya ölçümleme & analiz) web sitesi verilebilir. İş Dünyası kategorisindeki sayı azlığı, sosyal medyada yüksek takipçi sayısına sahip bu tarz profilin fazla yaygın olmamasıdır. Aşağıdaki Tablo 7'de bu kategorilerde bulunan kişi sayıları gösterilmiştir.

Tablo 7. Kategorilerin Kişi Sayıları

KATEGORİ	Kamu ve Siyaset	Basın-Yayın	Kültür Sanat	Spor	İş Dünyası	Fenomenler	TOPLAM
KİŞİ SAYISI	75	84	77	60	37	45	378

Toplam belirlediğimiz 6 kategori için atılan tweetlerin ve siber zorbalık tespitlerinin dağılımı da Tablo 8'de gösterilmiştir.

Tablo 8. Kategorilerin Tweet Dağılımı ve Siber Zorbalık Tespit Oranları

KATEGORİ	TWEET SAYISI	SİBER ZORBALIK TESPİT SAYISI	SİBER ZORBALIK ORANI
Basın-Yayın	7205	533	%7
Kültür Sanat	6116	637	%10
Kamu ve Siyaset	5794	505	%9
Spor	4214	371	%9
Fenomenler	3261	453	%14
İş Dünyası	2608	127	%5
TOPLAM	**29198**	**2626**	**%9**

Tablo 8'deki değerler incelendiğinde Fenomenler için siber zorbalık ihtimalinin daha yüksek olduğu görülebilmektedir. Bunun yanı sıra İş Dünyası kategorisindeki düşük oran da dikkat çekmiştir. Diğer kategorilerdeki oranların ise toplam orana yakın olduğu görülmüştür.

29198 adet tweetin okunması ve etiketlenmesi tamamen manuel olarak yapılmıştır. Tweetler, 1-7 Şubat 2021 tarihleri arasındaki en son ve popüler tweetler olarak çekilmiş ve MS Excel dosyasına aktarılmıştır. Siber zorbalık tespiti yapılırken Xu ve ark. [4] çalışmasında yapıldığı gibi öfkeli, utanç verici, empati yoksunu, korkutucu, gurur kırıcı, üzüntü verici vb. ifadelerin yer almasına özen gösterilmiştir.

Tweetlerin çekildiği tarihlerde derbi müsabakası, Boğaziçi Üniversitesi gündemi gibi siber zorbalığa götürecek fanatiklik ve taraflı olmayı arttıracak gündemlerin olması ya da MEB yüz yüze eğitim hakkında karar açıklanması vb. gibi çokça konuşulacak gündemlerin olması oranı arttırabilmektedir.

Çalışmamızda özellikle "Text" ve "Kategori" bilgisinin siber zorbalık tespitinde etkili olup olmadığı araştırılmıştır. Ayrıca tweetlerin atıldığı kaynağa göre Mobil ve Web olarak iki "kaynak cihaz" bilgisi elde edilmiştir. Bu kaynak cihaz bilgisinin de siber zorbalığın tespitinde etkili olan bir bilgi olup olmadığı da araştırılmıştır.

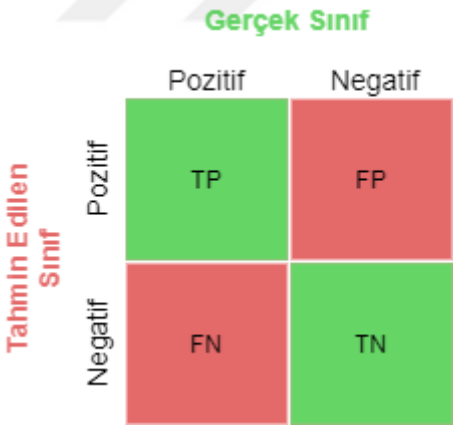
Çalışmamızda çok fazla tweet olması ve siber zorbalık içerenlerin oranca düşük kalması sebebiyle elde ettiğimiz 2626 adet siber zorbalık içerikli veri ile diğerleri arasından rastgele seçilmiş aynı sayıda yani 2626 adet siber zorbalık içermeyen veri modele verilmiş ve sınıflandırma için girdi sağlanmıştır. Aksi takdirde modellerimiz %100'e yakın bir

performansla çalışacak ancak siber zorbalık tespiti ile alakalı gerçekçi ve kullanılabilir bilgiler vermeyecektir.

5.1. Auto Model Tahmin Modellerinin Sonuçları

Bu bölümde RapidMiner Stuio programının Auto Model üzerinden gerçekleştirdiğimiz sınıflandırma ve tahmin modellerinin kıyaslanması ve detaylarının açıklanmasını ele alacaktır. 4.3.1 Sınıflandırma ve Tahmin bölümünde de detaylı olarak anlatıldığı gibi Auto Model; RapidMiner Studio'nun, model oluşturma ve modeli uygulama konusunda kolaylık sağlayan ve hız kazandıran bir özelliğidir. Tahmin, Kümeleme ve Aykırı değerleri tespit etme olarak 3 temel işlevi bulunmaktadır. Çalışmamızda kullandığımız ise Tahmin işlevidir. Auto Model verileri otomatik olarak değerlendirir ve problem çözümü için modeller sunar ve hesaplamalar tamamlandıktan sonra bu modellerin sonuçlarını karşılaştırmada yardımcı olur.

Makine öğrenmesinde; modellerin performansını değerlendirmek, karşılaştırmak ve analiz etmek için kullanılan, tahmin değerleriyle gerçek değerlerin karşılaştırıldığı hata matrisi (confusion matrix) sıkça kullanılmaktadır. Bu matris Şekil 27'de gösterilmiştir. Çalışmamızda, model sonuçları değerlendirildiğinde bu matristen de faydalanılmıştır.



Şekil 27. Hata Matrisi [90]

- TP - True Positive (Doğru Pozitif) : Pozitif değerler içerisinden, pozitif olarak doğru bir şekilde tahmin edilenlerin sayısıdır.
- TN – True Negative (Doğru Negatif) : Negatif değerler içerisinden, negatif olarak doğru bir şekilde tahmin edilenlerin sayısıdır.
- FP - False Positive (Yanlış Pozitif) : Negatif değerler içerisinden, pozitif olarak yanlış bir şekilde tahmin edilenlerin sayısıdır.

- FN – False Negative (Yanlış Negatif) : Pozitif değerler içerisinden, negatif olarak yanlış bir şekilde tahmin edilenlerin sayısıdır.

Çalışmamıza uygulanan Auto Model işlemleri neticesinde elde edilen tüm modelleri incelemeden önce model sonuçlarında detayları verilen hata türleri ve performans metrikleri açıklanmıştır. İkili sınıflandırma yöntemlerinde modelin performansını gösteren duyarlılık (recall), kesinlik (precision) ve f-ölçeği (f-measure) değerleri Tablo 9'da açıklanmıştır. Hata türlerinin sembol ve tanımları da Tablo 10'da gösterilmiştir. Tablo 9'daki sembolleri daha açıklayıcı olarak göstermek adına Şekil 27'deki şema verilmiştir. Seçilen değerler arasında yeşil ile boyalı olan üstteki kısım doğru şekilde sınıflandırılmış pozitifleri gösterirken kırmızı ile boyalı üstteki kısım yanlış sınıflandırılmış pozitifleri göstermektedir. Buradaki üst yeşil kısmın tüm seçilmiş pozitif kısma (yeşil üst kısım + kırmızı üst kısım) oranı kesinlik (precision) değerini vermektedir. Benzer şekilde Şekil 27'deki yeşil üst kısmın (doğru şekilde sınıflandırılmış pozitifler) seçilmemiş olan ama gerçekte pozitif olan değerlere oranı da duyarlılık (recall) değerini göstermektedir.

Tablo 9. Sınıflandırma Performans Metriği

Değer	Sembolü	Tanımı
Recall (Duyarlılık)	$\frac{DoğruPozitifler}{DoğruPozitifler + YanlışNegatifler}$	Doğru sınıflandırılmış pozitiflerin, tüm gerçek pozitiflere oranıdır.
Precision (Kesinlik)	$\frac{DoğruPozitifler}{DoğruPozitifler + YanlışPozitifler}$	Doğru sınıflandırılmış pozitiflerin, tüm pozitif şekilde sınıflandırılmış örneklere oranıdır.
F-Measure (F Ölçeği)	$\frac{2 * duyarlılık * kesinlik}{duyarlılık + kesinlik}$	Sınıflandırma analizinde bir performans ölçüsüdür. Kesinlik ve duyarlılık değerlerinin harmonik ortalamasıdır.
Doğruluk (Accuracy)	$\frac{DoğruPozitifler + Doğru\ Negatifler}{Pozitifler + Negatifler}$	Doğru olarak sınıflandırılmış pozitif ve negatiflerin, tüm diğer pozitif ve negatiflerin toplamına oranıdır. Doğru tahmin adedinin toplam adede oranıdır.

Tablo 10. Hata Türü Sembol ve Tanımları

Hata Türü	Sembolü	Tanımı
Kök Ortalama Kare Hata **(Root Mean Square Error - RMSE)**	$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Y_i - X_i)}$	Bir makine öğrenmesi modelinin, tahminleyicinin tahmin ettiği değerler ile gerçek değerleri arasındaki uzaklığın bulunmasında sıklıkla kullanılan, hatanın büyüklüğünü ölçen kuadratik bir metriktir. Tahmin hatalarının standart sapmasıdır. Başka bir deyişle, verilere en iyi uyan çizgi etrafında o verilerin ne kadar yoğun olduğunu söyler.
Mutlak Hata **(Mean Absolute Error - MAE)**	$\text{MAE} = \frac{\sum_{i=1}^{n}\lvert y_i - x_i\rvert}{n}$	Ortalama mutlak hata iki sürekli değişken arasındaki farkın ölçüsüdür. MAE, her gerçek değer ile veriye en iyi uyan çizgi arasındaki ortalama dikey mesafedir. MAE aynı zamanda her veri noktası ile en iyi uyan çizgi arasındaki ortalama yatay mesafedir.
Göreceli Hata **(Relative Error – RE)**	RE = MAE / Gerçek değer	Ortalama göreceli hata, tahminin gerçek değerden mutlak sapmasının ortalamasının gerçek değere bölünmesidir. Yüzde olarak ifade edilir.
Ortalama Kare Hata **(Mean Squared Error - MSE)**	$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - X_i)^2$	Basitçe, ortalama kare hata bir regresyon eğrisinin bir dizi noktaya ne kadar yakın olduğunu söyler. MSE, bir makine öğrenmesi modelinin, tahminleyicinin performansını ölçer, her zaman pozitif değerlidir ve MSE değeri sıfıra yakın olan tahminleyicilerin daha iyi bir performans gösterdiği söylenebilir.
Korelasyon (r)	$r = \frac{\sum(xy) - (\sum x)(\sum y)/n}{\sqrt{(\sum x^2 - (\sum x)^2/n)\,(\sum y^2 - (\sum y)^2/n)}}$	Korelasyon iki ya da daha fazla değişken arasındaki doğrusal ilişkiyi gösterir. Auto Model, etiket ve tahmin özellikleri arasındaki korelasyon katsayısını döndürür.

Tablo 10'da gösterilen değerler makine öğrenmesi modellerinin tahmin ettiği değerlerin performansı ifade eden metriklerdendir. Çalışmamızda modellerin değerlendirilmesinde bu metrikler de kullanılmıştır.

Çalışmamızda kullandığımız Auto Model'in sunduğu sonuçlar aşağıdaki başlıklar altında gösterilmiştir. Bu modeller; 4.6.1 Sınıflandırma ve Tahmin başlığı altında da anlatılan Naive Bayes, Genelleştirilmiş Doğrusal Model, Derin Öğrenme, Karar Ağacı, Random Forest, Gradient Boosted Trees ve Support Vector Machine modelleridir. Modellerin kullanılması aşamasında RapidMiner Auto Model özelliğinin sunduğu ön tanımlı değerler ile set edilmiştir. Bu parametreler her bir model için optimize edilerek daha iyi sonuçlar elde edilebilir. Bu ayrı bir çalışma konusudur.

5.1.1. Naive Bayes Model Sonuçları

Verimizin ön işlemeden geçirilmesi ve etiketlenmesinin ardından Auto Model çalıştırılmıştır. RapidMiner Studio Auto Model'in sunduğu sınıflandırma modellerinden biri olan Naive Bayes Modeli kullanılmıştır. Elde edilen sonuçlar Tablo 11'de gösterilmiştir.

Tablo 11. Naive Bayes Model Sonuçları

MODEL	Naive Bayes (with RapidMiner Studio Auto Model)
Toplam Süre	1 dk 26 sn
Doğruluk	90,25%
Sınıflandırma Hatası	8,89%
RMSE	0,185
MAE	0,098
RE	7,30%
MSE	0,041
Korelasyon	0,077
Precision	0,908
Recall	0,878
F-Measure	0,893

Hata Matrisi

		GERÇEK	
		POZİTİF	NEGATİF
TAH	POZİTİF	(TP) 2129	(FP) 216

	NEGATİF	(FN) 296	(TN) 2611

Tablo 11'de görüldüğü gibi RapidMiner Auto Model yoluyla Naive Bayes sınıflandırma modeli kullanılarak yapılan sınıflandırmada %90,25 oranında bir doğruluk elde edilmiştir. Bu sonuç sınıflandırma problemleri için yeterli ve memnun edici bir sonuçtur. Modelin işlem süresi 1 dk 26 sn sürmüştür. Hata matrisindeki değerler incelendiğinde pozitif olarak sınıflandırılan (tahmin edilen) 2345 adet verinin 2129 tanesi doğru tahmin edilmişken 216 tanesi gerçeğinden farklı olarak yanlış sınıflandırılmıştır. Aynı şekilde negatif olarak sınıflandırılan toplam 2907 adet verinin 2611 tanesi doğru tahmin edilmişken 296 tanesi gerçeğinden farklı olarak yanlış sınıflandırılmıştır.

Hata matrisindeki değerler ile bulunan precision (kesinlik), recall (duyarlılık) ve f-measure (f-ölçeği) metriklerinin değerleri Tablo 11'de gösterilmiştir. Bu değerlere göre f-ölçeği değeri 0,893 bulunmuştur.

5.1.2. Genelleştirilmiş Doğrusal Model Sonuçları

Verimizin ön işlemeden geçirilmesi ve etiketlenmesinin ardından Auto Model çalıştırılmıştır. RapidMiner Studio Auto Model'in sunduğu sınıflandırma modellerinden biri olan Genelleştirilmiş Doğrusal Model kullanılmıştır. Elde edilen sonuçlar Tablo 12'de gösterilmiştir.

Tablo 12'de görüldüğü gibi RapidMiner Auto Model yoluyla Genelleştirilmiş Doğrusal Model kullanılarak yapılan sınıflandırmada %80,73 oranında bir doğruluk elde edilmiştir. Bu sonuç diğer sınıflandırma modellerine göre daha düşük performanslıdır. Model 4 dk. 38 sn. süre ile çalışmıştır.

Hata matrisindeki değerler incelendiğinde pozitif olarak sınıflandırılan (tahmin edilen) 2566 adet verinin 1984 tanesi doğru tahmin edilmişken 582 tanesi gerçeğinden farklı olarak yanlış sınıflandırılmıştır. Aynı şekilde negatif olarak sınıflandırılan toplam 2686 adet verinin 2256 tanesi doğru tahmin edilmişken 430 tanesi gerçeğinden farklı olarak yanlış sınıflandırılmıştır.

Tablo 12. Genelleştirilmiş Doğrusal Model Sonuçları

MODEL	Genelleştirilmiş Doğrusal Model (with RapidMiner Studio Auto Model)
Toplam Süre	4 dk 38 sn
Doğruluk	80,73%
Sınıflandırma Hatası	19,27%
RMSE	0,222
MAE	0,113
RE	%11,10
MSE	0,005
Korelasyon	0,045
Precision	0,773
Recall	0,822
F-Measure	0,797

Hata Matrisi

		GERÇEK	
		POZİTİF	NEGATİF
TAHMİN	POZİTİF	(TP) 1984	(FP) 582
	NEGATİF	(FN) 430	(TN) 2256

Hata matrisindeki değerler ile bulunan precision (kesinlik), recall (duyarlılık) ve f-measure (f-ölçeği) metriklerinin değerleri Tablo 12'de gösterilmiştir. Bu değerlere göre f-ölçeği değeri 0,797 bulunmuştur.

Model sonuçları incelendiğinde (bazı) "text", "kategori", ""to-user" ve "kaynak cihaz" niteliklerinin sonuç için ne kadar etkili ve sonuçla ne kadar ilgili modele özel bir sonuç vermiştir. Bu sonuç da Tablo 13'te gösterilmiştir.

Bir kısmı gösterilen Tablo 13'te yer alan bilgilere göre çalışmamızda araştırmaya çalıştığımız konulardan biri olan Kategori siber zorbalık tespitinde etkisi olup olmadığı konusunda verdiği sonuçlar incelenmiştir. Kamu ve Siyaset kategorisinde atılan tweetlerin daha çok siber zorbalık içerme eğiliminde olduğu görülmüştür. Buradan hareketle devlet büyüklerine hakaret konusunda ayrıca ve detaylı bir çalışma yapılabileceği öngörülmüştür.

Tablo 13. Genelleştirilmiş Doğrusal Modelinde Niteliklerin Sonuca Etkisi

ÖZNİTELİK	KATSAYI
KATEGORİ	
Kategori: Basın-Yayın	-0,062
Kategori: Fenomenler	-0,032
Kategori: Kamu ve Siyaset	0,043
Kategori: Kültür Sanat	0
Kategori: Spor	0
Kategori: İş Dünyası	0
KAYNAK CİHAZ	
Kaynak Cihaz: Mobil	0,017
Kaynak Cihaz: Web	-0,032
TEXT	
Text: Defol	0,980
Text: Ulan	0,950
Text: Hain	0,936
Text: Terör	0,727
Text: Rezil	0,690
…	…

Yine benzer şekilde tweetin atıldığı kaynağın mobil mi yoksa web mi olduğu bilgisi de siber zorbalık tespitinde incelenmiştir. Mobil kaynaktan atılan tweetlerin daha çok siber zorbalık içerikli olduğu ve web kaynağından atılan tweetlerin ise siber zorbalık içermeme eğiliminde olduğu sonucuna varılmıştır.

Son olarak ise tweetin içerdiği metin bilgisinin beklenen şekilde siber zorbalık tespitinde kullanılabilir olduğu görülmüştür.

5.1.3. Derin Öğrenme Model Sonuçları

Verimizin ön işlemeden geçirilmesi ve etiketlenmesinin ardından Auto Model çalıştırılmıştır. RapidMiner Studio Auto Model'in sunduğu sınıflandırma modellerinden biri olan Derin Öğrenme Modeli kullanılmıştır. Elde edilen sonuçlar Tablo 14'te gösterilmiştir.

Tablo 14'te görüldüğü gibi RapidMiner Auto Model yoluyla Genelleştirilmiş Doğrusal Model kullanılarak yapılan sınıflandırmada %88,19 oranında bir doğruluk elde edilmiştir. Bu sonuç diğer sınıflandırma modellerine kıyasla orta düzeyde bir performans göstermiştir. Model 1 dk. 39 sn. süre ile çalışmıştır.

Tablo 14. Derin Öğrenme Model Sonuçları

MODEL	Derin Öğrenme (with RapidMiner Studio Auto Model)
Toplam Süre	1 dk 39 sn
Doğruluk	88,19%
Sınıflandırma Hatası	11,81%
RMSE	0,219
MAE	0,134
RE	%13,40
MSE	0,048
Korelasyon	0,078
Precision	0,861
Recall	0,893
F-Measure	0,877

Hata Matrisi

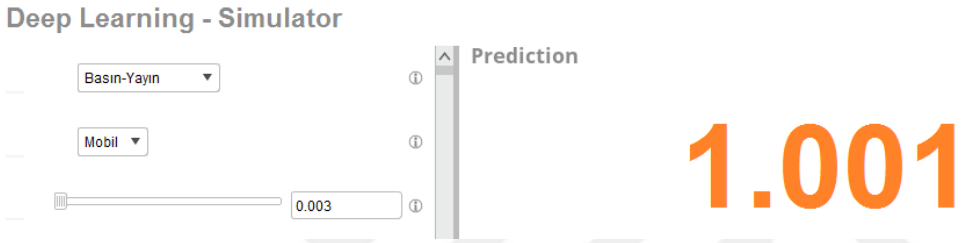
		GERÇEK	
		POZİTİF	NEGATİF
TAHMİN	POZİTİF	(TP) 2204	(FP) 357
	NEGATİF	(FN) 263	(TN) 2428

Hata matrisindeki değerler incelendiğinde pozitif olarak sınıflandırılan (tahmin edilen) 2561 adet verinin 2204 tanesi doğru tahmin edilmişken 357 tanesi gerçeğinden farklı olarak yanlış sınıflandırılmıştır. Aynı şekilde negatif olarak sınıflandırılan toplam 2691 adet verinin 2428 tanesi doğru tahmin edilmişken 263 tanesi gerçeğinden farklı olarak yanlış sınıflandırılmıştır.

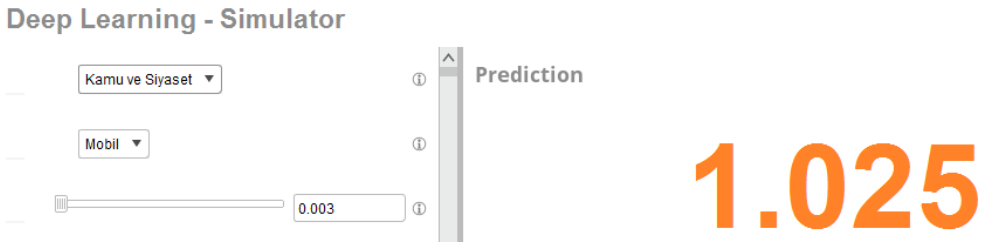
Hata matrisindeki değerler ile bulunan precision (kesinlik), recall (duyarlılık) ve f-measure (f-ölçeği) metriklerinin değerleri Tablo 14'te gösterilmiştir. Bu değerlere göre f-ölçeği değeri 0,877 bulunmuştur.

Modelin "Simülatör" sonucunda yer alan değerler incelendiğinde de bazı sonuçlara varmak mümkün olacaktır. Simülatör ekranında; sol panelde tüm öznitelikler için giriş ayarları değiştirilebilirken, sağ panelde bu değişikliklerin sonuca nasıl yansıdığını hesaplamaktadır ve gerçek zamanlı olarak göstermektedir.

İlk olarak; kategori özniteliğindeki değişimlerin sonuca nasıl yansıdığı araştırılmıştır. Bunun için; text ve kaynak cihaz değerleri sabit tutularak kategori alanındaki değerler değiştirilmiştir ve Simülatör sonucundaki değişimler göz önünde bulundurularak kategorinin siber zorbalık tespitindeki etkisi anlaşılmıştır. Örnek olarak; tüm diğer öznitelikler sabitken kategori özniteliği Basın-Yayın olarak seçilmiştir. Buradaki tahmin sonucu Şekil 28'de gösterilmiştir. Ardından tüm diğer öznitelikler sabitken kategori özniteliği Kamu ve Siyaset olarak seçilmiştir. Buradaki tahmin Şekil 29'da gösterilmiştir. Her iki sonuca bakıldığında Şekil 29'daki tahmin değerinin daha büyük olduğu görülmüştür. Bu da kategori özniteliğinde Kamu ve Siyaset değerinin Basın-Yayın değerine göre sonuca etkisinin daha fazla olduğunu göstermektedir. Bu işlemler tüm Kategori değerleri için tekrarlanmıştır.



Şekil 28. Deep Learning Simulatör (Basın-Yayın)



Şekil 29. Deep Learning Simulatör (Kamu ve Siyaset)

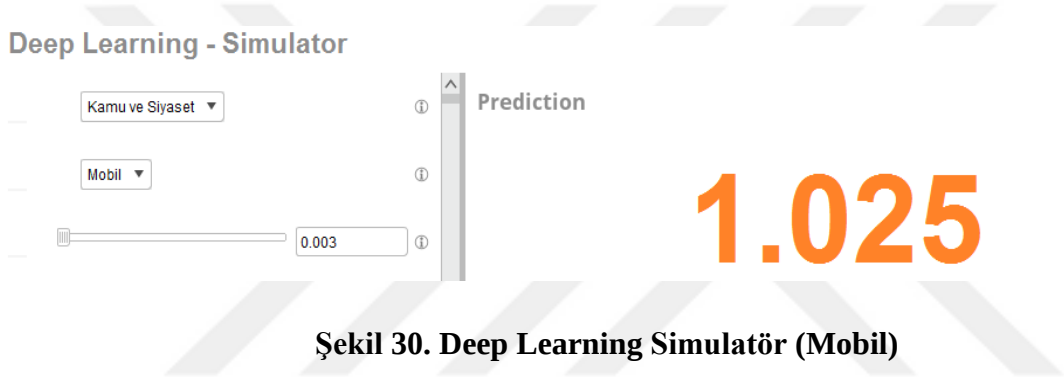
Deep Learning-Simulatör yukarıda 2 tane örneği verilen şekilde diğer tüm kategoriler için çalıştırıldığında Kamu ve Siyaset kategorisinin yine en etkili kategori olduğu görülmüştür. Diğer kategorileri de sıralayacak olursak sonuç şu şekilde olmaktadır.

- Kamu ve Siyaset > Spor > Kültür Sanat = Fenomenler > Basın-Yayın > İş Dünyası

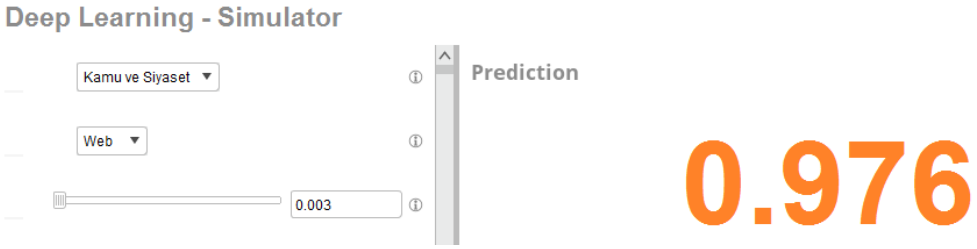
Kamu ve Siyaset kategorisinden sonra ise en etkili olan kategorinin Spor olduğu görülmüştür. Kültür Sanat ile Fenomenler kategorisinin eşit olduğu bu sonuçta ise peşinden Basın-Yayın ve İş Dünyası kategorileri gelmektedir. Bu anlamda bakıldığında Kategori

bilgisinin siber zorbalık tespiti için etkili olduğu ve özellik olarak kullanılabileceği görülmüştür.

İkinci olarak; Kaynak Cihaz değişikliğinin sonuca etkisi incelenmiştir. Bunun için; text ve kategori değerleri sabit tutularak Kaynak Cihaz özniteliğindeki değerler değiştirilmiştir ve Simülatör sonucundaki değişimler göz önünde bulundurularak kaynak cihazın siber zorbalık tespitindeki etkisi anlaşılmıştır. Diğer tüm değerler sabitken kaynak cihaz değeri Mobil olarak seçilmiştir. Bu değerlere göre çıkan tahmin sonucu Şekil 30'da gösterilmiştir. Ardından yine tüm değerler sabitken kaynak cihaz değeri Web olarak seçilmiştir ve sonucu Şekil 31'degösterilmiştir.



Şekil 30. Deep Learning Simulatör (Mobil)



Şekil 31. Deep Learning Simulatör (Web)

Şekil 30 ve Şekil 31'deki sonuçlar incelendiğinde, Şekil 30'daki tahmin değerinin daha büyük olduğu görülmüştür. Bu da Mobil kaynaktan atılan tweetlerin siber zorbalık içermesi konusunda daha etkili olduğu anlamına gelmektedir. Bu Simülatör işlemi yapılırken Kategori özniteliğindeki değer Kamu ve Siyaset olarak sabitlenmiştir. Simülatör işlemi diğer kategoriler için de kaynak cihaz bilgisi değiştirilerek incelendiğinde sonucun değişmediği ve Mobil'in her sonuçta yine Web'ten daha etkili olduğu görülmüştür.

5.1.4. Karar Ağacı Model Sonuçları

Verimizin ön işlemeden geçirilmesi ve etiketlenmesinin ardından Auto Model çalıştırılmıştır. RapidMiner Studio Auto Model'in sunduğu sınıflandırma modellerinden biri olan Karar Ağacı Modeli kullanılmıştır. Elde edilen sonuçlar Tablo 15'te gösterilmiştir.

Tablo 15'te görüldüğü gibi RapidMiner Auto Model yoluyla Karar Ağacı Modeli kullanılarak yapılan sınıflandırmada %89,20 oranında bir doğruluk elde edilmiştir. Bu sonuç diğer sınıflandırma modellerine kıyasla orta düzeyde bir performans göstermiştir. Model 1 dk. 2 sn. süre ile çalışmıştır.

Hata matrisindeki değerler incelendiğinde pozitif olarak sınıflandırılan (tahmin edilen) 2773 adet verinin 2494 tanesi doğru tahmin edilmişken 279 tanesi gerçeğinden farklı olarak yanlış sınıflandırılmıştır. Aynı şekilde negatif olarak sınıflandırılan toplam 2479 adet verinin 2191 tanesi doğru tahmin edilmişken 288 tanesi gerçeğinden farklı olarak yanlış sınıflandırılmıştır. Hata matrisindeki değerler ile bulunan precision (kesinlik), recall (duyarlılık) ve f-measure (f-ölçeği) metriklerinin değerleri Tablo 15'te gösterilmiştir. Bu değerlere göre f-ölçeği değeri 0,898 bulunmuştur.

Tablo 15. Karar Ağacı Model Sonuçları

MODEL	Karar Ağacı Modeli (with RapidMiner Studio Auto Model)
Toplam Süre	1 dk 2 sn
Doğruluk	89,20%
Sınıflandırma Hatası	10,80%
RMSE	0,777
MAE	0,061
RE	%6,20
MSE	0,031
Korelasyon	0,024
Precision	0,899
Recall	0,896
F-Measure	0,898

Hata Matrisi

		GERÇEK	
		POZİTİF	NEGATİF
TAHMİN	POZİTİF	(TP) 2494	(FP) 279
	NEGATİF	(FN) 288	(TN) 2191

5.1.5. Random Forest Model Sonuçları

Verimizin ön işlemeden geçirilmesi ve etiketlenmesinin ardından Auto Model çalıştırılmıştır. RapidMiner Studio Auto Model'in sunduğu sınıflandırma modellerinden biri olan Random Forest Modeli kullanılmıştır. Elde edilen sonuçlar Tablo 16'da gösterilmiştir.

Tablo 16. Random Forest Model Sonuçları

MODEL	Random Forest (with RapidMiner Studio Auto Model)
Toplam Süre	10 dk 21 sn
Doğruluk	82,31%
Sınıflandırma Hatası	17,69%
RMSE	0,177
MAE	0,063
RE	%6,40
MSE	0,031
Korelasyon	0,031
Precision	0,789
Recall	0,825
F-Measure	0,806

Hata Matrisi

		GERÇEK	
		POZİTİF	NEGATİF
TAHMİN	POZİTİF	(TP) 1935	(FP) 518
	NEGATİF	(FN) 411	(TN) 2388

Tablo 16'da görüldüğü gibi RapidMiner Auto Model yoluyla Karar Ağacı Modeli kullanılarak yapılan sınıflandırmada %82,31 oranında bir doğruluk elde edilmiştir. Bu sonuç diğer sınıflandırma modellerine kıyasla düşük düzeyde bir performans göstermiştir. Model 10 dk. 21 sn. süre ile çalışmıştır.

Hata matrisindeki değerler incelendiğinde pozitif olarak sınıflandırılan (tahmin edilen) 2453 adet verinin 1935 tanesi doğru tahmin edilmişken 518 tanesi gerçeğinden farklı olarak yanlış sınıflandırılmıştır. Aynı şekilde negatif olarak sınıflandırılan toplam 2799 adet verinin 2388 tanesi doğru tahmin edilmişken 411 tanesi gerçeğinden farklı olarak yanlış sınıflandırılmıştır.

Hata matrisindeki değerler ile bulunan precision (kesinlik), recall (duyarlılık) ve f-measere (f-ölçeği) metriklerinin değerleri Tablo 16'da gösterilmiştir. Bu değerlere göre f-ölçeği değeri 0,806 bulunmuştur.

Bu modelde benzer şekilde Simülatör çalıştırıldığında kategori değerleri ile değişimler izlenmiş ve aşağıdaki gibi bir sıralama yapılmıştır.

- Kamu ve Siyaset > Kültür Sanat > Spor > Basın-Yayın > Fenomenler > İş Dünyası

Bu sonuçlara göre Kamu ve siyaset kategorisi ilk sırada iken Kültür ve Sanat kategorisi onu takip etmiştir. Kategori özelliğinin yine siber zorbalık tespitinde önemli olduğu anlaşılmıştır. Bu model de Mobil kaynak, Web kaynağından iyi sonuç vermiş olsa da Web örneğinin arttırılması gerektiği de göz önünde bulundurulmalıdır.

5.1.6. Gradient Boosted Trees Model Sonuçları

Verimizin ön işlemeden geçirilmesi ve etiketlenmesinin ardından Auto Model çalıştırılmıştır. RapidMiner Studio Auto Model'in sunduğu sınıflandırma modellerinden biri olan Gradient Boosted Trees Modeli kullanılmıştır. Elde edilen sonuçlar Tablo 17'de gösterilmiştir.

Tablo 17'de görüldüğü gibi RapidMiner Auto Model yoluyla Gradient Boosted Trees Modeli kullanılarak yapılan sınıflandırmada %81,13 oranında bir doğruluk elde edilmiştir. Bu sonuç diğer sınıflandırma modellerine kıyasla düşük düzeyde bir performans göstermiştir. Model 29 dk. 23 sn. süre ile en uzun süre çalışan model olmuştur.

Hata matrisindeki değerler incelendiğinde pozitif olarak sınıflandırılan (tahmin edilen) 2494 adet verinin 1975 tanesi doğru tahmin edilmişken 519 tanesi gerçeğinden farklı olarak yanlış sınıflandırılmıştır. Aynı şekilde negatif olarak sınıflandırılan toplam 2758 adet verinin 2286 tanesi doğru tahmin edilmişken 472 tanesi gerçeğinden farklı olarak yanlış sınıflandırılmıştır.

Tablo 17. Gradient Boosted Trees Model Sonuçları

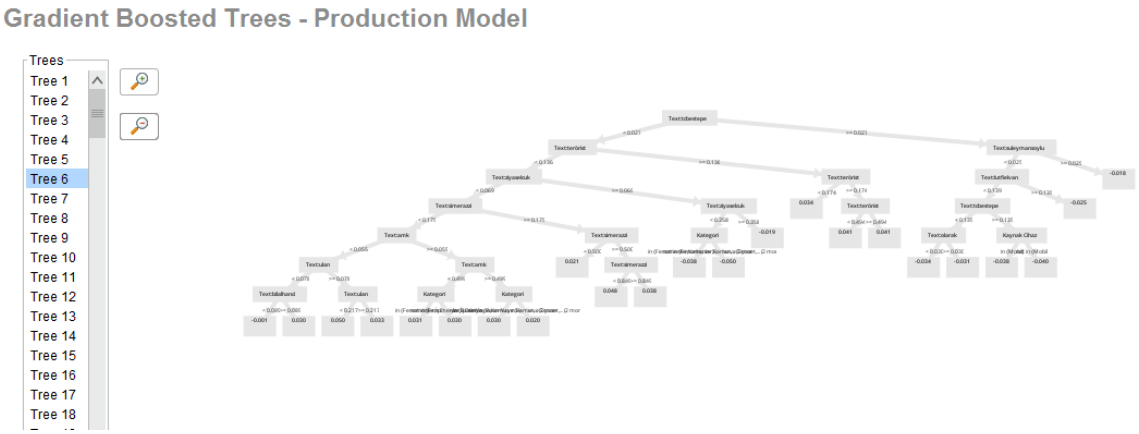
MODEL	Gradient Boosted Trees (with RapidMiner Studio Auto Model)
Toplam Süre	29 dk 23 sn
Doğruluk	81,13%
Sınıflandırma Hatası	18,87%
RMSE	0,179
MAE	0,061
RE	%6,30
MSE	0,032
Korelasyon	0,073
Precision	0,792
Recall	0,807
F-Measure	0,799

Hata Matrisi

		GERÇEK	
		POZİTİF	NEGATİF
TAHMİN	POZİTİF	(TP) 1975	(FP) 519
	NEGATİF	(FN) 472	(TN) 2286

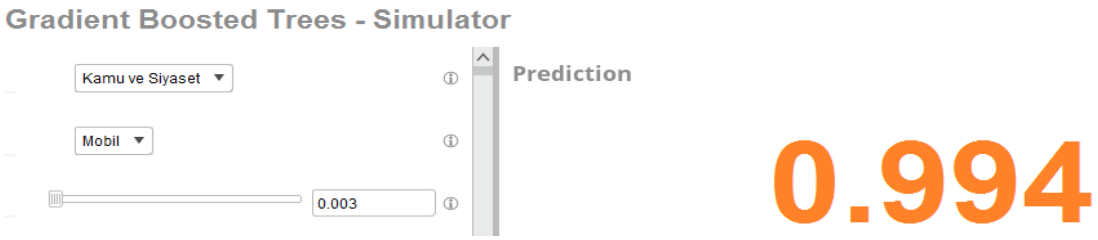
Hata matrisindeki değerler ile bulunan precision (kesinlik), recall (duyarlılık) ve f-measere (f-ölçeği) metriklerinin değerleri Tablo 17'de gösterilmiştir. Bu değerlere göre f-ölçeği değeri 0,799 bulunmuştur.

Model sonucunda oluşan örnek ağaçlar listelenebilmektedir ve bir örneği Şekil 32'de gösterilmiştir. Siber zorbalık tespitini yaparken her bir adımda hangi niteliklerin ne kadar etkili olduğunu anlatan bir şekildir. Karar ağacının iyileştirilmiş ve tekrarlanmış halidir.

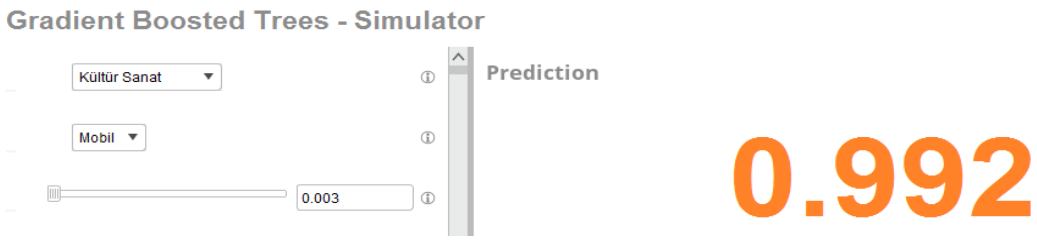


Şekil 32. Gradient Boosted Tree Model Sonucu Örnek Ağaç

Bu modelde de Simülatör çalıştırılmıştır ve ilk olarak; kategori özniteliğindeki değişimlerin sonuca nasıl yansıdığı araştırılmıştır. Bunun için; text ve kaynak cihaz değerleri sabit tutularak kategori alanındaki değerler değiştirilmiştir ve Simülatör sonucundaki değişimler göz önünde bulundurularak kategorinin siber zorbalık tespitindeki etkisi anlaşılmıştır. Örnek olarak; tüm diğer öznitelikler sabitken kategori özniteliği Kamu ve Siyaset olarak seçilmiştir. Buradaki tahmin sonucu Şekil 33'te gösterilmiştir. Ardından tüm diğer öznitelikler sabitken kategori özniteliği Kültür Sanat olarak seçilmiştir. Buradaki tahmin Şekil 34'te gösterilmiştir. Her iki sonuca bakıldığında Şekil 33'teki tahmin değerinin daha büyük olduğu görülmüştür. Bu da kategori özniteliğinde Kamu ve Siyaset değerinin Kültür Sanat değerine göre birbirlerine çok yakın da olsa sonuca etkisinin daha fazla olduğunu göstermektedir. Bu işlemler tüm Kategori değerleri için tekrarlanmıştır.



Şekil 33. Gradient Boosted Tree Simulatör (Kamu ve Siyaset)



Şekil 34. Gradient Boosted Tree Simulatör (Kültür Sanat)

Tüm kategorilere tek tek bakıldığında çıkan sonuçlara göre ise sıralama şu şekildedir;

- Kamu ve Siyaset > Kültür Sanat > Spor > Fenomenler > Basın-Yayın > İş Dünyası

Bu sonuçlara göre Kamu ve siyaset kategorisi ilk sırada iken Kültür ve Sanat kategorisi onu takip etmiştir. İş Dünyası yine son sıradadır. Modelin %6,3 göreceli hata ile çalıştığı da göz önünde bulundurulduğunda Kategori özelliğinin yine siber zorbalık tespitinde önemli olduğu anlaşılmıştır. Bu model de Mobil kaynak Web kaynağından iyi sonuç vermiştir.

5.1.7. Support Vector Machine Model Sonuçları

Verimizin ön işlemeden geçirilmesi ve etiketlenmesinin ardından Auto Model çalıştırılmıştır. RapidMiner Studio Auto Model'in sunduğu sınıflandırma modellerinden biri olan Random Forest Modeli kullanılmıştır. Elde edilen sonuçlar Tablo 18'de gösterilmiştir.

Tablo 18'de görüldüğü gibi RapidMiner Auto Model yoluyla Gradient Boosted Trees Modeli kullanılarak yapılan sınıflandırmada %91,11 oranında bir doğruluk elde edilmiştir. Bu sonuç diğer sınıflandırma modellerine kıyasla en iyi performansı göstermiştir. Model 5 dk. 7 sn. süre ile çalışmıştır.

Hata matrisindeki değerler incelendiğinde pozitif olarak sınıflandırılan (tahmin edilen) 2723 adet verinin 2484 tanesi doğru tahmin edilmişken 239 tanesi gerçeğinden farklı olarak yanlış sınıflandırılmıştır. Aynı şekilde negatif olarak sınıflandırılan toplam 2529 adet verinin 2301 tanesi doğru tahmin edilmişken 228 tanesi gerçeğinden farklı olarak yanlış sınıflandırılmıştır.

Tablo 18. Support Vector Machine Model Sonuçları

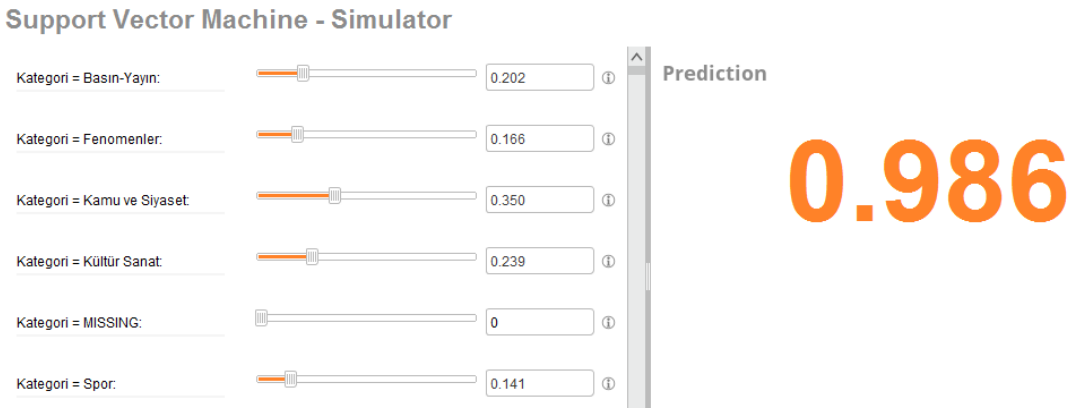
MODEL	Support Vector Machine (with RapidMiner Studio Auto Model)
Toplam Süre	5 dk 7 sn
Doğruluk	91,11%
Sınıflandırma Hatası	8,89%
RMSE	0,178
MAE	0,006
RE	%6,10
MSE	0,032
Korelasyon	0,025
Precision	0,912
Recall	0,916
F-Measure	0,914

Hata Matrisi

		GERÇEK	
		POZİTİF	NEGATİF
TAHMİN	POZİTİF	(TP) 2484	(FP) 239
	NEGATİF	(FN) 228	(TN) 2301

Hata matrisindeki değerler ile bulunan precision (kesinlik), recall (duyarlılık) ve f-measere (f-ölçeği) metriklerinin değerleri Tablo 18'de gösterilmiştir. Bu değerlere göre f-ölçeği değeri 0,914 bulunmuştur.

Bu model sonucunda Şekil 35'te kategorilerin ağırlıkları verilmiştir. Bu ağırlıklara bakıldığında kategorilerin siber zorbalık tespiti için sonucu ne derecede etkilediği görülmüştür.



Şekil 35. Support Vector Machine Simülatör

Şekil 35'teki sonuçlara göre ise sıralama şu şekildedir;

- Kamu ve Siyaset > Kültür Sanat > Basın-Yayın > Fenomenler > Spor > İş Dünyası

Bu sonuçlara göre Kamu ve siyaset kategorisi ilk sırada iken Kültür ve Sanat kategorisi onu takip etmiştir. İş Dünyası yine son sıradadır. Modelin %6,1 göreceli hata ile çalıştığı da göz önünde bulundurulduğunda az bir farkla da olsa en az hata ile çalışan model olduğu görülmüştür. Kategori özelliğinin yine siber zorbalık tespitinde önemli olduğu anlaşılmıştır. Bu model de Mobil kaynak, Web kaynağından iyi sonuç vermiştir.

5.2. Modellerin Karşılaştırılması

Modelleri öncelikle hata türüne göre ve sonra da F-Measure değerlerine göre karşılaştıran tablo ve grafikler sunulmuştur.

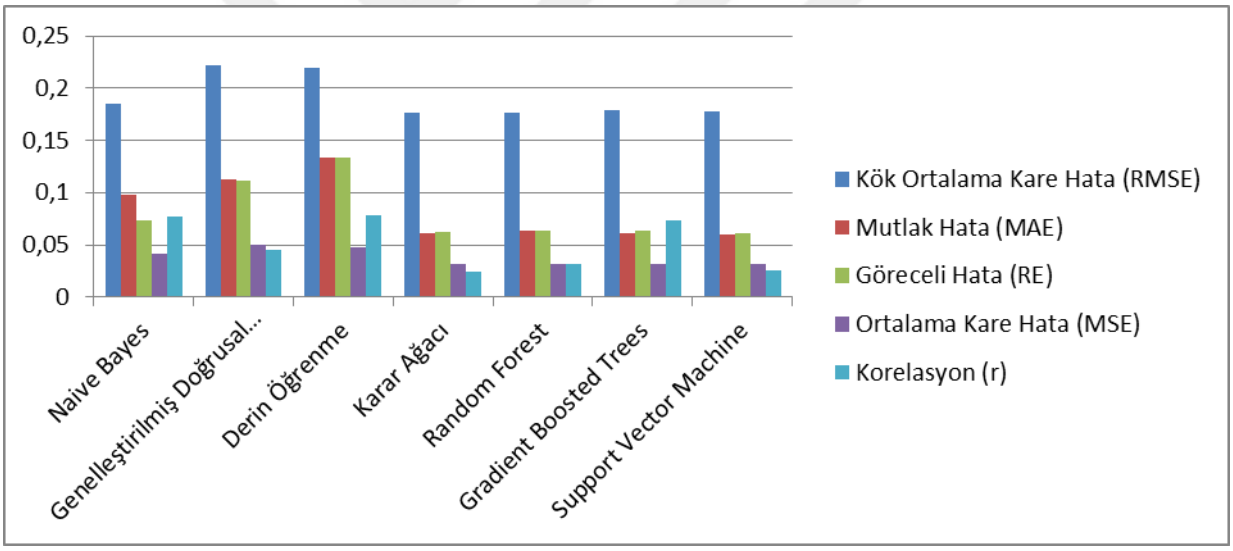
Tablo 19'da her bir hata türü için modellerin sonuçları verilmiştir. Tabloda yer alan verilere göre en iyi performansı; RMSE hata türüne göre Karar Ağacı ile Random Forest modelleri; MAE ve RE hata türlerine göre Support Vector Machine modeli; MSE hata türüne göre Karar Ağacı ile Random Forest modelleri; Korelasyon katsayısına göre ise Derin Öğrenme modeli sergilediği görülmüştür. Tüm sonuçlara göre ise hata türleri bazında ikişer kez en iyi performans sergileyen Karar Ağacı, Random Forest ve Support Vector Machine modelleri olduğu görülmüştür.

Tablo 19. Modellerin Hata Türüne Göre Kıyaslanması

MODEL	Kök Ortalama Kare Hata (RMSE)	Mutlak Hata (MAE)	Göreceli Hata (RE)	Ortalama Kare Hata (MSE)	Korelasyon (r)
Naive Bayes	0,185	0,098	%7,3	0,041	0,077
Genelleştirilmiş Doğrusal Model	0,222	0,113	% 11,1	0,05	0,045
Derin Öğrenme	0,219	0,134	% 13,4	0,048	0,078
Karar Ağacı	0,177	0,061	% 6,2	0,031	0,024
Random Forest	0,177	0,063	% 6,4	0,031	0,031

Gradient Boosted Trees	0,179	0,061	% 6,3	0,032	0,073
Support Vector Machine	0,178	0,06	% 6,1	0,032	0,025

Tablo 19'daki değerlerin kıyaslanması için Şekil 36'daki sütun grafik gösterilmiştir. Şekil 36'daki değerlere göre RMSE hata türüne göre Karar Ağacı modeli ve Random Forest Modelleri daha iyi performans sergilediği görülmüştür. MAE ve RE hata türlerine göre SVM modeli en iyi performansı göstermiştir. MSE hata türüne göre ise Karar Ağacı ve Random Forest modeli daha iyi performansı göstermişlerdir. Korelasyon katsayısına göre ise Derin Öğrenme modeli sergilediği görülmüştür. Tüm sonuçlara göre ise hata türleri bazında ikişer kez en iyi performans sergileyen Karar Ağacı, Random Forest ve Support Vector Machine modelleri olduğu görülmüştür.



Şekil 36. Modellerin Hata Türüne Göre Kıyaslanması

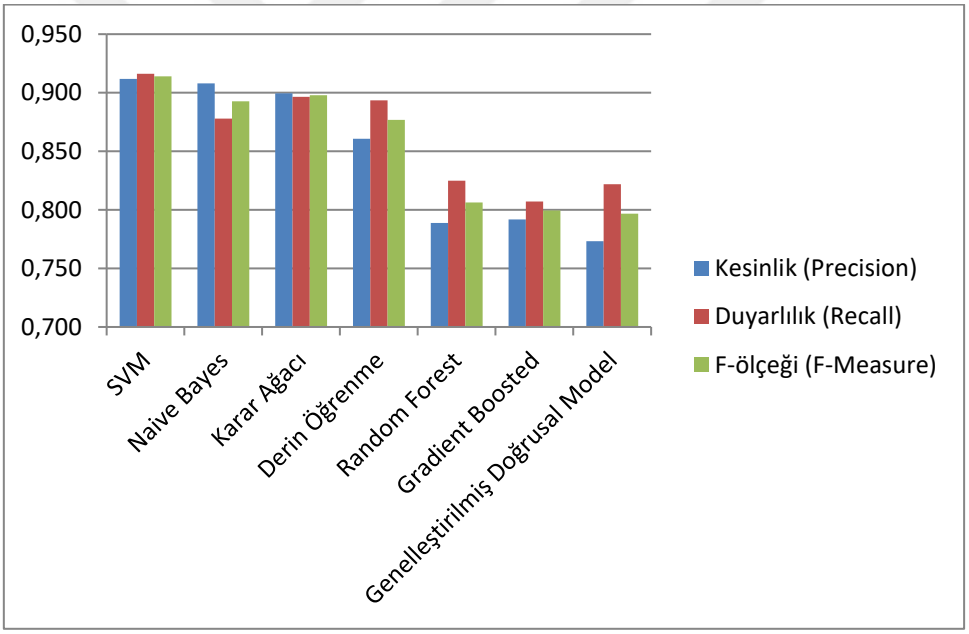
Modeller, f-ölçeğine göre kıyaslanmıştır. Bütün modellerin f-ölçeği değerleri Tablo 20'de bir arada gösterilmiştir. En yüksek f-ölçeğine sahip olan SVM Modeli aynı zamanda en yüksek doğruluk oranına da sahiptir.

Tablo 20. Modellerin F-Ölçeği ve Doğruluk Değerleri

Modeller	p	r	f	Doğruluk
SVM	0,912	0,916	0,914	91,11%
Naive Bayes	0,908	0,878	0,893	90,25%

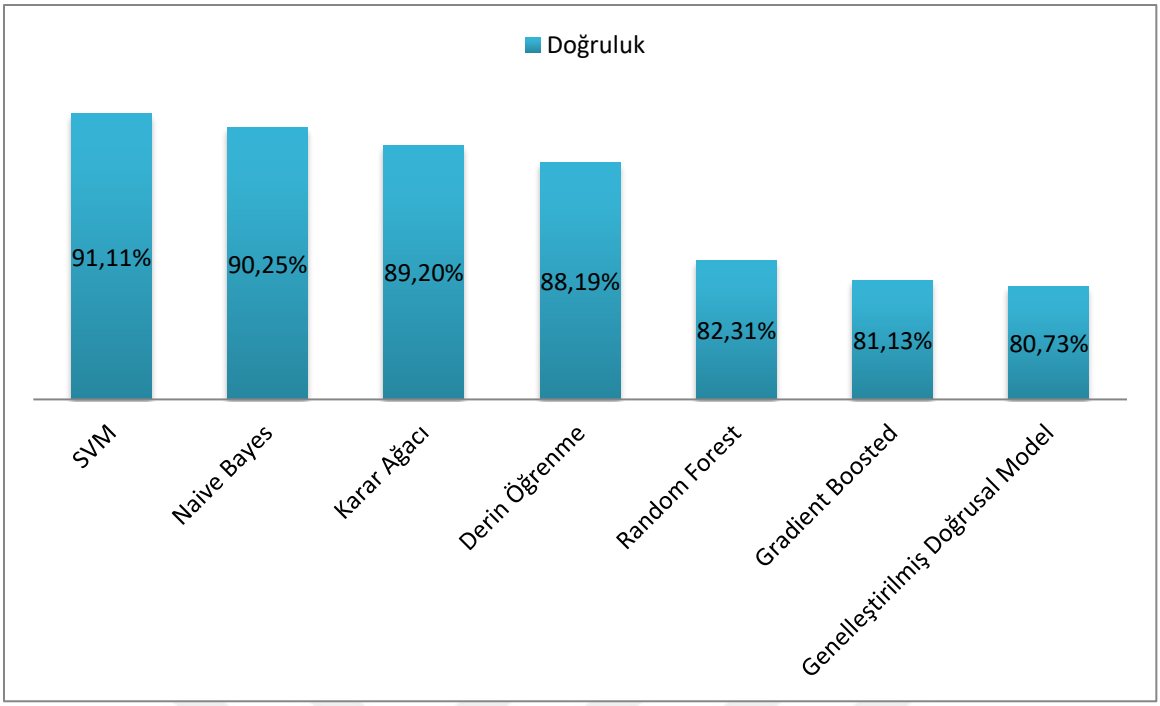
Karar Ağacı	0,899	0,896	0,898	89,20%
Derin Öğrenme	0,861	0,893	0,877	88,19%
Random Forest	0,789	0,825	0,806	82,31%
Gradient Boosted Trees	0,792	0,807	0,799	81,13%
Genelleştirilmiş Doğrusal Model	0,773	0,822	0,797	80,73%

Tablo 20'de gösterilen performans metriklerin sütun grafik üzerinde Şekil 37'de gösterilmiştir. Değerlere göre en iyi f-ölçeğine Support Vector Machine (Destek Vektör Makineleri) Modeli göstermiştir. Model ayrıca Precision (Kesinlik), Recall (Duyarlılık) değerlerinde de en iyi değerleri göstermiştir.



Şekil 37. Modellerin F1 Ölçeği ve Doğruluk Değerleri

Doğruluk değerlerini sırasıyla gösteren grafik Şekil 38'teki gibidir. Şekil 38 incelendiğinde en iyi performansı SVM modeli ve ikinci en iyi performansı da Naive Bayes modelinin sergilediği görülmüştür.



Şekil 38. Modellerin Doğruluk Değerleri

6. SONUÇLAR

Bu çalışmamızda Türkçe yazılmış metin içeriklerinden siber zorbalığın otomatik olarak tespit edilmesi için makine öğrenmesi yöntemlerin geliştirilmesi ve bunların başarısı araştırılmıştır. Bu kapsamda veri kaynağı olarak Twitter seçilmiştir ve buradan mükerrer veriler ayıklandığında 29198 adet tweet elde edilmiştir. Bu veri seti Türkçe yazılmış metinler için kullanılmış olan en büyük veri seti olma özelliği taşımaktadır.

Veri setimizde, ilgili tweetlerin kime atıldığı (to-user) ve bu kişinin hangi kategoride olduğunu (kategori), atılan tweetin kimin tarafından atıldığı (from-user), hangi cihaz kaynağından atıldığı (kaynak cihaz) ve kaç defa retweet edildiği bilgisi de bulunmaktadır. Burada siber zorbalık tespiti yapılırken sınıflandırmayı etkileyecek olanlar belirlenip RapidMiner Auto Model içerisinde incelenmiştir. Bu şekilde "To-User", "Kategori", "Kaynak Cihaz" ve "Text" özellikleri Auto Model ile incelenmiş ve hangi özelliklerin ne derece etkili olduğu tespit edilmiştir.

Çalışmamızda veri setinin insan eliyle tek tek incelenip metnin anlamı yorumlanarak etiketlenmesi çalışmamın başarısını arttırmaktadır. Metnin küfürlü ya da kötü kelimeler içermesi siber zorbalık içermesi için yeterli olmadığı anlaşılmaktadır.

RapidMiner Auto Model sonuçlarında hem hata türlerine göre hem de f-ölçeği performans metrikleri hesaplanmıştır. Bu değerler ile kıyaslama yapılmıştır. Tablo 19'da görüldüğü gibi RMSE hata türüne en iyi performansı Random Forest ve Karar Ağacı modelleri sergilemiştir. Diğer modeller ise yakın performans sergilemiştir. MAE ve RE hata türlerine bakıldığında Support Vector Machine modeli en iyi performansı sergilemiştir ve ardından da Karar Ağacı modeli gelmektedir. Son olarak MSE hata türünde en iyi performansları Karar Ağacı ve Random Forest beraber sergilemektedir.

Tablo 20'deki sonuçlarda görüldüğü gibi f-ölçeği performans metriğine göre ise en iyi performansı Support Vector Machine modeli sergilerken ikinci olarak Karar Ağacı gelmektedir, üçüncü olarak ise Naive Bayes modelinin olduğu görülmektedir. Tablo 20'de gösterilen bir diğer performans metriği olarak ise doğruluk değerleri kullanılmıştır. F-ölçeği değerlerine göre olan sonuca yakın bir sonuç çıkmasına rağmen sıralamada değişiklik göstermektedir. Bu değerlere göre en iyi performansı yine Support Vector Machine modeli sergilemiştir. İkinci olarak Naive Bayes modeli yer alırken üçüncü model Karar Ağacı olmuştur.

Modeller araştırılırken RapidMiner Stuio Auto Model özelliği kullanılmıştır ve parametre değerleri Auto Model sunduğu ön tanımlı değerler kullanılmıştır. Bu parametre değerleri her bir değer için optimize edildikçe daha iyi sonuçlar elde edilebilecektir. Örnek olarak Random Forest modelinin Tablo 20'deki değerlerden daha yüksek performans sergilemesi beklenirken ön tanımlı değerler performansını düşürmüş olabileceği tahmin edilmektedir. Ayrı bir çalışma konusu olarak parametrelerin optimize değerleriyle modellerin sonuçları iyileştirilebilmektedir.

Tablo 8 ve Tablo 13'te Kategori özelliği siber zorbalık tespitinde önemli bir etken olduğu görülmektedir. Özellikle Kamu ve Siyaset kategorisine atılan tweetlerin siber zorbalık tespitinde etkili olduğu görülmektedir. Bu sonucun etkilerinin ve nedenlerinin araştırılması ayrı bir çalışma konusudur. Çalışma öncesinde Spor kategorisinin de siber zorbalık konusunda sonuçta etkili olacağı tahmin edilmekteydi ancak belirgin bir etkisi olduğu görülmemiştir. Kategori özelliği, sosyal medya araştırmaları için kurulmuş olan web siteler yardımıyla listeler tespit edilerek kullanıma hazırlanmıştır. Örnek olarak çalışmamızda da kullandığımız BoomSocial.com (sosyal medya ölçümleme& analiz) web sitesi verilebilir.

Çalışmamızda sonuca en çok etki eden özellik, beklenildiği gibi "Text" özelliği olmuştur. Benzer çalışmalarda da incelenen "text" özelliği konusunda diğer çalışmalara kıyasla

tezimizin başarılı olduğu görülmüştür. Bunun en önemli sebepleri arasında veri setimizin büyüklüğü, veri setimizin güncel olması, ön işleme başarısı, etiketleme başarısı olduğu görülmektedir. Diğer çalışmalara kıyasla veri setinin büyük olduğu da görülmektedir. Daha önce de belirtildiği gibi haftalar süren bir çalışma olarak yaklaşık otuzbin tweetin elde edilmesi ve incelenmesi çalışma sonucunun verimli olması için etkili olmuştur.

Kaynak cihazlar arasında bir kıyaslama yapıldığında mobil kaynaktan atılan tweetlerin siber zorbalık içermede daha etkili olduğu görülmüştür. Bunun sebebi olarak, mobil kullanımın daha yaygın olması ve anlık olarak sosyal medyanın takip edilebiliyor olmasının, bildirimlere açık olmasının, herhangi bir tweetin diğer haberleşme uygulamaları üzerinden yaygınlaşabiliyor ve erişiminin artabiliyor olmasının etkili olabileceği düşünülmektedir. Buna rağmen "Web" kaynağından atılan örneğin arttırılması gerektiği görülmüştür. Çünkü veri setimizdeki örneklerde Web kaynağından atılan örnek sayısının az olduğu görülmektedir.

Çalışmalarda kategori özelliğinin tek tek ve detaylı incelenmesi önerilir. Ayrıca günümüzde en önemli sorunlardan biri olarak da gösterilen siber zorbalık konusunun özellikle çocuk yaşlardaki etkisinin ölçülebilmesi için veri setindeki kullanıcıların (from-user ve to-user) yaş tespitinin de yapılmış olarak modellerde kullanılması önerilir. Siber zorbalık tespiti konusunda gelecekte çalışmaların artmasını temenni etmekteyiz. Yapılan çalışmaların artması ve kolaylaştırması için hazırlanan veri setlerinin artması ve açık kaynak olarak yaygınlaştırılması faydalı olacaktır. Bu da Türkçe metinlerde siber zorbalık tespiti için gerekli görülmektedir.

KAYNAKLAR

[1] "Siber Zorbalık: Gençlerimizi Bekleyen Yeni Tehlike | Osman Tolga Arıcak | Makale | genclikbirikimi.org". http://genclikbirikimi.org/kunye-31603-siber-zorbalik-genclerimizi-bekleyen-yeni-tehlike-makale (erişim Oca. 24, 2021).

[2] D. Yin, Z. Xue, L. Hong, B. Davison, A. Edwards, ve L. Edwards, "Detection of harassment on Web 2.0", Oca. 2009.

[3] K. Dinakar, R. Reichart, ve H. Lieberman, "Modeling the Detection of Textual Cyberbullying", 2011.

[4] "Learning from bullying traces in social media | Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies". https://dl.acm.org/doi/10.5555/2382029.2382139 (erişim Şub. 04, 2021).

[5] "[PDF] Careful what you share in six seconds: Detecting cyberbullying instances in Vine | Semantic Scholar". https://www.semanticscholar.org/paper/Careful-what-you-share-in-six-seconds%3A-Detecting-in-Rafiq-Hosseinmardi/039cfab52407e5bb3f3e0b16dfd99d7a72479d12 (erişim Şub. 04, 2021).

[6] K. Burn-Thornton ve T. Burman, "The Use of Data Mining to Indicate Virtual (Email) Bullying", içinde *2012 Third Global Congress on Intelligent Systems*, Kas. 2012, ss. 253-256, doi: 10.1109/GCIS.2012.107.

[7] "(PDF) Detecting cyberbullying: Query terms and techniques". https://www.researchgate.net/publication/262238159_Detecting_cyberbullying_Query_terms_and_techniques (erişim Şub. 04, 2021).

[8] V. Nahar, S. AL Maskari, X. Li, ve C. Pang, "Semi-supervised Learning for Cyberbullying Detection in Social Networks", Tem. 2014, ss. 160-171, doi: 10.1007/978-3-319-08608-8_14.

[9] M. Munezero, C. Suero Montero, T. Kakkonen, E. Sutinen, M. Mozgovoy, ve V. Klyuev, "Automatic detection of antisocial behaviour in texts", *Informatica (Slovenia)*, c. 38, ss. 3-10, Mar. 2014.

[10] B. S. Nandhini ve J. I. Sheeba, "Online Social Network Bullying Detection Using Intelligence Techniques", *Procedia Computer Science*, c. 45, ss. 485-492, Oca. 2015, doi: 10.1016/j.procs.2015.03.085.

[11] R. Zhao, A. Zhou, ve K. Mao, "Automatic Detection of Cyberbullying on Social Networks based on Bullying Features", Oca. 2016, doi: 10.1145/2833312.2849567.

[12] H. Sanchez ve S. Kumar, "Twitter Bullying Detection", s. 7.

[13] Y. Chen, Y. Zhou, S. Zhu, ve H. Xu, "Detecting Offensive Language in Social Media to Protect Adolescent Online Safety", içinde *2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing*, Eyl. 2012, ss. 71-80, doi: 10.1109/SocialCom-PASSAT.2012.55.

[14] M. Dadvar, D. Trieschnigg, ve F. de Jong, "Experts and Machines Against Bullies: A Hybrid Approach to Detect Cyberbullies", May. 2014, c. 8436, doi: 10.1007/978-3-319-06483-3_25.
[15] M. Dadvar, D. Trieschnigg, ve F. de Jong, "Expert knowledge for automatic detection of bullies in social networks", Kas. 2013.
[16] P. J. C. Pérez, C. J. L. Valdez, J. P. S. Barrera, ve P. F. Pérez, "MISAAC: Instant messaging tool for Ciberbullying Detection", s. 4.
[17] L. Del Bosque ve S. Garza, "Aggressive Text Detection for Cyberbullying", Kas. 2014, ss. 221-232, doi: 10.1007/978-3-319-13647-9_21.
[18] N. Potha ve M. Maragoudakis, "Cyberbullying Detection using Time Series Modeling", içinde *2014 IEEE International Conference on Data Mining Workshop*, Ara. 2014, ss. 373-382, doi: 10.1109/ICDMW.2014.170.
[19] A. Squicciarini, S. Rajtmajer, Y. Liu, ve C. Griffin, "Identification and characterization of cyberbullying dynamics in an online social network", Ağu. 2015, ss. 280-285, doi: 10.1145/2808797.2809398.
[20] U. Bretschneider, T. Wöhner, ve R. Peters, "Detecting Online Harassment in Social Networks", 2014.
[21] "(PDF) Improving Cyberbullying Detection with User Context". https://www.researchgate.net/publication/249012795_Improving_Cyberbullying_Detection_with_User_Context (erişim Şub. 04, 2021).
[22] M. Dadvar ve F. de Jong, "Cyberbullying Detection; A Step Toward a Safer Internet Yard", program adı: WWW'12 - Proceedings of the 21st Annual Conference on World Wide Web Companion, Nis. 2012, doi: 10.1145/2187980.2187995.
[23] S. M. Serra ve H. S. Venter, "Mobile cyber-bullying: A proposal for a pre-emptive approach to risk mitigation by employing digital forensic readiness", içinde *2011 Information Security for South Africa*, Ağu. 2011, ss. 1-5, doi: 10.1109/ISSA.2011.6027507.
[24] P. Galán-García, J. G. de la Puerta, C. L. Gómez, I. Santos, ve P. G. Bringas, "Supervised machine learning for the detection of troll profiles in twitter social network: application to a real case of cyberbullying", *Logic Journal of the IGPL*, c. 24, sy 1, ss. 42-53, Şub. 2016, doi: 10.1093/jigpal/jzv048.
[25] P. Jacobs, "Data Mining: What General Managers Need to Know", *Harvard Management Update*, c. 4, sy 10, s. 8, 1999.
[26] D. J. Hand, "Data Mining: Statistics and More?", *The American Statistician*, c. 52, ss. 112-118, 1998.
[27] F. Gorunescu (auth.), *Data Mining: Concepts, Models and Techniques*, 1. bs. Springer-Verlag Berlin Heidelberg, 2011.
[28] Jiawei Han, Kamber, Micheline, ve Pei, Jihan, *Data Mining Concepts and Tecniques*. The Morgan Kaufmann Series in Data Management Systems, 2016.
[29] C. C. Aggarwal, "Social Network Analysis", içinde *Data Mining: The Textbook*, C. C. Aggarwal, Ed. Cham: Springer International Publishing, 2015, ss. 619-661.
[30] U. Cakir ve S. Guldamlasioglu, "Text Mining Analysis in Turkish Language Using Big Data Tools", Haz. 2016, ss. 614-618, doi: 10.1109/COMPSAC.2016.203.
[31] L. Safko ve D. K. Brake, *The social media bible: tactics, tools, and strategies for business success*. Hoboken, N.J: John Wiley & Sons, 2009.
[32] G. Dean, P. Bell, ve J. Newman, "The dark side of social media: review of online terrorism", *Pakistan Journal of Criminology*, c. 3, sy 3, ss. 103-122, 2012.
[33] N. Bekmamedova ve G. Shanks, "Social Media Analytics and Business Value: A Theoretical Framework and Case Study", içinde *2014 47th Hawaii International Conference on System Sciences*, Oca. 2014, ss. 3728-3737, doi: 10.1109/HICSS.2014.464.
[34] B. Akın ve U. T. Gürsoy ŞiMşek, "SOSYAL MEDYA ANALİTİĞİ İLE DEĞER YARATMA: DUYGU ANALİZİ İLE GELECEĞE YÖNELİM", *Mehmet Akif Ersoy Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, ss. 797-811, Ara. 2018, doi: 10.30798/makuiibf.435804.
[35] A. M. Kaplan ve M. Haenlein, "Users of the world, unite! The challenges and opportunities of Social Media", *Business horizons*, c. 53, sy 1, ss. 59-68, 2010.

[36] D. Boyd ve N. Ellison, "Social Network Sites: Definition, History, and Scholarship", *J. Computer-Mediated Communication*, c. 13, ss. 210-230, Eki. 2007, doi: 10.1111/j.1083-6101.2007.00393.x.
[37] Z. Vural ve M. Bat, "YENİ BİR İLETİŞİM ORTAMI OLARAK SOSYAL MEDYA: EGE ÜNİVERSİTESİ İLETİŞİM FAKÜLTESİNE YÖNELİK BİR ARAŞTIRMA.", *Journal of Yasar University*, c. 5, sy 20, 2010.
[38] S. Tilley, *The art & science of online marketing*. Collingwood, Vic.; Farnham: SitePoint ; O'Reilly [distributor, 2009.
[39] L. Safko, *The Social Media Bible: Tactics, Tools, and Strategies for Business Success*, 3. bs. Wiley, 2012.
[40] "Digital 2020", *We Are Social*. https://wearesocial.com/digital-2020 (erişim May. 17, 2020).
[41] "Digital 2020: Turkey", *DataReportal – Global Digital Insights*. https://datareportal.com/reports/digital-2020-turkey (erişim May. 17, 2020).
[42] M. K. Taş, "Pazarlama iletişiminde sosyal medya kullanımı: sigorta pazarına bir uygulama", Master's Thesis, İstanbul Ticaret Üniversitesi, 2014.
[43] N. G. Akçal, "SOSYAL MEDYA PAZARLAMASININ TÜKETİCİ DAVRANIŞI VE SATIN ALMA ÜZERİNDEKİ ETKİSİ", Ankara Hacı Bayram Veli Üniversitesi, 2020.
[44] D. Cukul, "Moda Pazarlamasında Bir Araç Olarak Sosyal Medya: Mavi Jeans'in Başarısı", *Akdeniz Sanat*, c. 4, sy 8, Art. sy 8, Kas. 2011.
[45] J. Bollen, H. Mao, ve X.-J. Zeng, "Twitter mood predicts the stock market", *Journal of Computational Science*, c. 2, sy 1, ss. 1-8, Mar. 2011, doi: 10.1016/j.jocs.2010.12.007.
[46] E. Cülcüloğlu, "İnsan Kaynaklarında Yeni Dönem", *Türkiye Bilişim Derneği Bilişim Dergisi*, sy 153, s. 19, 2013.
[47] P. Yilmaz, "Sosyal Medyanın İnsan Kaynakları Yönetiminde İşe Alım Süreci Üzerindeki Etkisi", *Anadolu Bil Meslek Yüksekokulu Dergisi*, sy 41, Art. sy 41, Oca. 2016.
[48] B. Türkcan, "YERLİ TURİSTLERİN TURİSTİK KONAKLAMA TESİSİ TERCİHLERİNDE İNTERNET TEMELLİ SOSYAL AĞLARIN ROLÜ: İZMİR İLİ ÖRNEĞİ", *Ege Stratejik Araştırmalar Dergisi*, c. 8, sy 1, Art. sy 1, Oca. 2017.
[49] Ö. Duğan ve B. O. Aydin, "Sosyal Medyanın Turizmde Tanıtım Amaçlı Kullanımı: T.C. Kültür ve Turizm Bakanlığı Örneği", *Uluslararası Türk Dünyası Turizm Araştırmaları Dergisi*, c. 3, sy 1, Art. sy 1, Tem. 2018.
[50] G. Özbay ve A. Çekin, "Turizm, Ulaşım ve Sosyal Medya Etkileşimi: Doğu Ekspresi ve Kars'taki Konaklama İşletmelerine Yönelik Bir Araştırma", *Ahi Evran Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, c. 6, sy 1, Art. sy 1, Mar. 2020, doi: 10.31592/aeusbed.646499.
[51] F. Aksoy ve G. Çalişir, "KAMU KURUMLARINDA SOSYAL MEDYA KULLANIMI: KASTAMONU İL KÜLTÜR VE TURİZM MÜDÜRLÜĞÜ ÖRNEĞİ", *Alınteri Sosyal Bilimler Dergisi*, c. 3, sy 1, Art. sy 1, Tem. 2019, doi: 10.30913/alinterisosbil.503286.
[52] İ. Giritlioğlu, "Türkiye'de Yerel Turizm Ofislerinin Web Site İçeriklerinin Değerlendirilmesine Yönelik Bir Araştırma", *Kahramanmaraş Sütçü İmam Üniversitesi Sosyal Bilimler Dergisi*, c. 11, sy 1, Art. sy 1, Haz. 2014.
[53] H. Altinbaş, "Sporun Pazarlaması ve Pazarlama İletişiminde Spor", *Selçuk İletişim*, c. 5, sy 1, Art. sy 1, Kas. 2013, doi: 10.18094/si.69455.
[54] A. Smith, *Introduction to sport marketing*. Amsterdam; Oxford: Elsevier Butterworth-Heinemann, 2008.
[55] F. Peltekoğlu ve P. Hürmeriç, "TÜRK FUTBOL TAKIMLARINDA PAZARLAMA HALKLA İLİŞKİLERİ ARACI OLARAK SOSYAL MEDYA", *Selçuk İletişim*, c. 7, sy 2, Art. sy 2, Kas. 2013, doi: 10.18094/si.46150.
[56] "Spor Sektöründe Sosyal Medya İstatistiği", *Sosyal Medya Ölçümleme Raporlama Analiz*, 2020. https://www.boomsocial.com/ (erişim Eki. 12, 2020).
[57] "The Impact of Polices on Government Social Media Usage: Issues, Challenges, and Recommendations | Request PDF", *ResearchGate*. https://www.researchgate.net/publication/238638171_The_Impact_of_Polices_on_Government_Social_Media_Usage_Issues_Challenges_and_Recommendations (erişim Eyl. 20, 2020).
[58] A. Y. Dara, "KAMU YÖNETİMİNDE HALKLA İLİŞKİLER VE SOSYAL MEDYA", *Selçuk İletişim*, c. 7, sy 1, Art. sy 1, Kas. 2013, doi: 10.18094/si.57021.

[59] A. Yildirim, “Devlet-Vatandaş İlişkisinde Kamu Web Siteleri ve Sosyal Medyanın Etkisi”, *İletişim Çalışmaları Dergisi*, c. 1, sy 1, Art. sy 1, Eki. 2015.
[60] M. Z. Sobaci ve Ö. Köseoğlu, “OMBUDSMAN ve SOSYAL MEDYA: KAMU DENETÇİLİĞİ KURUMU’NUN TWITTER KULLANIMINA İLİŞKİN AMPİRİK BİR ANALİZ”, *Ankara Üniversitesi SBF Dergisi*, c. 71, sy 1, Art. sy 1, Mar. 2016, doi: 10.1501/SBFder_0000002386.
[61] “The Health Communicator’s Social Media Toolkit”, s. 59.
[62] A. B. Dari, “Sosyal Medya ve Sağlık”, *21. Yüzyılda Eğitim Ve Toplum Eğitim Bilimleri Ve Sosyal Araştırmalar Dergisi*, c. 6, sy 18, Art. sy 18, Ara. 2017.
[63] A. Culotta, “Towards detecting influenza epidemics by analyzing Twitter messages”, içinde *Proceedings of the First Workshop on Social Media Analytics - SOMA ’10*, Washington D.C., District of Columbia, 2010, ss. 115-122, doi: 10.1145/1964858.1964874.
[64] S. Savaş ve N. Topaloglu, “Siber Güvenlikte Yeni Bir Boyut: Sosyal Medya İstihbaratı A New Dimension In Cyber Security: Social Media Intelligence”, Şub. 2016.
[65] S. Savaş ve N. Topaloglu, “Data analysis through social media according to the classified crime”, *TURKISH JOURNAL OF ELECTRICAL ENGINEERING & COMPUTER SCIENCES*, c. 27, ss. 407-420, Oca. 2019, doi: 10.3906/elk-1712-17.
[66] S. Mansour, “Social Media Analysis of User’s Responses to Terrorism Using Sentiment Analysis and Text Mining”, *Procedia Computer Science*, c. 140, ss. 95-103, Oca. 2018, doi: 10.1016/j.procs.2018.10.297.
[67] M. Chi, “Big data in national security: online resource”.
[68] M. Cheong ve V. C. S. Lee, “A microblogging-based approach to terrorism informatics: Exploration and chronicling civilian sentiment and response to terrorism events via Twitter”, *Inf Syst Front*, c. 13, sy 1, ss. 45-59, Mar. 2011, doi: 10.1007/s10796-010-9273-x.
[69] “Library Genesis: Jo, Taeho - Text mining: concepts, implementation, and big data challenge”. http://libgen.rs/book/index.php?md5=669C2C62A9A83A24C0736558E33A6439 (erişim Eki. 31, 2020).
[70] G. Miner, Delen D., Elder J., A. Fast, T. Hill, ve R. Nisbet, *Practical Text Mining and Statistical analysis for Non-Structured Text Data Applications*. Waltham, USA: Elsevier Science & Technology.
[71] Liu Bing, *Sentiment analysis : mining opinions, sentiments, and emotions*. New York: Cambridge University Press, 2016.
[72] M. Farhadloo ve E. Rolland, “Fundamentals of Sentiment Analysis and Its Applications”, 2016, ss. 1-24.
[73] E. Filatova, “Irony and Sarcasm: Corpus Generation and Analysis Using Crowdsourcing”, içinde *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC’12)*, Istanbul, Turkey, May. 2012, ss. 392-398, Erişim: Ara. 02, 2020. [Çevrimiçi]. Erişim adresi: http://www.lrec-conf.org/proceedings/lrec2012/pdf/661_Paper.pdf.
[74] O. Grljevic ve Z. Bošnjak, “Sentiment analysis of customer data”, *Strategic Management*, c. 23, ss. 38-49, Oca. 2018, doi: 10.5937/StraMan1803038G.
[75] J. Juvonen ve E. Gross, “Extending the School Grounds?-Bullying Experiences in Cyberspace”, *The Journal of school health*, c. 78, ss. 496-505, Eki. 2008, doi: 10.1111/j.1746-1561.2008.00335.x.
[76] Q. Li, “A Cross-Cultural Comparison of Adolescents’ Experience Related to Cyberbullying”, *Educational Research - EDUC RES*, c. 50, ss. 223-234, Eyl. 2008, doi: 10.1080/00131880802309333.
[77] J. Patchin ve S. Hinduja, “Bullies Move Beyond the Schoolyard A Preliminary Look at Cyberbullying”, *Youth Violence and Juvenile Justice*, c. 4, ss. 148-169, Nis. 2006, doi: 10.1177/1541204006286288.
[78] R. Slonje ve P. K. Smith, “Cyberbullying: another main type of bullying?”, *Scand J Psychol*, c. 49, sy 2, ss. 147-154, Nis. 2008, doi: 10.1111/j.1467-9450.2007.00611.x.
[79] N. E. Willard, *Cyberbullying and Cyberthreats: Responding to the Challenge of Online Social Aggression, Threats, and Distress*, 2nd edition. Champaign, Ill: Research Press, 2007.
[80] R. Kowalski ve S. Limber, “Electronic Bullying Among Middle School Students”, *The Journal of adolescent health : official publication of the Society for Adolescent Medicine*, c. 41, ss. S22-30, Oca. 2008, doi: 10.1016/j.jadohealth.2007.08.017.

[81] G. Gredler, “Olweus, D. (1993).Bullying at school: What we know and what we can do. Malden, MA: Blackwell Publishing, 140 pp., $25.00”, *Psychology in The Schools - PSYCHOL SCH*, c. 40, ss. 699-700, Kas. 2003, doi: 10.1002/pits.10114.
[82] “Cyberbullying: An Emerging Threat to the ‘Always On’ Generation – Bill Belsey”. http://www.billbelsey.com/?p=1827 (erişim Oca. 24, 2021).
[83] Prof. Dr. O. T. ARICAK, “Psychiatric Symptomatology as a Predictor of Cyberbullying among University Students”, *Egitim Arastirmalari - Eurasian Journal of Educational Research*, c. 34, ss. 167-184, Ara. 2009.
[84] C. R. Center, “Bullying Beyond the Schoolyard: Preventing and Responding to Cyberbullying (2nd edition)”, *Cyberbullying Research Center*. http://cyberbullying.org/bullying-beyond-schoolyard-preventing-responding-cyberbullying-2nd-edition (erişim Oca. 24, 2021).
[85] Erhan ÖZTÜRK, “CYBERBULLYING DETECTION USING TEXT CLASSIFICATION FOR TURKISH LANGUAGE”, Yüksek Lisans Tezi, ÇUKUROVA ÜNİVERSİTESİ.
[86] “Rapidminer Nedir? RapidMiner ne işe yarar?” https://rapidminer.dqturkiye.com/rapidminer-nedir.html (erişim Şub. 23, 2021).
[87] “ahmetax/trstop”, *GitHub*. https://github.com/ahmetax/trstop (erişim Mar. 09, 2021).
[88] J. A. Nelder ve R. W. M. Wedderburn, “Generalized Linear Models”, *Journal of the Royal Statistical Society. Series A (General)*, c. 135, sy 3, ss. 370-384, 1972, doi: 10.2307/2344614.
[89] L. Deng, “Deep Learning: Methods and Applications”, *FNT in Signal Processing*, c. 7, sy 3-4, ss. 197-387, 2014, doi: 10.1561/2000000039.
[90] J. Mohajon, “Confusion Matrix for Your Multi-Class Machine Learning Model”, *Medium*, Eyl. 09, 2020. https://towardsdatascience.com/confusion-matrix-for-your-multi-class-machine-learning-model-ff9aa3bf7826 (erişim May. 10, 2021).

ÖZGEÇMİŞ

Furkan Zahit Atay; İlkokul ve ortaokul öğrenimini Antalya'da tamamlayan Atay, Lise öğrenimini Özel Antalya Envar Fen Lisesi'nde okumuştur.

2011 yılında Sakarya Üniversitesi Mühendislik Fakültesi Endüstri Mühendisliği'ni kazandıktan sonra 2015 yılında bölümü tamamlamıştır.

2016 yılında TUBİTAK BİLGEM'de Bilgi Sistemleri Birimi'nde E-Dönüşüm Uygulamaları Bölümü'nde iş hayatına başlamıştır. Halen burada çalışmaya devam etmektedir.

2018 yılında Bilgi Güvenliği Mühendisliği bölümünde yüksek lisans eğitimine başlamıştır. Evli ve 1 kız babasıdır.

EK

Ek 1: Twitter hedef listesi

Genel Sıra No	Kategori Sıra No	HESAP	@username	Kategori
1	**1**	Recep Tayyip Erdoğan	@RTErdogan	Kamu ve Siyaset
2	**2**	Abdullah Gül	@cbabdullahgul	Kamu ve Siyaset
3	**3**	Kemal Kılıçdaroğlu	@kilicdarogluk	Kamu ve Siyaset
4	**4**	Fahrettin Koca	@drfahrettinkoca	Kamu ve Siyaset
5	**5**	Ahmet Davutoğlu	@Ahmet_Davutoglu	Kamu ve Siyaset
6	**6**	Muharrem İnce	@vekilince	Kamu ve Siyaset
7	**7**	Devlet Bahçeli	@dbdevletbahceli	Kamu ve Siyaset
8	**8**	Ekrem İmamoğlu	@ekrem_imamoglu	Kamu ve Siyaset
9	**9**	Süleyman Soylu	@suleymansoylu	Kamu ve Siyaset
10	**10**	Mansur Yavaş	@mansuryavas06	Kamu ve Siyaset
11	**11**	Meral Akşener	@meral_aksener	Kamu ve Siyaset
12	**12**	Melih Gökçek	@06melihgokcek	Kamu ve Siyaset
13	**13**	Ziya Selçuk	@ziyaselcuk	Kamu ve Siyaset
14	**14**	Bülent Arınç	@bulent_arinc	Kamu ve Siyaset
15	**15**	Vasip Şahin	@vasipsahin	Kamu ve Siyaset
16	**16**	Binali Yıldırım	@BY	Kamu ve Siyaset
17	**17**	Numan Kurtulmuş	@NumanKurtulmus	Kamu ve Siyaset
18	**18**	AK Parti	@Akparti	Kamu ve Siyaset
19	**19**	İbrahim Kalın	@ikalin1	Kamu ve Siyaset
20	**20**	Mehmet Şimşek	@memetsimsek	Kamu ve Siyaset
21	**21**	MHP	@MHP_Bilgi	Kamu ve Siyaset
22	**22**	Kadıköy Belediyesi	@kadikoybelediye	Kamu ve Siyaset
23	**23**	Sağlık Bakanlığı	@saglikbakanligi	Kamu ve Siyaset
24	**24**	İçişleri Bakanlığı	@TC_icisleri	Kamu ve Siyaset
25	**25**	Milli Savunma Bakanlığı	@tcsavunma	Kamu ve Siyaset

26	**26**	Kadir Toptaş	@Kadir__Topbas	Kamu ve Siyaset
27	**27**	Selahattin Demirtaş	@hdpdemirtas	Kamu ve Siyaset
28	**28**	MEB	@tcmeb	Kamu ve Siyaset
29	**29**	Mevlüt Çavuşoğlu	@MevlutCavusoglu	Kamu ve Siyaset
30	**30**	Özgür Özel	@eczozgurozel	Kamu ve Siyaset
31	**31**	Bekir Bozdağ	@bybekirbozdag	Kamu ve Siyaset
32	**32**	Şamil Tayyar	@samiltayyar27	Kamu ve Siyaset
33	**33**	Hüseyin Avni Mutlu	@Valimutlu	Kamu ve Siyaset
34	**34**	CHP	@herkesicinCHP	Kamu ve Siyaset
35	**35**	Mustafa Balbay	@mustafabalbay	Kamu ve Siyaset
36	**36**	Dışişleri Bakanlığı	@TC_Disisleri	Kamu ve Siyaset
37	**37**	Binali Yıldırım İletişim Ofisi	@BA_Yildirim	Kamu ve Siyaset
38	**38**	Ömer Çelik	@omerrcelik	Kamu ve Siyaset
39	**39**	Şafak Pavey	@safakpavey	Kamu ve Siyaset
40	**40**	Fatma Şahin	@FatmaSahin	Kamu ve Siyaset
41	**41**	Abdulhamit Gül	@abdulhamitgul	Kamu ve Siyaset
42	**42**	TSK	@TSKGnkur	Kamu ve Siyaset
43	**43**	İBB	@istanbulbld	Kamu ve Siyaset
44	**44**	Aykut Erdoğdu	@aykuterdogdu	Kamu ve Siyaset
45	**45**	Egemen Bağış	@EgemenBagis	Kamu ve Siyaset
46	**46**	Faruk Çelik	@farukcelikcomtr	Kamu ve Siyaset
47	**47**	Savcı Sayan	@SavciSayan	Kamu ve Siyaset
48	**48**	İletişim Başkanlığı	@iletisim	Kamu ve Siyaset
49	**49**	Mustafa Sarıgül	@M_Sarigul	Kamu ve Siyaset
50	**50**	Gürsel Tekin	@gurseltekin34	Kamu ve Siyaset
51	**51**	Tunç Soyer	@tuncsoyer	Kamu ve Siyaset
52	**52**	HDP	@HDPgenelmerkezi	Kamu ve Siyaset

53	**53**	Fahrettin Altun	@fahrettinaltun	Kamu ve Siyaset
54	**54**	Fuat Oktay	@fuatoktay	Kamu ve Siyaset
55	**55**	Burhan Kuzu	@BurhanKuzu	Kamu ve Siyaset
56	**56**	Temel Karamollaoğlu	@T_Karamollaoglu	Kamu ve Siyaset
57	**57**	Sezgin Tanrıkulu	@MSTanrikulu	Kamu ve Siyaset
58	**58**	Canan Kaftancıoğlu	@Canan_Kaftanci	Kamu ve Siyaset
59	**59**	Ali Babacan	@alibabacan	Kamu ve Siyaset
60	**60**	Zehra Zümrüt Sekçuk	@ZehraZumrutS	Kamu ve Siyaset
61	**61**	Mahir Ünal	@mahirunal	Kamu ve Siyaset
62	**62**	Osman Baydemir	@baydemirosman	Kamu ve Siyaset
63	**63**	Mustafa Varank	@varank	Kamu ve Siyaset
64	**64**	İyi Parti	@iyiparti	Kamu ve Siyaset
65	**65**	Suat Kılıç	@suatkilic	Kamu ve Siyaset
66	**66**	Abdullatif Şener	@senerabdullatif	Kamu ve Siyaset
67	**67**	Pervin Buldan	@PervinBuldan	Kamu ve Siyaset
68	**68**	Hasip Kaplan	@HasipKAPLAN	Kamu ve Siyaset
69	**69**	Sırrı Sakık	@sakiksirri	Kamu ve Siyaset
70	**70**	Mustafa Şentop	@MustafaSentop	Kamu ve Siyaset
71	**71**	Koray Aydın	@korayaydintr	Kamu ve Siyaset
72	**72**	Murat Ongun	@Mrt_Ongun	Kamu ve Siyaset
73	**73**	Naci Ağbal	@naci_agbal	Kamu ve Siyaset
74	**74**	Engin Altay	@enginaltaychp	Kamu ve Siyaset
75	**75**	Sebahat Tuncel	@tuncelsebahat	Kamu ve Siyaset
76	**1**	Cem Yılmaz	@CMYLMZ	Kültür Sanat
77	**2**	Okan Bayülgen	@okanbayulgen	Kültür Sanat
78	**3**	Demet Akalın	@DemetAkalin	Kültür Sanat
79	**4**	Ata Demirer	@atademirer	Kültür Sanat
80	**5**	Sıla Gençoğlu	@silagencoglu	Kültür Sanat
81	**6**	Levent Üzümcü	@LeventUzumcu	Kültür Sanat
82	**7**	Haluk Levent	@haluklevent	Kültür Sanat

83	**8**	Yıldız Tilbe	@YildizzTilbee	Kültür Sanat
84	**9**	Gülben Ergen	@gulbenergen	Kültür Sanat
85	**10**	Yılmaz Erdoğan	@yilmazerdogan	Kültür Sanat
86	**11**	Hülya Avşar	@hulyavsar	Kültür Sanat
87	**12**	Ahmet Kural	@AHMTKURAL	Kültür Sanat
88	**13**	Mustafa Ceceli	@mustafaceceli	Kültür Sanat
89	**14**	Erdil Yaşaroğlu	@erdilyasaroglu	Kültür Sanat
90	**15**	Tarkan	@tarkan	Kültür Sanat
91	**16**	Eser Yenenler	@eseryenenler	Kültür Sanat
92	**17**	Hande Yener	@handeyener	Kültür Sanat
93	**18**	Selçuk Erdem	@selcukerdem	Kültür Sanat
94	**19**	Hadise Açıkgöz	@Hadise	Kültür Sanat
95	**20**	Yalın	@YALINonline	Kültür Sanat
96	**21**	Emre Aydın	@emreaydin	Kültür Sanat
97	**22**	Sertab Erener	@sertaberener	Kültür Sanat
98	**23**	Mehmet Ali Alabora	@memetalialabora	Kültür Sanat
99	**24**	Murat Dalkılıç	@MuratDalkilic	Kültür Sanat
100	**25**	Nil Karaibrahimgil	@Niltakipte	Kültür Sanat
101	**26**	Kenan Doğulu	@kenandogulu	Kültür Sanat
102	**27**	Tolga Çevik	@TolgaCevik	Kültür Sanat
103	**28**	Burcu Esmersoy	@besmersoy	Kültür Sanat
104	**29**	Şahan Gökbakar	@sgokbakar	Kültür Sanat
105	**30**	Ferhat Göçer	@ferhatgocer	Kültür Sanat
106	**31**	Hayko Çepkin	@HAYKOCPQN	Kültür Sanat
107	**32**	Bengü	@bengu	Kültür Sanat
108	**33**	Doğa Rutkay	@Doarutkay	Kültür Sanat
109	**34**	Ece Erken	@eceerken	Kültür Sanat
110	**35**	Demet Evgar	@dmtevgar	Kültür Sanat
111	**36**	Ebru Gündeş	@EbruGundes	Kültür Sanat
112	**37**	Barış Atay	@barisatay	Kültür Sanat
113	**38**	Oğuzhan Koç	@oguzhankoc	Kültür Sanat
114	**39**	Ezgi Mola	@ezgimola	Kültür Sanat
115	**40**	Kıvanç Tatlıtuğ	@kivanctatlitug	Kültür Sanat
116	**41**	Ozan Doğulu	@OzanDogulu	Kültür Sanat
117	**42**	Elif Şafak	@Elif_Safak	Kültür Sanat
118	**43**	Ahmet Mümtaz Taylan	@taylan1789	Kültür Sanat
119	**44**	Gani Müjde	@GaniMujde	Kültür Sanat
120	**45**	Çağatay Ulusoy	@cagatayulusoyy	Kültür Sanat
121	**46**	Enver Aysever	@enveraysevera	Kültür Sanat
122	**47**	Metin Üstündağ	@metinustundag	Kültür Sanat
123	**48**	Gökhan Özoğuz	@gokhanozoguz	Kültür Sanat
124	**49**	Gülşen	@gulsen	Kültür Sanat
125	**50**	Ajda Pekkan	@pekkan1	Kültür Sanat
126	**51**	İbrahim Büyükak	@ibrahimbuyukak	Kültür Sanat
127	**52**	Oğuzhan Uğur	@OguzhanUgur	Kültür Sanat
128	**53**	Burak Özçivit	@burakozcivit	Kültür Sanat
129	**54**	Ceyhun Yılmaz	@ceyhunyilmaz	Kültür Sanat

130	**55**	Sarp Apak	@sarpapak81	Kültür Sanat
131	**56**	Ziynet Sali	@ziynetsali	Kültür Sanat
132	**57**	Alp Kırşan	@alpkirsan	Kültür Sanat
133	**58**	Murat Cemcir	@muratcemcir	Kültür Sanat
134	**59**	Sermiyan Midyat	@SermiyanMidyat	Kültür Sanat
135	**60**	Gülse Birsel	@gulseyazar	Kültür Sanat
136	**61**	Sinan Akçıl	@sinanakcil	Kültür Sanat
137	**62**	Sunay Akın	@SunayAkin	Kültür Sanat
138	**63**	Fırat Tanış	@firattanis	Kültür Sanat
139	**64**	Uğur Işılak	@ugur_isilak	Kültür Sanat
140	**65**	Bekir Develi	@BekirDeveli	Kültür Sanat
141	**66**	Serdar Tuncer	@Yaziyor	Kültür Sanat
142	**67**	Berna Laçin	@bernalacin35	Kültür Sanat
143	**68**	İrem Derici	@iremderici	Kültür Sanat
144	**69**	Sagopa Kajmer	@Sagopakajmerrap	Kültür Sanat
145	**70**	Engin Altan Düzyatan	@eadksk	Kültür Sanat
146	**71**	Demet Özdemir	@dmtzdmr	Kültür Sanat
147	**72**	Ahmet Ümit	@baskomsernevzat	Kültür Sanat
148	**73**	Atalay Demirci	@Atalaydemirci	Kültür Sanat
149	**74**	Nihat Doğan	@NihatDoganVevo	Kültür Sanat
150	**75**	Atilla Taş	@AtillaTasNet	Kültür Sanat
151	**76**	Serenay Sarıkaya	@SrnySrkyResmi	Kültür Sanat
152	**77**	Aleyna Tili	@aleynatilki	Kültür Sanat
153	**1**	Mesut Özil	@MesutOzil1088	Spor
154	**2**	Ultraaslan	@ultrAslan	Spor
155	**3**	Çarşı	@forzabesiktas	Spor
156	**4**	Arda Turan	@ArdaTuran	Spor
157	**5**	12numaraorg	@12numaraorg	Spor
158	**6**	Gökhan Gönül	@_gokhangonul	Spor
159	**7**	Burak Yılmaz	@yilmazburak17	Spor
160	**8**	Nuri Şahin	@nurisahin	Spor
161	**9**	Gökhan Töre	@Gokhan_Tore7	Spor
162	**10**	Ahmet Çakar	@ahmetcakar1	Spor
163	**11**	Fatih Terim	@fatihterim	Spor
164	**12**	Selçuk İnan	@8Selcukinan	Spor
165	**13**	Emre Çolak	@emrecolak1010	Spor
166	**14**	Volkan Demirel	@1VolkanDEMIREL	Spor
167	**15**	Tuncay Şanlı	@tuncaysanli17	Spor
168	**16**	Caner Erkin	@CanerErkin	Spor
169	**17**	Semih Kaya	@semihkaya_26	Spor
170	**18**	Hidayet Türkoğlu	@hidoturkoglu15	Spor
171	**19**	Sinan Güler	@sinanguler	Spor
172	**20**	Genç Fenerbahçeliler	@gencfborg	Spor
173	**21**	Salih Uçan	@salihucan48	Spor
174	**22**	Alper Potuk	@AlperPOTUK	Spor
175	**23**	Selçuk Şahin	@_selcuksahin	Spor
176	**24**	Uğur Karakullukçu	@ukarakullukcu	Spor

177	**25**	Cenk Tosun	@CenkTosun_	Spor
178	**26**	Olcay Şahan	@olcay_sahan10	Spor
179	**27**	Sabri Sarıoğlu	@sbrsarioglu	Spor
180	**28**	Ersun Yanal	@ersunyanal	Spor
181	**29**	Cedi Osman	@cediosman	Spor
182	**30**	Hamit Altıntop	@FHamitAltintop	Spor
183	**31**	Oğuzhan Özyakup	@Ozyakup	Spor
184	**32**	Bekir İrtegün	@bekirtegun	Spor
185	**33**	Necati Ateş	@neco_ates	Spor
186	**34**	Merih Demiral	@Merihdemiral	Spor
187	**35**	Olcay Adın	@Olcanadin	Spor
188	**36**	Cengiz Ünder	@cengizunder	Spor
189	**37**	Ersan Gülüm	@GlmErsan	Spor
190	**38**	Ozan Tufan	@Ozan	Spor
191	**39**	Yusuf Yazıcı	@yaziciyusuf97	Spor
192	**40**	Sinan Gümüş	@sinanguemues18	Spor
193	**41**	Çağlar Söyüncü	@Syncaglar	Spor
194	**42**	Veli Kavlak	@velikavlak08	Spor
195	**43**	Cenk Gönen	@GonenCenk	Spor
196	**44**	Volkan Ballı	@volkanballi	Spor
197	**45**	Okay Yokuşlu	@Okayokuslu	Spor
198	**46**	Ozan Kabak	@ozankabak4	Spor
199	**47**	Alen Markaryan	@AlenMarkaryan	Spor
200	**48**	Melih Mahmutoğlu	@melihmahmutoglu	Spor
201	**49**	Mustafa Pektemek	@mpektemek11	Spor
202	**50**	Hasan Ali Kaldırım	@HAK3	Spor
203	**51**	Ümit Karan	@ukaran1976	Spor
204	**52**	Ömer Bayram	@omerbayram19_	Spor
205	**53**	Uğurcan Çakır	@cakirugurcan96	Spor
206	**54**	Aytaç Kara	@aytayt35	Spor
207	**55**	Ufuk Sarıca	@UfukSarica15	Spor
208	**56**	Ertuğrul Sağlam	@esaglamofficial	Spor
209	**57**	Ümit Özat	@UmitOzatOnline	Spor
210	**58**	Mert Hakan Yandaş	@MHYandas	Spor
211	**59**	Ertuğrul Taşkıran	@ErtugrulTSKRN	Spor
212	**60**	Erkan Zengin	@ErkanZ10	Spor
213	**1**	Gaf Ebesi	@gafebesi	Fenomenler
214	**2**	Danla Biliç	@danlabilic	Fenomenler
215	**3**	Odun Herif	@OdunHerif	Fenomenler
216	**4**	Özgür Demirtaş	@ProfDemirtas	Fenomenler
217	**5**	Aykut Elmas	@onudabitirin	Fenomenler
218	**6**	Pucca	@PuCCaa	Fenomenler
219	**7**	Furkan Sarıkaya	@Overlokcu12	Fenomenler
220	**8**	Fırat	@firatinblogu	Fenomenler
221	**9**	Mahfi Eğilmez	@mahfiegilmez	Fenomenler
222	**10**	Oğuz	@oguzatbas	Fenomenler
223	**11**	Prof Nadir	@NadirOfc	Fenomenler

224	**12**	Semih Varol	@Himesovski	Fenomenler
225	**13**	Turgut Uyar	@byTurgutUyar	Fenomenler
226	**14**	Emrö	@Emredb	Fenomenler
227	**15**	İronik Adam	@ironikadamm	Fenomenler
228	**16**	Atakan Özyurt	@atakannozyurt	Fenomenler
229	**17**	Kıvanç	@Evseksisi	Fenomenler
230	**18**	Önder Şeren	@onderseren	Fenomenler
231	**19**	Fırat Çetin	@pakiteysis	Fenomenler
232	**20**	Meriç Keskin	@otekidunyali	Fenomenler
233	**21**	Psk. Eray Efe	@Argostroloji	Fenomenler
234	**22**	Kuralsız	@KuRaLsiz_	Fenomenler
235	**23**	Taci Kalkavan	@taci_kalkavan	Fenomenler
236	**24**	Hakan Hepcan	@hakanhepcan	Fenomenler
237	**25**	Bilal Hancı	@BilalHanci	Fenomenler
238	**26**	Emre Tuncer	@Hayat_Felsefesi	Fenomenler
239	**27**	Ömür Özdemir	@ceriLevis	Fenomenler
240	**28**	Fatih Yasin	@Fatihyasinim	Fenomenler
241	**29**	Halil İbrahim Göker	@gokeroloji	Fenomenler
242	**30**	Ayarsız Herif	@AyarsizHerif	Fenomenler
243	**31**	Buse Terim Bahçekapılı	@BuseTerim	Fenomenler
244	**32**	Kerem Cevahir	@kerem_cevahir	Fenomenler
245	**33**	Saptroloji	@Saptroloji	Fenomenler
246	**34**	Şefik	@sefikdiyin	Fenomenler
247	**35**	Celal Can Algül	@celalcanalgul	Fenomenler
248	**36**	Huysuz Ayı	@HUYSUZ_AYI	Fenomenler
249	**37**	Simerazzi	@simerazzi	Fenomenler
250	**38**	Talha Başaran	@TalhaBasaran	Fenomenler
251	**39**	Pislik Adam	@BasaranOrhun	Fenomenler
252	**40**	Orkun Işıtmak	@Orkunisitmak	Fenomenler
253	**41**	Jahrein	@jahreindota	Fenomenler
254	**42**	Zeynep	@nebilicen	Fenomenler
255	**43**	Kemal Ekşioğlu	@KemalEksioglu	Fenomenler
256	**44**	Barış Özcan	@BarisOzcan	Fenomenler
257	**45**	Tahsin Hasoğlu	@tahsinhasoglu	Fenomenler
258	**1**	Fatih Portakal	@fatihportakal	Basın-Yayın
259	**2**	Cüneyt Özdemir	@cuneytozdemir	Basın-Yayın
260	**3**	Acun Ilıcalı	@acunilicali	Basın-Yayın
261	**4**	Ahmet Hakan	@ahmethc	Basın-Yayın
262	**5**	Can Dündar	@candundaradasi	Basın-Yayın
263	**6**	Uğur Dündar	@ugurdundarsozcu	Basın-Yayın
264	**7**	İsmail Küçükkaya	@KucukkayaIsmail	Basın-Yayın
265	**8**	Metin Uca	@MetinUca	Basın-Yayın
266	**9**	Penguen	@penguendergi	Basın-Yayın
267	**10**	İrfan Değirmenci	@degirmencirfan	Basın-Yayın
268	**11**	Ece Temelkuran	@Etemelkuran	Basın-Yayın
269	**12**	Ayşe Arman	@ArmanAyse	Basın-Yayın
270	**13**	Banu Güven	@banuguven	Basın-Yayın

271	**14**	İsmail Saymaz	@ismailsaymaz	Basın-Yayın
272	**15**	Ayşe Özyılmazel	@ayseozyilmazel	Basın-Yayın
273	**16**	Özgür Mumcu	@ozgurmumcu	Basın-Yayın
274	**17**	Nevsin Mengü	@nevsinmengu	Basın-Yayın
275	**18**	Yılmaz Özdil	@yilmazsozcu	Basın-Yayın
276	**19**	Saba Tümer	@saba_t	Basın-Yayın
277	**20**	Mesut Yar	@mesutyar	Basın-Yayın
278	**21**	Soner Yalçın	@hsoneryalcin	Basın-Yayın
279	**22**	Mirgün Cabas	@MirgunCabas	Basın-Yayın
280	**23**	Bekir Coşkun	@cosknbekr2	Basın-Yayın
281	**24**	Şirin Payzin	@siring	Basın-Yayın
282	**25**	Tuncay Özkan	@ATuncayOzkan	Basın-Yayın
283	**26**	Yekta Kopan	@yektakopan	Basın-Yayın
284	**27**	Vedat Milör	@vedatmilor	Basın-Yayın
285	**28**	Barış Yarkadaş	@barisyarkadas	Basın-Yayın
286	**29**	Armağan Çağlayan	@Armagan_caglaya	Basın-Yayın
287	**30**	Cem Seymen	@cemseymen	Basın-Yayın
288	**31**	Mehmet Demirkol	@mehmetdemirkolb	Basın-Yayın
289	**32**	Fatih Altaylı	@fatihaltayli	Basın-Yayın
290	**33**	Umut Katırcı	@umutkatirci	Basın-Yayın
291	**34**	Can Ataklı	@can_atakli_	Basın-Yayın
292	**35**	Nedim Şener	@nedimsener2010	Basın-Yayın
293	**36**	Cevo	@CEVDETKAVLAK	Basın-Yayın
294	**37**	Hakan Demir	@suyorumcusu	Basın-Yayın
295	**38**	Ahmet Şık	@sahmetsahmet	Basın-Yayın
296	**39**	Rahşan Gülşan	@rahsangulsan	Basın-Yayın
297	**40**	İhsan Şenocak	@ihsansenocak	Basın-Yayın
298	**41**	Ömür Gedik	@omurgedik	Basın-Yayın
299	**42**	Ali Türkşen	@aliturksen	Basın-Yayın
300	**43**	Nihat Genç	@nihadagenc	Basın-Yayın
301	**44**	Abdurrahman Dilipak	@aDilipak	Basın-Yayın
302	**45**	Hasan Cemal	@hsncml	Basın-Yayın
303	**46**	Dr. Barış Esen	@barisesen	Basın-Yayın
304	**47**	Dr. Hulki Cevizoğlu	@mhulkicevizoglu	Basın-Yayın
305	**48**	Ezgi Başaran	@ezgibasaran	Basın-Yayın
306	**49**	Ruşen Çakır	@cakir_rusen	Basın-Yayın
307	**50**	Esra Erol	@ErolEsra	Basın-Yayın
308	**51**	Zeki Kayahan Coşkun	@zekikayahan	Basın-Yayın
309	**52**	Nihat Sırdar	@nihatsirdar	Basın-Yayın
310	**53**	Emin Çölaşan	@ecolasan	Basın-Yayın
311	**54**	Banu Avar	@AvarBanu	Basın-Yayın
312	**55**	Emre Kongar	@emrkongar	Basın-Yayın
313	**56**	Sedef Kabas	@SedefKabas	Basın-Yayın
314	**57**	Ersin Düzen	@ersinduzen	Basın-Yayın
315	**58**	Nureddin Yıldız	@nurettinyildiz	Basın-Yayın
316	**59**	Emre Uslu	@EmreUslu	Basın-Yayın
317	**60**	Nihal Bengisu Karaca	@nibenka	Basın-Yayın

318	**61**	Emin Çapa	@ecapa_aklinizi	Basın-Yayın
319	**62**	İhsan Eliaçık	@ihsaneliacik	Basın-Yayın
320	**63**	Hilal Kaplan	@hilal_kaplan	Basın-Yayın
321	**64**	Cengiz Semercioğlu	@cengizsemercio	Basın-Yayın
322	**65**	Ali Ece	@aliececom	Basın-Yayın
323	**66**	Feyzullah Doğru	@feyzullahdogru	Basın-Yayın
324	**67**	Mehmet Baransu	@mehmetbaransu	Basın-Yayın
325	**68**	Ali Keskin	@alikeskin_tr	Basın-Yayın
326	**69**	Açıl Sezen	@aksezen	Basın-Yayın
327	**70**	Koray Çalışkan	@koraycaliskan	Basın-Yayın
328	**71**	Sinan Meydan	@SMEYDAN	Basın-Yayın
329	**72**	Yusuf Kaplan	@yenisafakwriter	Basın-Yayın
330	**73**	İbrahim Karagül	@ibrahimkaragul	Basın-Yayın
331	**74**	Süleyman Özışık	@slymnoz	Basın-Yayın
332	**75**	Mustafa Armağan	@mustafarmagan	Basın-Yayın
333	**76**	Fehmi Koru	@fkoru	Basın-Yayın
334	**77**	Rasim Ozan Kütahyalı	@RasimOzan_K	Basın-Yayın
335	**78**	Ahmet Ercanlar	@ahmetercanIar	Basın-Yayın
336	**79**	Hakan Bayrakçı	@ahakanbayrakci	Basın-Yayın
337	**80**	Erkan Tan	@ERKANTAN_	Basın-Yayın
338	**81**	İlber Ortaylı	@ILBERORTAYLIGSU	Basın-Yayın
339	**82**	Deniz Ülke Arıboğan	@DenizUlke	Basın-Yayın
340	**83**	Dücane Cündioğlu	@ducane	Basın-Yayın
341	**84**	Nevzat Tarhan	@drnevzattarhan	Basın-Yayın
342	**1**	Selçuk Bayraktar	@Selcuk	İş Dünyası
343	**2**	Abdurrahim Albayrak	@_Aalbayrak	İş Dünyası
344	**3**	Nevzat Aydın	@zagortenay76	İş Dünyası
345	**4**	Umut Yildiz	@umutayildiz	İş Dünyası
346	**5**	Sina Afra	@SinaAfra	İş Dünyası
347	**6**	Ümit Nazlı Boyner	@umitnazliboyner	İş Dünyası
348	**7**	Erkin Şahinöz	@ErkinSahinoz	İş Dünyası
349	**8**	Rifat Hisarcıkoğlu	@Rhisarcikliogl u	İş Dünyası
350	**9**	Ali Yalçın	@_aliyalcin_	İş Dünyası
351	**10**	Burak Buyukdemir	@burakbuyukdemir	İş Dünyası
352	**11**	Demet Sabancı	@demetsabanci	İş Dünyası
353	**12**	Cem Boyner	@cem_boyner	İş Dünyası
354	**13**	Dr Kerem Kınık	@drkerem	İş Dünyası
355	**14**	Necdet Ünüvar	@necdetunuvar	İş Dünyası
356	**15**	Doç. Dr. Mustafa Aydın	@drmaydin	İş Dünyası
357	**16**	Erdal Kaplanseren	@Kaplanseren	İş Dünyası
358	**17**	Bilal Ekşi	@BilalEksiTHY	İş Dünyası
359	**18**	Burak Elmas	@burakelmas_tr	İş Dünyası
360	**19**	Talip Geylan	@TalipGeylan06	İş Dünyası
361	**20**	Ali Rıza Babaoğlan	@alibabaoglan	İş Dünyası
362	**21**	EnverYücel	@EnverYucel	İş Dünyası
363	**22**	Nazım Salur	@NazimSalur	İş Dünyası
364	**23**	Dr. Serdar Çam	@serdar_cam	İş Dünyası

365	**24**	Leyla Alaton	@AlatonLeyla	İş Dünyası
366	**25**	Boğaçhan Göksu	@bogachangoksu	İş Dünyası
367	**26**	İsfendiyar Zülfikari	@IsfendiyarZ	İş Dünyası
368	**27**	İlhan Taşçı	@ilhantasci	İş Dünyası
369	**28**	Salih Turhan	@AGDSalihTURHAN	İş Dünyası
370	**29**	Recep Konuk	@recep_konuk	İş Dünyası
371	**30**	Fettah Tamince	@Ftamince	İş Dünyası
372	**31**	Kaan Terzioğlu	@Kaan_Terzioglu	İş Dünyası
373	**32**	Adil Sani Konukoğlu	@askonuk	İş Dünyası
374	**33**	Önder Kahveci	@OnderKahveci	İş Dünyası
375	**34**	Hacı Sabancı	@hacisabanci	İş Dünyası
376	**35**	Nail Olpak	@NailOlpak	İş Dünyası
377	**36**	Murat Erkan	@Murat__Erkan	İş Dünyası
378	**37**	Abdurrahman Kaan	@AKaanTR	İş Dünyası