T.C.
ERCİYES ÜNİVERSİTESİ
MÜHENDİSLİK FAKÜLTESİ

SOSYAL MEDYA TEHDİT ANALİZİ

Hazırlayan

Melisa DEMİR
1030510331
Duygu Gözde KAYABAŞI
1030510338

Danışman
Doç. Dr. Özkan Ufuk NALBANTOĞLU

Bilgisayar Mühendisliği
Bitirme Ödevi

Ocak 2025
KAYSERİ

"Sosyal Medya Tehdit Analizi" adlı bu çalışma, jürimiz tarafından Erciyes Üniversitesi Bilgisayar Mühendisliği Bölümünde Bitirme Ödevi olarak kabul edilmiştir.

11/10/2024

JÜRİ :

Danışman : Doç. Dr. Özkan Ufuk NALBANTOĞLU

Üye :

Üye :

ONAY :

Yukarıdaki imzaların, adı geçen öğretim elemanlarına ait olduğunu onaylarım.

.... / .... /20...
Prof. Dr. Veysel ASLANTAŞ
Bilgisayar Müh. Bölüm
Başkanı

ÖNSÖZ / TEŞEKKÜR

Bu proje, sosyal medya üzerinden toplum güvenliğini tehdit eden unsurların tespit edilmesine yönelik farkındalığı artırmak ve bu alandaki tehditlerin etkin bir şekilde analiz edilmesine katkı sağlamak amacıyla hayata geçirilmektedir. Projemizin ana hedefi, sosyal medyada yer alan verilerin yapay zekâ ve doğal dil işleme teknikleriyle analiz edilerek potansiyel tehditleri önceden belirlemek ve bu konuda stratejik kararlar alınmasına destek olmaktır.

Bu çalışmada, X (Twitter), Reddit ve Telegram gibi popüler sosyal medya platformlarından elde edilen veriler üzerine metin madenciliği ve makine öğrenmesi yöntemleri uygulanmıştır. Analiz sonuçları ile tehditlerin erken tespit edilebilmesi planlanmaktadır.

Projemizin başarıyla bu aşamaya gelmesinde emeği geçen danışman hocamız Doç. Dr. Özkan Ufuk Nalbantoğlu'na rehberliği ve destekleri için en içten teşekkürlerimizi sunarız. Ayrıca, bu projeye katkı sağlayan yapay zekâ modellerini geliştiren araştırmacılara ve kullanılan kütüphanelerin yazılım geliştiricilerine de teşekkür ederiz.

Saygılarımızla,

SOSYAL MEDYA TEHDİT ANALİZİ

Melisa DEMİR
Duygu Gözde KAYABAŞI
Erciyes Üniversitesi, Bilgisayar Mühendisliği
Danışman : Doç. Dr. Özkan Ufuk NALBANTOĞLU

ÖZET

Projemiz, sosyal medya platformlarından toplanan büyük verinin analizi aracılığıyla potansiyel tehdit unsurlarını tespit etmeyi hedefleyen yapay zeka destekli bir sistemdir. Bu sistem, doğal dil işleme (NLP) teknikleri kullanarak kullanıcı paylaşımlarını analiz eder ve metinlerdeki duygusal tonu belirlemek için duygu analizi uygular. Toplanan veriler, istatistiksel analiz yöntemleriyle işlenerek anlamlı bilgilerin elde edilmesine olanak tanır. Verilerin güvenilirliğini artırmak için, proje sürecinde veri ön işleme aşaması gerçekleştirilir; bu aşama, verilerin temizlenmesi ve düzenlenmesini içerir. Makine öğrenmesi algoritmaları, tehdit içeren içerikleri tanımlamak ve sınıflandırmak için kullanılacaktır; böylece sistem, gerçek zamanlı olarak ortaya çıkan tehditleri belirleyerek toplumsal güvenliği artıracaktır. Algoritmanın etkinliği, sürekli öğrenme yeteneği sayesinde zamanla daha da gelişecektir.

Anahtar Kelimeler : Yapay zeka, Doğal dil işleme, Makine öğrenmesi, Büyük veri, İstatistiksel analiz, Duygu analizi, Tehdit unsuru, Veri ön işleme, Algoritma

- **Yapay Zeka:** Bilgisayarların insan benzeri görevleri yerine getirmesini sağlayan teknolojilerdir. Yapay zeka, karar verme süreçlerini otomatikleştirme, öğrenme ve problem çözme yetenekleri sunar. [1]
- **Doğal Dil İşleme:** Bilgisayarların insan dilini anlama, yorumlama ve işleme yeteneğidir. Bu teknoloji, metin madenciliği, dil çevirisi ve duygu analizi gibi görevlerde kullanılır. [2]
- **Makine Öğrenmesi:** Bilgisayarların deneyim yoluyla öğrenmesini ve gelişmesini sağlayan algoritmalardır. Bu süreç, verilerden otomatik olarak öğrenip gelecekteki olayları tahmin etmeyi içerir. [3]

- **Büyük Veri:** Geleneksel veri işleme yöntemleriyle işlenemeyen, büyük ve karmaşık veri setleridir. Bu veri setleri, iş süreçlerinde analiz edilerek karar vermeyi optimize etme potansiyeli taşır. [4]
- **İstatistiksel Analiz:** Veri setlerini inceleyerek anlamlı bilgiler çıkarma, örüntüleri tanıma ve sonuçlar elde etme sürecidir. İstatistiksel analiz, verilerin davranışlarını belirlemek için hipotez testlerinden faydalanır. [5]
- **Duygu Analizi:** Metinlerdeki duygusal tonların ve ifadelerin belirlenmesi sürecidir. Duygu analizi, sosyal medya paylaşımlarındaki olumlu ya da olumsuz duyguları anlamak için kullanılabilir. [6]
- **Tehdit Unsuru:** Bir sistem, süreç, varlık veya birey üzerinde zarar verme potansiyeline sahip olay, durum veya eylemleri ifade eder.
- **Veri Ön İşleme:** Analiz öncesi verilerin temizlenmesi ve düzenlenmesi işlemleri.
- **Algoritma:** Belirli bir problemi çözmek için izlenen adımlar dizisi.

SOCIAL MEDIA THREATS ANALYSIS

Melisa DEMİR
Duygu Gözde KAYABAŞI
Erciyes University, Computer Engineering
Supervisor: (Associate Professor) Özkan Ufuk NALBANTOĞLU

ABSTRACT

Our project is a system supported by artificial intelligence that targeting to detect all potential threat element through analysis of the big data that collected social media platforms. This system analyzes all posts of users using natural language processing techniques and applies sentiment analysis to identifying the sentimental tone in the texts. It gives you a chance to get meaningful information processing collected data, using by methods of statistical analysis. To increases reliability of those datas, the step of Data preprocessing should be executed. This step includes data cleaning and data filtering. Machine Learning algorithms should be using on to determine all contents that including threat elements and to classify them . So this way , the system will be increase the social security by determining the threats that showing up as in real time. The effectiveness of the algorithms will be develops itself by time cause of ability to learn continuously

Keywords: Artificial Intelligence, Natural Language Processing, Machine Learning, Big Data, Statistical Analysis, Sentiment Analysis, Threat Element, Data Preprocessing, Algorithm .

- **Artificial Intelligence:** The technologies that enable computers to perform human-like tasks. Artificial Intelligence presents talents like the problem solving, learning and automating decision-making processes. [1]
- **Natural Language Processing:** The ability of computers to understand, interpret and process human language. This technology is used in tasks such as text mining, language translation and sentiment analysis. [2]
- **Machine Learning:** The algorithms that allow computers to learn and improve through experience. This process involves predicting future events by automatically learning from data. [3]

- **Big Data:** These are large and complex data sets that cannot be processed with traditional data processing methods. These datasets have the potential to optimize decision-making by being analyzed in business processes. [4]
- **Statistical Analysis:** It is the process of examining data sets to extract meaningful information, recognize patterns and obtain results. Statistical analysis uses hypothesis testing to determine the behavior of data. [5]
- **Sentiment Analysis:** It is the process of determining emotional tones and expressions in texts. Sentiment analysis can be used to understand positive or negative sentiment in social media posts. [6]
- **Threat Element:** A Reference to events, situations, or actions that have the potential to cause harm to a system, process, asset, or individual.
- **Data Preprocessing:** Process of cleaning and organizing data before analysis.
- **Algorithm:** A sequence of steps followed to solve a particular problem.

İÇİNDEKİLER

SOSYAL MEDYA TEHDİT ANALİZİ

TABLOLAR LİSTESİ

ŞEKİLLER LİSTESİ

KISALTMALAR

AI	:	Artificial Intelligence
NLP	:	Natural Language Processing
DD	:	Doğal Dil İşleme
$CVFDT$	:	Concept-adapting Very Fast Decision Tree
$DT - Part$	:	Decision Tree Partitioning
MLP	:	Multilayer Perceptron
SVM	:	Support Vector Machines
NLP	:	Doğal Dil İşleme
$PRAW$	:	PythonRedditAPIWrapper
API	:	Application Programming Interface
CSV	:	Comma Seperated Values
ID	:	Identification (Kimlik)
TF	:	Term Frequency
IDF	:	Inverse Document Frequency
SVM	:	Support Vector Machine
ROC	:	Receiver Operating Characteristic
AUC	:	Area Under The Curve

GİRİŞ

Sosyal medya tehdit analizi projesi , bireylerin günlük hayatında sıkça kullandığı bazı sosyal medya platformlarının verilerinin analizini içermektedir.

İnsanlar çeşitli konularda görüşlerini ve önerilerini sosyal medya platfomrları aracılığı ile paylaşmaktadırlar. Kimi zaman bu paylaşımların içerikleri toplum veya anayasa kurallarına aykırı olabilmektedir. Günümüzde çokça gözlemlemiş olduğumuz bu içerikler tehdit oluşturabilmektedir.

Yapay zekâ aracılığı ile tehdit içeren paylaşımların tespit edilmesi sağlanacaktır. Tespit edilen tehdit verileri de bir uygulama aracılığı ile görselleştirilerek kullanıcılara sunulacaktır. Toplum veya birey zararına olabilecek tehditlerin erken tespiti sağlanması planlanmaktadır.

Literatür Özeti

1. Sosyal Medyanın Etkileri ve Analizleri:

    - Jean Baudrillard'ın teorileri, sosyal medyanın bireyler ve toplum üzerindeki etkilerini açıklamak için kullanılmıştır. Bu bağlamda sosyal medyanın, bireylerin ihtiyaçlarına göre sürekli evrildiği ve iktidarlar açısından manipülasyon aracı olarak önem taşıdığı belirtilmiştir. [7]
    - Dini ve etnik temelli nefret söylemlerinin X (Twitter) üzerindeki varlığı, sosyal medya platformlarının nefret söyleminin yayılmasına olanak sağladığını ortaya koymaktadır. Bu söylemler toplumsal huzuru tehdit etmektedir. [8]

2. Yapay Zekâ Tabanlı Sosyal Medya Analizleri:

    - Turistik mekanlara yönelik sosyal medya verilerinin yapay zekâ yöntemleriyle analiz edilmesi, duygu analizi ve görsel içerik işleme gibi yöntemlerle gerçekleştirilmiştir. Bu analizler, müşteri memnuniyetinin ve ziyaretçi deneyimlerinin anlaşılmasına yardımcı olmuştur. [9]
    - Sahte haberlerin tespiti için doğal dil işleme (NLP) yöntemleri kullanılarak haberlerin gerçek veya sahte olduğunu belirlemek üzere bir yapay zekâ modeli önerilmiştir. [10]

3. Yapay Zekâ ve Siber Tehdit İstihbaratı:

    - Makine öğrenmesi ve yapay zekâ, siber tehditlerin tespitinde ve istihbarat süreçlerinde önemli bir rol oynamaktadır. Büyük veri analizi ile tehditlerin hızlı ve doğru şekilde belirlenmesi sağlanmaktadır. [11]
    - Bir başka çalışmada, X (Twitter) verileri üzerinden risk yönetimi ve çevrimiçi itibar analizi için yapay zekâ modelleri geliştirilmiştir. Bu modeller, olumsuz tweetlerin sınıflandırılması ve kurumsal itibarın korunmasında kullanılmaktadır. [12]

4. Yapay Zekâ ve İstihbarat Analizi:

    - Yapay zekâ, istihbarat süreçlerinde veri toplama, analiz ve bilgi üretim aşamalarında etkin bir şekilde kullanılmaktadır. Kategorileştirme ve büyük veri yığınlarının işlenmesi, yapay zekâ sistemleriyle kolaylaştırılmaktadır. [13] [14]

Projenin Amacı ve Önemi

Sosyal Medya Tehdit Analizi Projesi, bireylerin günlük yaşamlarında sıkça kullandıkları sosyal medya platformlarında yapılan paylaşımların analiz edilmesini hedeflemektedir. İnsanlar sosyal medya aracılığıyla görüş ve önerilerini paylaşırken, kimi zaman bu içerikler toplum veya anayasa kurallarına aykırı olabilmektedir. Bu tür paylaşımlar tehdit oluşturabileceğinden, projenin temel amacı, yapay zekâ destekli analizlerle tehdit içeren içeriklerin erken tespit edilmesini sağlamaktır.

Proje kapsamında, tehdit oluşturan paylaşımlar tespit edilip, kullanıcı dostu bir uygulama aracılığıyla görselleştirilerek sunulacaktır. Böylece, toplum ya da birey zararına olabilecek durumların erken fark edilmesi ve önlem alınması hedeflenmektedir.

Projeyi özgün kılan ve önemli kılan noktalar şunlardır:

Türkçe Dil Modellemesi: Sosyal medya tehdit analizi projelerinde genellikle Türkçe dil desteği bulunmamaktadır. Bu projede, Türkçenin yapısına uygun doğal dil işleme modelleri geliştirilerek özgün bir yaklaşım sunulmaktadır.

Gerçek Zamanlı Tehdit Takibi: Sosyal medya platformlarındaki tehdit eğilimlerinin anlık analiz edilip görselleştirilmesi, tehditlerin hızlı ve etkili bir şekilde izlenmesini sağlamaktadır.

Görselleştirilmiş Tehdit Haritaları: Özgün grafikler ve haritalarla tehditlerin görselleştirilmesi, verilerden hızlı anlam çıkarılmasına olanak tanımaktadır. Bu yaklaşım, Türkiye'de sosyal medya analizinde nadir kullanılan bir yöntemdir.

Platforma Özgü Modeller: X (Twitter), Reddit ve Telegram platformları için ayrı analiz modelleri geliştirilerek, her platformun kendine özgü etkileşim biçimleri dikkate alınmaktadır. Bu yaklaşım, analizlerin doğruluğunu ve etkisini artırmaktadır.

Bu proje, sosyal medya tehditlerinin tespitinde yenilikçi yöntemler sunarak toplumsal güvenlik ve bilinçlenme açısından önemli bir adım niteliğindedir.

Tezin Organizasyonu

Bu çalışmanın düzenlemesi şu şekildedir.

1. Bölüm, VERİ TOPLAMA , TEMİZLEME VE ETİKETLEME SÜRECİ

2. Bölüm, MODEL EĞİTİMİ VE GERÇEK ZAMANLI VERİ ANALİZİ

3. Bölüm, TARTIŞMA, SONUÇ VE ÖNERİLER

1. BÖLÜM

VERİ TOPLAMA , TEMİZLEME VE ETİKETLEME SÜRECİ

1.1. Veri Kaynağı Seçimi

Sosyal medya tehdit analizi projesi için içerik tipleri farklı olan 3 adet sosyal medya platformu seçilmiştir. Bunlar :

- **Telegram :** Veri tipi olarak kullanıcıların gönderdiği mesajları içerir. Telegram platformunda haber kanalları bulunmaktadır. Bu haber kaynaklarında haber olarak paylaşılan içerikler bulunmaktadır. Bu içerikler kimi zaman taraflı veya paylaşılması yasak olan, sansürsüz, verilerden oluşabilmektir. Bu verilerin tespiti amacıyla Telegram platformu veri kaynaklarından biri olarak seçilmiştir.
- **X (Twitter) :** Son zamanlarda sıkça kullanılan sosyal medya platformlarında birisi olan X (Twitter), milyonlarca insanın fikirlerini sansürsüz şekilde paylaşabilmesine olanak tanımaktadır. Bu özelliği ile tehdit içerikli bir çok veri bulunmaktadır. Veri kaynaklarından birisi olarak seçilmesinin en büyük nedeni de fazla sayıda ihtiyacımız olan veri içermesidir.
- **Reddit:** Bu platformda kullanıcılar çoğunlukla anonimdir. Anonimlik görüşleri belirtmede insanlara daha fazla cesaret vermektedir. Reddit'te kullanıcılar çeşitli konularda topluluklara ayrılmışlardır. Milyonlarca kullanıcının paylaşımlarından oluşan veriler proje için uygun bir veri kaynağı olarak seçilmesini sağlamıştır.

Tablo 1.1. Platformlara Göre Veri Dağılımı ve Özellik Tablosu.

Platform	Veri Türü	İçerik Uzunluğu	Format
Twitter (X)	Post , Repost	280 Karakter	JSON
Telegram	Mesaj , Dosya	Uzun / Orta	Düz Metin / JSON
Reddit	Gönderi , Yorum	Uzun / Orta	JSON

1.2. Veri Çekme

1.2.1. Twitter'dan Veri Çekme

Projemizin başlangıç aşamasında, Twitter (X) platformundan veri toplamak amacıyla Twitter API'sini kullanmayı planladık. Ancak bu süreçte, API'nin veri çekme limitleriyle ilgili bir değişiklik yaşandı. Twitter (X) API'si, ücretsiz hesaplar için veri çekme limitini 100 ile sınırlamıştır. Bu durum, projemizde ihtiyaç duyduğumuz büyük miktarda veriyi toplama sürecimizi ciddi şekilde etkiledi. Bu platform için yeterli miktarda veri sayısına ulaşabilmek için, daha önce Twitter(X) platformundan çekilmiş ve hazır hale getirilmiş açık veri setlerinden faydalanmayı tercih ettik. Bu hazır veri setleri, çeşitli araştırmacılar ve veri bilimciler tarafından sağlanmış olup, proje kapsamında ihtiyaç duyduğumuz 30.000 adet veriyi sağlamıştır.

1.2.2. Reddit'ten Veri Çekme

1. **Gerekli Kütüphanelerin ve Bilgilerin Tanımlanması:**

   PRAW(PythonRedditAPIWrapper), RedditAPI ile etkileşim kurmak için kullanılan bir Python kütüphanesidir. Reddit'ten veri çekme ve gönderi yönetimi gibi işlemleri kolaylaştırmaktadır. Reddit API platformunda oturum açılarak elde edilen kimlik doğrulama bilgileri (Client_id, Client_secret ve User_agent) kullanılarak API'ye erişim sağlanmaktadır.

2. **Reddit Platformundan Veri Çekme İşlemi:**

   Veri çekmek istenen subreddit olarak bir konu seçilmektedir. Aynı zamanda çekilecek veri limiti ve veri içerik formatında belirlenir. Veriler : Başlık, gönderi içeriği, yorum sayısı, beğeni sayısı ve gönderi tarihi başlıkları altında çekilir. Çekilen veriler istenilen başlıklar kolon isimleri olarak belirlendikten sonra bir csv dosyasına kaydedilir.

Reddit platformundan 10000 adet veri çekilmiştir.

1.2.3. Telegram'dan Veri Çekme

1. **Gerekli Kütüphanelerin ve Bilgilerin Tanımlanması**

   Telegram'dan veri çekmek için Telethon kütüphanesi kullanılır. Bu kütüphane, Telegram API ile iletişim kurmayı sağlayan bir araçtır. Kullanıcı doğrulama için API ID, API Hash ve telefon numarası gibi bilgiler gereklidir. api_id ve api_hash Telegram Developer hesabından alınır ve API'ye erişim sağlar. phone_number ise Telegram hesabına giriş yapılacak telefon numarasıdır.

2. **Telegram İstemcisi Tanımlama**

   Telegram API ile iletişim kurmak için bir istemci oluşturulur. Bu istemci, API ile yapılacak tüm işlemlerin temelidir.

   - TelegramClient: Telethon'un temel sınıfıdır ve bir Telegram istemcisi oluşturur.
   - session_name: Oturum bilgilerini temsil eder ve bir dosyada saklanır. Bu, tekrar giriş yapılmasını gereksiz kılar.

```
client = TelegramClient( session: 'session_name', api_id, api_hash)
```

Şekil 1.1. İlgili kod parçası (1)

3. **Grup Mesajlarını Çekme** Telegram gruplarının kimlikleri (ID) telegram botları aracılığı ile alınır. Belirtilen grup ID'sine ait mesajlar veri çekme fonksiyonu ile çekilerek bir csv dosyasına kaydedilir.

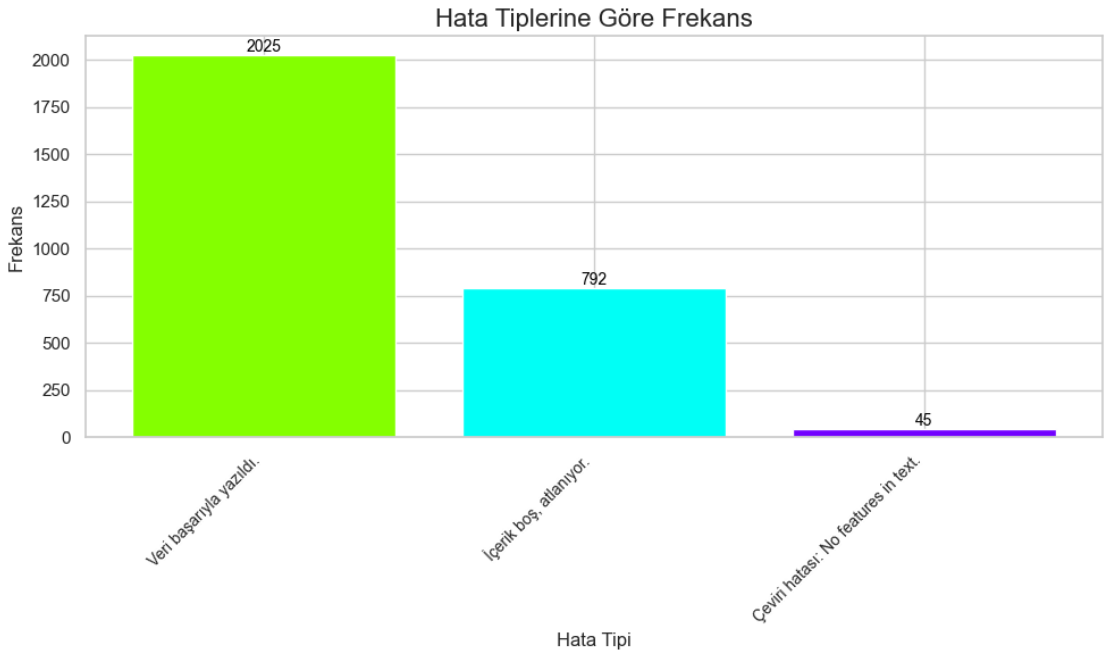
Telegram platformundan 10000 adet veri çekilmiştir.

1.2.4. Veri çekme işlemi sırasında hata analizi

Elde edilen verilerin çekilmesi sırasında oluşan hata tiplerinin frekansı aşağıdaki şekilde görselleştirilmiştir:

- "Veri başarıyla yazıldı."
- "İçerik boş, atlanıyor."
- "Çeviri hatası: No features in text."

Grafik, veri çekme işlemi sırasında hata türlerinin sıklığını görselleştirerek, veri çekme sürecinde hangi tip hataların daha yaygın olduğunu göstermektedir.



Şekil 1.2. Hata tiplerine göre frekans grafiği

1.3. Veri Temizleme Süreci

Veri temizleme süreci, ham verilerin analiz edilebilir ve modelleme için uygun hale getirilmesi amacıyla yapılan bir adımdır. Bu süreçte çekilen verilerdeki boş sütunlar silinir, gereksiz karakterler silinir ve veri tek tip formata getirilir.

Veri temizleme aşamasında tüm platformlar için kullanılan kodun işleyişi ve veri temizleme süreci :

1.3.1. Veri Yükleme

```python
import pandas as pd
import re

# Veri setini okur
df = pd.read_csv('data.csv')
```

Şekil 1.3. İlgili kod parçası (2)

Kaydedilen veriler pandas kütüphanesi sayesinde bir pandas DataFrame nesnesine aktarılır, böylece işlenmesi kolay hale gelir.

1.3.2. Gereksiz Sütunların ve Boş Değerlerin Kaldırılması

```python
# Gereksiz sütunları kaldırma
df = df.drop(columns=['media'], errors='ignore')

# Boş değerleri kaldırır
df = df.dropna()
```

Şekil 1.4. İlgili kod parçası (3)

Daha az karmaşık bir veri seti oluşturmak amacıyla "media" gibi modelleme ve analiz için gereksiz olan sütunlar veri kümesinden çıkarılır. Ayrıca boş hücreler analiz sürecinde hatalara neden olabileceğinden bu işlemle eksik veriler temizlenir.

1.3.3. Metin Temizleme

```python
# Metin temizleme fonksiyonu
def temizle(metin):
    metin = metin.lower()
    metin = re.sub(r'[^\w\s]', '', metin)
    metin = re.sub(r'\s+', ' ', metin).strip()
    return metin

# Metni temizle
df['Gönderi İçeriği'] = df['Gönderi İçeriği'].apply(temizle)
```

Şekil 1.5. İlgili kod parçası (4)

Gürültülü metinler sade ve analiz edilebilir bir forma dönüştürülmesi için metni küçük harfe dönüştürerek tutarlılık sağlanır, noktalama işaretleri ve özel karakterler regex kullanılarak kaldırılır ve gereksiz boşluklar temizlenerek ve metin sıkıştırılır.

1.3.4. Temizlenmiş Verileri Kaydetme

```python
# Temizlenmiş veriyi kaydet
df.to_csv('cleared_data.csv', index=False)
```

Şekil 1.6. İlgili kod parçası (5)

Temizlenmiş veriler, yeniden kullanılmak üzere yeni bir CSV dosyasına kaydedilir.

```
Başlık,Gönderi İçeriği,Yorum Sayısı,Upvote Sayısı,Gönderi Tarihi
"SEJ, Reddit'in pazarlama ve ürün liderleriyle bir AMA düzenliyor - Reddit'in markanız için nasıl çalışacağını öğrenmek için bu fırsatı

[Search Engine Journal](https://www.reddit.com/user/SearchEngineJournal/) (SEJ), Reddit'in pazarlama, içgörüler ve ürün liderleriyle bir

[Katılmak için buraya kaydolun](https://www.searchenginejournal.com/webinar-lp-official-reddit-ama-exclusive-chance-to-ask-reddit-your-c

Sunucular ve moderatör şunlardır:

https://preview.redd.it/a7kl0vlxp9yd1.png?width=1100&format=png&auto=webp&s=e1172672fb2b995c1ba07c04c8c571ea86877a0e

Düzenleme: Reddit, soru toplamaya başlamak için resmi AMA başlığını yayınladı, bu nedenle Reddit'e sormak istediğiniz bir şey varsa mutl
[https://www.reddit.com/r/RedditforBusiness/comments/1gjmu3d/sej\_is\_hosting\_an\_ama\_with\_reddits\_global/](https://www.reddit.com/r
Facebook Etkileyicilerini Bulmak İçin İyi Bir Platform Mu?,"Sadece Facebook etkileyicilerini bulup onlarla bağlantı kurmak için iyi bir
```

Şekil 1.7. İlgili kod parçası (6)

1.3.5. Veri Temizleme İşlemi Sonuçları

1. **Telegram** Telegram platformundan çekilen veriler üzerinde veri temizleme işlemi sonucunda 2031 adet veri model eğitimi için hazır duruma getirilmiştir.

2. **Twitter(X)** Twitter(X) platformu için hazırlanan veri seti üzerinde yapılan veri temizleme işlemi sonucunda 33910 adet veri model eğitimi için hazır duruma getirilmiştir.

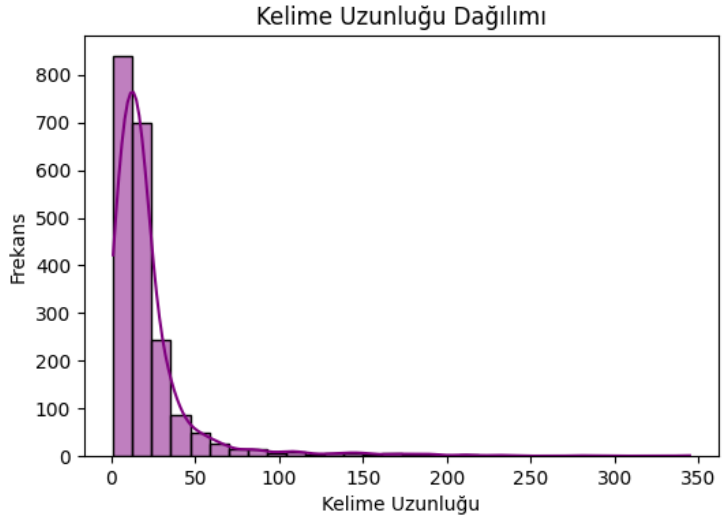
3. **Reddit** Reddit platformundan çekilen veriler üzerinde yapılan veri temizleme işlemi sonucunda 4557 adet veri model eğitimi için hazır duruma getirilmiştir.

1.4. VERİ ÖN İŞLEME VE ETİKETLEME

1.4.1. Veri Ön İşleme

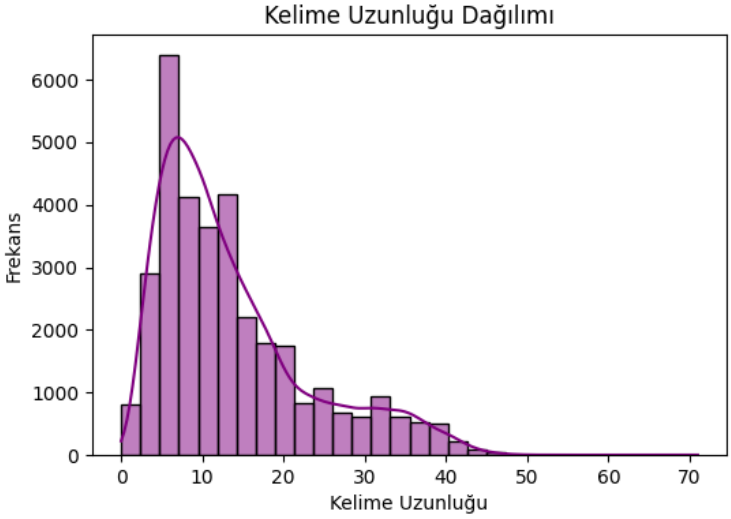
1.4.1.1. Kelime Uzunluğu Dağılımı Grafikleri

Kelime uzunluğu dağılımı, veri setlerindeki metinlerin genel yapısını anlamak için kullanılmıştır. Her bir sosyal medya platformunun (Twitter(X), Reddit, Telegram) kendine özgü dil ve yazım tarzları olduğu için, kelimelerin uzunluğu da farklılık gösterebilmektedir. Bu grafikler, veri setindeki metinlerin hangi uzunluktaki kelimelerle yoğunlaştığını, kısa veya uzun kelimelerin sıklığını görselleştirmiştir.



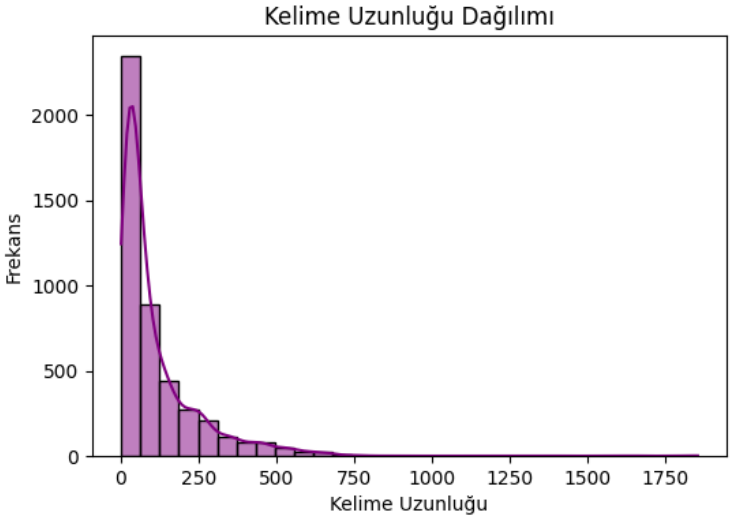
Şekil 1.8. Telegram platformu için kelime uzunluğu dağılımı grafiği

Telegram daha çok mesaj dilinin kullanıldığı bir platform olduğundan verilerde kullanılan kelime uzunlukları değişkenlik gösterir. Grafiğe göre bu veri setinde kelimeleri orta uzunluktaki kelimelerin frekansının daha yüksek olduğu görülmektedir.



Şekil 1.9. Twitter (X) platformu için kelime uzunluğu dağılımı grafiği

Twitter(X) 'da gönderilerin karakter sınırlaması olduğundan kelime uzunluğu aralığı diğer platformlara göre daha küçüktür ve daha kısa kelimeler sıklıkla kullanılır. Grafikte de en çok kullanılan kelimelerin kısa kelimeler olduğu gözlemlenmektedir.



Şekil 1.10. Reddit platformu için kelime uzunluğu dağılımı grafiği

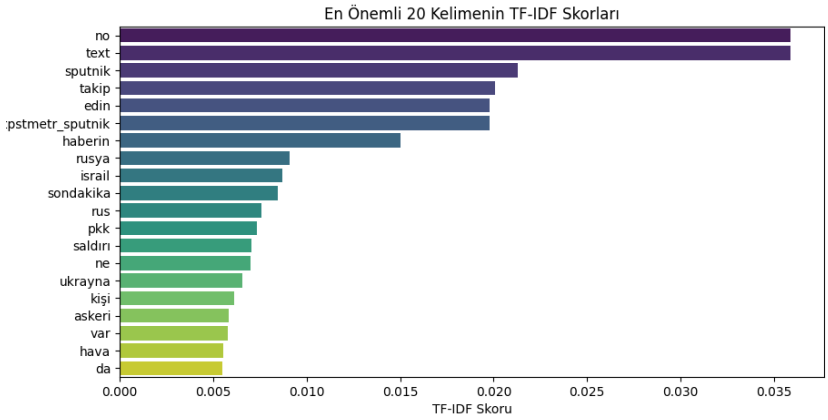
Reddit platformunda gönderilerin karakter sınırlaması olmadığından uzun içeriklere sıklıkla rastlanmaktadır. Grafikte diğer platformlara göre büyük bir farkla daha fazla karakter içeren kelimelerin veri setinde çokça bulunduğu gözlemlenmektedir.

1.4.1.2. En Önemli 20 Kelimenin TF-IDF Skorları

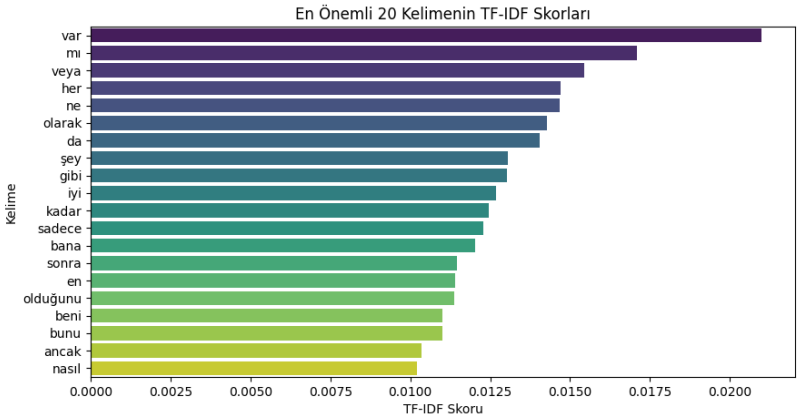
TF (Term Frequency): Bir kelimenin bir belgede ne kadar sık geçtiğini ölçmektedir. Sık geçen kelimeler genellikle o belgenin bağlamında önemli kabul edilmektedir.

IDF (Inverse Document Frequency): Bir kelimenin tüm belgeler içinde ne kadar nadir olduğunu ölçmektedir. Yaygın kelimeler düşük IDF skoruna sahip olur, çünkü ayırt edici değillerdir.

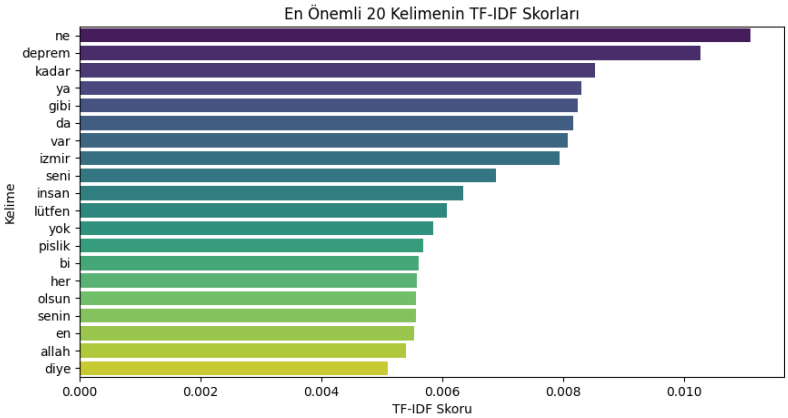
TF-IDF Skoru: TF VE IDF metriklerinin çarpımıyla hesaplanmaktadır. Skoru yüksek olan kelimeler, hem belirli bir belgede sık geçer hem de diğer belgelerde nadir görülür, dolayısıyla o belgenin içeriği hakkında daha fazla bilgi taşımaktadırlar. [15]



Şekil 1.11. Telegram platformu için en önemli 20 kelimenin bir listesi



Şekil 1.12. Reddit platformu için en önemli 20 kelimenin bir listesi



Şekil 1.13. Twitter (X) platformu için en önemli 20 kelimenin bir listesi

1.5. VERİ ETİKETLEME SÜRECİ

Veri etiketleme metin verilerinin analiz için anlamlı kategorilere ayrılmasını sağlamaktadır. Veri etiketleme, doğal dil işleme (NLP) alanında model geliştirmede önemli bir rol oynamaktadır. NLP modellerinin eğitilmesi ve gerçek dünyada kullanılabilir hale getirilmesi için etiketi doğru verilmiş veriler gerekmektedir. NLP modelleri gözetimli öğrenme yaklaşımında metni anlamlandırmak için etiketli verilere ihtiyaç duymaktadırlar. Sosyal medya tehdit analizi projesinde veriler "Olumsuz" ve "Tarafsız" olmak üzere 2 kategori olarak etiketlenmektedirler. Çıkış noktamız tehdit içeriğini tespit etmek olduğu için verileri normalin aksine positive, negative ve neutral şelinde ayırmak yerine olumsuz ögeyi daha öne çıkararak sadece tarafsız ve olumsuz kavramları altında değerlendirdik. Bu etiketleme işlemi, sınıflandırma modellerinin hangi kelime, ifade veya dil kalıplarının belirli kategorilere ait olduğunu öğrenmesine olanak tanımaktadır.

1.6. Veri Etiketeleme Aşamaları

1.6.1. Olumsuz Kelime Listesi

Şekil 1.14. İlgili kod parçası (7)

Potansiyel tehdit veya risk içeren metinleri belirlemek için bir olumsuz kelime listesi oluşturulmuştur. Bu listedeki veriler , veri kümesindeki metinlerde aranarak içerik sınıflandırması yapılmaktadır.

1.6.2. Etiketleme Fonksiyonu

```python
# Etiketleme fonksiyonu
def etiketle(metin):
    metin = metin.lower()  # Metni küçük harfe çevir
    olumsuz_var = any(kelime in metin for kelime in olumsuz_kelimeler)

    if olumsuz_var:
        return "Olumsuz"

    else:
        return "Tarafsız"
```

Şekil 1.15. İlgili kod parçası (8)

- **Metni Küçük Harfe Çevirme:** Büyük / küçük harf duyarlılığını ortadan kaldırarak doğru eşleşmeler sağlanmaktadır.
- **Kelime Kontrolü**: Metin, olumsuz kelimeler listesinde bulunan herhangi bir kelimeyi içeriyorsa "Olumsuz", aksi durumda "Tarafsız" olarak sınıflandırılmaktadır.

1.6.3. Verilerin Etiketlenmesi

```python
# Veriyi etiketler
df['etiket'] = df['Gönderi İçeriği'].apply(etiketle)
```

Şekil 1.16. İlgili kod parçası (9)

Veri kümesinin bulunduğu CSV dosyasına bir "etiket" sütunu eklenmektedir ve her gönderi için sınıf bilgisi kaydedilmektedir.

1.6.4. Sonuçların Kaydedilmesi

```python
# Etiketlenmiş veriyi csv'ye kaydet
df.to_csv('data.csv', index=False)
```

Şekil 1.17. İlgili kod parçası (10)

Etiketlenmiş veriler, yeni bir CSV dosyasına kaydedilmektedir.

Bu sürecin uygulanması, Sosyal Medya Tehdit Analizi Projesi için önemli bir adım olup tehdit içeren içeriklerin hızlı bir şekilde tespit edilmesini sağlamaktadır.

1.7. Veri toplama platformları özelinde etiketleme süreçleri

Telegram, Reddit ve X platformlarından elde edilen verilerin etiketlenmesi, bu projenin doğruluğu ve etkinliği için kritik öneme sahiptir. Ancak, her platformun kendine özgü veri yapıları, içerik özellikleri ve zorlukları bulunmaktadır. Her platform için veri etiketleme sürecinin genel hatları ve karşılaşılan zorluklar şu şekildedir:

1.7.1. X (Twitter) Verileri Etiketleme Süreci

- **Veri Yapısı:** X (Twitter)'te metin uzunluğu sınırlıdır (280 karakter), bu da verilerin kısa, genellikle özlü veya argoya dayalı olmasını sağlamaktadır.
- **Etiketleme Adımları:**
  1. **Veri Toplama:** Gerekli anahtar kelimeler veya hashtag'ler ile API aracılığıyla veri çekilmektedir. Bu platform için 10000 adet veri çekilmiştir.
  2. **Dil İşleme:** Dil algılama (Türkçe metinleri filtrelemek), özel karakterlerin temizlenmesi ve postların yapılandırılması gerçekleştirmektedir.
  3. **Kelime Bazlı Etiketleme:** Olumsuz içerikler (tehdit, saldırı, nefret söylemi gibi) belirlemek için anahtar kelime eşlemesi kullanılmaktadır.
  4. **Sınıflandırma:** Postlar "Olumsuz" ve "Tarafsız" olmak üzere 2 sınıfa ayrılmaktadır.
- **Zorluklar:**
  - **Kısaltmalar ve Emojiler:** Emojiler ve kısaltmalar bağlamı değiştirebilir, bu nedenle ekstra özen gerekmektedir.
  - **Dil Çeşitliliği ve Çok Anlamlılık:** Aynı kelime farklı bağlamlarda farklı anlamlar taşıyabilmektedir.

1.7.2. Telegram Verileri Etiketleme Süreci

- **Veri Yapısı:** Telegram grupları ve kanallarında uzun metinler, medya dosyaları ve bağlantılar gibi çeşitli içerikler bulunmaktadır. Genellikle, kullanıcıların anonimliği nedeniyle dil kullanımı daha az resmi olabilmektedir.
- **Etiketleme Adımları:**
  1. **Veri Toplama:** Telegram API aracılığıyla grup sohbetleri ve kanallarından veri çekilmektedir. Bu platform için 10000 adet veri toplanmıştır.
  2. **Metin Ön İşleme:** Medya bağlantıları, emojiler, HTML etiketleri ve gereksiz bilgiler çıkarılmaktadır. Mesajlar temizlenmektedir.
  3. **Anahtar Kelime Eşlemesi:** Önceden belirlenmiş kelime listeleri ile mesajlardaki tehdit içerikli ifadeler tespit edilmektedir.
  4. **Sınıflandırma:** Her mesaj, içerdiği potansiyel tehdit unsurlarına göre "olumsuz" veya "Tarafsız" olarak sınıflandırılmaktadır.
- **Zorluklar:**
  - **Anonimlik ve Özgür Dil Kullanımı:** İnsanların daha rahat veya aşırı ifadeler kullanabilmesi veri temizleme ve analiz aşamasını zorlaştırabilmektedir.
  - **Bağlamsal Tehditler:** Kelime bazlı eşleştirme yetersiz kalabilmektedir, bağlam analizi gerekli olabilmektedir.

1.7.3. Reddit Verileri Etiketleme Süreci

- **Veri Yapısı:** Reddit'teki içerikler genellikle forum tarzında olup, başlıklar ve uzun içerik metinleri ile zengin bir yapı sunmaktadır. Ayrıca yorumlar, konunun anlaşılmasına büyük katkı sağlamaktadır.

- **Etiketleme Adımları:**

  1. **Veri Toplama:** Subreddit'lerden veri çekmek için Reddit API ve PRAW (Python Reddit API Wrapper) kullanılmaktadır. Reddit için 10000 adet veri toplanmıştır.
  2. **Metin Temizliği:** URL'ler, medya bağlantıları ve özel karakterler çıkarılmaktadır.
  3. **Sözlük Tabanlı Etiketleme:** Anahtar kelime eşleştirmesi veya tehdit modelleriyle olumsuz içerik tespiti yapılmaktadır.
  4. **Başlık ve İçerik Analizi:** Başlıklar ve gönderi içeriği birlikte analiz edilerek içerik sınıflandırılmaktadır.
  5. **Sınıflandırma ve İstatistik:** Her gönderi, tehdit veya tarafsız olarak sınıflandırılmakta ve istatistiksel veriler çıkarılmaktadır.

- **Zorluklar:**

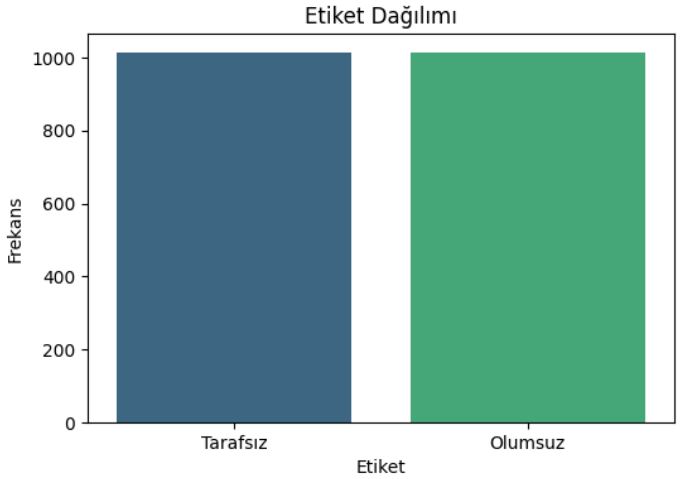
  - **Uzun Metinler:** Reddit'te uzun içerikler daha fazla bağlamsal analiz gerektirmektedir.

Platformların Ortak Zorlukları

- **Spam ve Gereksiz Veriler:** Üç platformda da spam veya alakasız içerik bulunma ihtimali yüksektir. Bu tür verilerin temizlenmesi gerekmektedir.
- **Gerçek Zamanlı İşleme:** Özellikle X (Twitter) ve Telegram gibi dinamik platformlarda veri hızla değişmektedir. Gerçek zamanlı tehdit analizi yapacak bir sistem oluşturmak zordur.
- **Çok Anlamlılık :** Kelimeler veya ifadeler bağlama bağlı olarak farklı anlamlar taşıyabilmektedir, bu da yanlış sınıflandırmaya yol açabilmektedir.

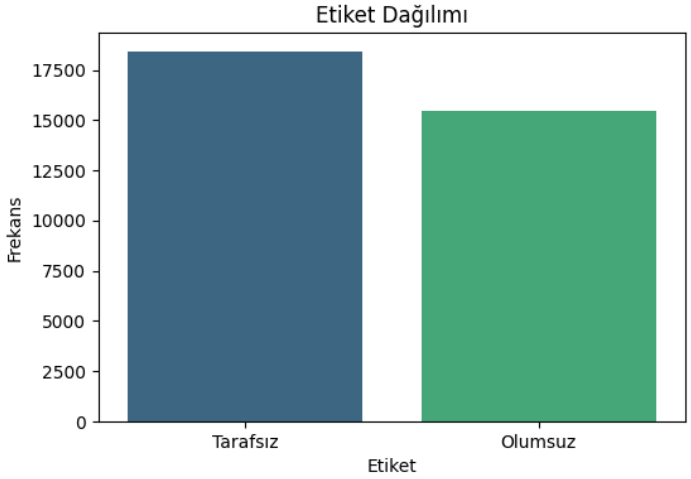
Sonuç olarak, her platformun kendine özgü veri yapısına uygun teknikler geliştirilmesi ve bağlam analizi yapılması gerekmektedir. Doğal dil işleme modelleri oluştururken etiketleme sürecinin doğru yapılması, bu platformlardan gelen verilerin daha verimli bir şekilde kullanılmasını sağlamaktadır.

1.8. Etiket Dağılımları



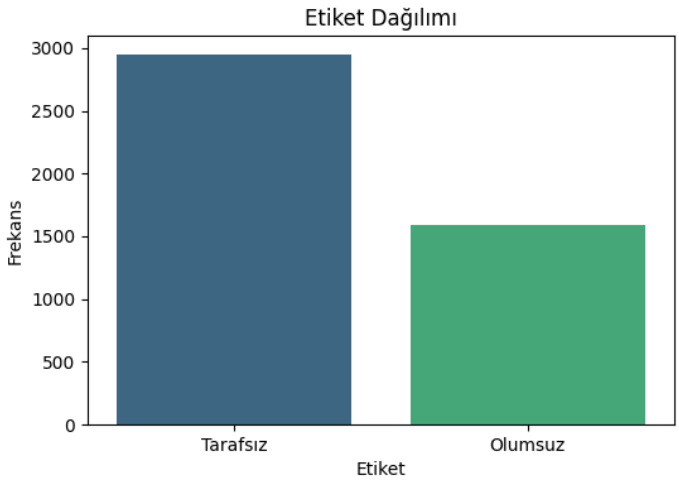
Şekil 1.18. Telegram platformu için etiket dağılımı grafiği

Telegram modeli eğitmek için hazırlanan veri setinde olumsuz ve tarafsız olarak etiketlenen veri miktarları eşit ve dengelidir. Modelin performansını artırmak için veri seti dengeli hale getirilmiştir.



Şekil 1.19. Twitter (X) platformu için etiket dağılımı grafiği

Twitter (X) veri setinde tarafsız ve olumsuz etiketleri dengelidir ve modelin performansını olumlu yönde etkilemektedir.



Şekil 1.20. Reddit platformu için etiket dağılımı grafiği

Reddit veri setinde tarafsız etiketli veri sayısı olumsuz etiketli verilerden daha fazladır. Model eğitimi için sınıfları dengelemek gerekmektedir. Dengesiz sınıflar modelin performansının fazla sayıda yer alan sınıf için yüksek olmasını sağlamaktadır.

2. BÖLÜM

MODEL EĞİTİMİ VE GERÇEK ZAMANLI VERİ ANALİZİ

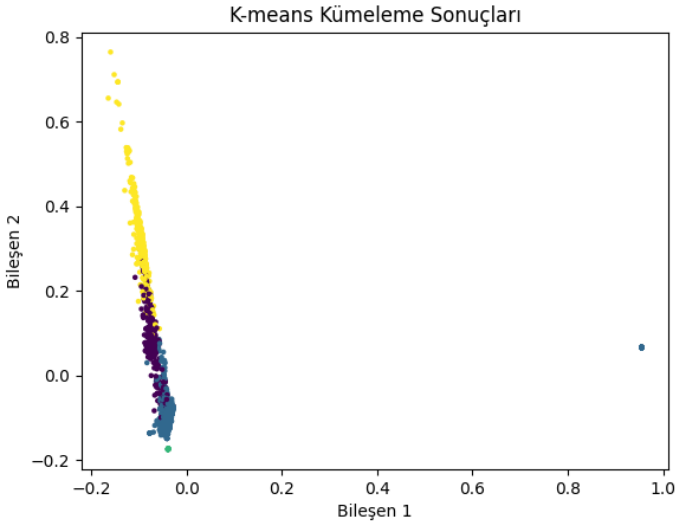
2.1. K-means algoritması

K-Means algoritması, gözetimsiz bir makine öğrenimi algoritması olup, veri kümelerini önceden tanımlanan K sayısı kadar - ki bu sayı optimum bir değer alana kadar değiştirebilir - gruba ayırmak için kullanılır. Bu algoritma, her bir veri noktasını en yakın küme merkezine atayarak kümelerin homojenliğini artırmayı hedefler. [16]

K-Means algoritmasının başarıyla uygulanması, veri setinin anlamlı ve homojen kümelere ayrılmasını sağlar. Bu sayede, tehdit içeren mesajlar ile içermeyen mesajlar arasındaki farklar daha belirgin hale gelir ve modelin öğrenme süreci daha verimli olur. Kümeleme sonrası elde edilen gruplar, tehdit algılama sisteminin daha doğru tahminler yapmasını sağlar. Ayrıca, her bir kümenin temsil ettiği anahtar özellikler, modelin hangi faktörlerin tehdit içeren mesajları belirlemede daha etkili olduğunu anlamamıza yardımcı olur. Bu, tehdit algılama sisteminin gelecekteki gelişimleri için önemli bir adım teşkil etmektedir.

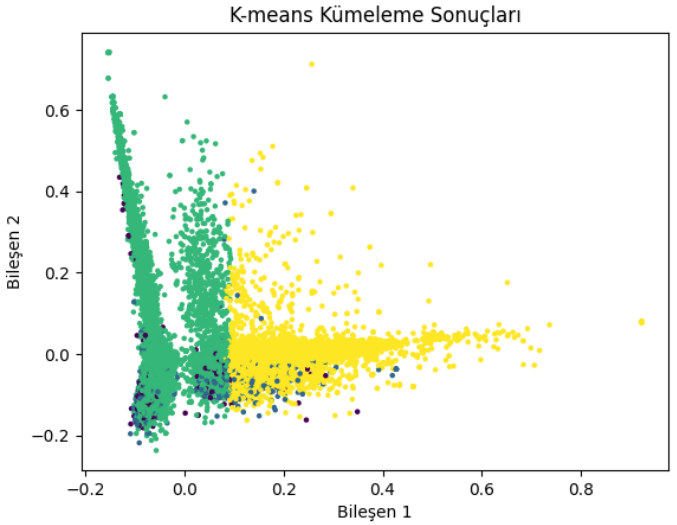
Bu proje için , veri etiketleme sürecinin ardından K-Means algoritması kullanılarak verilerin kümelenmesi gerçekleştirilmiştir. K-Means algoritması, etiketleme işlemi tamamlanmış verileri gruplandırarak, tehdit içeren ve içermeyen mesajları belirli kümelere ayırmıştır. Bu süreç, model eğitimi için daha düzenli bir veri seti oluşturmayı ve modelin performansını artırmayı amaçlamaktadır.

2.1.1. Kümeleme Grafikleri



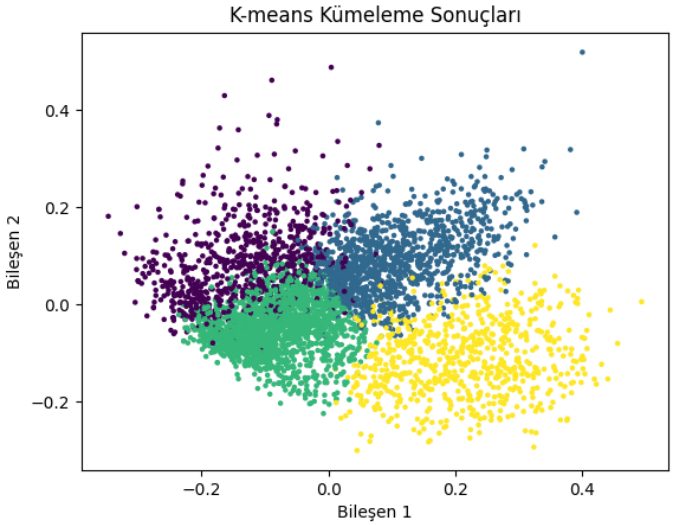
Şekil 2.1. Telegram platformu için kümeleme sonuçları grafiği

Telegram veri seti için K değeri 4 olarak seçilmiştir. Grafikte Küme ayrışmasının düşük olduğu gözlemlenmiştir. Kümeler sıkışık ve ayrışma net değildir. Küme yoğunluğu yüksek olmasına rağmen kümelerin birbirine çok yakın olduğu görülmektedir. Bu yapı telegram veri setinin içeriğinin tek tip mesajlar içermesinden kaynaklı olduğu düşünülmektedir.



Şekil 2.2. Twitter (X) platformu için kümeleme sonuçları grafiği

Twitter veri seti için kümele grafiğinde K sayısı 4 olarak belirlenmiştir. Kümeler arasındaki ayrışma telegram veri setine göre daha belirgin ve kümeler geniş bir alana yayılmıştır. Ancak bazı geçiş bölgelerindeki veri noktaları belirsizlik taşımaktadır. Küme yoğunluğu daha az ve kümeler geniş bir alana yayılmıştır. Ancak bu yayılımın bir kısmı fazla geniş olduğu için bazı noktaların aykırı olabileceği gözlemlenmektedir.

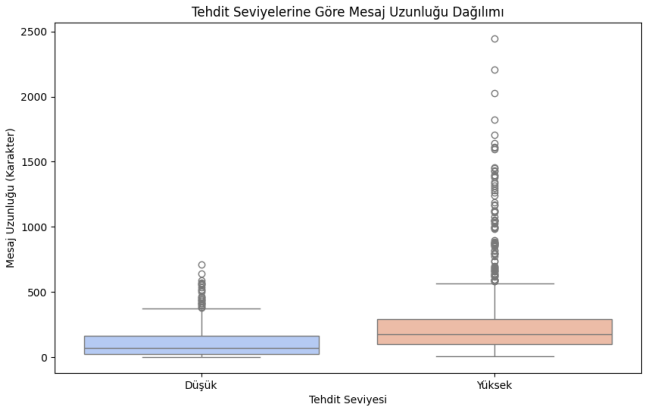


Şekil 2.3. Reddit platformu için kümeleme sonuçları grafiği

Reddit veri setinde en uygun K değeri 4 olarak seçilmiştir. Kümeler daha dengeli bir şekilde dağıtılmış ve aralarındaki örtüşmelerin diğer grafiklere göre daha az olduğu görülmektedir. Özellikle dört küme de net bir şekilde tanımlanmıştır. Dengeli bir yoğunluk var ve aykırı değer sayısı çok azdır. Bu veri setinde böyle bir yapının olma nedeni reddit verilerinin konu başlıklarıyla çekilmiş olmasıdır.

2.1.2. Tehdit Seviyelerine Göre Mesaj Uzunluğu Dağılımı

Tehdit Seviyelerine Göre Mesaj Uzunluğu Dağılım analizi, tehdit içeren ve içermeyen mesajların uzunluk açısından farklılık gösterip göstermediğini belirlememizi sağlamaktadır. Grafiklerde, her tehdit seviyesi için mesaj uzunluğunun ortalama değerleri ve dağılımı görsel olarak sunulmaktadır. Bu dağılım, tehdit içeren mesajların genellikle daha uzun olup olmadığı veya kısa mesajların daha belirgin tehdit içerip içermediği konusunda bilgi vermektedir.

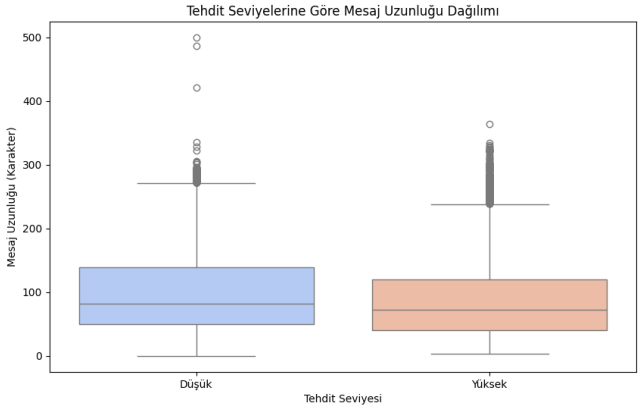


Şekil 2.4. Telegram platformu için mesaj uzunluğu dağılımı grafiği

Düşük tehdit seviyesindeki mesajlar, genellikle 200-400 karakter arasında yoğunlaşmakta olup, medyan uzunlukları ortalama 300 karakterdir. Ancak bu seviyedeki mesajlarda minimum ve maksimum uzunlukların kutu sınırlarının dışında yer aldığı gözlemlenmiştir.

Yüksek tehdit seviyesindeki mesajlar, 800-1600 karakter arasında dağılım göstermekte ve medyan uzunlukları ortalama 1200 karakter olarak hesaplanmıştır. Benzer şekilde, bu seviyedeki mesajlarda da minimum ve maksimum uzunlukların kutu sınırlarının dışında olduğu görülmüştür.

Grafikteki daireler, her iki tehdit seviyesinde de aşırı uç değerleri temsil etmektedir.



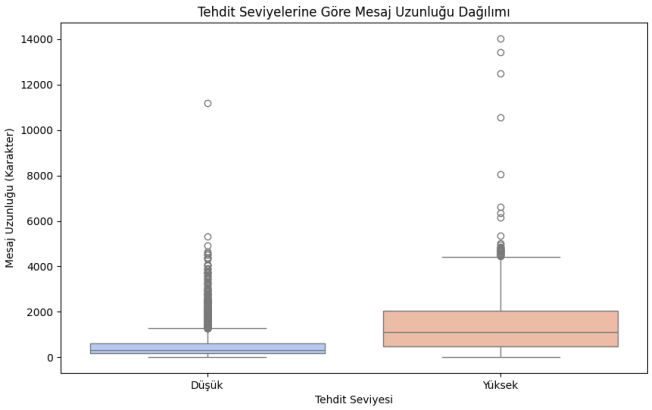
Şekil 2.5. Twitter (X) platformu için mesaj uzunluğu dağılımı grafiği

Düşük tehdit seviyesindeki mesajlar, genellikle 100 karakterin altında

yoğunlaşmakta olup, medyan uzunlukları ortalama 75 karakterdir. Ancak bu seviyedeki mesajlarda minimum ve maksimum uzunlukların kutu sınırlarının dışında yer aldığı gözlemlenmiştir.

Yüksek tehdit seviyesindeki mesajlar, 200-400 karakter arasında dağılım göstermekte ve medyan uzunlukları ortalama 300 karakter olarak hesaplanmıştır. Benzer şekilde, bu seviyedeki mesajlarda da minimum ve maksimum uzunlukların kutu sınırlarının dışında olduğu görülmüştür.

Grafik üstündeki daireler, her iki tehdit seviyesinde de aşırı uç değerleri temsil etmektedir.



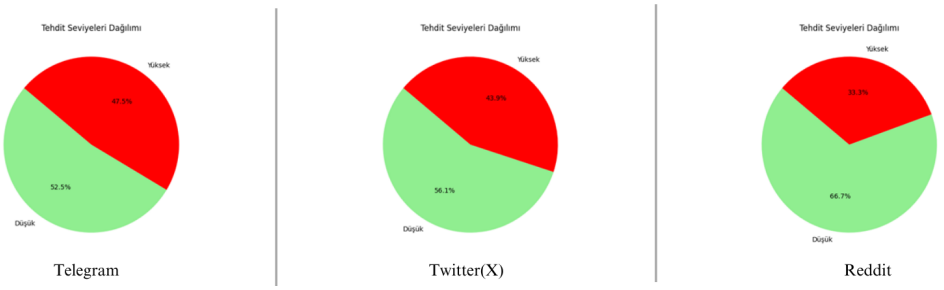
Şekil 2.6. Reddit platformu için mesaj uzunluğu dağılımı grafiği

Düşük tehdit seviyesindeki mesajlar, genellikle 200-600 karakter arasında yoğunlaşmakta olup, medyan uzunlukları ortalama 400 karakterdir. Ancak bu seviyedeki mesajlarda minimum ve maksimum uzunlukların kutu sınırlarının dışında yer aldığı gözlemlenmiştir.

Yüksek tehdit seviyesindeki mesajlar, 2000-8000 karakter arasında dağılım göstermekte ve medyan uzunlukları ortalama 4000 karakter olarak hesaplanmıştır. Benzer şekilde, bu seviyedeki mesajlarda da minimum ve maksimum uzunlukların kutu sınırlarının dışında olduğu görülmüştür.

Grafik üstündeki daireler, her iki tehdit seviyesinde de aşırı uç değerleri temsil etmektedir.

2.1.3. Tehdit seviyeleri dağılımı



Şekil 2.7. Bütün platformlar için tehdit seviyeleri dağılımı grafiği

Yukarıdaki görselde her platform için tehdit seviyelerinin oranları pasta grafik olarak gösterilmiştir. Veri setlerindeki tehdit oranları her platform için farklı olduğu gözlemlenmektedir.

2.2. Model Seçimi

Model seçim sürecinde SVM, Random Forest ve Derin Öğrenme gibi farklı model yaklaşımları üzerinde çalışılmıştır. Bu modellerin doğruluk, Precision, Recall ve F1 Skoru gibi temel performans metrikleri karşılaştırılmıştır. Yapılan değerlendirmeler sonucunda, SVM modeli sosyal medya tehdit analizinde en iyi sonucu sağladığından dolayı nihai model olarak seçilmiştir.

2.2.1. SVM Modelinin Özellikleri

- **Gözetimli Öğrenme Algoritması Olması:** SVM, sınıflandırma ve regresyon problemlerinde kullanılan bir gözetimli öğrenme algoritmasıdır. Sosyal medya tehdit analizindeki sınıflandırma görevine uygun bir yapıdadır.
- **Optimum Sınır Belirleyebilmesi :** SVM, sınıflar arasındaki en geniş marjı (margin) sağlayacak bir karar sınırı (hyperplane) belirler. Bu, modelin genel performansını artırmaktadır.
- **Karmaşık Verilerle Etkili Olması :** SVM, doğrusal ayrıştırılamayan veriler için çekirdek (kernel) fonksiyonlarını kullanarak veriyi daha yüksek boyutlu bir uzaya taşır. Bu, sosyal medya verilerindeki karmaşıklığı ve ilişkileri daha iyi modellemesini sağlar.

- **Aşırı Öğrenmeye Dirençli Olması :** SVM, yüksek boyutlu verilerle çalışırken aşırı öğrenme (overfitting) riskini minimize eder.
- **Dengesiz Verilerde Performansı :** Tehdit seviyelerinin dengesiz dağıldığı veri setlerinde bile SVM iyi performans gösterebilir. [17]

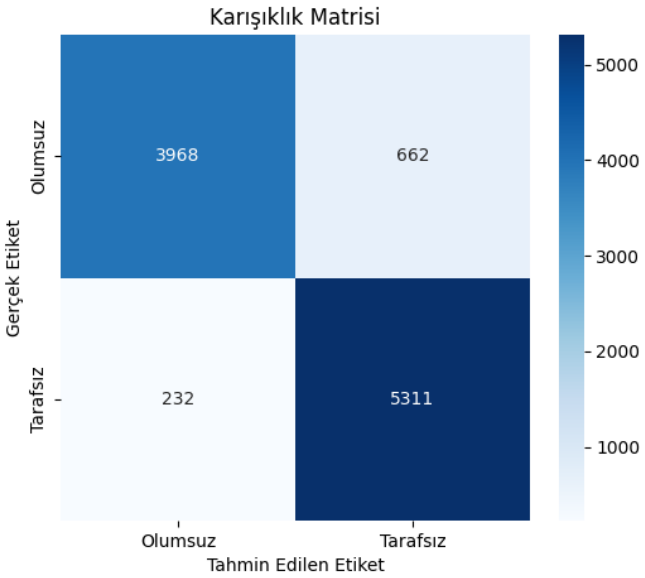
2.2.2. Diğer Modellerle Karşılaştırma

Random Forest modeli veri açıklanabilirliği açısından avantajlıdır; ancak, sosyal medya verisinin yüksek boyutluluğu karşısında SVM kadar etkili sonuç verememiştir. Derin öğrenme modelleri ise yüksek doğruluk oranları sunabilse de eğitim sürecinin karmaşıklığı ve donanım gereksinimi nedeniyle proje kapsamında tercih edilmemiştir.

2.3. Model Performansı ve Değerlendirme

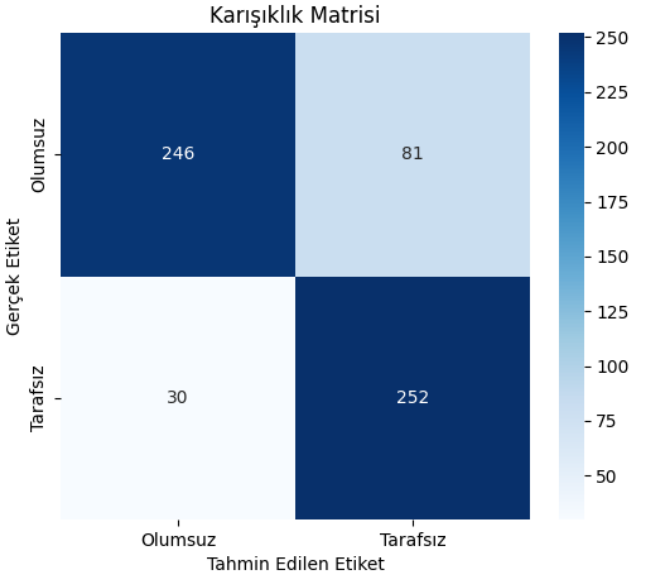
2.3.1. Karmaşıklık Matrisi (Confusion Matrix)

Confusion matrix, modelin doğru ve yanlış sınıflandırmalarını gösteren bir tablodur. Bu projede confusion matrix, her bir platform için yapılan model eğitimleri sırasında twitter(X), telegram ve reddit için ayrı ayrı elde edilmiştir. Confusion Matrix tabloları ve bunlarla ilgili açıklamalara aşağıda yer verilmiştir.



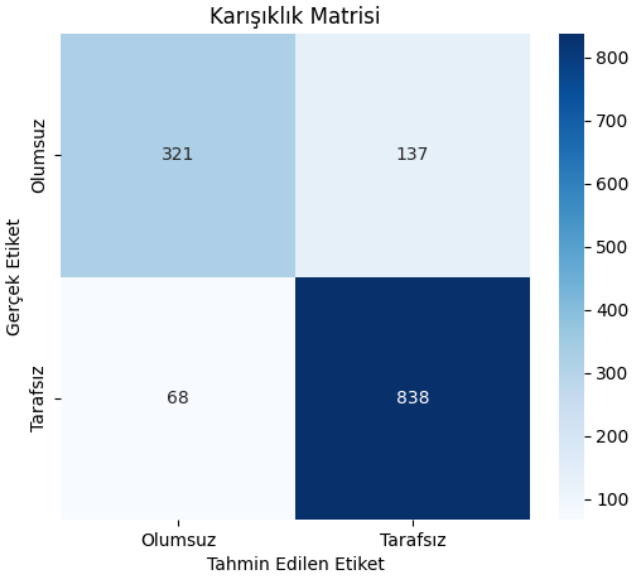
Şekil 2.8. Twitter (X) platformu için karmaşıklık matrisi grafiği

Matristeki doğru sınıflandırma (3968 + 5311) toplamda 9279'dir. Yanlış sınıflandırma sayısı ise (662 + 232) toplamda 894 olarak hesaplanmıştır. Modelin genel doğruluk oranının (accuracy) (9279/10173) %91,21 olduğu sonucuna varılmıştır. Modelin genel doğruluğunun bu verilere dayanarak oldukça yüksek olduğu görülmektedir.



Şekil 2.9. Telegram platformu için karmaşıklık matrisi grafiği

Matristeki doğru sınıflandırma (246 + 252) toplamda 498. Yanlış sınıflandırma sayısı ise (81 + 30) toplamda 111 olarak hesaplanmıştır. Modelin genel doğruluk oranının (accuracy) (498/609) %81,7 olduğu sonucuna varılmıştır. Bu, modelin sınıflandırma işlemini oldukça iyi bir şekilde gerçekleştirdiğini göstermektedir.

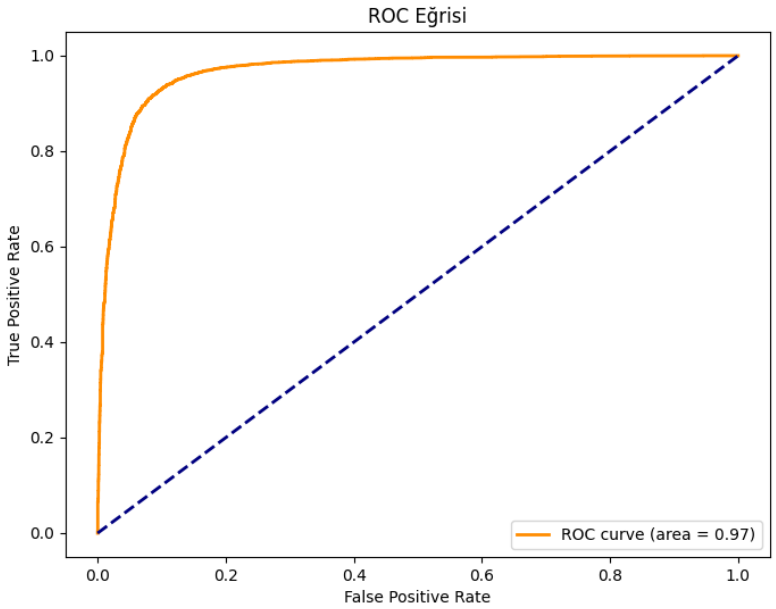


Şekil 2.10. Reddit platformu için karmaşıklık matrisi grafiği

Matristeki doğru sınıflandırma (321 + 838) toplamda 1159. Yanlış sınıflandırma sayısı ise (137 + 68) toplamda 205 olarak hesaplanmıştır. Modelin genel doğruluk oranının (accuracy) (1159/1364) %84,97 olduğu sonucuna varılmıştır. Model, sınıflandırma görevini yüksek bir doğrulukla yerine getirmektedir.

2.3.2. ROC Eğrisi

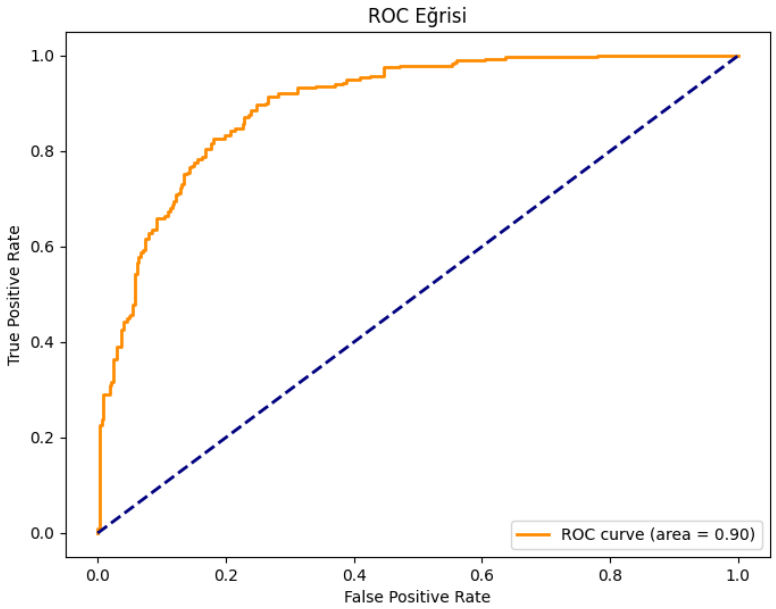
ROC eğrisi, modelin sınıflandırma eşik değerlerine göre doğruluk ve hata oranlarını grafiksel olarak gösterir. Eğrinin altında kalan alan (AUC), modelin genel performansını ölçen bir metriktir. Aşağıda her platformun ROC eğrisi grafikleri gösterilmiştir.



Şekil 2.11. Twitter (X) platformu için ROC eğrisi grafiği

Bu grafikte ROC eğrisi ve AUC değerinin %97 olması, modelin mükemmele yakın bir sınıflandırma performansı sergilediğini göstermektedir.

Model, pozitif sınıfları doğru bir şekilde tespit edebilme (yüksek duyarlılık) ve negatif sınıfları yanlış pozitiflere dönüştürmeme (yüksek özgüllük) konusunda son derece etkilidir.

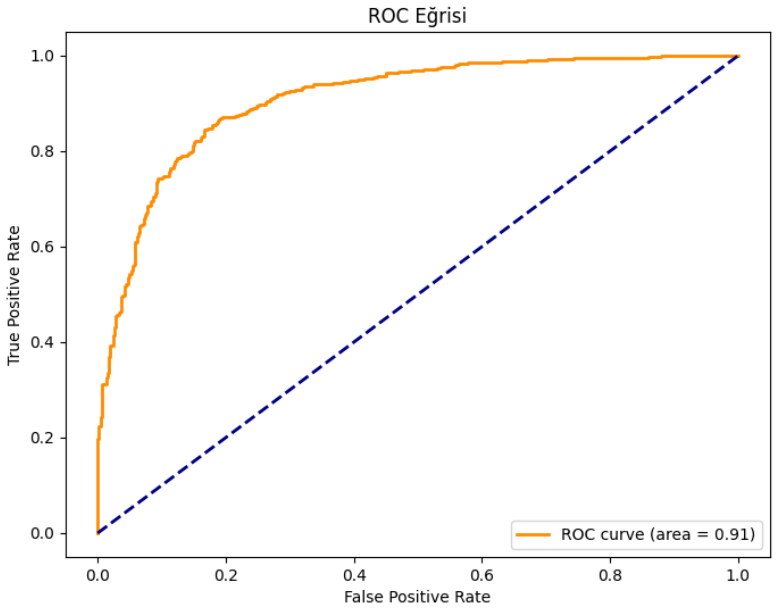


Şekil 2.12. Telegram platformu için ROC eğrisi grafiği

Yukarıdaki ROC eğrisi, modelin tehdit sınıflandırma görevinde oldukça iyi bir performans sergilediğini göstermektedir.

Model, hem pozitif sınıfları doğru şekilde tespit edebilmekte hem de negatif sınıfları yanlış pozitiflere dönüştürmemektedir.

AUC değerinin %90 olduğu grafikte görülmektedir. Bu, modelin dengeli bir sınıflandırma yaptığını ortaya koymaktadır.



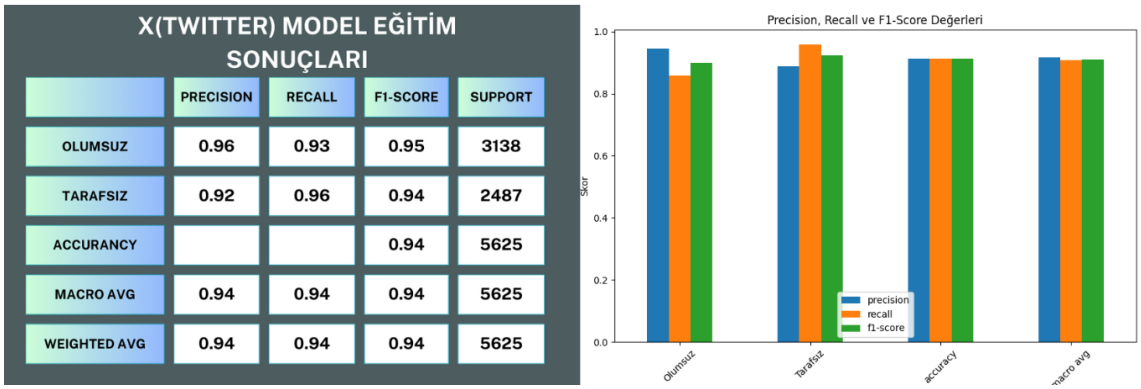
Şekil 2.13. Reddit platformu için ROC eğrisi grafiği

Reddit modelinin eğitimi sırasında ortaya çıkan bu ROC eğrisi, modelin tehdit sınıflandırma görevinde yüksek bir performans sergilediğini göstermektedir.

Model, hem pozitif sınıfları doğru şekilde tespit edebilmekte hem de negatif sınıfları yanlış pozitiflere dönüştürmemektedir. AUC değeri %91 olduğu grafiğin sağ alt köşesinde gösterilmiştir. Bu, modelin dengeli bir sınıflandırma yaptığını göstermektedir.

2.3.3. Model Eğitim Sonuçlarının Değerlendirilmesi

Aşağıdaki tablo ve grafiklerde, eğitilen modellerin Precision, Recall ve F1-Score değerleri farklı sınıflar (Olumsuz, Tarafsız) için ayrı ayrı değerlendirilmiştir.



X(TWITTER) MODEL EĞİTİM SONUÇLARI

	PRECISION	RECALL	F1-SCORE	SUPPORT
OLUMSUZ	0.96	0.93	0.95	3138
TARAFSIZ	0.92	0.96	0.94	2487
ACCURANCY			0.94	5625
MACRO AVG	0.94	0.94	0.94	5625
WEIGHTED AVG	0.94	0.94	0.94	5625

Şekil 2.14. Twitter (X) platformu için model eğitim sonuçları

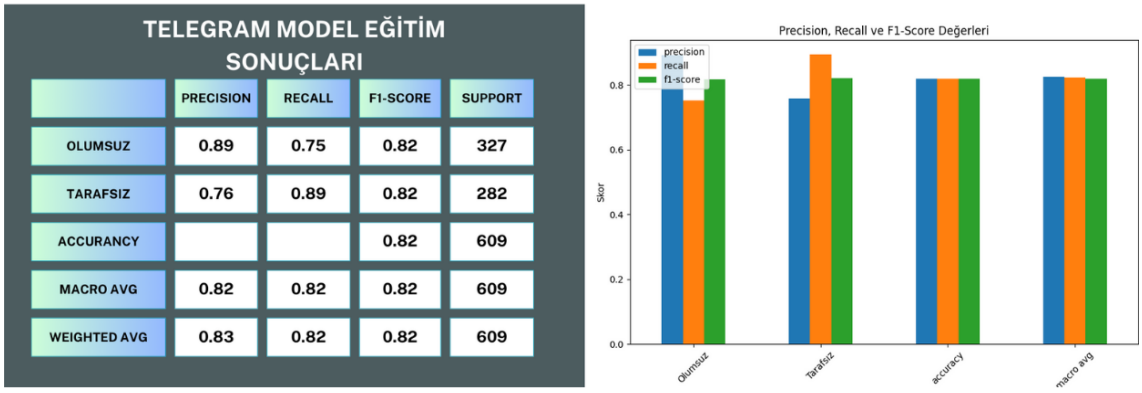
Precision (Kesinlik): Olumsuz sınıf için %96, tarafsız sınıf için %92 olarak ölçülmüştür. Bu, modelin yanlış pozitifleri (yanlış tehdit algısı) minimize etme konusunda oldukça başarılı olduğunu göstermektedir.

Recall (Duyarlılık): Tarafsız sınıf için %96 ile en yüksek değer ölçülmüştür. Bu da modelin gerçek pozitifleri tespit etme konusunda güçlü bir performans sergilediğini göstermektedir.

F1-Score: Precision ve Recall değerlerinin dengeli bir şekilde birleştirilmesiyle elde edilen F1-Score, tüm sınıflar için %94 ile %95 arasında değişmektedir. Bu durum, modelin genel performansının oldukça dengeli olduğunu göstermektedir.

Support: Her bir sınıf için veri setindeki örnek sayısını ifade eder. Model, 3138 olumsuz, 2487 tarafsız örnek üzerinde eğitilmiştir.

Genel Performans: Grafikten de görülebileceği gibi, Precision, Recall ve F1-Score değerleri arasında küçük farklılıklar bulunmakla birlikte, modelin performansı her bir sınıfta tutarlı ve yüksektir. Macro Average ve Weighted Average değerleri de modelin genel başarısını desteklemektedir.



Şekil 2.15. Telegram platformu için model eğitim sonuçları

Precision (Kesinlik): Olumsuz sınıf için %89, tarafsız sınıf için %76 olarak ölçülmüştür. Bu, özellikle olumsuz sınıfta yanlış pozitiflerin düşük olduğunu ve modelin bu sınıfta daha güvenilir sonuçlar üretebildiğini göstermektedir.

Recall (Duyarlılık): Tarafsız sınıf için %89 ile daha yüksek bir değer elde edilmiştir. Bu durum, modelin tarafsız sınıfa ait örnekleri daha doğru bir şekilde tanıyabildiğini ortaya koymaktadır. Ancak olumsuz sınıfta %75 olan Recall değeri, modelin bir miktar gerçek olumsuz mesajları kaçırabileceğine işaret etmektedir.

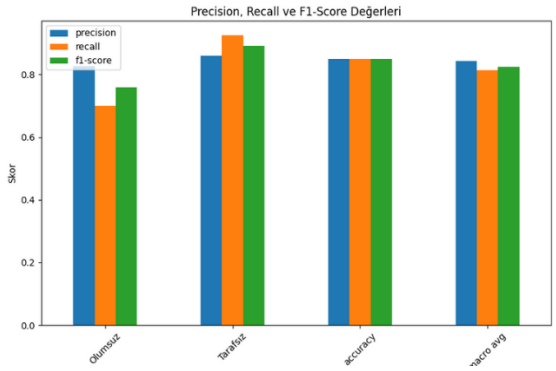
F1-Score: Her iki sınıf için de %82 olarak ölçülen F1-Score, modelin genel performansının dengeli olduğunu göstermektedir. Precision ve Recall değerleri arasındaki denge, sınıflar arasında farklılık göstermektedir.

Support: Veri setindeki örnek sayıları, modelin performansını değerlendirmede önemli bir faktördür. Model, 327 olumsuz ve 282 tarafsız mesaj üzerinde eğitilmiştir. Veri dengesizliği, performansa etkide bulunmuş olabilir.

Genel Performans: Precision, Recall ve F1-Score değerlerinden görüldüğü üzere model, özellikle olumsuz sınıfı sınıflandırmada daha iyi sonuçlar vermektedir. Ancak tarafsız sınıfta Recall değeri daha yüksek olsa da precision'ın düşük olması tarafsız örneklerde yanlış pozitiflerin artabileceğini göstermektedir.

REDDİT MODEL EĞİTİM SONUÇLARI

	PRECISION	RECALL	F1-SCORE	SUPPORT
OLUMSUZ	0.83	0.70	0.76	458
TARAFSIZ	0.86	0.92	0.89	906
ACCURANCY			0.85	1364
MACRO AVG	0.84	0.81	0.82	1364
WEIGHTED AVG	0.85	0.85	0.85	1364



Şekil 2.16. Reddit platformu için model eğitim sonuçları

Precision (Kesinlik): Olumsuz sınıf için %83, tarafsız sınıf için %86 olarak ölçülmüştür. Bu, tarafsız ve olumsuz sınıfların her ikisinde de yanlış pozitiflerin nispeten az olduğunu göstermektedir.

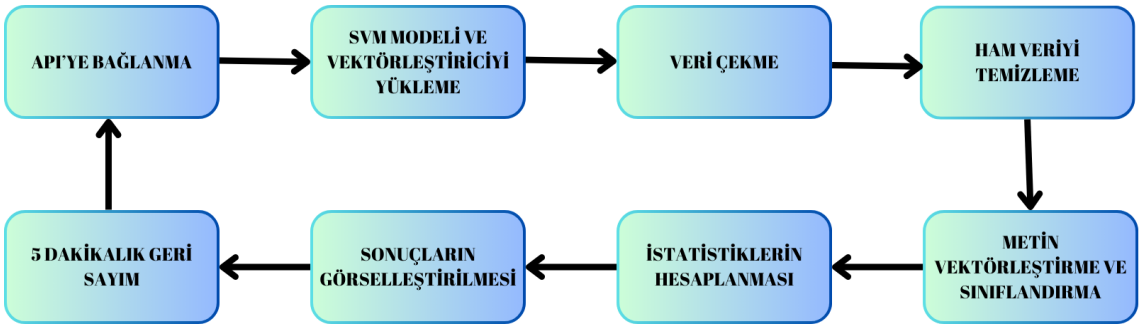
Recall (Duyarlılık): Tarafsız sınıf için %92 ile daha yüksek bir değer elde edilmiştir. Bu durum, modelin tarafsız sınıfa ait örnekleri daha doğru bir şekilde tanıyabildiğini ortaya koymaktadır. Ancak olumsuz sınıfta %70 olan Recall değeri, modelin bir miktar gerçek olumsuz mesajları kaçırabileceğine işaret etmektedir.

F1-Score: Olumsuz sınıf için de %76 tarafsız sınıf için %89 olarak ölçülen F1-Score, modelin genel performansının tarafsız sınıfta daha yüksek olduğunu göstermektedir.

Support: Veri setindeki örnek sayıları, modelin performansını değerlendirmede önemli bir faktördür. Model, 458 olumsuz ve 906 tarafsız mesaj üzerinde eğitilmiştir. Veri dengesizliği, performansı etkilemiştir.

2.4. Gerçek Zamanlı Veri Analizi

2.4.1. Sistem Mimarisi

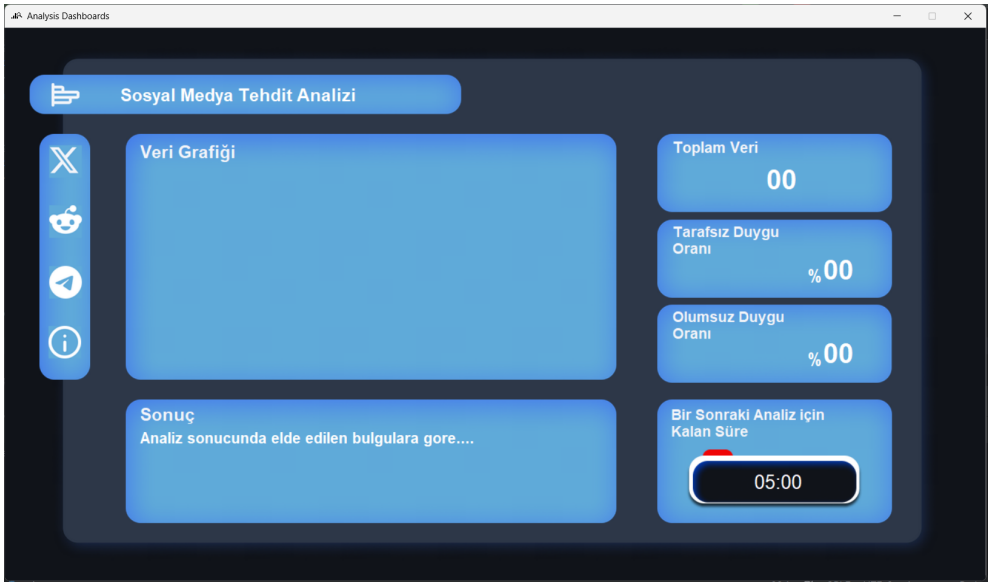


Şekil 2.17. Gerçek Zamanlı Çalışma Döngüsü

Yukarıdaki gösterilen sistem mimarisinde de görüldüğü üzere gerçek zamanlı veri analizi için API'ye bağlanma adımıyla başlayan süreç oluşturulan modelleri yükleme, verileri API'den çekme işlemi, veri temizleme ve ön işleme, çekilen verileri model sayesinde olumsuz veya tarafsız olarak sınıflandırma, sınıflandırma işleminden elde edilen istatistiklerin hesaplanması adımı ve son olarak verilerin görselleştirilmesiyle tamamlanmaktadır. Bu döngü her 5 dakikada bir tekrar ederek gerçek zamanlı veri analizi sağlanmaktadır.

Ancak bu işlem Twitter (X) platformu için API'ye bağlanma adımını içermemektedir. Twitter (X) platformu için veriler önceden hazırlanmış olan bir CSV dosyasından rastgele olarak çekilmektedir. Bunun sebebi daha önce de belirtildiği gibi Twitter (X) API'nin veri çekme limitlerini düşürmesi ve bu projede geliştirilen uygulamanın çalışması için yeteri kadar veri sağlayamamasıdır.

2.4.2. Tehdit Tespiti ve Görselleştirme



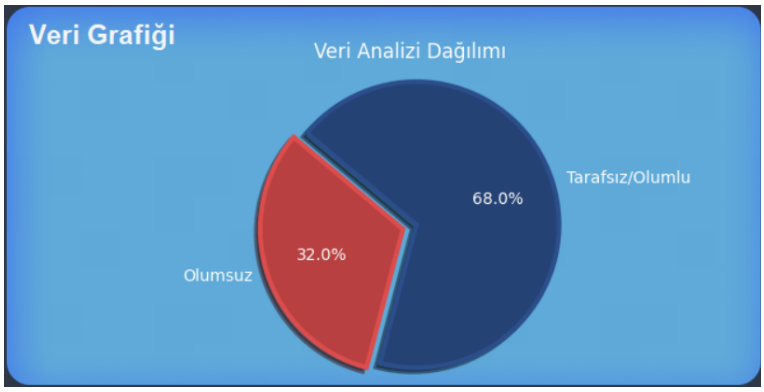
Şekil 2.18. Arayüz tasarımına genel bakış

Sosyal medya verilerinin analizi sırasında elde edilen tahminlerin yalnızca sayısal sonuçlar olarak kalması yerine, bu verilerin kullanıcı tarafından kolay anlaşılabilir olması için görselleştirilmektedir. Bu bağlamda, tehdit tespiti için modellenen SVM modeli analiz edilen metinlerin olumsuz veya tarafsız olarak sınıflandırılmasını sağlamaktadır. Bu sınıflandırma sonuçları, görselleştirme teknikleri ile desteklenerek kullanıcıya sunulmaktadır.

2.4.2.1. Grafik Gösterim Alanı

Görselleştirme işlemi için pasta grafik seçilmiştir. Pasta grafik, olumsuz ve tarafsız veri oranlarını net bir şekilde görselleştirmektedir. Analiz raporlarının net şekilde görselleştirilmesi sonuçların kolay yorumlanabilmesini de sağlamaktadır.

Aşağıdaki görselde görselleştirilmiş bir analiz sonucu gösterilmektedir:



Şekil 2.19. Arayüz tasarımına genel bakış

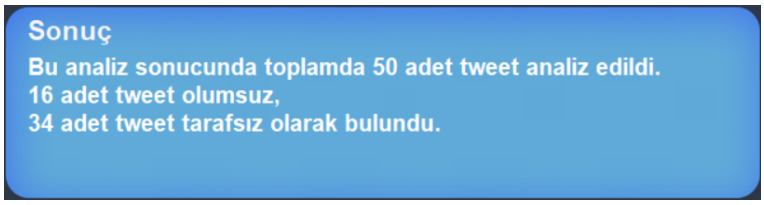
2.4.2.2. Oranların Metin Olarak Gösterilmesi İçin Ayrılan Alanlar

Verilerin görselleştirilmesi işleminde arayüzde gösterilen bir diğer görsel unsur da analiz sonucunda elde edilen olumsuz duygu oranı, tarafsız duygu oranı ve toplam çekilen veri sayısının yazdığı metin alanlarıdır. Bu alanlar kullanıcı deneyimini artırmak amacıyla kullanıcıya sunulmaktadır.

Şekil 2.20. Arayüz tasarımına genel bakış

2.4.2.3. Sonuç Alanı

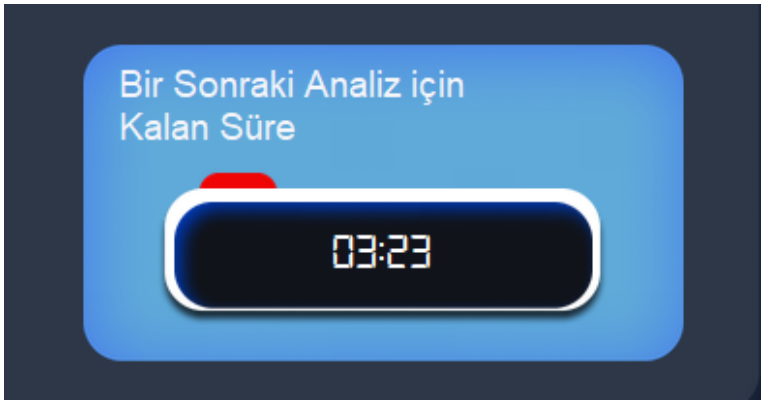
Bu alanda kullanıcıya analiz sonucunda elde edilen bulgular verilerek kullanıcının analiz sonucunu daha net anlamlandırabilmesi sağlanmaktadır. Ayrıca bu alanda olumsuz olarak sınıflandırılan verilerin içerisinde hangi tehdit kelimelerinin bulunduğu da gösterilmektedir.



Şekil 2.21. Arayüz tasarımına genel bakış

2.4.2.4. Geri Sayım Alanı

Uygulama analizleri her 5 dakikada bir tekrar etmektedir. Kullanıcının bu süreyi kolay bir şekilde takip edebilmesi için arayüz ekranında bir geri sayım alanına yer verilmiştir.



Şekil 2.22. Arayüz tasarımına genel bakış

2.4.2.5. Menü Çubuğu

Bu projede Twitter (X), Reddit ve Telegram platformları için tehdit analizi modelleri geliştirilmiştir. Bu kısımda analiz işlemine bir başla komutu verebilmek için platformlar nezlinde –ki bunlar Twitter butonu, Reddit butonu ve Telegram butonudur – 3 buton oluşturulmuştur. Butonların görevi o platform için analiz işlemini başlatıp arayüze yansıtma görevini gerçekleştirmektir. Platformların logoları ile bütünleşmiş bu butonlar kullanıcı için platformlar arasında tercih yapabilme şansı tanımıştır. Butona tıklanılmasıyla beraber platformdan veri çekme işlemi başlar, analiz işlemini gerçekleştirir ve arayüze sonucu yansıtır. Kullanıcı başka bir platform butonuna tıklayana kadar en son seçtiği platform için analiz yapılmaya devam edecektir.

Şekil 2.23. Arayüz tasarımına genel bakış (6)

3. BÖLÜM

TARTIŞMA, SONUÇ ve ÖNERİLER

3.1. Tartışma

3.1.1. Modellerin Performans Değerlendirmesi

Sosyal medya tehdit analizi projesinde kullanılan sınıflandırma modellerinin performansı, farklı metrikler üzerinden değerlendirilmiştir. Elde edilen sonuçlar, modellerin genel olarak sosyal medya platformlarındaki tehdit içeriklerini başarılı bir şekilde tespit edebildiğini göstermektedir. Ancak performans değerlerinin, kullanılan veri setlerinin yapısına, platformların dil özelliklerine ve sınıflar arasındaki veri dengesizliğine bağlı olarak değişiklik gösterdiği görülmüştür.

3.1.1.1. Performans Metrikleri

Precision (Kesinlik): Modeller, özellikle tehdit içeren mesajların yanlış pozitiflerini en aza indirmede etkili olmuş, tarafsız mesajlar üzerinde ise kısmen daha düşük bir başarı göstermiştir. Precision değerleri, olumsuz sınıflarda daha yüksek olurken, tarafsız sınıflarda modelin daha seçici davranması gerektiği gözlemlenmiştir.

Recall (Duyarlılık): Tehdit içeriklerini doğru bir şekilde tespit etme kapasitesinin, tarafsız sınıflarda daha yüksek, olumsuz sınıflarda ise daha sınırlı olduğu görülmüştür. Bu durum, özellikle tehdit mesajlarının eksik tespit edilebileceğine işaret etmektedir.

F1-Score: Precision ve Recall arasında bir denge sağlayan bu metrik, modellerin genel performansını ortaya koymuştur. Tüm sınıflar arasında dengeli bir performans sergilense de tarafsız sınıflarda daha yüksek değerler elde edilmiştir.

3.1.1.2. Veri Seti Etkisi

- Kullanılan veri setindeki sınıflar arasındaki dengesizlik, modellerin bazı sınıflar üzerinde daha iyi performans göstermesine neden olmuştur. Özellikle olumsuz sınıfta veri eksikliği, Recall değerlerinin düşmesine ve bu sınıfta tehdit içeriklerinin daha az tespit edilmesine yol açmıştır.
- Platformlara göre dil yapısı ve mesaj içeriklerinin farklılık göstermesi, modellerin başarı oranlarını doğrudan etkilemiştir. Örneğin, kısa mesajların yaygın olduğu Twitter'da model daha kısa ve net tehdit mesajlarını tespit etmede daha başarılı olurken, uzun mesajların bulunduğu Reddit'te daha karmaşık yapılar üzerinde çalışmak zorunda kalmıştır.

3.2. Sonuç

Bu çalışmanın temel amacı, sosyal medya platformlarında tehdit içeriklerinin tespitine yönelik etkili bir analiz sistemi geliştirmek olmuştur. Özellikle, kullanıcıların hızla ve geniş çapta içerik paylaşabildiği platformlarda, tehdit içeriklerinin gerçek zamanlı olarak tespit edilmesi hem bireysel hem de toplumsal güvenliği sağlamak açısından kritik bir önem taşımaktadır. Bu proje kapsamında geliştirilen sistem, farklı sosyal medya platformlarından elde edilen veri setleriyle eğitilmiş ve tehdit içeriklerini belirlemek için doğal dil işleme teknikleriyle desteklenmiş sınıflandırma modelleri kullanılmıştır. Modellerin performansı, çeşitli metriklerle ölçülmüş ve genel olarak tehdit tespiti açısından başarılı sonuçlar elde edilmiştir. Ancak sınıflar arasındaki veri dengesizlikleri ve platformlar arası dil farklılıkları gibi bazı zorlukların modellerin performansını etkilediği gözlemlenmiştir.

Bu çalışmanın sonuçları, şu anda geliştirilen tehdit tespit sisteminin sosyal medya platformlarında etkili bir şekilde uygulanabileceğini göstermektedir. Projenin şu ana kadar ulaştığı noktada:

- Gerçek zamanlı tehdit tespiti için temel bir yapı oluşturulmuş,
- Olumsuz ve tarafsız içeriklerin ayrımında makul bir doğruluk oranı yakalanmış,
- Her platform için kendi dil yapısına özgü modeller eğitilmiştir.

Sonuç olarak, bu proje sosyal medya tehdit analizi alanında önemli bir başlangıç sunmuş ve gelecekte bu tür sistemlerin daha geniş veri setleri ve daha karmaşık modellerle güçlendirilmesi durumunda çok daha etkin hale gelebileceğini ortaya koymuştur. Geliştirilen bu sistemin, gerçek zamanlı sosyal medya analizine entegre edilerek hem bireyler hem de toplumlar için daha güvenli bir dijital ortam oluşturulmasına katkı sağlayacağı öngörülmektedir.

3.3. Öneriler

3.3.1. Veri Çeşitliliği ve Büyüklüğü

Projede kullanılan veri setleri, analiz edilen platformlara özgü kelime ve cümle yapılarından oluşmaktadır. Daha büyük ve çeşitli veri setleri ile çalışılarak, modelin doğruluk oranı ve genelleme yeteneği artırılabilir. Ayrıca, platformlar arasında daha dengeli veri dağılımı sağlanmalıdır.

3.3.2. Model Optimizasyonu

Kullanılan modellerin performansını artırmak için hiperparametre optimizasyonuna daha fazla odaklanılabilir.

3.3.3. Çoklu Dil Desteği

Proje şu anda Türkçe veriler üzerinde işlem yapmaktadır. Ancak, sosyal medyada genellikle birden fazla dilde içerik paylaşımı yapılmaktadır. Modelin çok dilli içerikleri analiz edebilmesi için dil tespiti ve çeviri araçlarıyla desteklenebilir

3.3.4. Sahte İçerik Tespiti

Tehdit içeriklerinin yanında, sahte haber veya manipülatif bilgilerin tespiti de kritik öneme sahiptir. Sisteme bu tür içerikleri belirleyecek bir modül eklenebilir.

3.3.5. Kapsamlı Görselleştirme

Tehdit analiz sonuçlarının görselleştirilmesi, kullanıcıların sistemi daha iyi anlamasına yardımcı olur. Verilerin coğrafi dağılımı, zaman serisi analizi ve diğer metriklerin görselleştirilmesi için daha kapsamlı araçlar kullanılabilir.

3.3.6. Etik ve Hukuki Düzenlemeler

Sosyal medya analizleri sırasında veri gizliliği ve etik kuralların ihlal edilmemesi önemlidir. Bu kapsamda, yasal düzenlemelere uygun bir yapı benimsenmeli ve kişisel verilerin korunmasına dikkat edilmelidir.

3.3.7. Daha Gelişmiş Tehdit Sınıflandırması

Şu anki model yalnızca tehdit ve tarafsız içerikleri sınıflandırmaktadır. Tehdit içeriklerini daha detaylı kategorilere ayırmak (örneğin, fiziksel tehdit, siber saldırı tehdidi, nefret söylemi vb.) sistemi daha işlevsel hale getirebilir.

KAYNAKLAR

1. Field, A., 2013. Discovering Statistics Using IBM SPSS Statistics (4th ed.), **Sage Publications**.

2. Jurafsky, D. & Martin, J.H., 2019. Speech and Language Processing (3rd ed.), **Prentice Hall**.

3. Liu, B., 2012. Sentiment Analysis and Opinion Mining, **Morgan & Claypool Publishers**.

4. Marr, B., 2015. Big Data: Using SMART Big Data, Analytics and Metrics to Make Better Decisions and Improve Performance, **Wiley**.

5. Murphy, K.P., 2012. Machine Learning: A Probabilistic Perspective, **MIT Press**.

6. Russell, S., .N.P., 2010. Artificial Intelligence: A Modern Approach (3rd ed.), **Prentice Hall**.

7. Osman METİN & Şeref KARAKAYA, 2017. Jean Baudrillard Perspektifinden Sosyal Medya Analizi Denemesi, **Afyon Kocatepe Üniversitesi Sosyal Bilimler Dergisi / Cilt: 19, Sayı: 2**.

8. Hasan KARAMAN & Şevki IŞIKLI, 2016. Twitter'daki Dini ve Etnik Temelli Nefret Söylemlerinin Analizi, **AJIT-E: Academic Journal of Information Technology / Cilt: 7 Sayı: 25**.

9. Aksu, Muhammed Çağrı & Karaman, E., 2022. Turistik mekanlara yönelik sosyal medya paylaşımlarının yapay zekâ yöntemleriyle değerlendirilmesi: Artvin ili örneği, **Journal of Tourism and Gastronomy Studies**.

10. Mesut Toğaçar & Kamil Abdullah Eşidir & Burhan Ergen, 2022. Yapay Zekâ Tabanlı Doğal Dil İşleme Yaklaşımını Kullanarak İnternet Ortamında Yayınlanmış Sahte Haberlerin Tespiti, **Journal of Intelligent Systems: Theory and Applications**.

11. Beyza Özdemir, 2024. Siber Tehdit İstihbaratında Yapay Zeka ve Makine Öğrenmesi, **The Journal of Defence and Security Research**.

12. Ayşe Simin Kara & Satı Kaya, 2023. YAYIN ÇAĞINDAN SOSYAL YAYIN ÇAĞINA: RİSK YÖNETİMİNDE BİR YAPAY ZEKÂ MODELİ ÖNERİSİ, **İstanbul Ticaret Üniversitesi Sosyal Bilimler Dergisi**.

13. Y. Furkan Şen & Doğanay Yurtoğlu, 2020. The Importance of Artificial Intelligence in Intelligence Analysis in the Context of Technology and Security Relationship, **Journal of Security Studies**.

14. Cem BAYDOĞAN & Bilal ALATAŞ, 2021. Çevrimiçi Sosyal Ağlarda Nefret Söylemi Tespiti için Yapay Zeka Temelli Algoritmaların Performans Değerlendirmesi, **Fırat Üniversitesi Müh. Bil. Dergisi**.

15. Manning, C. D., R.P..S.H., 2008. Introduction to Information Retrieval, **Cambridge University Press**.

16. MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations, **Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability** .

17. Burges, C.J.C., 1998. SVM models, **A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery**.

ÖZGEÇMİŞ

KİŞİSEL BİLGİLER

Adı, Soyadı	:	Melisa DEMİR
Uyruğu	:	Türkiye (T.C.)
Doğum Tarihi ve Yeri	:	15.05.2002 - ERZURUM
Telefon	:	0 539 452 93 98
E-posta	:	1030510331@erciyes.edu.tr melisademir02525@gmail.com
Adres	:	Mevlana Mah. Batuhan Sok. Mevlana Apartmanı No : 4 Kat : 1 İç Kapı No : 1 38039, Talas KAYSERİ TÜRKİYE

EĞİTİM

Derece	**Kurum**	**Mez.Yılı**
Lise	Oltu Fen Lisesi, ERZURUM	2020
Ortaokul	Mehmet Akif ERSOY Ortaokulu, ERZURUM	2016

ÖZGEÇMİŞ

KİŞİSEL BİLGİLER

Adı, Soyadı	:	Duygu Gözde KAYABAŞI
Uyruğu	:	Türkiye (T.C.)
Doğum Tarihi ve Yeri	:	18.09.2002 - MERSİN
Telefon	:	0 535 055 33 06
E-posta	:	1030510338@erciyes.edu.tr duygugozde33@gmail.com
Adres	:	Limonluk Mah. 2430. Sok. Selenevler Apt. No : 8 Kat : 10 / Daire : 20 33200, Yenişehir MERSİN TÜRKİYE

EĞİTİM

Derece	**Kurum**	**Mez.Yılı**
Lise	75. Yıl Fen Lisesi, MERSİN	2020
Ortaokul	Çankaya Ortaokulu, MERSİN	2016