T.C.
ERZİNCAN BİNALİ YILDIRIM ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
YAPAY ZEKA VE ROBOTİK ANABİLİM DALI

TWİTTER'DA YANLIŞ BİLGİ VE DEZENFORMASYON TESPİTİ İÇİN YAPAY ZEKA TEKNİKLERİNİN KULLANILMASI VE KARŞILAŞTIRILMASI

Omar Raad Mahmood MAHMOOD

Danışman: Dr. Öğr. Üyesi Funda AKAR

TEZ JÜRİ ÜYELERİ
Doç. Dr. Durmuş ÖZDEMİR
Dr. Öğr. Üyesi Funda AKAR
Dr. Öğr. Üyesi İsmail AKGÜL

YÜKSEK LİSANS TEZİ
ERZİNCAN, 2024

Kabul ve Onay Sayfası

Dr. Öğr. Üyesi Funda AKAR danışmanlığında, Omar Raad Mahmood MAHMOOD tarafından hazırlanan bu çalışma 27/08/2024 tarihinde aşağıdaki jüri tarafından Yapay Zeka ve Robotik Anabilim Dalı'nda Yüksek Lisans Tezi olarak kabul oybirliği ile kabul edilmiştir.

Başkan:	Doç.Dr.Durmuş ÖZDEMİR	İmza:
Üye :	Dr.Öğr.Üyesi Funda AKAR	İmza:
Üye :	Dr.Öğr.Üyesi İsmail AKGÜL	İmza:

Yukarıdaki sonuç Enstitü Yönetim Kurulunun …. / …. / 20.... tarih ve …../…….. sayılı kararı ile onaylanmıştır.

Doç. Dr. Kemal Volkan ÖZDOKUR
Enstitü Müdür V.

Not: Bu tezde kullanılan özgün ve başka kaynaklardan yapılan bildirişlerin, şekil ve tabloların kaynak olarak kullanımı, 5846 sayılı Fikir ve Sanat Eserleri Kanunundaki hükümlere tabidir.

Bilimsel Etiğe Uygunluk Sayfası

"TWİTTER'DA YANLIŞ BİLGİ VE DEZENFORMASYON TESPİTİ İÇİN YAPAY ZEKA TEKNİKLERİNİN KULLANILMASI VE KARŞILAŞTIRILMASI" isimli "Yüksek Lisans" tezim tarafımca intihal tespit programı ile incelenmiştir. Buna göre tezimde bilimsel etik ihlali ve intihal olarak nitelendirilebilecek herhangi bir durum olmadığını taahhüt ederim.

Bu çalışmadaki tüm bilgilerin, akademik ve etik kurallara uygun bir biçimde elde edildiğini; aynı zamanda bu kural ve davranışların gerektirdiği gibi, bu çalışmanın özünde olmayan tüm materyal ve sonuçları tam olarak aktardığımı ve referans gösterdiğimi beyan ederim. 27/08/2024

Omar Raad Mahmood MAHMOOD

ÖZET

TWİTTER'DA YANLIŞ BİLGİ VE DEZENFORMASYON TESPİTİ İÇİN YAPAY ZEKA TEKNİKLERİNİN KULLANILMASI VE KARŞILAŞTIRILMASI

Omar Raad Mahmood MAHMOOD

Yüksek Lisans Tezi, Erzincan Binali Yıldırım Üniversitesi, Fen Bilimleri Enstitüsü, Yapay Zeka ve Robotik Anabilim Dalı
Danışman: Dr.Öğr.Üyesi Funda AKAR
2024, 84 sayfa

Bu çalışma, sosyal medya platformu Twitter'daki yanlış bilgileri tespit etmek için bir dizi yapay zeka (AI) tekniğini incelerek yanlış bilgi ve sahte haberlerin kamusal söylem üzerindeki potansiyel etkisine ilişkin yaygın bir sorunu ele almaktadır. Aldatıcı ve güvenilir içeriği ayırt etmek için Uzun Kısa Süreli Bellek (LSTM), Destek Vektör Makinesi (SVM), Rastgele Orman (RF), Gauss Naive Bayes (GaussianNB), Gradyan Arattırma (GBoost), Karar Ağacı (DT), Lojistik Regresyon (LR), Aşırı Gradyan Arattırma (XGBoost) ve K-En Yakın Komşular (K-NN) dahil olmak üzere çeşitli modeller kullanılmıştır. Analizde doğal dil işleme (NLP), derin öğrenme (DL) ve makine öğrenimi (ML) teknikleri kullanılarak 23,481 sahte tweet ve yaklaşık 21,418 gerçek tweetten oluşan bir veri seti kullanılarak her bir modelin yanlış bilgi kalıplarını tespit etmedeki etkinliği gösterilmiştir. Çalışma, özellikle doğruluk, verimlilik ve ölçeklenebilirliğe odaklanarak, bu AI sistemlerinin güçlü yönlerinin ve sınırlamalarının kapsamlı bir değerlendirmesini sunmaktadır. Sonuçlar, XGBoost'un %99.82 doğruluk ve %99.81 F1 puanı ile en iyi performansı sergilediğini göstermektedir. Bunu %99.63 doğruluk ve %99.62 F1 puanı gösteren (GBoost) ve %99.61 doğruluk ve %99.59 F1 puanı ile (DT) takip etmektedir. Sonuçlar bu modellerin en etkili modeller olduğunu göstermektedir. Diğer modeller %99.31 ile %81.63 arasında değişen doğruluk gösterirken, sonuçlar ana modellerin performansına dair fikir vererek dezenformasyonla mücadeleye ve bilginin güvenilirliğine önemli katkı sağlamaktadır.

Anahtar Kelimeler: sahta haber, sosyal medya, twitter, **y**anlış bilgi, yapay zeka

ABSTRACT

USING AND COMPARISON OF ARTIFICIAL INTELLIGENCE TECHNIQUES TO DETECT MISINFORMATION AND DISINFORMATION ON TWITTER

Omar Raad Mahmood MAHMOOD

Master's Thesis, Erzincan Binali Yıldırım University, Institute of Science and Technology,
Department of Artificial Intelligence and Robotics
Advisor: Assist Prof. Funda AKAR
2024, 84 pages

This study addresses a widespread problem of the potential impact of misinformation and fake news on public discourse by examining a set of artificial intelligence (AI) techniques to detect misinformation on the social media platform Twitter. A variety of models, including Long Short-Term Memory (LSTM), Support Vector Machine (SVM), Random Forest (RF), Gaussian Naive Bayes (GaussianNB), Gradient Search (GBoost), Decision Tree (DT), Logistic Regression (LR), Extreme Gradient Search (XGBoost), and K-Nearest Neighbors (K-NN), are used to distinguish deceptive and trustworthy content. The analysis uses natural language processing (NLP), deep learning (DL), and machine learning (ML) techniques to demonstrate the effectiveness of each model in detecting misinformation patterns using a dataset consisting of 23,481 fake tweets and approximately 21,418 genuine tweets. The study provides a comprehensive assessment of the strengths and limitations of these AI systems, particularly focusing on accuracy, efficiency, and scalability. The results show that XGBoost performs best with 99.82% accuracy and 99.81% F1-score. This is followed by (GBoost) with 99.63% accuracy and 99.62% F1-score and (DT) with 99.61% accuracy and 99.59% F1-score. The results show that these models are the most effective models. While the other models show accuracy ranging from 99.31% to 81.63%, the results provide insight into the performance of the main models, contributing significantly to the fight against disinformation and the reliability of information.

Keywords: fake news, social media, twitter, misinformation, artificial intelligence

TEŞEKKÜR

Öncelikle, tüm akademik çabalarımda başarılı olmamı sağlayan sağlık ve esenlik nimetleri için Yüce Allah'a şükrediyorum. Daha sonra, başından sonuna kadar doğru rehberlik, yararlı yorumlar ve katılım sağlama konusunda uzmanlığı paha biçilmez olan tez danışmanım Dr. Öğretim Üyesi Funda AKAR'a, en derin minnettarlığımı ifade etmek isterim. Beni sürekli cesaretlendirdi ve süreç boyunca mümkün olan her şekilde bana yardım etmeye her zaman hazır ve istekliydi, bana karşı bu kadar nazik olmasaydı bu tezi bitirmem hiç de kolay olmazdı.

Ayrıca dezenformasyon ve yalan haber alanında bilgi sahibi olmamı sağlayan Dr. Bassim Al Khashab'a beni işyerine davet ettiği ve aramızdaki sürekli ve verimli iletişimi için teşekkür etmek istiyorum.

Tüm sınıf arkadaşlarıma ve meslektaşlarıma bilgilerini benimle paylaştıkları için teşekkür etmek istiyorum. Onların tavsiyeleri ve teşvikleri beni doğru yolda tutmak için hayati önem taşıyordu.

Bana parlak bir gelecek dilediği için sevgili anneme de teşekkür etmek istiyorum. Ve en karanlık günlerimde bana ışık olduğu için sevgili babama. Bu araştırmayı tamamlamamda büyük destekleri olan arkadaşlarıma, akrabalarıma ve komşularıma da çok teşekkür ederim.

Omar Raad Mahmood MAHMOOD
Ağustos, 2024

İÇİNDEKİLER

Sayfa

ÖZET ..... **i**
ABSTRACT ..... **ii**
TEŞEKKÜR ..... **iii**
ŞEKİLLER LİSTESİ ..... **vii**
TABLOLAR LİSTESİ ..... **viii**
SİMGELER ve KISALTMALAR ..... **ix**
1. GİRİŞ **1**
1.1. Problem Bildirimi 2
1.2. Araştırmanın Amacı 3
1.3. Araştırmanın Önemi 3
2. KAVRAMSAL ÇERÇEVE VE İLGİLİ ÇALIŞMALAR **5**
2.1. Sahte Haberin Tanımı 5
2.2. Makine Öğrenmesi Teorik Temel 6
2.2.1. Lineer regresyon sınıflandırıcı 9
2.2.2. Karar ağacı sınıflandırıcı 10
2.2.3. GBoost sınıflandırıcısı 11
2.2.4. Rastgele orman sınıflandırıcı 11
2.2.5. XGBoost sınıflandırıcı 12
2.3. Derin Öğrenme Teorik Temel 13
2.3.1. Uzun kısa süreli bellek 13
2.3.2. Evrişimsel sinir ağları 15
2.4. Doğal Dil İşleme Teorik Temel 15
2.4.1. Destek vektör makinesi 16
2.4.2. K-en yakın komşu 17
2.4.3. Gaussian naive bayes 18
2.5. Sahte Haber Tespitine Yönelik Çalışmalar 20
3. YÖNTEM **24**
3.1. Kullanılan Yazılım Araçları 24
3.2. Kullanılan Donanım Araçları 26
3.3. Veri Kümesi 26

3.4. Veri Ön İşleme 28
3.4.1. Noktalama işaretleri ve özel karakterler kaldırılması 28
3.4.2. Durdurma sözcüklerinin kaldırılması 29
3.5. Metin temsili 30
3.5.1. Terim sıklığı-ters belge sıklığı (TF-IDF) 30
3.5.2. Vektörleştirme (Word2Vec) 31
3.6. Yapay Zeka Modelleri 32
3.6.1. Makine öğrenimi 33
3.6.1.1. Lojistik regresyon 33
3.6.1.2. Karar ağacı 34
3.6.1.3. Gradiant boosting 36
3.6.1.4. Rastgele orman 38
3.6.1.5. Aşırı gradyan artırma 40
3.6.2. Derin Öğrenme 42
3.6.2.1. Uzun kısa süreli bellek 42
3.6.3. Doğal Dil İşleme (NLP) 46
3.6.3.1. K-en yakın komşu 46
3.6.3.2. Gaussian naive bayes 48
3.6.3.3. Destek vektör makinesi 50
3.7. Değerlendirme Metrikleri 53
3.7.1. Doğruluk 53
3.7.2. Kesinlik 53
3.7.3. Duyarlılık 53
3.7.4. F1- Puan 54
3.7.5. Karışıklık Matrisi 54
3.7.6. ROC eğrisi 54
4. BULGULAR 56
4.1. Makine Öğrenimi Yöntem Sonuçları 56
4.2. Derin Öğrenme Yöntem Sonuçları 57
4.3. Karşılaştırmalı Analiz 57

5. SONUÇ VE ÖNERİLER 62
KAYNAKLAR 64

ŞEKİLLER LİSTESİ

Sayfa

Şekil 1. Sosyal medyada kullanıcılar arasındaki etkileşimler 1
Şekil 2. Denetimli Öğrenme 7
Şekil 3. Denetimsiz Öğrenme 8
Şekil 4. LSTM mimarisi 14
Şekil 5. Çalışmanın Modeli 25
Şekil 6. Gerçek Haber Veri Seti Dağılımı 27
Şekil 7. Sahte Haber Veri Seti Dağılımı 28
Şekil 8. Noktalama işaretleri ve özel karakterler kaldırılması 29
Şekil 9. Durdurma kelimesi kaldırma kaynağı 30
Şekil 10. Özellik vektörü temsili 31
Şekil 11. Karışıklık Matrisi 54
Şekil 12. ROC Eğrisi 55
Şekil 13. (A) LR için karışıklık matrisi ve (B) ROC eğrisi 59
Şekil 14. (A) DT için karışıklık matrisi ve (B) ROC eğrisi 59
Şekil 15. (A) GBoost için karışıklık matrisi ve (B) ROC eğrisi 59
Şekil 16. (A) RF için karışıklık matrisi ve (B) ROC eğrisi 60
Şekil 17. (A) XGBoost için karışıklık matrisi ve (B) ROC eğrisi 60
Şekil 18. (A) LSTM için karışıklık matrisi ve (B) ROC eğrisi 60
Şekil 19. (A) GaussianNB için karışıklık matrisi ve (B) ROC eğrisi 61
Şekil 20. (A) K-NN için karışıklık matrisi ve (B) ROC eğrisi 61
Şekil 21. (A) SVM için karışıklık matrisi ve (B) ROC eğrisi 61

TABLOLAR LİSTESİ

Sayfa

Tablo 1. Uygulanan modellerin başarı oranları..... 58

SİMGELER ve KISALTMALAR

Simgeler

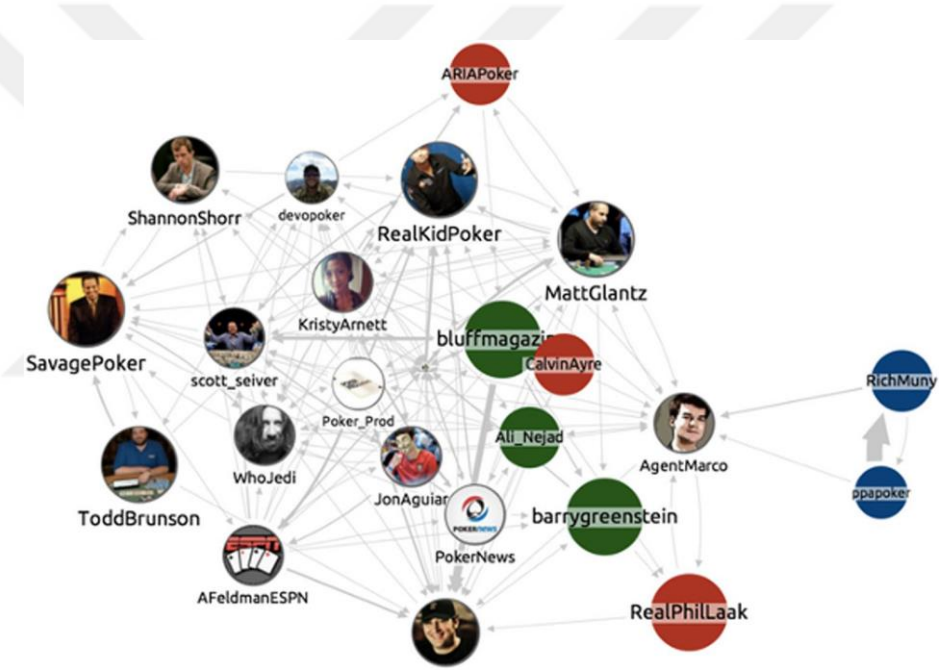
$\bar{X}$	Ortalama
$\%$	Yüzde
α	Güvenirlik Katsayısı
β	Regresyon katsayısı
B	Regresyon Sabiti
t	t-değeri
T^*	Optimal karar ağacıdır
f	Model fonksiyonu
W	Ağırlık matrisi
$\hat{y}$	Tahmin edilen sınıf (0 veya 1)
x_i	Giriş özellikleridir
σ	Sigmoid activision fonksiyonu

Kısıtlamalar

AI	Artificial Intelligence
ML	Machine Learning
DL	Deep Learning
NLP	Natural Language Processing
RF	Random Forest
DT	Decision Tree
LR	Logistic Regression
XGBoost	Extreme Gradient Boosting
GBoost	Gradient Boosting
RNN	Recurrent Neural Network
SVM	Support Vektor Machine
ROC	Receiver Operating Characteristic
K-NN	K-Nearest Neighbors
LSTM	Long Short-Term Memory
YZ	Yapay Zekâ
CNN	Convolutional Neural Network
GaussianNB	Gaussian Naive Bayes

1. GİRİŞ

Dijital dönüşüm çağında sosyal medya ve çeşitli haber kaynakları sınır tanımayan doğrudan iletişim sağlayarak günlük hayatın ayrılmaz bir parçası haline gelmiştir. Herkes bu sosyal medya platformlarında çok düşük bir maliyetle sahte haberler yayınlayabilir (Song et al., 2021). Pew Araştırma Merkezi, 2018 yılında Amerikalı yetişkinlerin %68'inin haberleri sosyal medyadan takip ettiğini belirterek, bu platformların bilginin yayılmasında oynadığı önemli rolü ortaya koymaktadır (Shu et al., 2019). Sosyal ağ siteleri, çeşitli konularda gerçek zamanlı fikir ve haber alışverişine olanak tanıyarak sanal etkileşimleri kolaylaştırmada önemli bir rol oynamaktadır. Şekil 1'de kullanıcı ve sosyal ağlar arasındaki ilişki gösterilmektedir.



Şekil 1. Sosyal medyada kullanıcılar arasındaki etkileşimler (Islam et al., 2020).

Son yıllarda özellikle Twitter, etkileşim yaratan ve kamusal konuşmaları etkileyen önemli bir mecra olarak ortaya çıkmıştır. Bu durum, sosyal ağların temel özelliklerinden, yani geniş erişilebilirlik, kullanım kolaylığı ve hızlı yayılma özelliklerinden kaynaklanmaktadır; bu da sahte haberlerin etkisinin daha da artmasına sebep olmaktadır (Imran et al., n.d.; Li & Lei, 2022). Özellikle sahte haber ve söylentiler şeklindeki yanlış bilgilendirme Facebook, Twitter ve Sina Weibo gibi internet üzerindeki sosyal ağlarda büyük bir sorun teşkil etmektedir (Limon & Jahan, n.d.). Haberler, görüşler ve trendler dijital ortamda olağanüstü bir hızla yayılmaktadır. Bununla birlikte, bu platformların her yerde bulunabilen doğası, onları yanlış bilgi ve

dezenformasyonun yaygın yayılımına da maruz bırakmakta ve bilgi ekosistemlerinin bütünlüğü için büyük bir zorluk teşkil etmektedir.

Sahte haber tespiti, farklı disiplinlerin üzerinde çalıştığı güncel bir araştırma alanıdır (George et al., 2021; Pennycook & Rand, 2021). Son yıllarda sahte haberler ile gerçek haberleri birbirinden ayırmak için birçok girişimde bulunulmuştur (B. Hu et al., 2024). Yapay zekânın ortaya çıkışı, aldatıcı içerikle mücadelede yeni bir dönemi temsil etmektedir. "Twitter'da Yanlış Bilgi ve Dezenformasyonu Tespit Etmek için AI Tekniklerinin Kullanılması ve Karşılaştırılması" başlıklı bu araştırma, gelişmiş AI metodolojileri ile Twitter'da yanlış anlatıların yayılmasıyla mücadeleye duyulan acil ihtiyaç üzerine ele alınmıştır. Çalışma, ML algoritmaları, NLP ve DL modelleri gibi en son AI tekniklerinin gerçek bilgiyi aldatıcı içerikten ayırt etmek için nasıl uygulanabileceğini araştırmaktadır. Bu çalışmanın karşılaştırmalı yönü ile yanlış bilgi tespiti bağlamında farklı AI yaklaşımlarının güçlü ve zayıf yönlerinin değerlendirilmesi sağlanmıştır. Algoritmanın doğruluğu, eğitim verilerinin kalitesi, haber hikayesinin karmaşıklığı ve sahte haber yaratıcısının karmaşıklığı gibi çeşitli faktörlere bağlıdır (Burgers et al., 2023). Araştırma performans metriklerini, doğruluk oranlarını ve hesaplama verimliliğini inceleyerek, sosyal medya platformlarını yanlış bilginin sinsi etkisine karşı güçlendirmeye yönelik devam eden söylemlere değerli içgörüler katmayı amaçlamaktadır.

Çalışma, mevcut zorlukları, çözümleri ve doğrulamaları anlamak amacıyla ML, DL ve NLP yöntemlerine dayalı yanlış bilgi tespiti üzerine yapılan mevcut araştırmaları kapsamlı bir şekilde değerlendirerek çeşitli stratejilerin avantaj ve dezavantajlarını incelemektedir.

1.1. Problem Bildirimi

Sosyal medyada kasıtlı olarak yanıltıcı nitelikte kullanıcı tarafından oluşturulan içeriklerin çokluğu nedeniyle sahte haberleri tespit etmek zorlu bir iştir (Yang et al., 2018). Bu durum toplum için olumsuz bir tehdit oluşturmaktadır. Sosyal istikrarı ve kamu güvenini zayıflattığı için olumsuz sonuçlar doğurmakta ve sahte haber tespitine olan talebin artmasına neden olmaktadır (L. Hu et al., 2022). Örneğin, Las Vegas'taki trajik silahlı saldırıdan birkaç saat sonra sosyal medyada yayılmaya başlayan sahte haberler birçok masum insanın zarar görmesine neden olmuştur (Cui et al., 2019).

Yanlış bilgiler genellikle gerçek hikayeler abartılarak veya dikkat çekmek için aldatıcı başlıklar ve sloganlar kullanılarak sosyal medyada yayılır. Nesnelliği korumak için, açıkça belirtilmedikleri sürece öznel değerlendirmelerden kaçınmak önemlidir. Açık, nesnel ve değerden bağımsız bir dil kullanılmalı, dikkatli kelime seçimi ve dilbilgisi doğruluğuna öncelik verilmelidir. Metin, tutarlı alıntı ve dipnot tarzı da dahil olmak üzere geleneksel yapı ve biçimlendirme özelliklerine uygun olmalıdır. Son olarak, ifadeler arasında nedensel bağlantılar kurarak mantıksal bir bilgi akışı sağlarken resmi bir dil kullanmak ve önyargılı bir dilden kaçınmak çok önemlidir. Bir başka risk de diğer elektronik medya kuruluşlarının bunu bir haber kaynağı olarak kullanması ve hikâyenin daha da yayılmasıdır. İnternet üzerinden bilgi ve içeriğin güvenilirliğini belirlemek zordur ve bu da yanlış bilginin kısa sürede yayılmasına yol açabilmektedir. İnternet üzerinde kaynakların güvenilirliğini değerlendirirken objektif olmak ve öznel değerlendirmelerden kaçınmak önemlidir.

1.2. Araştırmanın Amacı

Bu araştırmanın amacı, yalan haberleri tespit edebilen ve sınıflandırabilen bir model geliştirmek, en uygun mimariyi önermek. Çalışmanın amacı aşağıdaki şekilde ifade edilmiştir:

- Kullanılan modellerin doğruluğunu artırmak için, sahte haber tanımlamasının gömülü hale getirilmesi.
- Öğrenme oranı değiştikçe eğitim doğruluğunun değerlendirilmesi.
- Sahte haberlerin tespitine ilişkin bilimsel bir rapor hazırlanması.
- Farklı AI tekniklerinin, ve ML, DL ve NLP mimarilerinin performansını değerlendirmesi.
- Dezenformasyonu ve yanlış bilgileri, doğru bilgilerden ayırt edilmesi ve tespit edilmesi için en iyi modelin önerilmesi.

1.3. Araştırmanın Önemi

Bu çalışmanın amacı, yalan haberleri tespit edebilen ve sınıflandırabilen bir model geliştirmek, en uygun mimariyi önermek ve bilimsel bir rapor üretmektir. Bu hedef doğrultusunda gerçekleştirilecek olan çalışmanın önemi birkaç açıdan değerlidir:

1. **Toplumsal Güvenin Artırılması:** Sahte haberlerin tespit edilmesi, kamuoyunun doğru bilgilendirilmesine katkıda bulunur ve bireylerin güvenilir bilgi kaynaklarına erişimini sağlar. Bu, toplumda genel bir güven ortamının oluşmasına yardımcı olur.
2. **Bilgi Kirliliğinin Azaltılması:** Yanlış ve yanıltıcı bilgilerin yayılması, bilgi kirliliğine neden olur ve bireylerin doğru bilgiye ulaşmasını zorlaştırır. Geliştirilecek olan model, bilgi kirliliğini azaltarak, bilgi ekosisteminin temizlenmesine katkıda bulunur.
3. **Demokratik Süreçlerin Korunması:** Sahte haberler, seçimler gibi demokratik süreçleri olumsuz etkileyebilir ve halkın yanlış bilgilendirilmesine neden olabilir. Bu çalışma, sahte haberlerin tespiti yoluyla demokratik süreçlerin korunmasına ve adil bir ortamın sağlanmasına katkıda bulunabilir.
4. **Teknolojik Gelişmeler ve Yenilikler:** Çalışma kapsamında farklı ML, DL ve NLP mimarilerinin performansları değerlendirilecektir. Bu değerlendirme, yeni ve daha etkili algoritmaların geliştirilmesine olanak tanır ve bu alandaki teknolojik ilerlemelere katkıda bulunur.
5. **Akademik Katkılar:** Sahte haberlerin tespitine ilişkin hazırlanan bilimsel rapor, akademik literatüre önemli katkılar sağlar. Ayrıca, belirlenen kriterler ve önerilen en iyi tespit modeli, gelecekteki araştırmalara temel oluşturabilir ve bu alanda çalışan diğer araştırmacılara yol gösterici olabilir.
6. **Eğitim ve Farkındalık:** Çalışmanın sonuçları, eğitim kurumları ve medya kuruluşları için de değerli bilgiler sunar. Sahte haberlerin tanımlanması ve tespiti konusundaki bulgular, eğitim materyalleri ve medya okuryazarlığı programlarına entegre edilebilir.

Sonuç olarak, bu çalışma, yalan haberlerin tespit edilmesi ve sınıflandırılması konusundaki mevcut yaklaşımlara önemli katkılar sağlayacak, toplumda bilgi güvenliği ve doğruluğun artırılmasına yardımcı olacaktır. Geliştirilecek olan model ve öneriler, sadece teorik değil, pratik uygulamalar açısından da büyük bir öneme sahiptir.

2. KAVRAMSAL ÇERÇEVE VE İLGİLİ ÇALIŞMALAR

2.1. Sahte Haberin Tanımı

Yanlış bilgilendirme sorununu tartışırken araştırmacılar sahte haber, yanlış bilgilendirme ve bilgi düzensizliği gibi farklı terimler ve tanımlar kullanmaktadır. Ancak bu terimlerin evrensel bir tanımı bulunmamaktadır. Bu nedenle, akademisyenler tarafından sunulan sınıflandırmalar ve modeller yanlış bilginin farklı yönlerini kapsamaktadır. Makalelerde, "Sahte Haber" ifadesi araştırmacılar tarafından "okuyucuları yanıltmak için kasıtlı olarak uydurulan ve doğrulanabilen yalan haber" şekilde tanımlanmıştır (Allcott & Gentzkow, 2017; Wang et al., 2018). Diğer bir araştırmacı, dezenformasyon ile yanlış bilgi arasındaki ayrımın yazarın niyetinde yattığını vurgulamıştır (Jia, 2020). Bir başka araştırmacının tanımına göre ise sahte haber, yanıltıcı ve/veya yanlış ifadeler içeren yanıltıcı çevrimiçi bilgidir (Baptista & Gradim, 2022). Araştırmacıların bir başka tanımında sahte haberler; gerçek olaylarla ilgili olabilir veya olmayabilir ve okuyucuların dikkatini çekmek ve reklam gelirlerini, paylaşımları, tıklamaları ve/veya ideolojik kazanımları artırmak için fırsatçı bir niyetle haber kılığına sokulmuş belirli veya hayali bir kitleyi aldatmak ve/veya manipüle etmek için kasıtlı olarak yaratılabilir. (Scheufele & Krause, 2019; Schöpfel & Kergosien, n.d.). Sahte haberler bir tür yanlış bilgi olarak tanımlanırken, aynı zamanda yanlış bilginin bir alt kümesi olarak da sınıflandırılır. Yalan haber, yanlış veya yanıltıcı iddiaların (genellikle) kasıtlı olarak sunulmasıdır (Gelfert, 2018). Sahte haberin bir parçası, atıfta bulunduğu olayı doğru bir şekilde temsil etmeyen multimedya içeriği paylaşan bir haber yayınıdır (Cao et al., 2020). Farklı haber biçimlerini birbirinden ayırmak önemlidir. Haberler; tıklama tuzağı, aldatmaca, propaganda, hiciv ve parodi gibi çeşitli biçimlerde sınıflandırılabilir. Yanlış yönlendirilmekten kaçınmak için bu farklı formları tanımak önemlidir (Collins et al., 2021; Tandoc et al., 2018).

Henüz sahte haberin net ve onaylanmış evrensel bir tanımı yoktur (Shu et al., 2017; Zhou et al., 2019). Ancak ilgili literatürü tarayarak sahte haber konusu ve tanımı incelenmiştir. Sahte haberler, yanlış bilgiler arasındaki zarar ve dolandırıcılık boyutlarını belirleyerek tanımlanan üç bilgi türüne ayrılarak aralarındaki farklar aşağıdaki gibi açıklanabilir:

- **Yanlış bilgi (Mis-information)**: Söylenti gibi yanlış bir bilginin kasıtsız olarak paylaşılması. Yanlış bilgi, herhangi bir art niyet olmadan yanlış olan bilgi veya içeriktir. Bu tür bilgi, genellikle insanlar tarafından kasıtlı olmadan paylaşılır ve yanlış olduğu bilinmez.

Örneğin:

- Yanlış anlama veya eksik bilgi nedeniyle yanlış bilgi paylaşılması.
- Bir doğal afet hakkında hatalı bilgiler içeren sosyal medya gönderileri.
- İyi niyetli bir kişinin bir ünlünün ölümüyle ilgili bir söylentiyi gerçekliğini doğrulamadan paylaşması.

- **Dezenformasyon (Dis-information):** Dezenformasyon, bilerek ve kasıtlı olarak yanıltıcı veya yanlış bilgi yayma eylemidir. Bu tür bilgi, genellikle belirli bir amaca ulaşmak veya insanları yanıltmak için bilinçli olarak oluşturulur ve paylaşılır.

Örneğin:

- Politik bir figürü karalamak amacıyla sahte haberler yayımlanması.
- Sosyal medya üzerinden yaygınlaştırılan ve insanların kararlarını etkilemeyi amaçlayan yalan haberler.
- Kötü niyetli bir kişinin doğal bir felaketle ilgili bir aldatmaca yayarak paniğe ve kafa karışıklığına neden olması.

- **Kötü Niyetli Bilgi (Mal-information):** Kötü niyetli bilgi, doğru bilgilerin kasıtlı olarak bir kişiye, kuruluşa veya ülkeye zarar vermek amacıyla kullanılmasıdır. Bu, doğru bilginin bağlamından koparılarak yanıltıcı bir şekilde sunulmasını içerir.

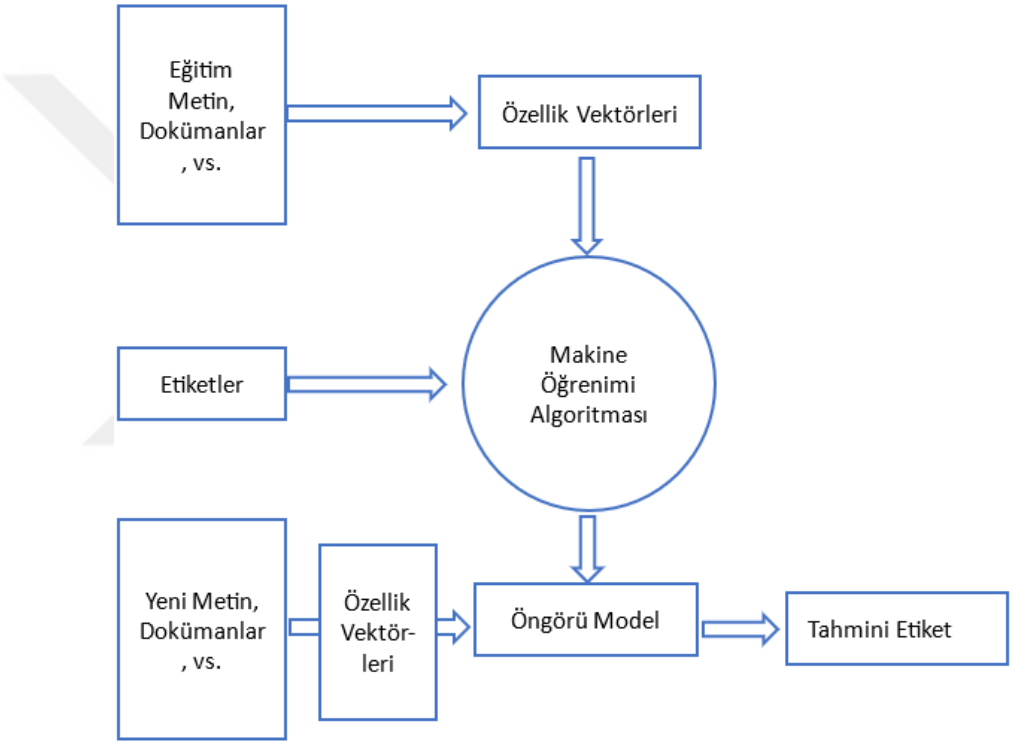
Örneğin:

- Gizli veya özel bilgilerin, bir kişinin itibarını zedelemek için ifşa edilmesi.
- Doğru olan bilgilerin, yanıltıcı bir şekilde sunularak yanlış bir algı oluşturulması.

2.2. Makine Öğrenmesi Teorik Temel

Makine öğrenmesi, verilerden öğrenme ve tahminler yapma yeteneği sağlayan bir AI dalıdır. Makine öğrenmesi, eğitim veri setlerine ve bu verilerin etiketlenme şekline bağlı olarak üç ana kategoriye ayrılır: denetimli öğrenme, yarı denetimli öğrenme ve denetimsiz öğrenme. (Al-Jarrah et al., n.d.; Fong, 2010).

- **Denetimli Öğrenme:** Bu, makine öğreniminde sınıflandırma için en yaygın çerçevedir. Her örneğin bir çıktı etiketi ile eşleştirildiği etiketli bir veri seti üzerinde bir modelin eğitilmesini içerir. Denetimli öğrenmede algoritma, çıktısını verilen girdi ile karşılaştırır ve daha doğru yanıt vermeyi öğrenir. Bu yönteme örneklerden öğrenme de denir (Alzubi et al., 2018). Örneğin, denetimli bir öğrenme algoritması, "gerçek" veya "gerçek değil" olarak etiketlenmiş twitter tweet'lerinden oluşan bir veri seti üzerinde eğitilebilir ve yeni tweet'leri buna göre sınıflandırmayı öğrenebilir. Şekil 2'de denetimli öğrenmenin işleyişi gösterilmektedir.



Şekil 2. Denetimli Öğrenme

Örnek Uygulamalar:

- **Sınıflandırma:** E-postaların spam veya spam olmadığını belirleme.
- **Regresyon:** Ev fiyatlarını tahmin etme.

Avantajları:

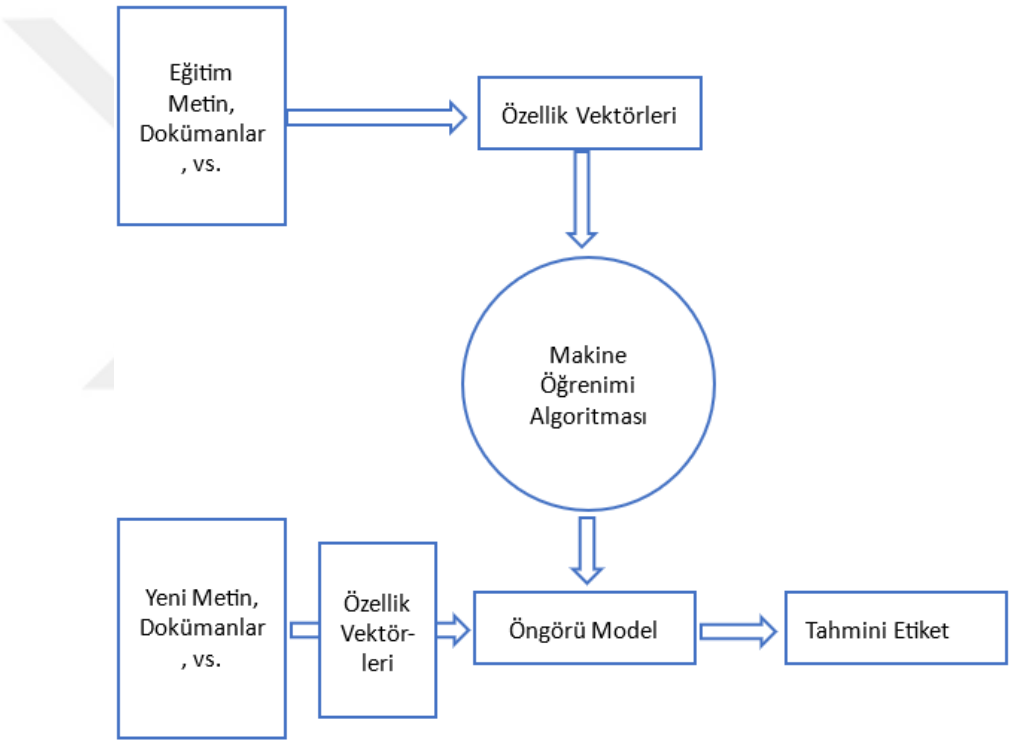
- Etiketlenmiş verilerden dolayı yüksek doğruluk.
- Tahminler genellikle daha güvenilir.

Dezavantajları:

- Büyük ve etiketlenmiş veri setlerine ihtiyaç duyar.

- Etiketleme işlemi zaman alıcı ve maliyetli olabilir.

- **Denetimsiz Öğrenme:** Denetimli öğrenmenin aksine, denetimsiz öğrenme, herhangi bir etiket olmadan yalnızca girdi verilerini kullanan bir öğrenme türüdür ve model, etiketlenmemiş verilerin kalıplarından ve özelliklerinden öğrenir ve verilerdeki yapıyı kendisinin bulması gerekir. (R. Sharma, 2020). Yaygın bir denetimsiz öğrenme görevi, algoritmanın veri noktalarını benzerliğe dayalı olarak kümeler halinde gruplandırdığı kümelemedir. Müşterileri satın alma davranışlarına göre farklı gruplara ayırmak buna bir örnek olabilir. Şekil 3'te denetimsiz öğrenmenin işleyişi gösterilmektedir.



Şekil 3. Denetimsiz Öğrenme

Örnek Uygulamalar:

- **Kümeleme (Clustering):** Müşteri segmentasyonu.
- **Boyut Azaltma (Dimensionality Reduction):** Verilerin görselleştirilmesi veya özellik çıkarımı.

Avantajları:

- Etiketlenmiş veri gerektirmez.
- Keşifsel veri analizi için kullanışlıdır.

Dezavantajları:

- Çıktılar yorumlanması zor olabilir.
- Model performansını değerlendirmek daha zordur.

- **Yarı-Denetimli Öğrenme:** Bu yaklaşım hem denetimli hem de denetimsiz öğrenmenin unsurlarını birleştirir. Daha büyük bir etiketsiz veri kümesinin yanında az miktarda etiketli veri kullanır. Algoritma, veri kümesinin yapısı hakkında bilgi edinmek ve etiketsiz veriler hakkında tahminlerde bulunmak için etiketli verilerden yararlanır. Bu, özellikle veri etiketlemenin pahalı veya zaman alıcı olduğu durumlarda faydalı olabilir. Bu yöntem, etiketlenmiş verilerin yüksek maliyetine karşılık, etiketlenmemiş verilerin bol bulunmasını avantaja çevirir. Model, etiketlenmiş verilerden öğrendiklerini, etiketlenmemiş verilerle destekleyerek genelleme yeteneğini artırır.

Örnek Uygulamalar:

- **Görüntü tanıma:** Az sayıda etiketlenmiş görüntü kullanarak büyük miktarda etiketlenmemiş görüntüyü sınıflandırma.
- **Doğal dil işleme:** Az miktarda etiketlenmiş metin kullanarak büyük miktarda metni anlamlandırma.

Avantajları:

- Daha az etiketlenmiş veri ihtiyacı.
- Büyük etiketlenmemiş veri setlerinin avantajlarından yararlanma.

Dezavantajları:

- Karmaşık modeller ve eğitim süreçleri gerektirebilir.
- Performans, kullanılan etiketlenmiş ve etiketlenmemiş veri miktarına bağlıdır.

2.2.1. Lineer regresyon sınıflandırıcı

Lineer regresyon (LR), özünde, özellikle ikili sınıflandırma görevleri için uyarlanmış denetimli öğrenme çerçevesinde çalışır. Regresyon analizi, bir yanıt değişkeni ile bir veya daha fazla açıklayıcı değişken arasındaki ilişkiyi tanımlamayı amaçlayan ve sıklıkla kullanılan istatistiksel bir yöntemdir (Smita, 2021). Birincil amacı, metin içeriğinden veya diğer ilgili veri kaynaklarından elde edilen bir dizi girdi özelliğine dayalı olarak, belirli bir girdinin gerçek veya aldatıcı haber makaleleri gibi belirli bir sınıfa ait olma olasılığını tahmin etmektir.

LR, girdi özelliklerini bir olasılık değerine dönüştürmek için lojistik bir fonksiyon kullanan ve çıktının 0 ila 1 aralığında olmasını sağlayan istatistiksel bir modeldir. Sınıflandırma eşiği genellikle 0,5 olarak belirlenir ve bunun üzerinde örnek bir sınıfa, altında ise diğer sınıfa atanır (Deepa R, 2024). Bu, girdinin pozitif sınıfa (örn. sahte haberler) ve negatif sınıfa (örn. gerçek haberler) ait olma olasılığını temsil eder. İkili sonuçları modellemek için istatistiksel analizde kullanılan bir tür genelleştirilmiş doğrusal modeldir. Optimizasyon süreci basittir ve hızlı model eğitimi ve çıkarımı sağlar. Bu, sahte haberleri tespit etmek ve yanlış bilginin yayılmasını azaltmak için çok önemlidir. Ancak LR, girdi özellikleri ile sonucun log-olasılığı arasında doğrusal bir ilişki olduğunu varsayar ve bu da gerçek dünya problemlerindeki doğrusal olmayan ilişkileri yakalayamayabilir. LR'un faydalarına rağmen, araştırmacılar ve uygulayıcılar onu gerçek dünya senaryolarına uygularken varsayımlarının ve sınırlamalarının farkında olmalıdır.

2.2.2. Karar ağacı sınıflandırıcı

Karar ağacı (DT) algoritması, denetimli öğrenme algoritmaları sınıfına aittir ve temel amacı, hedef değişkenlerin sınıfını veya değerini tahmin etmek için eğitim verilerinden türetilen karar kurallarını öğrenmek için kullanılabilecek bir eğitim modeli oluşturmaktır (Charbuty & Abdulazeez, 2021). Bir DT aslında çok sezgiseldir ve anlaşılması kolaydır. Temelde bir dizi evet/hayır ya da eğer/eğer sorularıdır. Aşağıda tipik bir DT şeması yer almaktadır. Kök düğümler ağaçtaki en üst ya da en alt düğümdür; çıplak düğümler akışın çeşitli isteğe bağlı akışlara dallandığı yerlerdir ve en önemlisi yaprak düğümler nihai çıktının bulunduğu son düğümlerdir.

DT Sınıflandırıcısı, minimum ön işleme ile hem sayısal hem de kategorik verileri işleyebilen çok yönlü bir algoritmadır. DT Sınıflandırıcısı, minimum ön işleme ile hem sayısal hem de kategorik verileri işleyebilen çok yönlü bir algoritmadır. Ayrıca yorumlanabilirliği ve özellik ölçeklendirme veya düzenlileştirmenin minimum etkisi ile bilinir. DT Sınıflandırıcısı, minimum ön işleme ile hem sayısal hem de kategorik verileri işleyebilen çok yönlü bir algoritmadır. Bununla birlikte, özellikle ağaç derinleştiğinde veya eğitim verileri gürültülü olduğunda aşırı uyuma karşı hassastır. Aşırı karmaşık ağaçlar sahte örüntüleri yakalayabilir ve bu da genelleme performansının düşük olmasına neden olur. Budama veya ağacın derinliğini sınırlama gibi teknikler, aşırı uyumu azaltmak için yaygın olarak kullanılır. Bu dezavantajlara rağmen, DT Sınıflandırıcısı basitliği, yorumlanabilirliği ve hem sınıflandırma hem de regresyon

görevlerini yerine getirmedeki etkinliği nedeniyle yaygın olarak kullanılmaya devam etmektedir. Sahte haber sınıflandırmasında, ayırt edici özellikleri belirlemek ve bilinçli kararlar almak için şeffaf bir çerçeve sağlar.

2.2.3. GBoost sınıflandırıcısı

GBoost algoritması denetimli öğrenme alanında çalışmaktadır. boost algoritması örnekleri ağırlıklandırır ve sınıflandırılması zor örnekler olarak bildirilen yanlış sınıflandırılmış örneklerin ağırlığını artırır (Nawar & Mouazen, 2017). Tahminler veya sınıflandırmalar yapmak için etiketli verilerden öğrenir. GBoost, güçlü bir tahmin modeli oluşturmak için genellikle karar ağaçları olmak üzere birden fazla zayıf öğrenicinin tahmin gücünü birleştiren bir topluluk öğrenme tekniğidir. Algoritma yinelemeli olarak çalışır, sonraki her model öncekiler tarafından yapılan hatalara odaklanır, böylece genel tahmin doğruluğunu artırır. Bu durumda, "sınıflandırılması zor" olarak belirtilen yanlış sınıflandırılmış verilere öncelik verilir. Yineleme, boosting tüm topluluk tahminlerinin ağırlıklı toplamını karşılayana kadar tekrarlanır (Aziz et al., 2020).

GBoost sınıflandırıcısı, kelime sıklığı, sözdizimsel kalıplar ve anlamsal ipuçları gibi metinden çıkarılan özelliklere dayalı olarak sahte haberleri sınıflandırmak için kullanılan bir algoritmadır. Bu yöntem, verilerdeki karmaşık, doğrusal olmayan ilişkileri ele almada etkilidir ve RF gibi diğer topluluk yöntemlerine kıyasla aşırı uyum sağlamaya daha az eğilimlidir. Bununla birlikte, GBoost sınıflandırıcısı, özellikle veri kümesi boyutu arttıkça yüksek hesaplama karmaşıklığına ve kaynak gereksinimlerine sahiptir. GBoost Sınıflandırıcısı gibi eğitim modelleri zaman alıcı ve hesaplama açısından zorlu olabilir. Bununla birlikte, GBoost Sınıflandırıcısı verilerdeki karmaşık örüntüleri tanımlamak için güçlü bir araçtır ve bu da onu gerçek ve sahte haber makalelerini ayırt etmek için ideal bir seçim haline getirir. GBoost Sınıflandırıcısını sahte haber sınıflandırması için kullanmak için, denetimli öğrenme içinde bir topluluk öğrenme tekniği olarak temellerini anlamak önemlidir.

2.2.4. Rastgele orman sınıflandırıcı

Rastgele orman (RF) genel tahmin performansını artırmak için birden fazla temel tahmin edicinin tahminlerini bir araya getiren bir topluluk öğrenme tekniğidir. Her biri eğitim veri kümesinden rastgele bir özellik alt kümesi kullanılarak oluşturulmuş birden fazla karar

ağacından oluşur. Eğitim süreci boyunca algoritma, bir durdurma kriteri karşılanana kadar bilgi kazancı veya safsızlık azaltma kriterlerine göre özellik uzayını özyinelemeli olarak bölümlere ayırarak ağaçları yinelemeli olarak büyütür.

Genel olarak, birden fazla sınıflandırıcının dahil edilmesi, özellikle kararsız sınıflandırıcılar söz konusu olduğunda varyansı azaltır ve daha güvenilir sonuçlar elde edilmesini sağlayabilir (Sheykhmousa et al., 2020). RF Sınıflandırıcı algoritmasının birincil avantajlarından biri, birden fazla ağacın tahminlerinin ortalamasını alarak makine öğreniminde yaygın bir tuzak olan aşırı uyumu azaltma yeteneğidir. RF modeli etkilidir çünkü nispeten ilişkisiz birçok modelden (ağaç) oluşan bir komite, herhangi bir bireysel kurucu modelden daha iyi performans gösterir (U. Sharma et al., 2020). RF Sınıflandırıcısı, gürültülü verilere ve aykırı değerlere karşı genelleme ve sağlamlığı artıran bir topluluk öğrenme modelidir. Geniş özellik uzaylarına sahip metin sınıflandırma görevleri için uygundur. Bununla birlikte, doğal karmaşıklığı nedeniyle yorumlanabilirlik eksikliği de dahil olmak üzere sınırlamaları vardır. Algoritmanın birleşik kararları daha iyi tahmin performansı ile sonuçlanır, ancak bireysel tahminlerin arkasındaki mantığı açıklamak zor olabilir. Ayrıca, RF Sınıflandırıcısı dengesiz sınıf dağılımlarına sahip veri kümeleri için uygun olmayabilir, çünkü çoğunluk sınıflarına öncelik verme eğilimindedir ve bu da önyargılı tahminlere neden olabilir. Bu sınırlamalara rağmen RF Sınıflandırıcısı, topluluk öğrenme ilkelerini kullanarak sahte haberleri sınıflandırmak için değerli bir araçtır.

2.2.5. XGBoost sınıflandırıcı

Bu sınıflandırma modeli bir tür denetimli öğrenmedir. XGBoost, öncelikle gradyan güçlendirme karar ağaçlarını kullanarak hızlı işleme ve performans için tasarlanmıştır (Dhaliwal et al., 2018). XGBoost, her aşamada zayıf öğreniciler oluşturarak metinsel verilerdeki kalıntıları düzelten oldukça paralelleştirilebilir bir algoritmadır. Performansı ağırlıklı bir toplamla belirlenir ve eksik verileri ele almada ve çeşitli amaç işlevlerini desteklemede çok yönlü ve etkilidir. Bununla birlikte, hiperparametre ayarı zaman alıcı ve hesaplama açısından pahalı olabilir. XGBoost'un sahte haber sınıflandırma görevleri gibi metin verileri üzerindeki performansı, özellik mühendisliği ve ön işleme için kullanılan tekniklere bağlı olarak değişebilir. Sahte haber sınıflandırmasında sınıflandırıcı, metinsel verilerden özellikler çıkarabilir ve etiketli sahte ve gerçek haber örnekleri üzerinde eğitilebilir. XGBoost

sınıflandırıcısı, topluluk öğrenme tekniklerini kullanarak sahte haberleri tanımlamak için sağlam bir çerçeve sunar ve bu da onu bu görev için cazip bir seçim haline getirir.

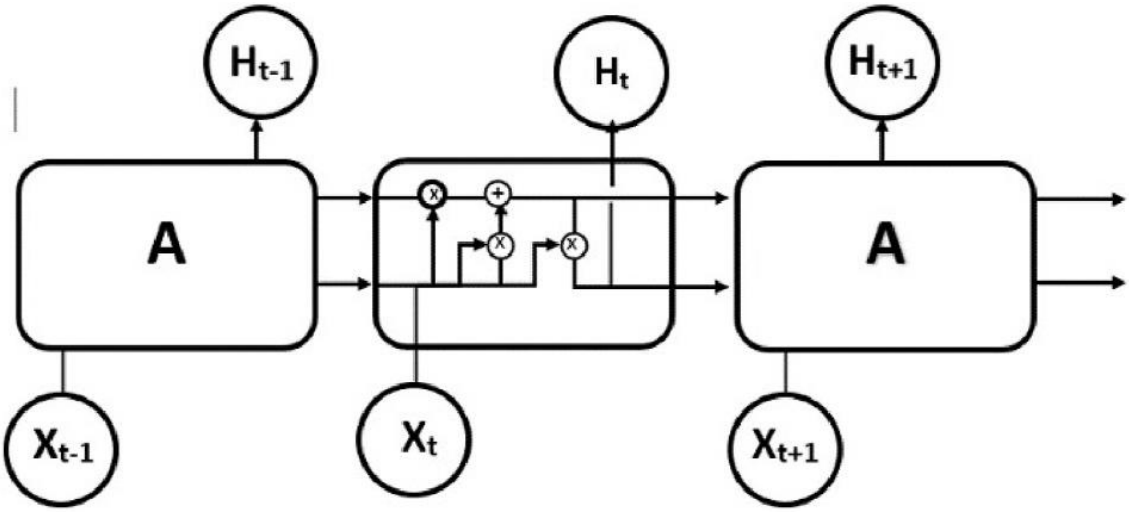
2.3. Derin Öğrenme Teorik Temel

DL'in teorik temeli, insan beyninin yapısal ve işlevsel organizasyonundan ilham alan yapay sinir ağları kavramına dayanmaktadır. DL, her katmanın girdi verilerini bir dizi ağırlıklı bağlantı ve doğrusal olmayan aktivasyon fonksiyonları aracılığıyla dönüştürdüğü çok sayıda birbirine bağlı düğüm veya nörondan oluşur. Bu katmanlar sıklıkla üç ana kategoride sınıflandırılır (girdi, gizli katmanlar ve çıktı katmanları). Özellikle, gizli katmanların derin ağlara dahil edilmesi, modelin girdi verilerinin karmaşık hiyerarşik temsillerini öğrenmesini sağlar ve böylece daha basit modeller tarafından gözden kaçabilecek ince desenlerin ve korelasyonların tespit edilmesini kolaylaştırır. Eğitim prosedürü, geriye yayılma ve gradyan inişi gibi tekniklerin kullanılması yoluyla bu ağırlıkların ayarlanmasını içerir. Bu süreçte ağ, beklenen ve gerçek sonuçlar arasındaki tutarsızlığı azaltmak amacıyla ağırlıklarda yinelemeli ayarlamalar yapar. DL'nin karmaşık, doğrusal olmayan ilişkileri temsil etme kapasitesi, onu özellikle görüntü ve ses tanıma, doğal dil işleme ve geleneksel makine öğrenimi tekniklerinin yetersiz kalabileceği diğer alanlarda avantajlı hale getirmektedir.

2.3.1. Uzun kısa süreli bellek

Uzun kısa süreli bellek (LSTM) tekrarlı sinir ağlarının (RNN) gelişmiş bir çeşididir (Shiri et al., 2023). Her veri noktası ile komşuları arasındaki zamansal ilişki önemlidir. LSTM, sıralı verilerdeki uzun vadeli bağımlılıkları yakalayabilir ve kaybolan gradyan probleminin neden geleneksel RNN'nin olumsuzluğunu önleybilir.

Bir LSTM bilgi akışını kontrol etmek için üç kapı kullanır: giriş, unutma ve çıkış. Giriş kapısı, mevcut girişe ve önceki duruma bağlı olarak dahili durumu değiştirir. Unutma kapısı, önceki durumun ne kadarının saklanacağına veya silineceğine karar verir. Çıkış kapısı, dahili durumun sistemi nasıl etkilediğini belirler (Fang et al., 2021). LSTM, ağ üzerinden bilgi akışını düzenlemek için geçit mekanizmaları ve zaman içinde hangi bilginin saklanacağına veya atılacağına karar vermek için sigmoid ve tanh aktivasyon fonksiyonları kullanır. Şekil 4'te bir LSTM'nin iç yapısını nasıl güncellediğini göstermektedir.



Şekil 4. LSTM mimarisi (Sudhakar & Kaliyamurthie, 2024).

Her zaman adımında, LSTM birimi mevcut zaman adımındaki girdiyi ve önceki zaman adımındaki çıktıyı alarak bir sonraki zaman adımına aktarılan bir çıktı oluşturur. LSTM ağları, hücre durumunu düzenlemek için bir geçit mekanizması kullanır, önemli bilgileri tutar ve alakasız bilgileri atar.

LSTM ağları, verilerin sıralı doğasını dikkate alarak ağ parametrelerini güncellemek için gradyan inişi ve geriye yayılma ilkelerini uygulayan Zaman İçinde Geriye Yayılma (BPTT) kullanılarak eğitilir. Metin halihazırda açık, özlü ve ifadeler arasında nedensel bağlantılar ile mantıksal bir yapı izlemektedir. Teknik terim kısaltmaları ilk kullanıldıklarında açıklanmıştır ve dil açık, nesnel ve değerden bağımsızdır. Metin stil kılavuzlarına uymakta, tutarlı atıflar kullanmakta ve tutarlı bir dipnot stili ve biçimlendirme özellikleri izlemektedir. Dil resmidir; kısaltmalardan, gündelik kelimelerden, gayri resmi ifadelerden ve gereksiz jargondan kaçınılır. Metinde dilbilgisi hataları, yazım hataları ve noktalama işaretleri bulunmamaktadır. Geliştirilmiş metnin içeriği kaynak metne mümkün olduğunca yakındır ve başka hiçbir unsur eklenmemiştir. Geliştirilmiş metnin içeriği kaynak metne mümkün olduğunca yakın ve başka bir unsur eklenmemiş.

LSTM, girdi verilerinin sıralı olduğu NLP, konuşma tanıma, zaman serisi tahmini ve duygu analizi gibi görevlerde sıklıkla kullanılır. Verilerin zamansal veya sıralı bir yapıya sahip olduğu

durumlarda özellikle kullanışlıdırlar. LSTM'lerin her tür veri için uygun olmadığını ve dikkatli kullanılması gerektiğini belirtmek önemlidir.

2.3.2. Evrişimsel sinir ağları

Evrişimsel Sinir Ağları (CNN) görüntüler gibi ızgara benzeri verileri işlemek için özel olarak tasarlanmış bir sinir ağı türüdür. Giriş verilerinin hiyerarşik temsillerini çıkarmak için konvolüsyonel ve havuzlama katmanları kullanırlar. CNN hem iyi özellik oluşturma hem de ayırt etme yeteneklerine sahiptir, bu da onları tipik ML sistemlerinde özellik oluşturma ve sınıflandırma için ideal hale getirir (A. Khan et al., 2020). CNN'lerin arkasındaki ana kavramlar şunlardır:

- Konvolüsyon işlemi, giriş verilerine bir filtre (çekirdek) uygular ve eleman bazında çarpma ve toplama yaparak özellikleri çıkarır. Evrişim katmanları, giriş verilerinin farklı özelliklerini yakalamak için birden fazla filtre kullanır.
- Havuzlama katmanları, konvolüsyonel katmanlar tarafından üretilen özellik haritalarını küçültmek için kullanılır ve önemli özellikleri korurken verilerin uzamsal boyutlarını azaltır. Maksimum havuzlama ve ortalama havuzlama tipik havuzlama işlemleridir.
- Özellik hiyerarşisi bir diğer önemli husustur. CNN girdi verilerinin hiyerarşik temsillerini öğrenir. Alt katmanlar kenarlar ve dokular gibi düşük seviyeli özellikleri yakalarken, üst katmanlar görevle ilgili daha soyut ve karmaşık özellikleri yakalar.
- CNN, giriş verilerinin farklı uzamsal konumlarında aynı filtre setinin uygulandığı ağırlık paylaşımından yararlanır. Bu özellik, CNN'lerinin çeviri değişmezliği elde etmesini sağlar, yani giriş verilerindeki konumlarından bağımsız olarak aynı özellikleri tanıyabilirler.
- CNN, konvolüsyonel bir katmandaki her nöronun girdi verilerinin küçük bir bölgesine bağlandığı yerel alıcı alanları kullanır. Bu yerel bağlantı, CNN'in girdi verilerindeki uzamsal bağımlılıkları verimli bir şekilde yakalamasına yardımcı olur.

2.4. Doğal Dil İşleme Teorik Temel

Doğal dil işleme (NLP), çeşitli disiplinlerden teoriler ve metodolojiler kullanır. Dilbilim, insan dilinin yapısını, gramerini ve anlamını analiz etmek için çerçeveler sağlayarak NLP'de merkezi bir rol oynar. Bu dilbilimsel anlayışlar, sözdizimsel ayrıştırma, anlamsal analiz ve söylem işleme gibi temel NLP görevlerini mümkün kılar. Olasılık teorisi ve istatistiksel yöntemler de

NLP'si için çok önemlidir, çünkü belirsizliğin modellenmesine ve büyük miktarda dilsel verideki kalıpların tanımlanmasına izin verirler. Gizli Markov modelleri, koşullu rastgele alanlar ve n-gram modelleri, konuşma parçası etiketleme, adlandırılmış varlık tanıma ve makine çevirisi gibi NLP görevlerinde yaygın olarak kullanılmaktadır. ML teknikleri, büyük veri kümeleri üzerinde modelleri eğitmek ve dil hakkında tahminler yapmak için verimli algoritmalar sağladığından NLP için çok önemlidir. ML, SVM ve GaussianNB gibi klasik tekniklerden sinir ağları ve dönüştürücülerle DL gibi modern gelişmelere kadar NLP'sini dönüştürmüştür. Bu durum duygu analizi, soru yanıtlama ve metin oluşturma gibi görevlerde çığır açılmasını sağlamıştır. Ayrıca, bilişsel bilim ve psikodilbilimden elde edilen bilgiler, insan benzeri anlayış ve etkileşimi taklit edebilen NLP sistemlerinin geliştirilmesini sağlamaktadır. Dilbilim, olasılık teorisi, ML ve bilişsel bilimi kapsayan bu teorik temeller ve metodolojiler, NLP'sinde yenilik ve ilerlemeyi teşvik etmek için birlikte çalışarak daha sofistike dil teknolojileri yaratır.

2.4.1. Destek vektör makinesi

Destek vektör makinesi (SVM), verileri sınıflandırmak için kullanılan güçlü bir denetimli öğrenme algoritmasıdır. Sahte haber tespiti bağlamında, SVM, hiper düzlem ile destek vektörleri olarak bilinen en yakın veri noktaları arasındaki mesafe olan marjı maksimize ederek farklı sınıfların veri noktalarını en iyi şekilde ayıran hiper düzlemi bulur. SVM, uygun çekirdek fonksiyonları kullanarak hem doğrusal olarak ayrılabilen hem de doğrusal olarak ayrılamayan verileri sınıflandırmak için etkilidir.

Sahte haber sınıflandırması için SVM kullanma süreci birkaç adımdan oluşmaktadır. İlk olarak, haber makalelerinden çıkarılan metinsel veriler, tokenizasyon, durak kelimesi kaldırma ve stemming gibi NLP teknikleri kullanılarak ön işleme tabi tutulur. Özellik vektörleri daha sonra kelime torbası veya Terim Frekansı-Ters Belge Frekansı (TF-IDF) gibi teknikler kullanılarak oluşturulur. Bu vektörler haber makalelerinin anlamsal içeriğini temsil eder. Word2Vec, sürekli bir vektör uzayında kelimelerin yoğun vektör temsilleri olan kelime katıştırmaları oluşturmak için kullanılan popüler bir NLP tekniğidir. Bu katıştırmalar kelimeler arasındaki anlamsal ilişkileri yakalayarak makinelerin dili daha etkili bir şekilde anlamasını sağlar. SVM algoritması, çıkarılan özelliklere dayalı olarak sahte ve gerçek haber makalelerini ayırt eden en uygun hiper düzlemi öğrenmek için etiketli bir veri kümesi üzerinde eğitilir.

SVM, yüksek boyutlu özellik uzaylarını verimli bir şekilde ele alması nedeniyle sahte haber sınıflandırması için avantajlıdır. Çok sayıda özelliğe sahip görevler için uygundur (yani, metindeki kelimeler veya ifadeler). SVM, çekirdek fonksiyonlarının kullanımı yoluyla doğrusal olmayan karar sınırlarını ele almada da etkilidir ve verilerdeki karmaşık ilişkileri yakalamasına olanak tanır.

Bununla birlikte, dezavantajları da vardır. Büyük veri kümeleriyle uğraşırken, SVM için eğitim süresi veri noktalarının sayısıyla kuadratik olarak ölçeklenir ve bu da onu hesaplama açısından pahalı hale getirir. Ek olarak, SVM'nin performansı, en iyi sonuçlar için dikkatli bir ayarlama gerektirebilecek düzenleme parametresi ve çekirdek türü gibi parametrelerin seçimine duyarlıdır. Optimum performansı sağlamak için bu faktörleri dikkatle değerlendirmek önemlidir. Ayrıca, bir sınıfın diğerine kıyasla önemli ölçüde az temsil edildiği dengesiz veri kümeleri, SVM için zorluk teşkil edebilir.

Bu diyagram, sahte haberlerin sınıflandırılması için SVM'nin çalışma yapısını göstermektedir. Süreç veri ön işleme ile başlar, ardından özellik çıkarma ve vektörleştirme yapılır. SVM algoritması daha sonra karar sınırını öğrenmek için etiketli veri kümesi üzerinde eğitilir. Test sırasında, yeni makaleler öğrenilen modele göre sahte veya gerçek olarak sınıflandırılmadan önce ön işleme tabi tutulur ve özellik vektörlerine dönüştürülür.

2.4.2. K-en yakın komşu

K-en yakın komşu (KNN) özellik uzayı içinde benzerlik metrikleri prensibiyle çalışan bu tür bir algoritmadır. Örnekler, en yakın komşularının fikir birliğine dayalı olarak sınıflandırılır. K-NN'nin teorik temelleri, doğrusal olmayan karar sınırlarını ele almadaki basitliğine ve etkinliğine dayanmaktadır ve bu da onu sahte haberlerle mücadelede uygun bir seçenek haline getirmektedir.

Bu nedenle, teknik terimlerin kullanımı tutarlı olmalı ve ilk kullanıldığında açıklanmalıdır. Algoritma, haber makalelerinden çıkarılan metinsel özellikleri ve bunların gerçekliğini gösteren ilgili etiketleri içeren bir veri kümesi alır. Metin dilbilgisi hataları, yazım hataları ve noktalama işaretlerinden arındırılmış olmalıdır. K-NN, yeni bir örnek ile eğitim setindeki tüm örnekler arasındaki benzerliği, tipik olarak Öklid veya kosinüs mesafesi gibi seçilen bir mesafe

metriği kullanarak hesaplayarak çalışan basit bir algoritmadır. Teknik bilgileri sunarken açık ve mantıklı bir yapının korunması önemlidir. Ayrıca, kullanılan dil açık, nesnel ve değerden bağımsız olmalı, önyargılı, duygusal, mecazi veya süslü dilden kaçınılmalıdır. Metne yeni içerik eklemekten kaçınmak ve orijinal anlamı mümkün olduğunca korumak önemlidir. Giriş örneği için K-NN daha sonra hesaplanan mesafelere göre belirlenir.

K-NN 'nin avantajları, sezgisel uygulamasında ve veri kümelerinin değişen karmaşıklıklarına uyum sağlama yeteneğinde yatmaktadır. Basitliği, hızlı prototip oluşturma ve denemeye olanak tanıyarak sahte haber sınıflandırmasında ilk keşifler için özellikle cazip hale getirir. Ayrıca, K-NN veri dağılımı konusunda hiçbir varsayımda bulunmaz, bu da onu gürültülü veya doğrusal olarak ayrılamayan verilere karşı dayanıklı hale getirir.

Ancak, K-NN'nin belirli sınırlamaları vardır. Algoritmanın her tahmin için mesafeleri yeniden hesaplaması gerektiğinden, hesaplama yükü daha büyük veri kümeleriyle artar. Ayrıca, K-NN'nin performansı büyük ölçüde sınıflandırma sırasında dikkate alınan komşu sayısını temsil eden hiperparametre 'K' seçimine bağlıdır. Uygun olmayan bir 'K' seçimi, algoritmanın etkinliğini tehlikeye atarak aşırı uyuma veya yetersiz uyuma yol açabilir.

K-NN'nin teorik temelleri, onu sahte haber sınıflandırma alanında kullanışlı kılmaktadır. Bununla birlikte, hesaplama talepleri ve hiperparametreler konusundaki hassasiyeti dikkatli bir uygulama ve iyileştirme gerektirmektedir. K-NN, günümüzün dijital ortamında yanlış bilginin yayılmasını belirlemeye ve azaltmaya yönelik kapsamlı bir yaklaşımın sadece bir parçasıdır.

2.4.3. Gaussian naive bayes

Gaussian Naive Bayes (GaussianNB) algoritması ile birlikte NLP teknikleri, verimlilikleri ve etkinlikleri nedeniyle sahte haber sınıflandırması alanında büyük ilgi görmüştür. GaussianNB sınıflandırıcısı, gözlemlenen bazı kanıtlar göz önüne alındığında bir hipotezin olasılığını tanımlayan Bayes teoremine dayanmaktadır. Haberlerin sahte veya gerçek olarak sınıflandırılması bağlamında, hipotez, haber içeriğinden çıkarılan çeşitli dilsel özelliklere göre değerlendirilir.

NLP GaussianNB algoritması birkaç temel adım içerir. İlk olarak, algoritma metni tokenize ederek, durak kelimeleri kaldırarak ve muhtemelen kelimeleri normalleştirmek için stemming veya lemmatization gerçekleştirerek metinsel verileri ön işleme tabi tutar. Bu ön işleme adımı özellik uzayı boyutluluğunu azaltır ve sınıflandırıcı verimliliğini artırır. Ön işlemden sonra algoritma, eğitim verilerindeki kelime veya özellik frekanslarına dayalı olasılıksal bir model oluşturur. Sınıf etiketleri (sahte veya gerçek haber) göz önüne alındığında kelime dağarcığındaki her bir kelimenin olasılığını hesaplar. Bu adım, özelliklerin (kelimelerin) koşullu olarak bağımsız olduğunu varsayar; bu da 'naif' varsayım olarak bilinen basitleştirici bir varsayımdır. Basitliğine rağmen, bu varsayım pratikte genellikle iyi çalışır ve algoritmanın verimli bir şekilde ölçeklenmesini sağlar.

Olasılıksal model eğitildikten sonra sınıflandırıcı, gözlemlenen özellikler göz önüne alındığında her bir sınıfın sonsal olasılığını hesaplamak için Bayes teoremini kullanır. En yüksek sonsal olasılığa sahip sınıf daha sonra girdi haber makalesine atanır. Başka bir deyişle, algoritma bir haberin sahte veya gerçek olma olasılığını içerdiği kelimelere dayanarak tahmin eder ve sınıf etiketini buna göre atar.

NLP GaussianNB algoritması, basitliği, ölçeklenebilirliği ve hesaplama verimliliği gibi çeşitli avantajlara sahiptir. Özellikle yüksek boyutlu özellik uzaylarına ve nispeten küçük eğitim veri kümelerine sahip metin sınıflandırma görevleri için kullanışlıdır. Dahası, algoritma genellikle sınırlı eğitim verileriyle bile iyi performans gösterir, bu da onu etiketli verilerin az olduğu senaryolar için uygun hale getirir.

Bununla birlikte, belirli sınırlamaları da vardır. Özellik bağımsızlığı varsayımı, özellikle kelimelerin genellikle karmaşık ilişkilere ve bağımlılıklara sahip olduğu doğal dil metinlerinde her zaman geçerli olmayabilir. Bu durum, özellikle son derece nüanslı veya bağlama bağlı dille uğraşırken optimumun altında bir performansa neden olabilir. Ayrıca algoritma, eğitim verilerinde iyi temsil edilmeyen kelime dağarcığı dışındaki kelimeler veya nadir kelimelerle ilgili zorluklarla karşılaşabilir ve bu da yanlış sınıflandırma hatalarına yol açabilir.

Özetle, NLP'nin GaussianNB algoritması sahte haberlerin sınıflandırılması için basit ve etkili bir yaklaşım sunmaktadır. Bununla birlikte, performansı eğitim verilerinin kalitesi ve temsil gücünün yanı sıra naif bağımsızlık varsayımının sınırlamalarından etkilenir.

2.5. Sahte Haber Tespitine Yönelik Çalışmalar

Yaygın sahte haber sorununu ele alan araştırmacılar, gerçek ve yanıltıcı bilgileri ayırt etmek için AI algoritmalarına yönelmişlerdir. Birçok çalışma, internet üzerinden sosyal ağlarda sahte haberlerin tespitini araştırmıştır, ancak yalnızca sınırlı sayıda kişi AI tekniklerini kullanarak sahte haberlerin otomatik olarak tespit edilmesine odaklanmıştır (Aïmeur et al., 2023). Bu literatür taraması, gerçek ve sahte haberlerin sınıflandırılmasına ilişkin bilimsel araştırmalarda ortaya çıkan çeşitli metodolojileri ve bulguları incelemektedir. Mevcut araştırmaların sentezi, yalnızca kullanılan çeşitli stratejileri aydınlatmayı değil, aynı zamanda boşlukları tespit etmeyi ve araştırma yaklaşımının geliştirilmesi için zemin hazırlamayı amaçlamaktadır.

AI algoritmalarının etkinliğini ortaya koyan çeşitli çalışmalar, sahte haber tespit sistemlerinin ilerlemesine önemli ölçüde katkıda bulunmuştur. Bir çalışmada yazarlar hibrit bir DL modeli ile üretken rakip ağlar tarafından desteklenen CNN ve RNN'yi entegre ederek %93.87 gibi etkileyici bir doğruluk oranına ulaşmıştır. (Hanshal et al., 2023). Benzer şekilde başka bir çalışmada, n-gram analizi ve ML algoritmalarını kullanarak Rastgele Arama Cross-Validation (CV) ile ince ayar yapmanın doğruluğu %94.2'ye çıkardığını gösterilmiştir (John & Journals, 2022). Bir başka çalışmada SVM nin GaussianNB ve NLP ile entegrasyonu sağlanarak özellikle Hindistan gibi bölgelerde etkili olan %93.6'lık dikkate değer bir doğruluk elde edilmiştir (Jain et al., 2019).

Sosyal medya alanında, bir çalışmada beş katlı çapraz doğrulamada ortalama %82.55 puan elde etmek için CNN-LSTM-SVM'yi birleştiren hibrit bir model sunulmuştur (Melvern et al., 2023). Bir başka çalışmada araştırmacılar, Word2vec katıştırmaları ile birlikte CNN ve RNN- LSTM kullanarak İngilizce ve Türkçe dil sistemlerinde yüksek doğruluk oranları elde etmişlerdir. (Güler & Gündüz, 2023). Başka bir araştırmada, hibrit bir CNN-RNN modeli önerilerek sahtekarlıkları tespit etmede %95,12'lik bir doğruluk elde edilmiştir (Asta & Budi Setiawan, 2023). Bir diğer çalışmada topluluk tabanlı bir DL çözümü önerilmiştir. Strateji iki temel model

kullanmıştır: metinsel analiz için bir Bi- LSTM -GRU-yoğun model ve LIAR veri setine dayalı olarak ek nitelikler için yoğun bir DL modeli. Çalışmada, 91,6% geri çağırma, 91,3% hassasiyet ve 91,4% F1-puanı ile 89,8% doğruluk elde edilmiştir (Aslam et al., 2021).

Dikkat tabanlı sinir ağlarını inceleyen araştırmacılar, kelime düzeyinde bellek ağı, geliştirme ve test setlerinde sırasıyla %8,4 ve %4,9 oranında diğer modellerden daha iyi performans gösterdiklerini ifade etmişlerdir (Trung Tin, 2018). Bir başka çalışmada araştırmacılar, Hong Kong protestoları sırasında Twitter'da sahte haberlerin yayılmasını incelemiş, otomatik kararlarda açıklanabilirliğin önemini vurgulayarak ve XLM-RoBERTa ile %99,3'e varan F1 puanı elde etmişlerdir (Zervopoulos et al., 2022).

WELFake modeli, sahte haberleri tespit etmede %96,73 gibi yüksek bir doğruluk oranına ulaşarak BERT ve CNN modelleri sırasıyla %1,31 ve %4,25 oranında geride bırakılmıştır. Dilsel özellikleri kelime gömme ile birleştirerek ve en doğru ML modeli olarak SVM'yi kullanıp, özellik seçimi ve topluluk oylama sınıflandırmasının etkinliğini göstermiştir (Verma et al., 2021).

Yanlış bilgilerin belirlenmesi farklı dilsel bağlamlarda incelenmiştir. Örneğin, Arapça sahte haberleri belirlemek için CNN mimarisi ile MARBERT kullanarak bir DL modeli geliştirilmiştir. Bu algoritma ile diğer yöntemleri geride bırakarak kayda değer bir doğruluk ve 95,6% F1 puanı elde edilmiştir (Alyoubi et al., 2023). Ayrıca, araştırmacılar sahte haber tespiti için 19 ML yaklaşımını karşılaştıran bir kıyaslama çalışması yürütmüştür. Bu yaklaşımlar arasında geleneksel öğrenme modelleri, geleneksel DL modelleri ve BERT gibi önceden eğitilmiş gelişmiş dil modelleri yer almıştır. Önceden eğitilmiş BERT tabanlı modeller, özellikle RoBERTa, en yüksek performansı göstererek birleştirilmiş derlem için 96% doğruluk oranlarına ulaşmıştır. N-gramlı Naïve Bayes de, özellikle yeterince büyük veri kümelerinde, sinir ağı tabanlı modellerle karşılaştırılabilir performansla umut verici sonuçlar göstermiştir (J. Y. Khan et al., 2021).

Yeni medya çağında sınırlı kaynaklara sahip bir dil olan Amharca'da sahte haber yayma sorununun üstesinden gelmeyi amaçlayan başka bir çalışmada araştırmacılar, Facebook'tan 12.000 açıklamalı haber makalesinden oluşan yeni bir veri seti elde ederek DL algoritmaları

GRU ve CNN uygulayıp sahte haberleri otomatik olarak tespit etmişlerdir. Çalışmada CNN modeli ile %93,92 doğruluk ve %94 f1 puanı elde edilmiştir (Hailemichael, 2021).

Başka bir çalışmada, ISOT veri seti kullanılarak SVM, Naive Bayes, RNN ve CNN dahil olmak üzere çeşitli yanlış haber tespit algoritmalarının performansı değerlendirilmiştir. CNN, doğrulama kümesinde %92 doğruluk, %90 kesinlik, %88 geri çağırma ve %90 F1 puanı ile en iyi sonuçları elde edilmiştir (Thiyagarajan S, Abinaya M, Viniyha G, 2024). Başka bir çalışmasında, boyutsallığı azaltmak için CNN, LSTM ve Ki-Kare'yi birleştiren hibrit bir Sinir Ağı mimarisi kullanılmış ve sahte haber tespitinde %97,8 doğruluk elde edilmiştir (Umer et al., 2020).

Farklı bir yaklaşımla, iki yönlü LSTM, meta veri ve söylenti yolu yayılımını kullanan hibrit bir model önererek %94,3'lük bir doğruluk elde edilmiştir. Bu model, çeşitli özelliklere dayalı olarak gerçek ve sahte haberleri ayırt etmek için çağdaş çözümlerden daha iyi performans göstermektedir (Mjaaland et al., 2020).

Başka bir çalışmada, LIWC aracı kullanılarak çıkarılan algoritmaları ve metin özelliklerini içeren birleşik bir ML yaklaşımı araştırılmış ve RF ve Perez-LSVM ile DS1 üzerinde %99 gibi etkileyici bir doğruluk elde edilmiştir (Ahmad et al., 2020).

Bir çalışmada Twitter'daki "2010 Şili Depremi" veri setini kullanarak doğruluğu artırmak için özellik mühendisliği kullanılmıştır. Çalışmada SVM ve Naïve Bayes üstün algoritmalar olarak ortaya çıkmakta ve Twitter dizilerindeki sahte haberlerin belirlenmesinde Count Vectors ve TF-IDF kullanılarak metinsel verilerin işlenmesinin önemini vurgulanmaktadır (Abdullah-All-Tanvir et al., 2019).

Bu başka araştırmada ise sahte haber tespiti için denetimli öğrenme algoritmalarını kullanan geleneksel ML modellerini kullanılarak XGBoost algoritması %75'in üzerinde en yüksek doğruluğu gösterirken, onu yaklaşık %73 doğrulukla SVM ve RF takip etmektedir (Khanam et al., 2021).

Sina Weibo hakkındaki yanlış söylentileri otomatik olarak tespit etmek için grafik çekirdeğini temel alan yeni bir hibrit SVM sınıflandırıcısını kullanana araştırmacıların modeli %91,3 doğruluk oranına ulaşmıştır. Özellikle sınıflandırma yöntemi, asılsız söylentilerin yayılmasından sonraki 24 saat içinde tespit etme konusunda %88 güven göstermiş, bunun da erken tespitte umut verici sonuçlara işaret ettiği ifade edilmiştir (Wu et al., 2015).

Bir başka araştırmacılar, dilsel özellik çıkarımı ve sıralı sinir ağı modelleri ile sözdizimsel, dilbilgisel, duygusal ve okunabilirlik özelliklerini kullanarak sahte haberlerin tespitini ele almaktadır. Birleştirilmiş dilsel özellik tabanlı sıralı sinir ağı ile %86 sınıflandırma doğruluğu elde edilmiştir. Bu yaklaşım, ML algoritmalarını kullanan ilk denemelerden daha iyi performans göstermiş ve topluluk yöntemleri kullanılarak maksimum %72 doğruluk elde edilmiştir (Choudhary & Arora, 2021).

Bir diğer çalışma, stilometrik analiz ve metin tabanlı kelime vektör temsilleri kullanarak yalnızca metinsel özelliklere odaklanarak sosyal medyada sahte haber tespitini araştırmaktadır. Boosting gibi topluluk yöntemleri en umut verici sonuçların sağladığı gösterilmiştir. Stilometrik ve kelime vektör özelliklerinin kombinasyonu, %95,49'a varan doğruluk oranıyla etkili olduğunu kanıtlarak bu yöntemlerin meta verilere veya kullanıcıyla ilgili bilgilere dayanmadan yüksek doğruluk elde etmedeki önemini ortaya koymaktadır (Reddy et al., 2020).

Literatürün bu kapsamlı incelemesi, sahte haberleri tespit etmek için kullanılan metodolojilerin zenginliğini ve çeşitliliğini vurgulayarak bu önemli alanda araştırma geliştirmek için sağlam bir temel sağlamaktadır. AI algoritmaları, dilbilimsel analiz ve yenilikçi modellerin bir araya gelmesi, sahte haber tespitinin gelişen manzarasını göstermekte ve daha fazla keşif ve iyileştirmeye davet etmektedir.

3. YÖNTEM

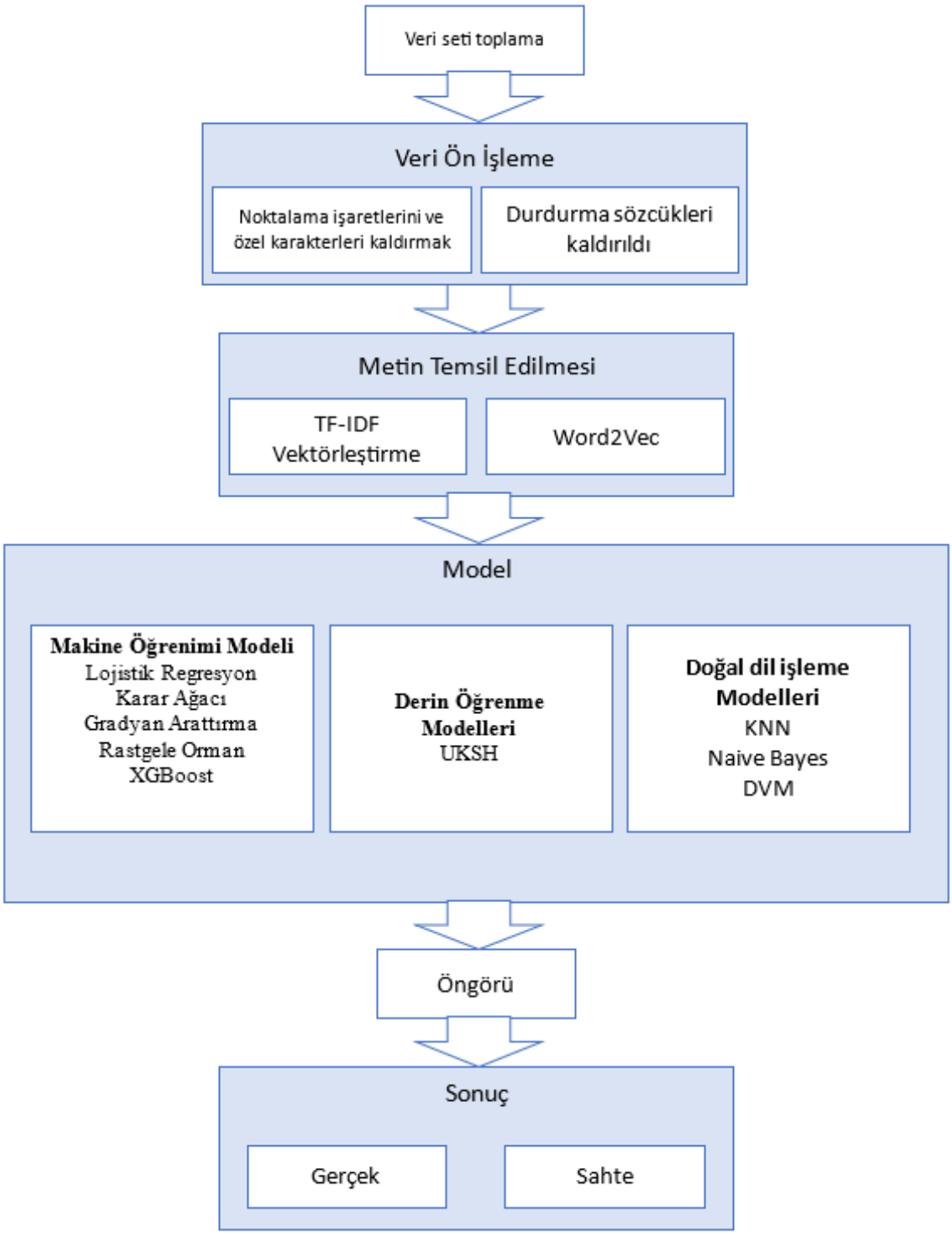
Araştırmada, gerçek ve hayali haber makalelerinden oluşan bir külliyatı analiz etmek için bir dizi AI modeli kullanılmış ve modellerin performansını artırmak için ön işleme teknikleri uygulanmıştır. Veri kümesi, Scikit-learn'in kelime alaka düzeyinin ölçülmesini kolaylaştıran TF-IDF Vectorizer kullanılarak ön işleme tabi tutulmuştur. Gensim Word2Vec modeli, anlamsal bağlantıların yakalanmasını kolaylaştıran kelime katıştırmalarının oluşturulması için kullanılmıştır. CountVectorizer, ham metin verilerini her bir belgedeki kelime sıklıklarını temsil eden sayısal bir formata dönüştürmek için kullanılmıştır. Veri seti, LR, DT, GBoost, RF, XGBoost, LSTM, SVM, K-NN ve GaussNB dahil olmak üzere dokuz farklı AI modeli kullanılarak değerlendirildi. Veri seti, 80% eğitim, 20% doğrulama olarak belirlenmiştir. Her modelin etkinliği değerlendirilerek en etkili modelin belirlenmesi kolaylaştırılmıştır. Deneysel tasarımda, modellerin etiketli veriler kullanılarak eğitildiği ve doğrulandığı denetimli bir öğrenme yaklaşımı kullanılmıştır. Hiperparametrelerin ayarlanması bir ızgara arama ve çapraz doğrulama tekniği kullanılarak gerçekleştirilmiştir. Kullanılan performans ölçütleri, en doğru modeli belirlemek amacıyla doğruluk ve F1 puanı dahil edilmiştir. Bu bölüm, ön işleme, model eğitimi ve değerlendirme aşamalarında üstlenilen prosedürleri tanımlayarak çalışmada kullanılan metodolojinin kapsamlı bir açıklamasını sunmaktadır. Amaç, benzer çabalar içinde olan gelecekteki araştırmacılar için açıklık ve tekrarlanabilirliği kolaylaştırmaktır. Şekil 5'te çalışmaya ait işlem basamakları gösterilmiştir.

3.1. Kullanılan Yazılım Araçları

Anaconda Navigator: Anaconda, paket ve ortam yönetimini daha az zahmetli hale getiren veri bilimi için bir paket dağıtımıdır. Farklı ortamlar oluşturmaya yardımcı olduğu için, farklı python ve paket sürümleri ile eğitim basit hale gelmektedir.

Spyder: Bilimsel Python Geliştirme IDE'si olarak da bilinen Spyder, bilim insanları, veri analistleri ve mühendisler için son derece özelleştirilebilir ve etkileşimli bir açık kaynaklı çapraz platform IDE olarak hizmet vermektedir.

Colab: Google'ın bulut üzerinde çalışan Jupyter notebook ortamı Colab, kaynak ve zaman tüketen DL kodlarının verimli bir şekilde yürütülmesini kolaylaştırmıştır. Açık paketi yüklemeye gerek kalmadan ücretli GPU'ları ve TPU'ları kullanmak için abone olundu.



Şekil 5. Çalışmanın Modeli

Python Kullanımı: Python, kullanım kolaylığı ve hızlı bir şekilde deney yapma yeteneği nedeniyle son yıllarda popülerlik kazanmıştır. Basitliği özellikle araştırma topluluğu için faydalıdır ve araştırmacıların sofistike terminoloji ve dokümantasyon öğrenmek yerine hedeflerine odaklanmalarına olanak tanır. Python'un araştırmacılar ve programcılar arasındaki popülerliği, yorumlanabilirliği ve daha yüksek üretkenliğinden kaynaklanmaktadır. Buna ek olarak, yüzlerce kararlı üçüncü taraf kütüphanesinin ve canlı bir topluluğun varlığı, onu çalışmak için mükemmel bir dil haline getirmektedir. Bu çalışma için Python'un kararlı sürümü (3.7) kullanılmıştır.

3.2. Kullanılan Donanım Araçları

Deneyler Intel® Core™ i5-5200U CPU @2.20GHz, 8 gigabayt fiziksel bellek, GPU: NVIDIA GeForce GTX 950M, 2 gigabayt RAM, 1 gigabayt sabit disk depolama kapasitesi ve 64 bit Windows 10 Pro işletim sistemi ile donatılmış bir kişisel bilgisayar kullanılarak gerçekleştirilmiştir. Bu donanım özellikleri, deneysel prosedürler için verimli yürütme ve güvenilir performans sağlamıştır.

3.3. Veri Kümesi

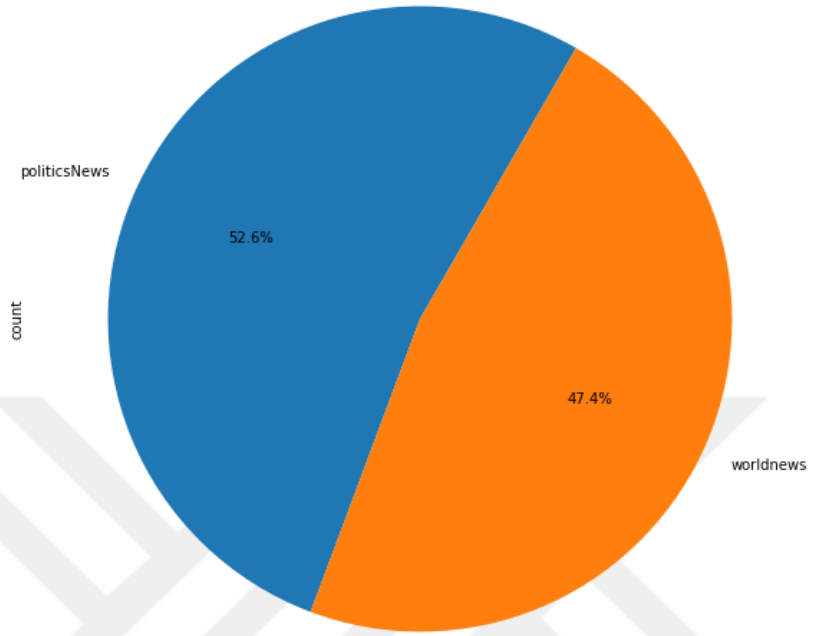
Bu araştırmada, Kaggle kaynak bağlantısından elde edilen iki farklı dosyadan oluşan kapsamlı bir veri seti kullanılmıştır (Kaggle, 2022)

ISOT veri kümesi, 5 Nisan 2022 tarihinde Reuters.com'dan alınan 44.898 haber ifadesinden oluşmaktadır. Gerçek haberler güvenilir kaynaklardan toplanırken, yanıltıcı bilgiler güvenilmez kaynaklardan elde edilmiştir. Bunlardan 21.417'si gerçek, 23.481'i ise sahtedir (Ali et al., 2022). ISOT veri seti, diğerleri de dahil olmak üzere birçok araştırmacı tarafından kullanılmıştır (Goldani, Momtazi, et al., 2021; Goldani, Safabakhsh, et al., 2021; Hakak et al., 2021; Samadi et al., 2021).

A: Gerçek Haberler Veri Seti

Gerçek Haber veri seti, gerçek haber olarak kategorize edilen 21417 makaleden oluşan bir derlemedir. Bu makaleler, %52,6 küresel siyasete ait ve geri kalan %47,4 dünya ile ilgili diğer

çeşitli konuları kapsayan bir konu yelpazesini kapsamaktadır (Şekil 6). Haberlerin yayınlanma tarihleri 13 Ocak 2016'dan 31 Aralık 2017'ye kadar uzanmaktadır.

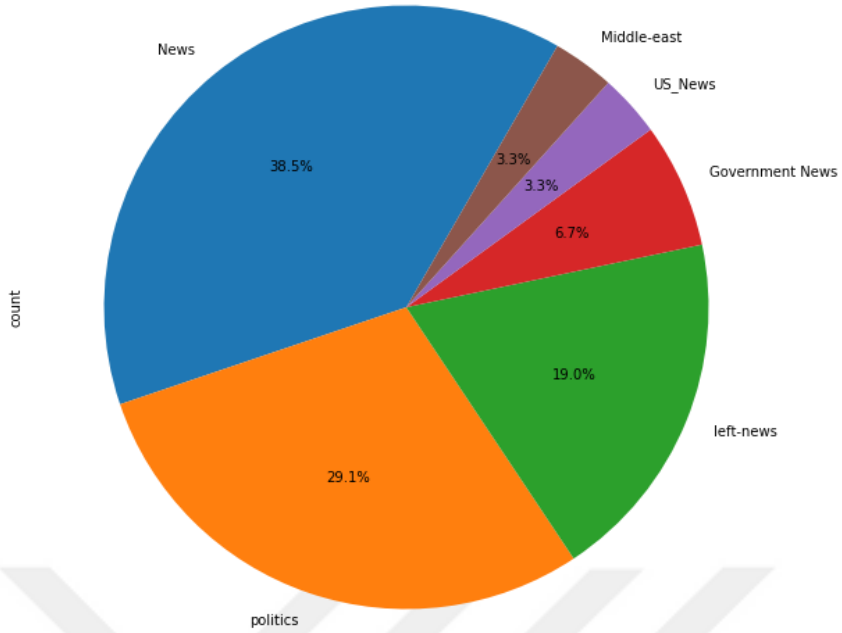


Şekil 6. Gerçek Haber Veri Seti Dağılımı

B: Sahte Haberler Veri Seti

Sahte haber veri seti, sahte haber kaynağı olarak tanımlanan 23481 makaleden oluşmaktadır. Bunların %38,5 genel haberlerle ilgiliyken, %29,1 siyasi içeriğe odaklanmakta, kalan %32,4 ise çeşitli konuları kapsamaktadır (Şekil 7). Bu sahte haber makalelerinin yayınlanma zaman çizelgesi 31 Mart 2015'ten 19 Şubat 2018'e kadar uzanmaktadır.

Bu veri seti, sahte ve gerçek haber makalelerinin sınıflandırılması ve tespitine yönelik modellerin araştırılması ve geliştirilmesi için değerli bir kaynak olarak hizmet vermektedir. Farklı konuların ve geniş bir zamansal aralığın dahil edilmesi, veri setinin uygulanabilirliğini artırmakta NLP ve ML alanlarındaki araştırmalar için sağlam bir temel sağlamaktadır.



Şekil 7. Sahte Haber Veri Seti Dağılımı

3.4. Veri Ön İşleme

Sosyal medya verileri çoğunlukla yapılandırılmamıştır ve Şekil 8'de gösterildiği gibi genellikle yazım hataları, argo, kötü dilbilgisi, sayılar ve semboller içeren gayri resmi konuşmalar içerir. Güvenilir performans sağlamak için, bilinçli kararlar alınmasına olanak tanıyan kaynak kullanım stratejileri geliştirmek gerekir. Tahmine dayalı modellemeden önce, üstün içgörüler elde etmek için verilerin temizlenmesi gerekir.

3.4.1. Noktalama işaretleri ve özel karakterler kaldırılması

Kod, veri çerçevesinin " metin " sütunundaki metin verilerini önceden işler ve ardından veri kümesini eğitim ve test kümelerine ayırır; bu, makine öğreniminde model performansını değerlendirmek için yaygın bir uygulamadır.

text.lower(): Metindeki tüm karakterleri küçük harfe dönüştürür.
re.sub('\[.*?\]', '', text): Köşeli parantezleri ve içeriklerini kaldırır.
re.sub("\\W", " ", text): Alfasayısal olmayan karakterleri boşlukla değiştirir.
re.sub('https?://\S+|www\.\S+', '', text): URL'leri kaldırır.

re.sub('<.*?>+', '', text): HTML etiketlerini kaldırır.

re.sub('[%s]'%re.escape(string.punctuation),'',text): Noktalama işaretlerini kaldırır.

re.sub('\n', '', text): Satırsonu karakterlerini kaldırır.

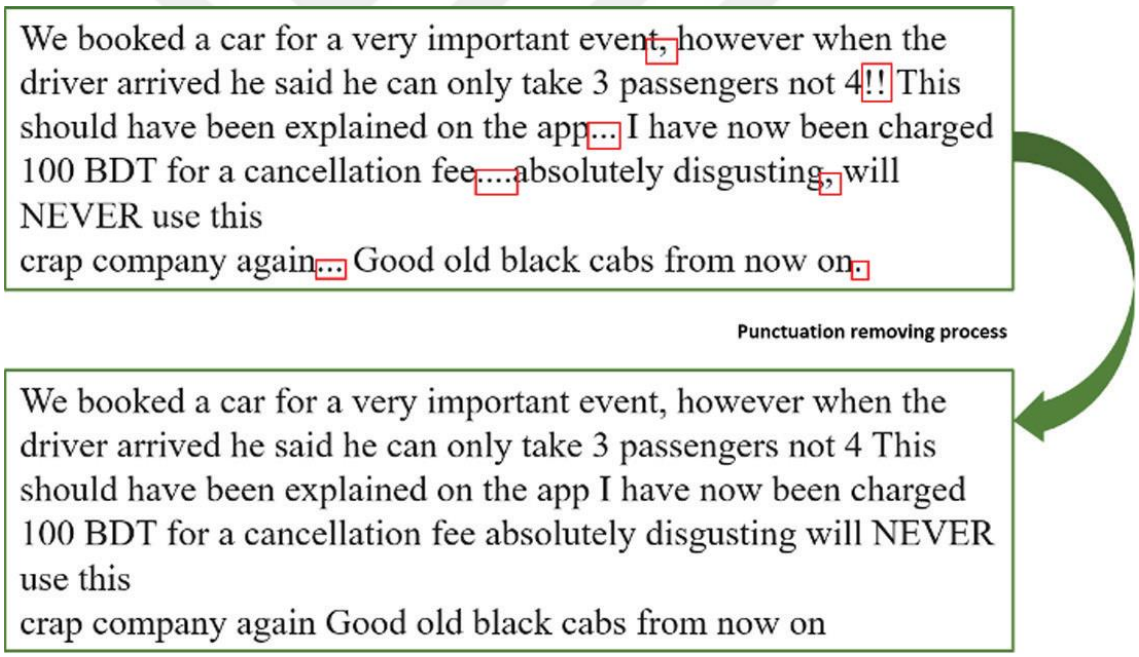
re.sub('\w*\d\w*', '', text): Rakam içeren kelimeleri kaldırır.

x = df_merge["text"]: df_merge veri çerçevesinin " metin" sütununu x değişkenine atar.

y = df_merge["class"]: df_merge veri çerçevesinin "class" sütununu y değişkenine atar.

train_test_split(x, y, test_size=0.20): Veri seti eğitim ve test kümelerine böler. x_train ve y_train eğitim için verilerinden %80 içerir, x_test ve y_test test için %20 içerir.

Çalışmada veri çerçevesinin "Metin" sütunundaki metin verilerini önceden işlendi ve temizlendi ve ardından veri setini, model performansını değerlendirmek için ML'de yaygın bir uygulama olan eğitim ve test setlerine ayırıldı.

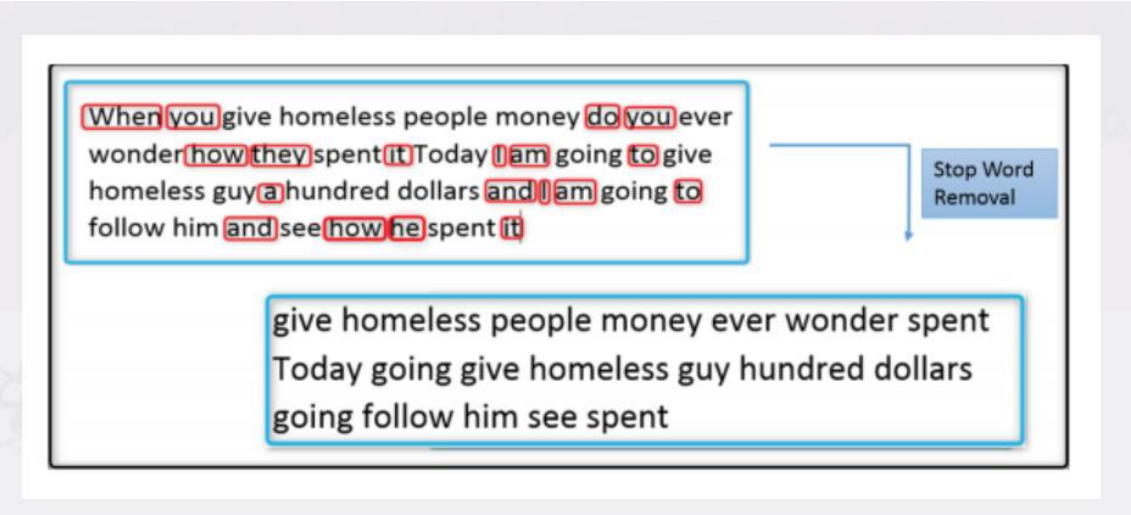


Şekil 8. Noktalama işaretleri ve özel karakterler kaldırılması

3.4.2. Durdurma sözcüklerinin kaldırılması

Duran kelimeler, "a", "be", "quite", "should" gibi bir dildeki en yaygın terimlerdir. Şekil 9'da gösterildiği gibi, genellikle anlamdan yoksundurlar ve bilgiyi zenginleştirmezler. Dahası, neredeyse her metinde görülürler. Bu nedenle, durak kelimelerin kaldırılmasının bazı

avantajları olabilir. Birincisi, büyük miktarda metni ortadan kaldırarak bellek kullanımını azaltır (ve böylece modellerimizi eğitmek için gereken özellik sayısını azaltır). Duran kelimeleri kaldırmanın bir diğer faydası da gürültüyü azaltması ve daha önemli özelliklere (iki sınıfı daha iyi ayırt eden özelliklere) odaklanmayı sağlamasıdır. Ancak, durak sözcükleri silmek her zaman iyi bir fikir değildir çünkü bazen önemli olan bilgileri de içerebilirler. Örneğin, dili çevirirken veya modellerken, genellikle tüm durak kelimelerini saklamak gerekir. Ancak bu durumda, metnin anlamına dayalı bir karar verilir. Dolayısıyla, durak sözcükleri güvenli bir şekilde kaldırabilir ve daha alakalı bağlam sözcüklerine odaklanılabilir.



Şekil 9. Durdurma kelimesi kaldırma kaynağı (Lakshmikumar et al., 2019)

3.5. Metin temsili

3.5.1. Terim sıklığı-ters belge sıklığı (TF-IDF)

TF-IDF vektörleştiricisi, ham metin verilerini her satırın bir belgeye ve her sütunun benzersiz bir terime karşılık geldiği bir matrise dönüştürür. Matristeki değerler, her belgedeki her terim için TF-IDF puanlarıdır.

- Terim Frekansı (TF):

$$TF(t,d) = \frac{d \text{ belgesinde görünen } t \text{ zaman dilimi sayısı}}{d \text{ belgesindeki toplam terim sayısı}} \quad (1)$$

- Ters Belge Frekansı (IDF):

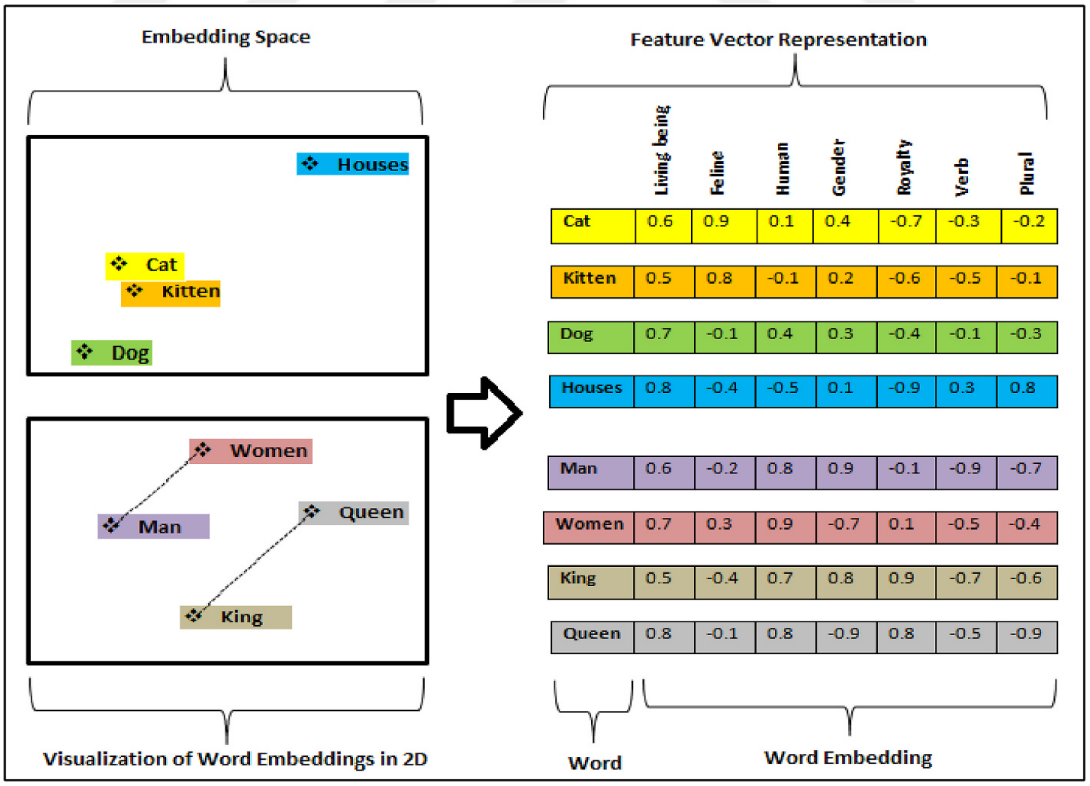
$$IDF(t, D) = \log\left(\frac{\text{Bütüncedeki toplam belge sayısı } |D|}{\text{Terim içeren belge sayısı } t}\right) \quad (2)$$

- TF-IDF Puanı:

$$\text{TF} - \text{IDF}(t, d, D) = TF(t, d) \times IDF\ (t, D) \quad (3)$$

3.5.2. Vektörleştirme (Word2Vec)

Word2Vec, bir metin verisi bütünceden kelime katıştırmaları oluşturmak için kullanılır. Word2Vec, bir hedef kelime verilen bağlam kelimelerini veya tam tersini tahmin ederek sürekli bir vektör uzayında kelimelerin dağıtılmış temsillerini öğrenir. Elde edilen sözcük katıştırmaları, sözcükler arasındaki anlamsal ilişkileri yakalar ve çeşitli NLP görevlerinde özellik olarak kullanılabilir. Şekil 10'da özellik vektörü temsil sürecini gösterilmektedir.



Şekil 10. Özellik vektörü temsili (Hamed et al., 2023).

Burada matematiksel olarak yapılan işlem şudur:

- Kelime Gösterimi: Word2Vec, bütüncedeki her kelimeyi yüksek boyutlu bir uzayda yoğun bir vektör olarak temsil eder. Her kelimeye benzersiz bir vektör atanır.
- Kelime Katıştırmalarını Öğrenme: Word2Vec, sağlanan derlem üzerinde eğitim yaparak bu kelime katıştırmalarını öğrenir. Bir hedef kelime (merkezi bir kelime) verildiğinde bağlam kelimelerini (metinde yakın kelimeler) veya tam tersini tahmin eder.
- Fonksiyon Amacı: Word2Vec'in eğitim amacı, tüm bütünce üzerinde bir hedef kelime (veya tam tersi) verildiğinde bağlam kelimelerini tahmin etme olasılığını en üst düzeye çıkarmaktır. Bu tipik olarak bir softmax fonksiyonu kullanılarak elde edilir.
- Eğitim Algoritması: Word2Vec, kelime vektörlerini iteratif olarak güncellemek için stokastik gradyan inişi (SGD) veya negatif örnekleme gibi optimizasyon algoritmaları kullanır ve tahmin göreviyle ilişkili kayıp fonksiyonunu en aza indirir.
- Vektör Alanı: Eğitimden sonra, her kelime belirtilen yüksek boyutlu vektör uzayında (bu durumda 100 boyutlu) yoğun bir vektörle temsil edilir. Bu vektörler, eğitim verilerindeki bağlamlarına dayalı olarak kelimeler arasındaki anlamsal ve sözdizimsel benzerlikleri yakalar.
- Kelime Dağarcığı Boyutu: Kelime dağarcığının boyutu (yani, benzersiz kelimelerin sayısı) Word2Vec modelinin kelime vektörlerinde bulunan anahtarlara (kelimeler) göre belirlenir.

3.6. Yapay Zeka Modelleri

Gerçek ve aldatıcı haber makalelerini ayırt etmek için sağlam metodolojiler geliştirme arayışında, otonom olarak üretilen sahte haberlerin gerçek zamanlı tespiti, ML ve NLP için önemli bir zorluktur (Gifu, 2023). Çalışmada AI algoritmalarından oluşan bir paket uygulanmıştır. Her algoritma, sahte haberleri tespit etme görevinin ortaya çıkardığı karmaşık zorlukları ele almak için titizlikle tasarlanmıştır. Bu araştırmada ele alınan modeller, ML, DL ve NLP tekniklerinden oluşan bir yelpazeyi kapsamakta ve her biri doğru sınıflandırma hedefine benzersiz bir bakış açısıyla katkıda bulunmaktadır.

3.6.1. Makine öğrenimi

3.6.1.1. Lojistik regresyon

Çalışmada LR, gerçek ve yanıltıcı haber makalelerini ayırt etmek için sağlam bir metodoloji sağlamak üzere TF-IDF ile birleştirilmiştir. Bir metin vektörleştirme tekniği olan TF-IDF, metin belgelerini sayısal vektörlere dönüştürerek derlem içindeki tek tek terimlerin önemini kapsar. Böylece her belge bir vektör olarak temsil edilir ve LR sınıflandırma için bu vektörleri kullanır. TF-IDF, bir belge içindeki terim sıklığını gösteren TF ve tüm derlemdeki terim önemini vurgulayan IDF hesaplar. TF ve IDF'nin çarpımı, TF-IDF puanını verir ve daha sonra metinsel içerik ile gerçek veya sahte haberlerin ikili sonucu arasındaki ilişkiyi lojistik fonksiyon aracılığıyla modellemek için LR tarafından kullanılan özellik vektörlerini oluşturur. Algoritma, her terimin TF-IDF puanı için ağırlıkları tahmin eder ve nihai karar sınırı, bu ağırlıklı özelliklerin doğrusal kombinasyonu yoluyla belirlenerek haber sınıflandırması için şeffaf ve yorumlanabilir bir çerçeve sağlar.

LR ve TF-IDF arasındaki sinerji, gerçek ve sahte haberlerle ilişkili ayırt edici dilbilimsel kalıpların yakalanmasını kolaylaştırdığı için önemlidir. TF-IDF, algoritmanın terimlere tek tek makalelerdeki ve veri setindeki alaka düzeylerine göre ağırlıklar atamasını sağlar ve böylece metin içeriğinin incelikli bir temsilini sunar. Bu yaklaşım, ince dilbilimsel ipuçlarının çok önemli olduğu haber sınıflandırma alanında çok önemlidir. LR'nin basitliği, TF-IDF tarafından oluşturulan bilgilendirici özellik temsili ile tamamlanarak modelin yorumlanabilirliğini artırır. Gerçek ve aldatıcı haber makaleleri arasında ayrım yaparken karar verme sürecini anlamak ve güvenmek isteyen paydaşlar için çok değerli hale getirir. Matematiksel işlem sırası şu şekildedir:

- **LR Fonksiyonu:**

$$P(Y = 1|X) = \frac{1}{1+e=-((\beta 0+\beta 1\ X1+\ \beta 2\ X2+,,,\beta_n\ f))} \quad (4)$$

Burada;
P(Y=1|X) çıktının (Y)1 (sahte haber) olma olasılığıdır.

X,TF-IDF vektörleştirmesinde elde edilen giriş özelliklerini temsil eder.
$\beta_0, \beta_1, \beta_2, \ldots, \beta_n$ eğitim sırasında öğrenilen katsayılardır.

- **Model Eğitimi:**

LR modeli, lojistik kayıp fonksiyonunu minimize ederek optimum değerleri bulmak üzere eğitilir.

$$\text{Logistic Loss} = \frac{1}{N} \sum_{n=1}^{N} [y_i \log(p_i) + (1 - y_i) \log\ (1 - p_i)] \quad (5)$$

Burada;
N eğitim numunelerinin sayısıdır.
y_i i inci eğitim örneği için gerçek sınıf etiketidir (doğru haber için 0, sahte haber için 1).
p_i i inci örneğin sınıf 1'e (sahte haber) ait olduğu tahmin edilen olasılıktır.

- **Tahmin:**

$$\text{tahmin} = \begin{cases} 1 \text{ ise } P\,(Y = 1|X) \geq 0.5 \\ 0 \text{ ise } P\,(Y = 1|X) < 0.5 \end{cases} \quad (6)$$

Bu süreç, metin verilerinin sayısal özelliklere dönüştürülmesini, bu özellikler üzerinde bir LR modelinin eğitilmesini ve yeni veriler üzerinde tahminler yapmak için eğitilmiş modelin kullanılmasını içerir. TF-IDF puanları, terimlerin sınıflar arasında ayrım yapmadaki önemini yakalar.

3.6.1.2. Karar ağacı

Veri sınıflandırması için karar ağaçları oluşturan bir ML algoritması olan DT sınıflandırıcısı, TF-IDF kullanarak gerçek ve sahte haberleri ayırt etme bağlamında kullanılmaktadır. Her belge bir kelime kümesi olarak ele alınır ve algoritma, belgeler arasında ayrım yapmadaki önemlerine göre ağırlıklar atar. TF-IDF, bir belgedeki bir kelimenin tüm belgelerdeki sıklığına göre önemini ölçer. DT bu bilgiyi veri kümesini özyinelemeli olarak bölmek için kullanır ve TF-IDF değerlerine dayalı kararlar verir. Algoritma, TF-IDF ağırlıklı özellikleri analiz ederek haber

makalelerinin gerçek veya sahte olarak sınıflandırılmasına katkıda bulunan anahtar terimleri tanımlar.

TF-IDF hesaplaması TF ve IDF kapsar. TF bir belge içindeki terim frekansı ölçerken, IDF tüm belgelerdeki terim nadirliğini ölçer. TF ve IDF'nin çarpımı TF-IDF puanını verir. Karar ağacının mimarisinde, her karar düğümünde algoritma özelliği (kelime) ve eşik TF-IDF değerini seçerek sınıf etiketlerine dayalı veri bölmeyi optimize eder. Bu süreç yinelemeli olarak bir ağaç yapısı üretir ve algoritmanın TF-IDF değerlerine dayalı olarak ağaçta gezinerek yeni belgeleri tahmin etmesini sağlar. DT yorumlanabilirliği, haber makalesi sınıflandırmasını etkileyen temel terimlerin anlaşılmasını kolaylaştırarak gerçek ve sahte haberler arasındaki ayırt edici özelliklere ilişkin içgörüler sağlar. Matematiksel işlem sırası şu şekildedir:

- **DT Fonksiyonu:**

Model işlevi, DT sınıflandırıcısının girdi özelliklerine dayalı olarak sahte ve gerçek haberleri nasıl ayırt ettiğini temsil eder. Bir özellik vektörü ile temsil edilen bir haber makalesi verildiğinde, DT sınıflandırıcısı matematiksel olarak şu şekilde gösterilebilir:

$$\mathrm{DT}\,(\mathrm{x}) = \begin{cases} 1 & \text{ise Düğüm}_{\text{yaprak}}\ \text{sahte haberi gösteren bir yaprak düğümdür} \\ 0 & \text{ise Düğüm}_{\text{yaprak}}\ \text{gerçek haberi gösteren bir yaprak düğümdür} \end{cases} \quad (7)$$

Burada; Düğüm_{yaprak} girdi haber makalesinin özelliklerine dayalı olarak karar ağacının kök düğümden çaprazlanmasıyla ulaşılan yaprak düğümü temsil etmektedir.

- **Model Eğitimi:**

DT sınıflandırıcısının eğitimi, etiketli haber makalelerinden oluşan bir veri seti kullanılarak ağacın oluşturulmasını içerir. Matematiksel olarak eğitim süreci, eğitim veri kümesi D üzerinde önceden tanımlanmış bir safsızlık ölçütü I'yi en aza indiren en uygun DT χ'in bulunmasını içerir. Bu şu şekilde gösterilebilir:

$$T^* = \arg\min_T I\,(D, T) \quad (8)$$

Burada;

T^*, optimal karar ağacıdır.

I(D,T), T ağacının ve *D* veri setinin safsızlık ölçüsüdür.

DT algoritması, seçilen bir safsızlık ölçüsüne (örneğin, Gini safsızlığı veya entropi) dayalı olarak her bir düğümdeki safsızlığı en aza indirmek için özellik uzayını özyinelemeli olarak böler.

- **Tahmin:**

DT eğitildikten sonra, yeni, görülmemiş haber makaleleri üzerinde tahminler yapmak için kullanılabilir.

Bir özellik vektörü x_{haber} ile temsil edilen yeni bir haber makalesi verildiğinde, tahmin süreci, karar ağacının kök düğümden yaprak düğümüne kadar'nin özelliklerine dayalı olarak geçmesini içerir. Matematiksel olarak bu aşağıdaki gibi gösterilebilir:

$$(x_{haber}) = \begin{cases} 1 & ise\ x_{haber}\ Sahte\ haberleri\ belirten\ bir\ yaprak\ düğümüne\ düşüyor \\ 0 & ise\ x_{haber}\ Gerçek\ haberleri\ belirten\ bir\ yaprak\ düğümüne\ düşüyor \end{cases} \quad (9)$$

Burada; DT x_{haber} yeni haber makalesi x_{haber} için öngörülen etikettir.

Sahte ve gerçek haber sınıflandırmasını matematiksel olarak tespit etme süreci, model fonksiyonunun tanımlanmasını, safsızlığı en aza indirmek için DT modelinin eğitilmesini ve ardından haber makalelerinin girdi özelliklerine dayalı tahminler yapmak için eğitilmiş modelin kullanılmasını içerir.

3.6.1.3. Gradiant boosting

Sağlam bir tahmin modeli oluşturmak için zayıf öğrenicileri birleştiren bir topluluk öğrenme algoritması olan GBoost Sınıflandırıcısı, TF-IDF kullanarak gerçek ve sahte haber sınıflandırması için kullanılır. Yinelemeli olarak, her bir zayıf öğrenici, birleştirilmiş tahminlerin artıklarına uyum sağlayarak önceki hataları düzeltir. Öğrenme süreci, belirli bir kayıp fonksiyonunu en aza indirerek genel tahmin doğruluğunu artırır. GBoost Sınıflandırıcısının scikit-learn uygulaması (random_state=1) sabit bir tohum aracılığıyla tekrarlanabilirliği sağlar.

TF-IDF gösteriminde algoritma, her satırın bir belgeye karşılık geldiği ve her sütunun benzersiz bir kelimeyi temsil ettiği bir özellik matrisi kullanır. Sayısal bir istatistik olan TF-IDF, bir belgedeki kelimenin önemini tüm belgelerdeki sıklığa göre değerlendirir. Algoritma, bu TF-IDF ağırlıklı özelliklere dayanarak gerçek ve sahte haberleri birbirinden ayırır. Matematiksel hesaplamalar, TF-IDF puanlarının hesaplanmasını, özellik matrisinin oluşturulmasını ve artırma yoluyla model parametrelerinin iteratif olarak optimize edilmesini içerir. Bu topluluk yaklaşımı, etkili gerçek ve sahte haber sınıflandırması için verilerdeki karmaşık örüntüleri yakalayarak algoritma performansını ve genelleştirmeyi geliştirir. Matematiksel işlem sırası şu şekildedir:

- **GBoost Fonksiyonu:**

GBoost için model işlevi, güçlü bir topluluk oluşturmak için birden fazla zayıf öğrenicinin (karar ağacı) tahminlerini birleştirmeyi içerir. Model fonksiyonu, bireysel zayıf öğrenicilerin tahminlerinin toplamı olarak ifade edilebilir:

$$f(X) = F_0(x) + \eta h_1(x) + \eta h_2(x) + \ldots + + \eta h_M(x) \quad (10)$$

Burada;

f(X), x özelliklerine sahip belirli bir haber makalesi için nihai tahmindir.

$F_0(x)$, başlangıç tahminidir (genellikle regresyon için ortalama veya sınıflandırma için log *odds*).

η her iterasyon sırasında adım boyutunu kontrol eden öğrenme oranıdır.

$h_1(x), h_2(x), \ldots h_M(x)$ negatife yerleştirilmiş zayıf öğrenicilerdir (karar ağaçları) kayıp fonksiyonunun gradineti.

- **Model Eğitimi:**

Eğitim süreci, kayıp fonksiyonunun negatif gradyanına bir dizi zayıf öğrenicinin uydurulmasını içerir. Her iterasyonda, öncekiler tarafından yapılan hataları düzeltmek için yeni bir zayıf öğrenici eklenir. Model, belirli bir kayıp fonksiyonunu en aza indirecek şekilde eğitilir. Sınıflandırma problemleri için yaygın kayıp fonksiyonu, negatif log-likelihood ile ilişkili olan sapmadır.

$$Kayıp = -\sum_{i=1}^{N}\left[y_i \log\left(\frac{e^{F_{m\ 1}(x_i)}}{1+e^{F_{m\ 1}(x_i)}}\right)\right] \quad (11)$$

Burada;

N örnek sayısıdır.

y_i i'inci örnek için gerçek etikettir (sahte için 1,gerçek için 0).

$F_{m-1}(x_i)$ topluluğunun (m-1)inci iterasyona kadar olan tahminidir.

- **Tahmin:**

Eğitimden sonra, topluluktaki tüm zayıf öğrenicilerin tahminleri birleştirilerek yeni örnekler için tahminler yapılır. Final tahmin, tahminlerin toplanması ve bir eşik değerinin uygulanmasıyla elde edilir.

$$Tahmin\ \hat{y} = \begin{cases} 1\ ise\ F\ (X) \geq 0.5 \\ 0\ ise\ F\ (X) < 0.5 \end{cases} \quad (12)$$

Burada;

$\hat{y}$ tahmin etiketidir (sahte için 1,gerçek için 0).

GBoost Sınıflandırıcısı, her bir ağacın öncekilerin hatalarını düzelttiği bir karar ağaçları topluluğu oluşturur. Model işlevi, bireysel zayıf öğrenicilerin toplamıdır ve eğitim süreci bir kayıp işlevinin en aza indirilmesini içerir. Tahminler, tüm zayıf öğrenicilerin tahminleri birleştirilerek ve örnekleri sahte veya gerçek olarak sınıflandırmak için bir eşik uygulanarak yapılır.

3.6.1.4. Rastgele orman

Giriş veri kümesinin farklı alt grupları üzerinde birçok DT kullanan ve tahmin edilen verimliliği en üst düzeye çıkarmak için sonuçların ortalamasını alan RF sınıflandırıcısı, sahte haberleri belirlemek için kullanılır (Kondamudi et al., 2023). DT ormanları oluşturan bir topluluk öğrenme algoritması olan RF Sınıflandırıcısı, TF-IDF kullanarak gerçek ve sahte haberleri sınıflandırmak için uygulanmıştır. RF Sınıflandırıcı (random_state=5), her bir ağacın rastgele bir veri alt kümesi üzerinde değiştirilerek (bootstrapping) eğitildiği bir karar ağaçları ormanı

oluşturur. Rastgele özellik seçimi aşırı uyumu azaltır ve model sağlamlığını artırır. random_state=5 parametresi tekrarlanabilirliği sağlar.

TF-IDF uygulamasında algoritma, gerçek ve aldatıcı makaleleri ayırt etmede kelime önemini yakalamak için TF-IDF gösteriminden yararlanır. TF-IDF, bir belgedeki kelime sıklığına göre tüm belgelerdeki sıklığa göre sayısal değerler atar. Her biri farklı özellik ve örnek alt kümelerine dayalı kararlar veren birden fazla DT oluşturulur ve tahminler çoğunluk oylamasıyla toplanır. Matematiksel hesaplamalar, TF-IDF matrisinin oluşturulmasını, her ağaç için rastgele alt kümelerin seçilmesini, bireysel karar ağaçlarının eğitilmesini ve sağlam ve doğru gerçek ve sahte haber sınıflandırması için çıktıların toplanmasını içerir. Matematiksel işlem sırası şu şekildedir:

- **RF Fonksiyonu:**

Model işlevi, girdi özelliklerinin tahminlere nasıl dönüştürüldüğünü tanımlar. Sahte ve gerçek haberlerin tespit edilmesi bağlamında, model işlevi f bir haber makalesinin özelliklerini temsil eden bir özellik vektörü X alır ve tahmin edilen etiket $\hat{y}$ çıktısını verir. Matematiksel olarak bu aşağıdaki gibi gösterilebilir:

$$\hat{y} = f\ (X) \tag{13}$$

Model fonksiyonu f LR, DT, destek vektör makineleri SVM, RF veya sinir ağları (NN) gibi çeşitli ML algoritmaları kullanılarak uygulanabilir. Her algoritmanın giriş özelliklerini tahminlere dönüştürmek için kendi matematiksel formülasyonu vardır.

- **Model Eğitimi:**

Modelin eğitilmesi, etiketli veriler kullanılarak f model fonksiyonunun parametrelerinin öğrenilmesini içerir. Her biri X_i özellik vektörüne ve y_i etiketine sahip n örnekten oluşan bir D veri seti verildiğinde amaç, seçilen bir kayıp fonksiyonunu L en aza indiren parametreleri bulmaktır. Matematiksel olarak bu bir optimizasyon problemi olarak gösterilebilir:

$$min_\theta\ \frac{1}{n}\sum_{i=1}^{n} L\ (f\ (X_i\ ;\ \theta), y_i) \tag{14}$$

Burada;

θ, f model fonksiyonunun parametrelerini temsil eder.

L, çapraz entropi kaybı veya menteşe kaybı gibi seçilen kayıp fonksiyonudur.

$f\ (X_i\ ;\ \theta)$, θ parametreli f model fonksiyonu kullanılarak i örnek için tahmin edilen etiketi temsil eder.

Optimizasyon problemi tipik olarak stokastik gradyan deseni (SGD) veya Adam gibi gradyan tabanlı optimizasyon algoritmaları kullanılarak çözülür.

- **Tahmin:**

Model eğitildikten sonra, yeni, görülmemiş veriler üzerinde tahminler yapmak için kullanılabilir. X_{haber} , özellik vektörüne sahip yeni bir haber makalesi verildiğinde, eğitilen model işlevi f, $\hat{y}_{haber}$ etiketini tahmin etmek için uygulanır. Matematiksel olarak bu şu şekilde gösterilebilir:

$$\hat{y}_{haber} = f\ (X_{haber}) \tag{15}$$

Tahmin edilen $\hat{y}_{haber}$ etiketi daha sonra uygulamaya ve seçilen karar eşiğine bağlı olarak haber makalesinin sahte $(\hat{y}_{haber} = 0)$ veya doğru $(\hat{y}_{haber} = 1)$, olduğunu gösterir şekilde yorumlanabilir.

Bir RF sınıflandırıcısında sahte ve gerçek haberleri tespit etmeye yönelik matematiksel süreç, bir model fonksiyonunun tanımlanmasını, etiketli verileri kullanarak parametrelerini öğrenmek için modelin eğitilmesini ve ardından yeni veriler üzerinde tahminler yapmak için eğitilmiş modelin kullanılmasını içerir. Bu süreç, model fonksiyonu için en iyi parametreleri bulmak üzere optimizasyon tekniklerine dayanır ve öğrenme sürecini yönlendirmek ve tahminleri yorumlamak için uygun kayıp fonksiyonlarının ve karar eşiklerinin seçilmesini içerir.

3.6.1.5. Aşırı gradyan artırma

GBoost ailesi içinde bir ML algoritması olan aşırı gradyan artırma (XGBoost) sınıflandırıcısı, TF-IDF kullanarak gerçek ve sahte haber sınıflandırması için kullanılır. Random_state=1

parametresine sahip XGBoost Sınıflandırıcısı, tekrarlanabilirlik için sabit bir tohumla sınıflandırma için XGBoost kullanımını gösterir. XGBoost sırayla bir karar ağaçları topluluğu oluşturur, önceki hataları telafi eder, düzenleme terimleri ve gradyan tabanlı bir optimizasyon süreci kullanır.

TF-IDF bağlamında, XGBoost Sınıflandırıcı, gerçek ve aldatıcı makaleleri ayırt etmede kelimenin önemini ölçmek için TF-IDF temsilini kullanır. TF-IDF, tüm belgelerdeki frekansa göre bir belgedeki kelime frekansına dayalı ağırlıklar atar. XGBoost, TF-IDF özelliklerini optimizasyon sürecine entegre eder ve bunları karar ağaçları oluşturmak için girdi olarak değerlendirir. Matematiksel hesaplamalar, TF-IDF ile tanımlanan bilgilendirici özelliklere odaklanarak genel kaybı en aza indiren karar ağaçlarının iteratif olarak oluşturulmasını içerir. Ağaçlar topluluğu, nüanslı veri modellerini yakalayarak gerçek ve sahte haberler arasında etkili bir ayrım yapan sağlam bir model oluşturur. XGBoost'taki artırma çerçevesi, karmaşık ilişkileri ele almada üstünlük sağlayarak tahmin performansını artırır. Matematiksel işlem sırası şu şekildedir:

- **XGBoost Fonksiyonu:**

XGBoost için model fonksiyonu bir karar ağaçları topluluğu olarak ifade edilebilir.

$$F(x) = \sum_{i=1}^{N} f_i\,(x) \tag{16}$$

Burada;
$F(x)$, topluluk tarafından yapılan nihai tahmindir.
N, topluluktaki ağaç sayısıdır..
$f_i\,(x)$ i DT tarafından yapılan tahmindir.

Her DT $f_i\,(x)$, x girdisinin özellik değerlerine dayalı bir dizi ikili karar olarak gösterilebilir.

- **Model Eğitimi:**

XGBoost'taki amaç fonksiyonu, bir kayıp fonksiyonunu bir düzenleme terimi ile birleştirir.

$$Obj = \sum_{i=1}^{n} \ell\,(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{17}$$

Burada;
ℓ , tahmin edilen $(\hat{y}_i)$ ile gerçek (y_i) değerleri arasındaki farkı ölçen kayıp fonksiyonudur.
$\Omega(f_k)$ bireysel ağaçların karmaşıklığını cezalandıran düzenlilik terimidir.
K, topluluktaki ağaç sayısıdır.

Optimizasyon, XGBoost, GBoost kullanarak amaç fonksiyonunu en aza indirir. Topluluğa sıralı bir şekilde yeni ağaçlar ekler ve her ağaç genel amaç fonksiyonunu azaltmayı hedefler.

- **Tahmin:**

Eğitilmiş XGBoost modelini kullanarak tahminler yapmak için, tüm bireysel ağaçların tahminleri toplanır:

$$\hat{y} = \sum_{i=1}^{N} f_i\ (x) \quad (18)$$

Burada;
$\hat{y}$ nihai tahmindir
$f_i\ (x)$ DT tarafından yapılan tahmindir.

Bu matematiksel çerçeve, XGBoost mevcut özelliklere dayanarak haber makalelerini sahte veya gerçek olarak etkili bir şekilde sınıflandırabileceğini kanıtlamaktadır.

3.6.2. Derin Öğrenme

3.6.2.1. Uzun kısa süreli bellek

Bu araştırma makalesinde sunulan DL modeli, Word2Vec katıştırmaları ve LSTM ağlarının bir kombinasyonunu kullanarak gerçek ve sahte haber makalelerini sınıflandırmak için tasarlanmıştır. Sıralı model mimarisi, gömme, sıralı işleme ve ikili sınıflandırma için uyarlanmış katmanlarla oluşturulmuştur. Sağlanan kod parçacığı, modelin hazırlanmasını ve eğitimini göstermektedir. Başlangıçta veri kümesi **sklearn.model_selection** modülündeki **train_test_split** işlevi kullanılarak eğitim ve test kümelerine ayrılır. **Test_size** parametresi 0.2 olarak ayarlanmıştır, böylece verilerin %20'si test için kullanılırken kalan %80'i eğitim için kullanılacaktır. Ayrıca, tekrarlanabilirliği sağlamak için random_state parametresi 42 olarak

ayarlanmıştır. Model**, X_train** ve **y_train** veri kümeleriyle fit yöntemi kullanılarak eğitilir. Doğrulama oranı 0,3 olarak belirlenmiştir, bu da eğitim verilerinin %30'unun eğitim sürecinde doğrulama için kullanılacağı anlamına gelmektedir. Model 20 epok üzerinden eğitilmiştir ve mimarisi bir gömme katmanı, bir LSTM katmanı ve bir yoğun katmandan oluşmaktadır. Giriş kelimeleri, gömme katmanında önceden eğitilmiş Word2Vec katıştırmaları kullanılarak yoğun vektörlere dönüştürülür ve bu da metinsel verilerdeki anlamsal ilişkilerin anlaşılmasına yardımcı olur. Gömme katmanından sonra, haber makalelerindeki sıralı kalıpları ve bağımlılıkları yakalamak için 128 birimlik bir LSTM katmanı dahil edilmiştir. Katmanın uzun diziler boyunca bağlamı koruma yeteneği, gerçek veya sahte içeriği gösteren ince nüansları tanımlamak için çok önemlidir. Model, bir olasılık puanı çıkarmak için sigmoid aktivasyon fonksiyonu kullanan yoğun bir katmanla son bulur. Bu puan, bir makalenin gerçek veya sahte olma olasılığını belirler ve öğrenilen özelliklere dayalı kesin tahminlere olanak saüğlar

Matematiksel olarak, gömme katmanı, aşağıda gösterildiği gibi eğitilmiş Word2Vec yerleştirmelerini kullanarak giriş kelimelerinin gömme gösterimini hesaplar.

Ön eğitimli Word2Vec katıştırmalarını kullanarak girdi kelimelerinin katıştırılmasını temsil etmek için katıştırma katmanını oluşturma adımları aşağıda gösterilmiştir:

- **Vektörleştirme- Word2Vec:**

LSTM ağı kullanarak sahte ve gerçek haberleri sınıflandırmak için Word2Vec gömülerinden yararlandı. Aşağıdaki adımlar izlenmiştir:

- **Kelime Gömme Katmanı Başlatma:** LSTM modelinin gömme katmanı önceden eğitilmiş Word2Vec gömmeleriyle başlatılır. Bu katman her bir kelime indeksini karşılık gelen Word2Vec gömme vektörüne eşler.
- **Veri Ön İşleme:** Giriş metin verisi, kelime indeksleri dizilerine dönüştürülür. Her kelime, kelime dağarcığındaki karşılık gelen dizin ile değiştirilir.
- **Dolgu Dizileri:** DL modellerinde toplu işlemeyi mümkün kılmak için tüm dizilerin aynı uzunlukta olmasını sağlamak gerekir. Bu, dizileri sıfırlarla doldurarak veya sabit bir uzunlukta keserek elde edilebilir.
- **Bir LSTM oluşturun:** bir veya daha fazla LSTM katmanını ve ardından sınıflandırma için bir veya daha fazla yoğun katmanı istifleyerek model mimarisi.

- **Model Derleme:** LSTM modeli, ikili sınıflandırma için ikili çapraz entropi gibi uygun bir kayıp fonksiyonu, Adam gibi bir optimize edici ve doğruluk gibi değerlendirme ölçütleri kullanılarak derlenir.
- **Eğitim:** LSTM modeli önceden işlenmiş eğitim verileri üzerinde eğitilmelidir. Gömme katmanının Word2Vec gömmeleri, kayıp fonksiyonunu en aza indirmek için diğer model parametreleriyle birlikte eğitim sırasında güncellenecektir.
- **Değerlendirme:** Eğitilen model, performansını değerlendirmek için ayrı bir doğrulama veri kümesi üzerinde değerlendirilir. Modelin sahte ve gerçek haberleri sınıflandırmadaki etkinliğini ölçmek için doğruluk, kesinlik, geri çağırma ve F1 puanı gibi metrikler izlenmelidir.
- **Çıkarım:** Eğitimden sonra model, yeni ve görülmemiş haber makalelerinin sonuçlarını tahmin etmek için kullanılabilir. Bunu yapmak için, girdi metni tokenize edilir, diziler tamponlanır ve tahminler elde etmek için eğitilmiş LSTM modelinden geçirilir.
- **Ayrıntılı ayarlama:** İsteğe bağlı olarak, Word2Vec katıştırmaları, katıştırma katmanı başlatılırken eğitilebilir = Doğru olarak ayarlanarak eğitim sırasında ince ayar yapılabilir. Bu, modelin göreve özgü verilere dayalı olarak katıştırmaları güncellemesini sağlar.

Word2Vec katıştırmalarını LSTM tabanlı bir DL modeline dahil ederek, katıştırmalar tarafından yakalanan anlamsal bilgilerden, özellikle sahte ve gerçek haber sınıflandırması gibi metin verilerini içeren görevlerde sınıflandırma performansını artırmak için yararlanılabilir. LSTM mimarisi, modelin metin dizilerindeki zamansal ve uzun menzilli bağımlılıkları yakalamasını sağlayarak bağlamı anlamaya ve doğru tahminler yapmaya yardımcı olabilir.Ardından, LSTM katmanı, zamansal bağımlılıkları etkili bir şekilde yakalamak için bellek hücrelerinden yararlanarak sıralı girdiyi işler. Gömme katmanı NLP yeteneklerini geliştirirken, LSTM katmanı zamansal bağımlılıkları ve bağlamı yakalar. Eğitim sırasında model parametreleri, ikili çapraz entropi kaybını en aza indirmek için geri yayılım ve optimizasyon teknikleri kullanılarak ayarlanır. Son olarak, yoğun katman, bir makalenin gerçek veya sahte olma olasılığını gösteren ikili bir çıktı üretir.

Modelin sonundaki yoğun katman, doğru tahminler yapılmasını sağlar. Matematiksel işlem sırası şu şekildedir:

- **LSTM Fonksiyonu:**

Model, kelime indeksleri olarak temsil edilen bir dizi kelimeyi girdi olarak alır ve haberin gerçek mi yoksa sahte mi olduğunu tahmin eder.

Giriş:

- **X_train**: Her satırın bir dizi kelime indeksine dönüştürülmüş bir haber makalesini temsil ettiği bir matris.
- **y_train**: Etiketleri içeren bir vektör (sahte için 0, gerçek için 1).

Model Mimarisi:

- **Gömme Katmanı**: Önceden eğitilmiş kelime katıştırmalarının Word2Vec kullanan eğitilemez katıştırma katmanı. Bu katman kelime indekslerini yoğun vektörlere dönüştürür. Matematiksel olarak şu şekilde gösterilir:

$$X_{gömmek} = Katıştırma(X_{endeksler}) \quad (19)$$

- **LSTM Katmanı:** LSTM katmanı sıralı bilgileri yakalar ve gömülü dizileri işler. Matematiksel olarak şu şekilde gösterilir:

$$h_t = LSTM\ (X_{endeksler}) \quad (20)$$

- **Dense Katmanı**: Haberin gerçek olma olasılığını gösteren bir olasılık puanı, sigmoid aktivasyon fonksiyonuna sahip tam bağlı bir katman tarafından çıkarılır. Matematiksel olarak şu şekilde gösterilir:

$$\hat{y} = \sigma\ (W.h_t + b) \quad (21)$$

Burada; σ sigmoid activision fonksiyonu, W ağırlık matrisi, h_t LSTM katmanının çıktısı ve b önyargıdır

Kayıp Fonksiyonu:

- **İkili Çapraz Entropi Kaybı:** Tahmin edilen olasılıklar ile gerçek etiketler arasındaki farkı ölçer.

$$Kayıp = -\frac{1}{N}\sum_{i=1}^{N}(y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)) \quad (22)$$

Burada;

$X_{gömmek}$ gömülü giriş dizilerini temsil eder.

h_t LSTM katmanının çıktısını temsil eder.

$\hat{y}$ haberin gerçek olduğuna dair tahmin edilen olasılığı temsil eder.

N örnek sayısını temsil eder.

- **Model Eğitimi:**

Model, eğitim verileri $(X_{eğitim}, y_{eğitim})$ kullanılarak eğitilir ve Adam optimizer kullanılarak optimize edilir.

Eğitim süreci, ikili çapraz entropi kaybını en aza indirmek için ağın ağırlıklarının ayarlanmasını içerir.

- **Tahmin:**

Eğitimden sonra model, test verisi X_{metin} üzerinde tahminler yapmak için kullanılır.

Tahminler olasılıklar $(\hat{y})$ olarak elde edilir.

Haberleri sahte (0) veya gerçek (1) olarak sınıflandırmak için 0,5'lik bir eşik uygulanır.

3.6.3. Doğal Dil İşleme (NLP)

3.6.3.1. K-en yakın komşu

Sınıflandırma görevleri için kullanılan popüler bir ML algoritması olan K-en yakın komşu (K-NN) sınıflandırıcısı, bir nesnenin özellik uzayında k-en yakın komşularının çoğunluk sınıfına göre sınıflandırıldığı yakınlık ilkesine göre çalışır. NLP kullanarak sahte haberleri tespit etme bağlamında, algoritma metin verilerini Word2Vec embeddings, lemmatization için WordNetLemmatizer ve özellik çıkarma için CountVectorizer kombinasyonunu kullanarak işler. Word2Vec kelimeleri anlamsal ilişkileri yakalayan yoğun vektörler olarak temsil eder; WordNetLemmatizer kelimeleri temel biçimlerine normalleştirir ve CountVectorizer kelimelerin oluşumlarını sayarak metni sayısal özelliklere dönüştürür. Bu işlem sonucunda elde edilen özellik vektörleri daha sonra K-NN modelini eğitmek için kullanılır. Bu, modelin haber

makalelerini yalnızca metinsel içeriklerine göre gerçek veya sahte olarak sınıflandırmasını sağlar.

Matematiksel hesaplama süreci, yüksek boyutlu uzaydaki özellik vektörleri arasındaki mesafenin (genellikle Öklid veya kosinüs mesafesi) hesaplanmasını içerir. K-NN algoritmasında, test aşamasında her bir test örneği için eğitim kümesindeki k-en yakın komşularına olan uzaklıklar hesaplanır. Test örneğinin sınıf etiketi daha sonra komşuları arasında oy çokluğuyla belirlenir. Bu algoritmanın mimarisi, metin verileri için ön işleme adımlarından (Word2Vec, WordNetLemmatizer, CountVectorizer), K-NN modelinden ve eğitim ve test sürecinden **(train_test_split)** oluşmaktadır. Test puanı, modelin test kümesini doğru bir şekilde sınıflandırmadaki doğruluğunu temsil eder ve algoritmanın gerçek ve sahte haber makalelerini ayırt etmedeki etkinliğini değerlendirmek için bir performans ölçütü sağlar.

- **K-NN Fonksiyonu:**

K-NN algoritması benzerlik ilkesine göre çalışır. Yeni, görülmemiş bir örnek sunulduğunda, bu örnek ile eğitim setindeki diğer tüm örnekler arasındaki mesafeyi (genellikle Öklid mesafesi) hesaplar. Ardından en yakın K komşuyu seçer (burada K, kullanıcı tarafından belirlenen bir hiper parametredir) ve bu komşular arasındaki çoğunluk sınıfına göre sınıf etiketini atar. Matematiksel olarak, sahte ve gerçek haberleri ayırt etmek gibi ikili bir sınıflandırma görevinde, sınıf tahmini aşağıdaki gibi gösterilebilir:

$$\hat{y} = argmax\left(\sum_{i=1}^{K} I\ (y_i)\right) \tag{23}$$

Burada;

$\hat{y}$ yeni örnek için tahmin edilen sınıftır.

K n yakın komşu sayısıdır.

y_i gösterge fonksiyonudur y_i doğruysa 1, aksi takdirde 0 döndürür.

- **Model Eğitimi:**

Eğitim aşamasında, K-NN algoritması herhangi bir parametre öğrenmez. Bunun yerine, eğitim setindeki her veri noktası için özellikleri ve ilgili sınıf etiketini depolayarak eğitim verilerini hafızaya alır. Bu nedenle, eğitim süreci tüm eğitim veri setinin depolanmasını içerir. Model,

sağlanan koddaki fit fonksiyonu kullanılarak eğitilir. Fonksiyon, derlemden elde edilen özellik matrisini ve haberin doğru veya yanlış olduğunu gösteren ilgili hedef etiketleri kabul eder.

- **Tahmin:**

Model eğitildikten sonra, yeni, görülmemiş veriler için sonuçları tahmin edebilir. Yeni bir örnek için tahminde bulunmak amacıyla algoritma, söz konusu örnek ile eğitim setindeki tüm örnekler arasındaki mesafeleri hesaplar. Daha sonra bu mesafelere göre en yakın K komşuyu seçer ve bu komşular arasındaki çoğunluk sınıfına göre sınıf etiketini atar. Bu süreç, model fonksiyonu açıklamasında anlatıldığı gibi matematiksel olarak gösterilir. K-NN sınıflandırma algoritması eğitim sırasında karmaşık matematiksel işlemler gerektirmez. Bunun yerine, tüm eğitim veri kümesini depolar ve en yakın K komşu arasındaki çoğunluk sınıfına göre sınıf etiketini belirlemek için tahmin sırasında mesafeleri hesaplar.

3.6.3.2. Gaussian naive bayes

Gaussian Naive Bayes (GaussianNB) kural seti, sınıf etiketi verildiğinde yeteneklerin koşullu olarak tarafsız olduğu "naif" varsayımıyla tamamen Bayes teoremine dayanan olasılıksal bir sınıf tekniğidir. Özellikle metin sınıfları için etkilidir ve spam filtreleme, duygu analizi ve kayıt kategorizasyonunda yaygın olarak kullanılır (Prameela et al., 2024).

Sağlanan kod parçacığı, NLP tekniklerini kullanarak sahte haber tespiti için bir Naive Bayes sınıflandırıcısının uygulanmasını göstermektedir. İlk adım, sklearn.model_selection modülündeki (train_test_split) işlevini kullanarak veri kümesini eğitim ve test kümelerine bölmeyi içerir. (X özellikler) ve (Y etiketler) olarak gösterilen veriler, eğitim (%80) ve test (%20) alt kümelerine ayrılır. Sonraki satırlar bir GaussianNB modelini örneklendirir ve eğitim verilerini (X_train ve y_train) kullanarak bu modeli eğitir. Eğitilen model daha sonra test kümesinin (X_test) etiketlerini tahmin etmek için kullanılır ve modelin doğruluğu puan yöntemi kullanılarak değerlendirilir.

GaussianNB algoritması, olasılıksal sınıflandırma için Bayes teoremine dayanır. Sahte haber tespiti bağlamında, belirli bir belgenin belirli bir sınıfa (gerçek veya sahte haber) ait olma olasılığını hesaplamak için sağlanan belgelerdeki (haber makaleleri) kelimelerin frekanslarını

kullanır. Algoritma, özelliklerin (kelimelerin) koşullu olarak bağımsız olduğunu varsayar, bu da hesaplamaları basitleştirir. CountVectorizer, ham metin verilerini her bir belgedeki kelime sıklıklarını temsil eden sayısal bir biçime dönüştürmek için kullanılır. Sağlanan kod Word2Vec ve WordNetLemmatizer'ın açık kullanımından yoksun olsa da, bu tekniklerin dahil edilmesi, anlamsal ilişkileri yakalayarak ve kelime varyasyonlarını azaltarak modelin performansını artırabilir. Bu NLP tekniklerinin ve GaussianNB algoritmasının birlikte kullanımı, eğitim verilerinde öğrenilen kalıplara dayalı olarak gerçek ve sahte haberlerin etkili bir şekilde sınıflandırılmasına olanak tanır.

Kod, kullanılan sistemde mevcut olandan daha fazla bellek (RAM) ayırmayı gerektirdiğinden, büyük veri kümeleriyle çalışmak özellikle binlerce veya milyonlarca veri noktası ve özellik ile uğraşabileceğiniz ML görevlerinde önemlidir.

Bu algoritmada, seyrek bir matrisi yoğun bir matrise dönüştürmek gerekir. Seyrek matrisler, birden fazla sıfır değerine sahip veri setinin bellek açısından verimli temsilleridir. GaussianNB algoritması, Bayes teoremini kullanarak girdi özelliklerine (makaledeki kelimeler) dayalı olarak her bir sınıfın (sahte veya doğru) olasılığını hesaplar. Matematiksel işlem sırası şu şekildedir:

- **GaussianNB Fonksiyonu:**

$$P\,(sınıf\,|özellikler) = \frac{P(özellikler|sınıf\,)\,\times P(sınıf)}{P\,(özellikler)} \tag{24}$$

Burada;
$P\,(sınıf\,|özellikler)$ Özellikler göz önüne alındığında sınıfın olasılığı (makaledeki kelimeler göz önüne alındığında sahte veya doğru haber).
$P(özellikler|sınıf\,)$ Sınıfa verilen özelliklerin olasılığı (haberin sahte veya doğru olduğu göz önüne alındığında kelimeleri gözlemleme olasılığı).
$P(sınıf)$ Sınıfın ön olasılığı (bir haber makalesinin sahte veya doğru olma olasılığı hakkındaki ön inanç).
$P\,(özellikler)$ Özelliklerin olasılığı (makaledeki kelimeleri gözlemleme olasılığı).

➢ **Model Eğitimi:**

- Eğitim aşamasında model, eğitim verilerinden olasılıkları ve önselleri öğrenir.
- Her bir sınıfın olasılıkları, o sınıfa ait belgelerdeki her bir sözcüğün sıklığı, bu belgelerdeki toplam sözcük sayısına oranlanarak tahmin edilir.
- Öncüller, eğitim setindeki her bir sınıftaki belgelerin oranı belirlenerek tahmin edilebilir.

➢ **Tahmin:**

Model eğitildikten sonra, Bayes teoremini kullanarak makaledeki kelimelere dayalı olarak her sınıfın olasılıklarını hesaplayarak yeni haber makalelerini tahmin edebilir. Makale için tahmin edilen sınıf, en yüksek olasılığa sahip olandır.

Matematiksel olarak tahmin süreci, özelliklere dayalı olarak her bir sınıfın olasılığının hesaplanmasını ve en yüksek olasılığa sahip sınıfın seçilmesini içerir.

$$\hat{y} = argmax_c \left(P(sınıf = c|özellikler)\right) \quad (25)$$

Burada;

$\hat{y}$ Tahmin edilen sınıf (sahte veya doğru).

c Sınıf (sahte veya doğru).

$P(sınıf = c|özellikler)$ Özellikler verilerine alındığında sınıfın olasılığıdır.

Son olarak, modelin doğruluğu, ayrı bir test seti kullanılarak tahmin edilen etiketlerin gerçek etiketlerle karşılaştırılmasıyla değerlendirilebilir.

GaussianNB algoritması, olasılıkları ve istatistiksel ilkeleri kullanarak haber makalelerini içeriklerine göre sahte veya gerçek olarak sınıflandırmak için basit ve etkili bir yöntem sunar.

3.6.3.3. Destek vektör makinesi

Geliştirilen algoritma, Python'da scikit-learn kütüphanesi kullanılarak uygulanan bir SVM sınıflandırıcısıdır. SVM, sınıflandırma ve regresyon görevleri için kullanılan bir algoritmadır. Sahte haber tespiti bağlamında, gerçek ve aldatıcı haber makalelerini ayırt etmede çok önemli

bir rol oynar. Algoritma, veri setinin (X), (train_test_split) fonksiyonu kullanılarak eğitim ve test kümelerine ayrıldığı bir eğitim ve test prosedürü izler. Daha sonra SVM modeli başlatılır ve fit yöntemi kullanılarak eğitim verileri (X_train, y_train) üzerinde eğitilir. Eğitimden sonra, modelin performansı puan yöntemi kullanılarak test verileri (X_test, y_test) üzerinde değerlendirilir ve test doğruluğu yazdırılır.

Modelin dil anlayışını geliştirmek için NLP teknikleri dahil edilmiştir. Ön işleme adımları, lemmatizasyon için WordNetLemmatizer, kelime yerleştirmeleri için Word2Vec ve metin verilerini sayısal vektörlere dönüştürmek için CountVectorizer kullanılmasını içerir. Word2Vec, anlamsal anlamları yakalayarak bağlamsal ilişkilerine dayalı olarak kelimelere vektör temsilleri atar. WordNetLemmatizer kelimeleri temel veya kök biçimlerine indirger. CountVectorizer işlenen metni sayısal bir biçime dönüştürerek kelime sayılarından oluşan bir matris oluşturur. SVM modeli daha sonra kalıpları öğrenmek ve tahminlerde bulunmak için bu sayısal temsillerden yararlanarak haber makalelerini gerçek veya sahte olarak etkili bir şekilde sınıflandırır. NLP teknikleri ve SVM'nin SVM bu kombinasyonu, algoritmanın ince dilsel nüansları ayırt etme ve güvenilir ve aldatıcı bilgileri ayırt etme doğruluğunu artırma yeteneğine katkıda bulunur.

Sonuç olarak, bu araştırmada sunulan çeşitli modeller toplu olarak sahte haber tespitine yönelik metodolojilerin ilerlemesine katkıda bulunmaktadır. Her model, LR'un yorumlanabilirliğinden RF ve XGBoost gibi topluluk yöntemlerinin sağlamlığına ve DL de LSTM sıralı işleme yeteneklerine kadar kendine özgü güçlü yönlerini ortaya koymaktadır. NLP tekniklerinin dahil edilmesi, modellerin ince dilsel ipuçlarını ayırt etme yeteneğini daha da zenginleştirmektedir. Takip eden tartışmalar, her modelin matematiksel temellerini ve uygulamalarını inceleyerek, doğru sahte haber sınıflandırmasının genel hedefine bireysel katkılarına ışık tutmaktadır. Matematiksel işlem sırası şu şekildedir:

- **SVM Fonksiyonu:**

SVM algoritması, özellik uzayında farklı sınıflara ait veri noktalarını en iyi şekilde ayıran hiper düzlemi bulmayı amaçlar. SVM için model fonksiyonu şu şekilde gösterilebilir:

$$f(x) = sign\ (\sum_{i=1}^{n} w_i\ .\ x_i + b) \tag{26}$$

Burada:

$f(x)$ verilen bir x giriş vektörü için sınıf etiketini tahmin eden karar fonksiyonudur.

n özellik sayısıdır.

w_i her bir özelliğe atanan ağırlıklardır.

x_i giriş özellikleridir.

b önyargı terimidir.

$sign$ girdi pozitifse +1, negatifse -1 ve sıfırsa 0 değerini döndüren işaret fonksiyonudur.

- **Model Eğitimi:**

Eğitim aşamasında, SVM algoritması, farklı sınıfların destek vektörleri (karar sınırına en yakın veri noktaları) arasındaki marjı maksimize eden hiper düzlemi optimize eder. Bu optimizasyon problemi kısıtlı bir optimizasyon problemi olarak formüle edilebilir:

$$minimize\ \frac{1}{2}\ ||w||^2 \tag{27}$$

Kısıtlamalara tabidir:

$$y_i(w_i\ .\ x_i + b) \geq 1\ \ for\ i = 1\ ,2\ ,\ .\ .\ .\ \ ,n \tag{28}$$

Burada, w ve b hiper düzlemin parametreleridir ve y_i , i^{th} eğitim örneğinin sınıf etiketidir.

- **Tahmin:**

SVM modeli eğitildikten sonra, yeni veri noktaları üzerinde tahminler yapmak için kullanılabilir. Yeni bir girdi vektörü x_{test} verildiğinde, karar fonksiyonunun işareti değerlendirilerek tahmin yapılır:

$$Tahmin = sign\ (\sum_{i=1}^{n} w_i\ .x_{test} + b) \tag{29}$$

Tahmin pozitif ise veri noktası bir sınıfa, negatif ise diğer sınıfa aittir.

Bu matematiksel çerçeve, gerçek veya sahte haberleri sınıflandırmak için en uygun hiper düzlemi bulmada SVM algoritmasına rehberlik eder.

3.7. Değerlendirme Metrikleri

3.7.1. Doğruluk

Veri setinin sınıf dağılımı dengeli olduğundan, doğruluk, modelin pozitif ve negatif sınıfları tespit etme becerisini değerlendirmek için uygun bir ölçüttür. Doğruluk, Gerçek Pozitif ve Gerçek Negatif toplamının veri kümesindeki toplam örnek sayısına bölünmesiyle hesaplanır.

$$Doğruluk = \frac{TP+TN}{TP+TN+FP+FN} \quad (30)$$

Burada;
TP Gerçek Pozitif sayısını temsil eder.
FP Yanlış Pozitif sayısını temsil eder.
FN Yanlış Negatif sayısını temsil eder.
TN Gerçek Negatif sayısını temsil eder.

3.7.2. Kesinlik

Kesinlik, pozitif sınıf sınıflandırmasının doğruluğunu ölçen bir metriktir. Doğru Pozitiflerin sayısının Doğru Pozitifler ve Yanlış Pozitiflerin toplamına bölünmesiyle hesaplanır.

$$Kesinlik = \frac{\boldsymbol{TP}}{\boldsymbol{TP+FP}} \quad (31)$$

3.7.3. Duyarlılık

Duyarlılık, bir sınıflandırma modelinin etkinliğini değerlendirmek için kullanılan istatistiksel bir ölçüdür. Modelin belirli bir veri kümesindeki ilgili tüm örnekleri doğru bir şekilde tanımlama ve bunları "gerçek" ya da "gerçek değil" olarak sınıflandırma yeteneğini değerlendirir.

$$Duyarlılık = \frac{TP}{TP+FN} \quad (32)$$

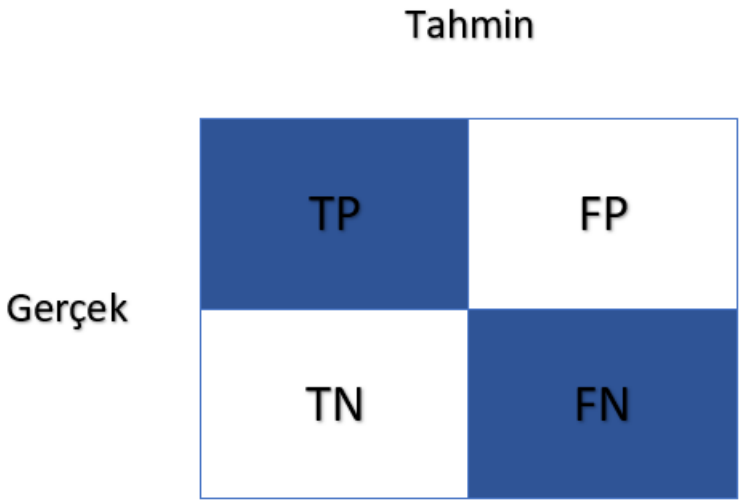
3.7.4. F1- Puan

F1-puanı, bir modelin hassasiyet ve geri çağırma değerlerinin harmonik ortalaması veya ağırlıklı ortalamasıdır. Yanlış negatifler ve yanlış pozitifler, doğru pozitifler ve doğru negatiflerden daha kritik olduğunda daha uygun bir doğruluk ölçüsüdür.

$$F1 = 2 \times \frac{Kesinlik \times Duyarlılık}{Kesinlik + Duyarlılık} \quad (33)$$

3.7.5. Karışıklık Matrisi

Algoritmaların sınıflandırma performansı karışıklık matrisi kullanılarak değerlendirilmiştir. Şekil 11'de gösterilen matris, modelin tahminlerinin ayrıntılı bir dökümünü sunmaktadır. Buradaki ifadeler (30) nolu denklemdeki ifadelerdir.



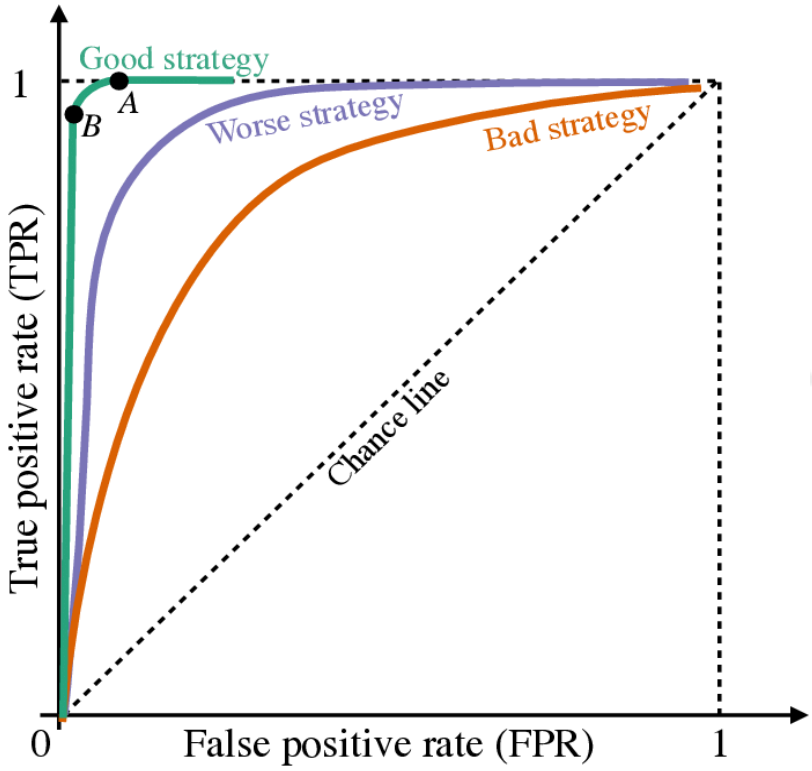
Şekil 11. Karışıklık Matrisi

3.7.6. ROC eğrisi

ROC eğrisi, gerçek pozitif oranı (TPR) ile yanlış pozitif oranı (FPR) karşılaştırarak ikili bir sınıflandırıcının etkinliğini değerlendirmek için kullanılan bir yöntemdir. TPR (y ekseni) doğru şekilde tanımlanan gerçek pozitiflerin oranını gösterirken, FPR (x ekseni) yanlış şekilde pozitif olarak sınıflandırılan gerçek negatiflerin oranını temsil eder. ROC eğrisi, çeşitli sınıflandırıcı stratejilerini temsil eden çizgiler içerir. Sol üst çeyrekte yer alan ve "iyi bir stratejiyi" temsil

eden bir çizgi, yüksek TPR ve düşük FPR'ye işaret ederek etkili bir sınıflandırıcıya işaret eder. Bu idealden sapan bir çizgi "daha kötü bir stratejiyi" temsil eder ve köşegeni takip eden bir çizgi, rastgele tahminden daha iyi performans göstermeyen bir sınıflandırıcıyı gösteren "kötü bir stratejiyi" temsil eder (Şekil 12).

ROC eğrisini analiz ederek, farklı sınıflandırıcıların etkinliği karşılaştırılabilir ve performanslarına göre en iyisi seçilebilir. Özetle, ROC eğrisinin bir modelin pozitif ve negatif sınıflar arasında ayrım yapma yeteneğini değerlendirir. Daha yüksek bir ROC daha iyi performansa işaret eder (Festus Ayetiran & Özgöbek, n.d.).



Şekil 12. ROC Eğrisi

4. BULGULAR

Gerçek ve sahte haberlerin sınıflandırılması için çeşitli ML ve DL modelleri ile yapılan deneylerden elde edilen sonuçlar bu bölümde sunulmakta ve tartışılmaktadır. Modeller arasında LSTM gibi DL modelleri, GaussianNB, K-NN, SVM gibi Doğal Dil İşleme NLP modelleri ve LR, DT, GBoost, RF ve XGBoost gibi ML modelleri bulunmaktadır.

4.1. Makine Öğrenimi Yöntem Sonuçları

Haber sınıflandırma alanında, bir dizi ML modeli performansları açısından değerlendirilmiş ve sonuçlar tüm modellerde yüksek düzeyde doğruluk ve F1 puanları göstermiştir. Özellikle LR modeli %98,63 bir doğruluk ve %98,57 bir F1 puanı göstererek görevlerin sınıflandırılmasında güvenilirliğini ve faydasını kanıtlamıştır. RF modeli de %98,92 doğruluk ve %98,87 F1 puanı ile övgüye değer bir performans sergilemiştir.

DT, GBoost ve XGBoost algoritmaları dahil olmak üzere toplu öğrenme yaklaşımları üstün performans ölçütleri sergilemiştir. DT modeli %99,61 bir doğruluk ve %99,59 bir F1 puanı sergilemiştir. GBoost modeli %99,63 doğruluk ve %99,62 F1 punaı ile küçük bir avantaj sergilemiştir. XGBoost modeli %99,81 F1 puanı ve %99,82 doğruluk oranıyla bu incelemede en iyi performansı göstermiştir. Bu örnek performans, modelin verilerdeki karmaşık örüntüleri etkili bir şekilde ayırt etme ve ele alma yeteneğini ve kapasitesini göstermektedir.

K-NN modeli %81,63 bir doğruluk ve %87,85 bir F1 punaı sergilemektedir ki bunlar NLP alanında gözlemlenen en yüksek değerler arasındadır. K-NN modeli tatmin edici bir performans sergilemesine rağmen, SVM ve GaussianNB gibi daha sofistike modeller daha üstün sonuçlar sergilemektedir. GaussianNB modeli %90,22 doğruluk ve %89,66 F1 puanı ile orta düzeyde bir performans sergilemektedir. Bu durum, modelin haber öykülerindeki karmaşık dilsel örüntüleri yakalama kapasitesinin sınırlı olduğunu göstermektedir.

Buna karşılık, SVM %98,40 doğruluk oranı ve %98,27 F1 puanı ile yüksek bir doğruluk derecesi sergilemektedir. Bu, gerçek ve sahte haberleri etkili bir şekilde ayırt etme kapasitesini göstermektedir. SVM modelinin üstün performansı, yüksek boyutlu verileri işleme kapasitesi,

marjları optimize etme, çekirdek hilesini kullanma, aşırı uyuma direnme ve destek vektörlerinden yararlanma gibi bir dizi temel özellik ve özelliğe atfedilebilir. Bu özellikler toplu olarak SVM'nin haber sınıflandırma alanındaki örnek performansına katkıda bulunmakta ve onu gerçek ve dezenformasyon arasında doğru ve kesin bir ayrım yapmak için son derece güvenilir bir model haline getirmektedir.

Bu modellerin, özellikle de XGBoost'un kayda değer başarısı, hassas haber sınıflandırması için gerekli olan ince dilsel nüansları yakalamada usta olan TF-IDF temsillerini kullanma kapasitelerine bağlanabilir. Başta XGBoost olmak üzere toplu öğrenme yöntemleri, karmaşık veri ilişkileriyle karşı karşıya kaldıklarında kayda değer avantajlar sergileyerek kesin ve güvenilir kategorizasyon gerektiren uygulamalar için en uygun hale gelirler.

4.2. Derin Öğrenme Yöntem Sonuçları

LSTM modeli, %99,31 doğruluk ve %99,29 F1 puanı elde ederek olağanüstü bir performans sergilemiştir. DL yaklaşımı, LSTM ağlarının ve Word2Vec katıştırmalarının sıralı işleme yeteneklerinden yararlanarak haber makalelerindeki karmaşık örüntüleri yakalamaktadır. Yüksek doğruluk ve F1 puanı, modelin gerçek ve aldatıcı haberler arasındaki farkı ayırt etmedeki etkinliğini vurgulamaktadır.

4.3. Karşılaştırmalı Analiz

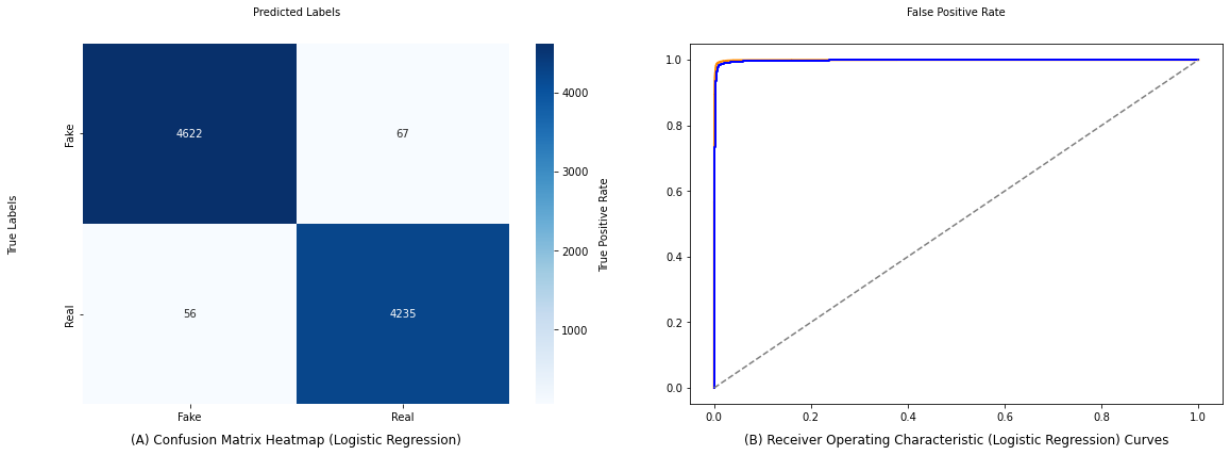
Tablo-1'deki gibi DL, NLP ve ML modelleri arasındaki sonuçlar karşılaştırıldığında, ML modellerinin, doğruluk ve F1 puanları açısından diğer yaklaşımlardan daha iyi performans gösterdiği açıktır. Yanlış haberlerin tespitine yönelik modeller incelendiğinde, ML yöntemlerinin, özellikle de XGBoost, GBoost ve DT gibi topluluk yöntemlerinin, diğer modellerle karşılaştırıldığında doğruluk ve F1 puanları açısından üstün performans gösterdiği görülmektedir. XGBoost en iyi performansı gösterirken, onu GBoost ve DT takip etmektedir. Bu birleşik modeller, çok sayıda yalın öğreniciyi entegre etme kapasiteleri ile ayırt edilir ve böylece metinsel verilerdeki karmaşık kalıpların yakalanmasını sağlar. TF-IDF temsillerinin kullanılması, gerçek ve sahte haberlerin ayırt edilmesinde kritik öneme sahip olan en bilgilendirici kelime ve ifadelere odaklanarak modelin etkinliğini artırmaya hizmet etmektedir.

Güçlü bir NLP modeli olan SVM de güçlü bir performans göstererek haber makalelerindeki dilsel kalıpları yakalamadaki etkinliğini ortaya koymuştur. DL yaklaşımlarından olan LSTM ağları da iyi sonuçlar elde etmiştir, bu daha fazla hesaplama kaynağı gerektirmesine rağmen DL modellerinin de oldukça etkili olabileceğini göstermektedir. Öte yandan K-NN ve GaussianNB gibi daha basit modellerin az gelişmiş olması ve K-NN performansının daha düşük olması, bu yöntemlerin sahte verilerin karmaşıklığına pek uygun olmayabileceğini düşündürmektedir. Bunun nedeni, modelin metinlerdeki karmaşık dilsel yakalama yeteneğinin sınırlı olmasıdır. Tablo1'deki veriler karşılaştırıldığında ML modellerinin daha iyi performans gösterdiği görülmektedir.

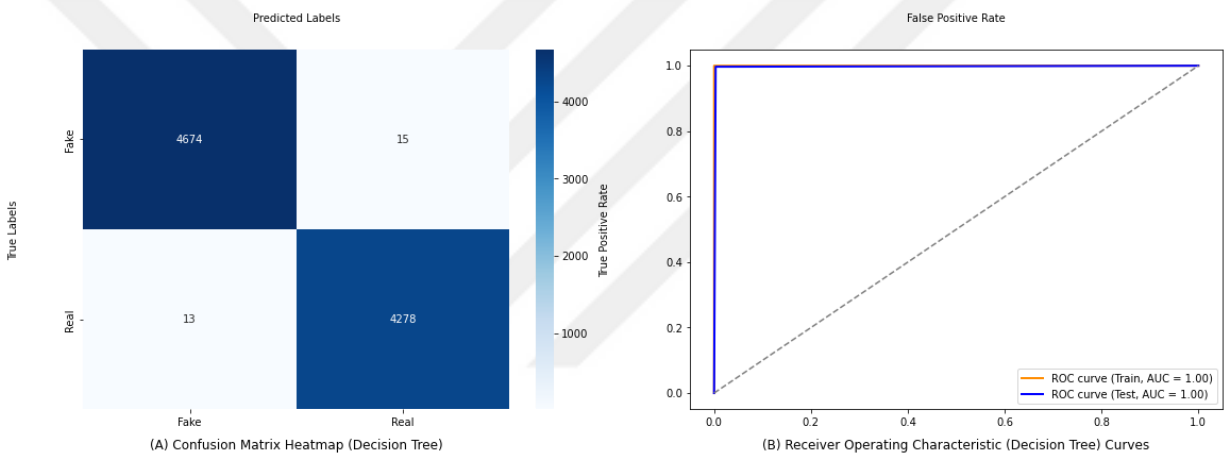
Tablo 1. Uygulanan modellerin başarı oranları

Model	Doğruluk	F1-puanı
LR	98.63%	98.57%
DT	99.61%	99.59%
GBoost	99.63%	99.62%
RF	98.92%	98.87%
XGBoost	**99.82%**	**99.81%**
LSTM	99.31%	99.29%
SVM	98.40%	98.27%
K-NN	81.63%	78.85%
GaussianNB	90.22%	89.66%

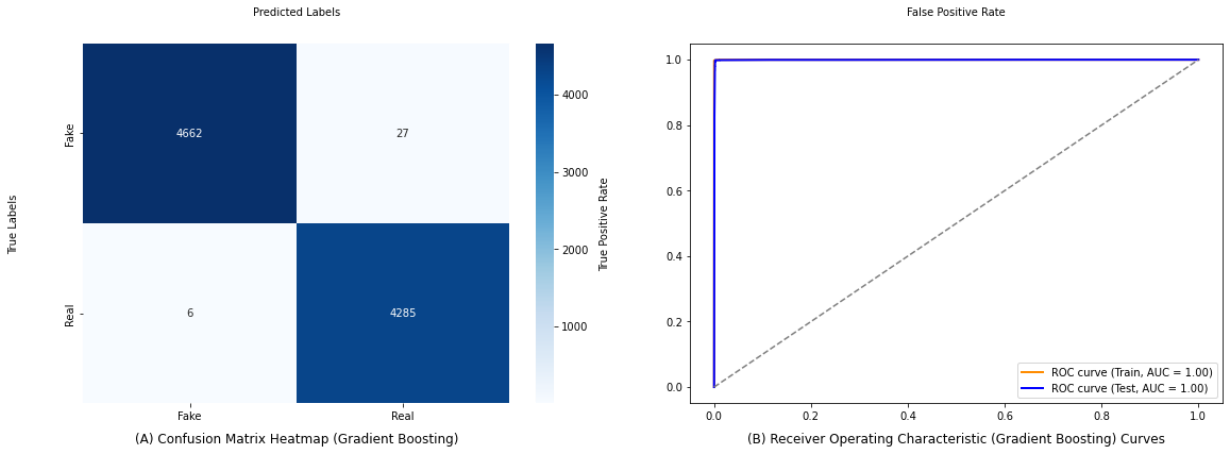
Özetle, deneysel sonuçlar, ML tekniklerinin, özellikle de toplu öğrenme modellerinin bir kombinasyonunun, gerçek ve sahte haberlerin sınıflandırılmasında üstün performans sağladığını göstermektedir. Daha fazla araştırma ile bu kritik görevde daha da yüksek doğruluk elde etmek için model yorumlanabilirliği, özellik önem analizi ve DL mimarilerindeki potansiyel geliştirmeler keşfedilebilir. Çalışılan modeller için karışıklık matrisi ve ROC eğrileri Şekil 13-21'de gösterilmiştir.



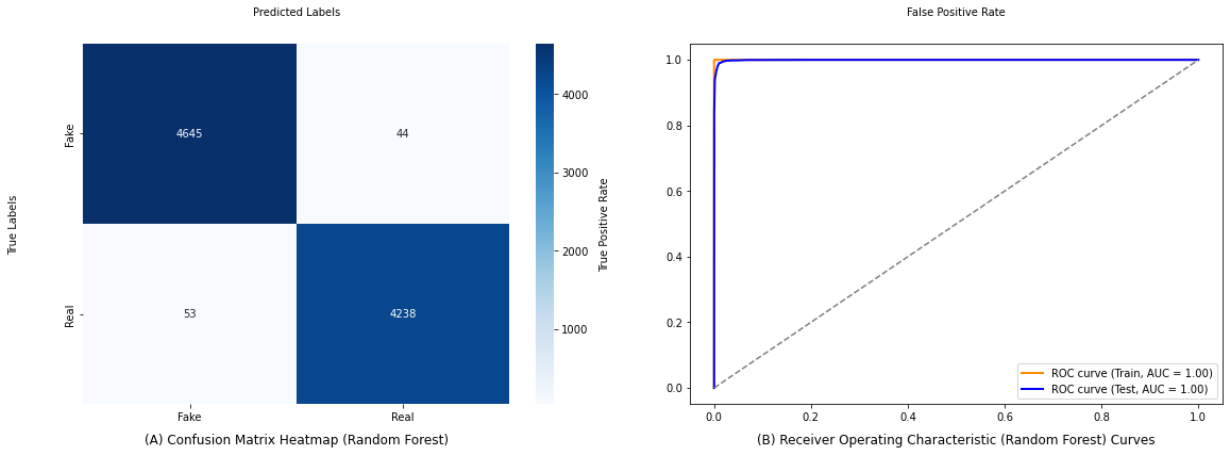
Şekil 13. (A) LR için karışıklık matrisi ve (B) ROC eğrisi



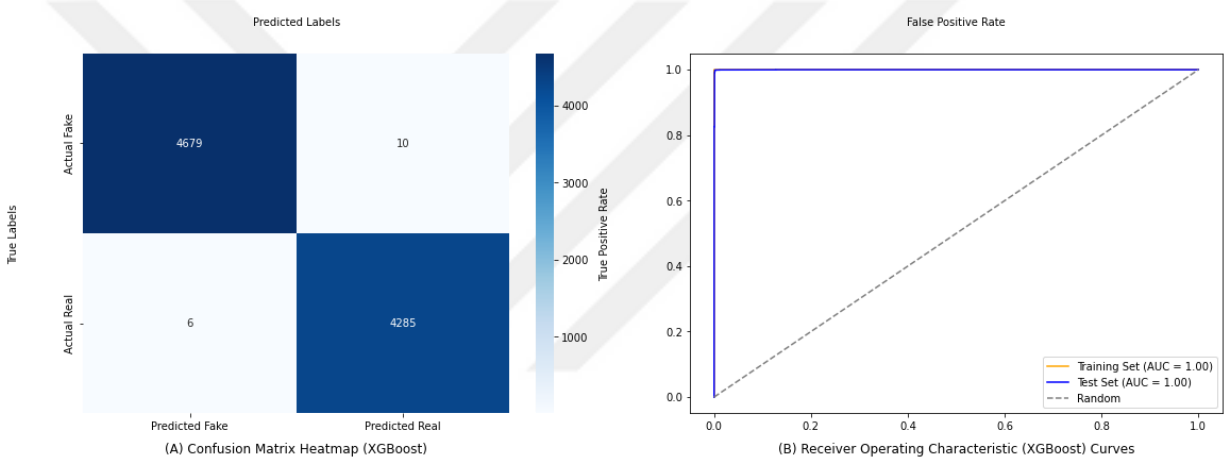
Şekil 14. (A) DT için karışıklık matrisi ve (B) ROC eğrisi



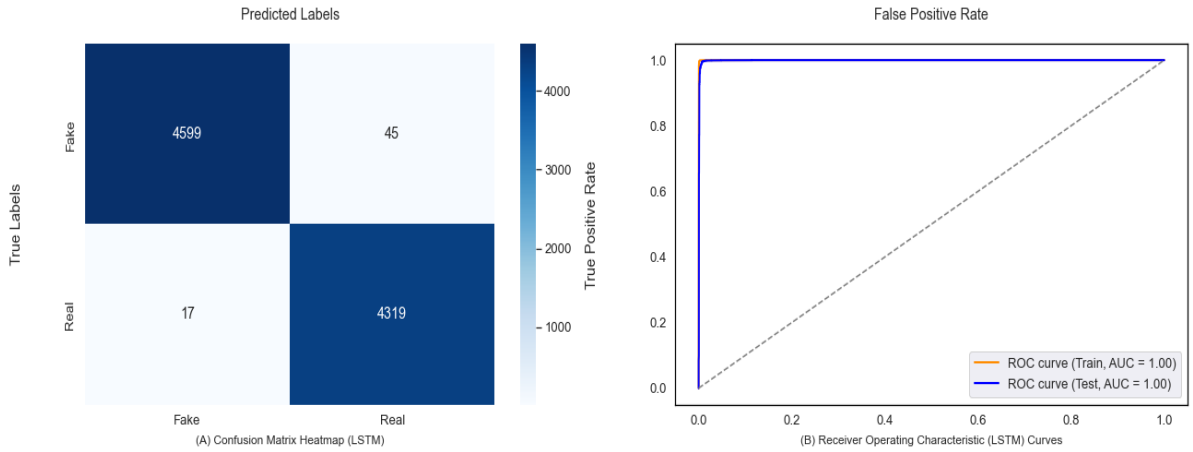
Şekil 15. (A) GBoost için karışıklık matrisi ve (B) ROC eğrisi



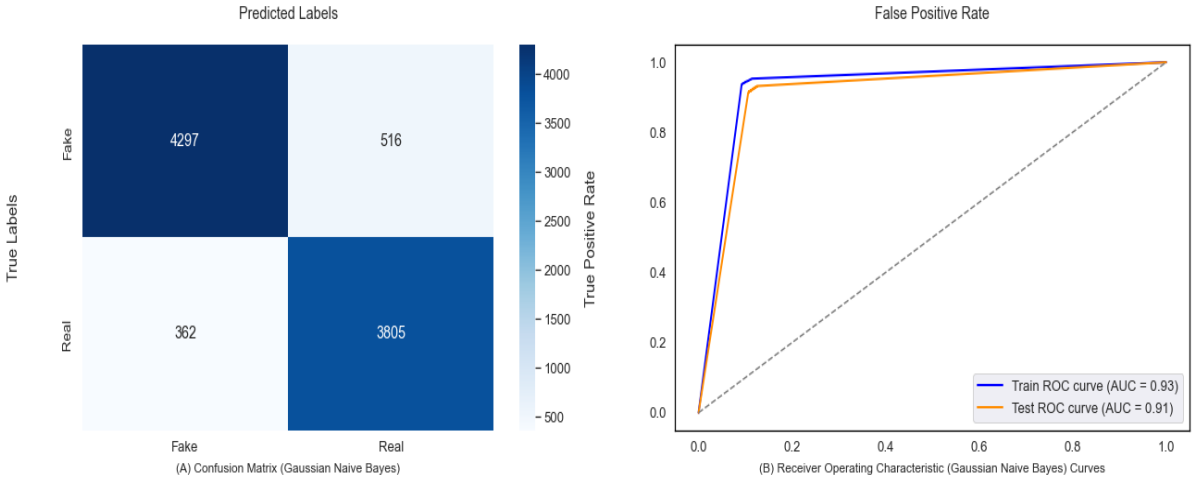
Şekil 16. (A) RF için karışıklık matrisi ve (B) ROC eğrisi



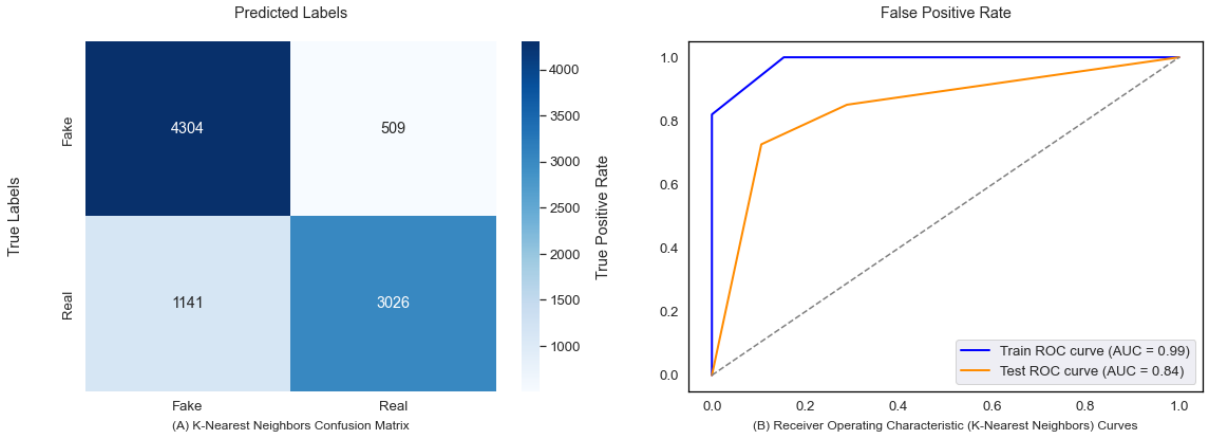
Şekil 17. (A) XGBoost için karışıklık matrisi ve (B) ROC eğrisi



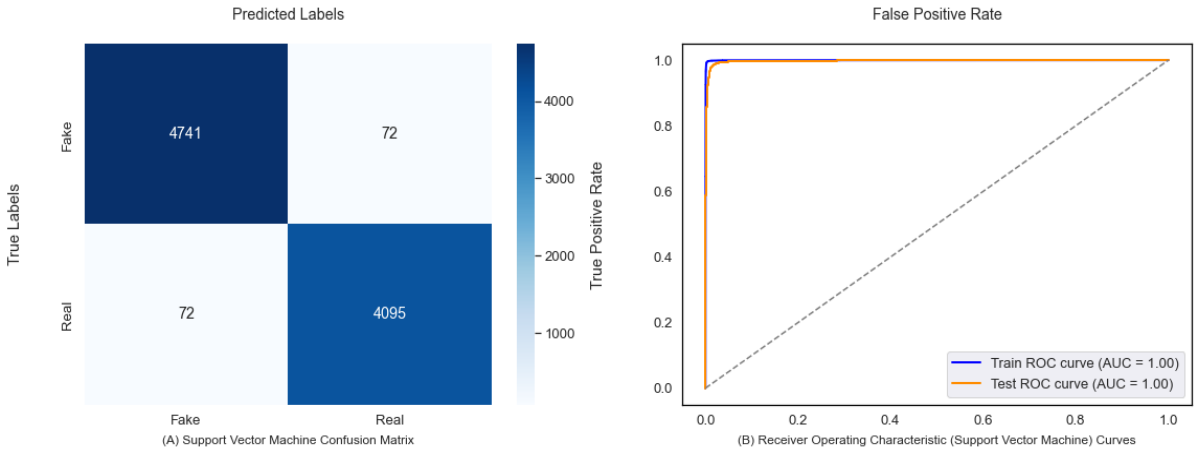
Şekil 18. (A) LSTM için karışıklık matrisi ve (B) ROC eğrisi



Şekil 19. (A) GaussianNB için karışıklık matrisi ve (B) ROC eğrisi



Şekil 20. (A) K-NN için karışıklık matrisi ve (B) ROC eğrisi



Şekil 21. (A) SVM için karışıklık matrisi ve (B) ROC eğrisi

5. SONUÇ VE ÖNERİLER

Bu araştırmada, Twitter'daki yanlış bilgilerin tespiti için çeşitli AI tekniklerinin kapsamlı bir incelemesini ve değerlendirmesini yapıldı. Çalışmada LSTM, SVM, RF Sınıflandırıcı, K-NN, Multinomial GaussianNB ve GBoost Sınıflandırıcı dahil olmak üzere farklı modeller kullanılmaktadır, aldatıcı içeriği güvenilir bilgilerden ayırt etmeyi amaçladı. Tweetler, retweetler ve kullanıcı etkileşimleri dahil olmak üzere Twitter verilerinin kapsamlı analizi yoluyla ve DL, NLP ve ML tekniklerini kullanarak, her bir AI modelinin güçlü yönlerini ve sınırlamalarını değerlendirildi. Bulgular, bu modellerin Twitter'da yanlış bilgi yayma, aldatıcı hesaplar ve yanlış anlatı kalıplarını belirlemede etkili olduğunu gösteriyor. AI tabanlı yanlış bilgi tespitini çevreleyen etik hususlar da ele alınmış ve mahremiyetin, önyargıların azaltılmasının ve ifade özgürlüğünün korunmasının önemi vurgulanmıştır. Yanlış bilgiyle mücadele etmek ve bireysel hakları korumak arasında bir denge kurmak, sorumlu AI uygulaması için çok önemlidir.

Araştırmada, Twitter yanlış bilgi tespiti için AI tekniklerinin performansına ilişkin değerli içgörüler sunmaktadır. Bu bulgular, AI ve sosyal medya analizi alanındaki akademisyenler, araştırmacılar ve uygulayıcılar için bir kaynak niteliğindedir. Ayrıca, Twitter'da paylaşılan bilgilerin bütünlüğünü korumayı taahhüt eden sosyal medya şirketlerine ve politika yapıcılara pratik rehberlik sağlarlar.

Bu araştırma, AI kullanarak Twitter yanlış bilgilerini anlama ve ele alma konusunda önemli adımlar atmış olsa da, gelecekte keşfedilecek ve geliştirilecek birkaç yol vardır:

- Twitter'daki yanlış bilginin değişen doğasına uyum sağlayabilecek modeller geliştirebilir.
- Modellerin aldatıcı taktiklerdeki hızlı değişiklikleri ele alma yeteneğini geliştirmek için gerçek zamanlı öğrenme mekanizmalarını dahil etmek.
- Çok modlu sahte haber tespiti üzerine yapılan araştırmalar; metinsel içerik, ses, video ve görüntü dosyalarının yanlış bilgi yayabileceğini ortaya koymaktadır. Gelecekteki araştırmalar, yanlış bilgi tespit hassasiyetini ve esnekliğini artırmak için ses işleme, NLP ve bilgisayarla görme tekniklerini kullanarak bu modaliteleri birleştirebilir (Algabri et al., 2024).

- Kullanıcı davranışlarını ve etkileşim modellerini anlamak için kullanıcı merkezli özellikler ve etkileşimler modellere dahil edilmelidir; bu da aldatıcı hesapların daha doğru bir şekilde tespit edilmesine katkıda bulunabilir.
- Yanlış bilgilendirmeyi çevreleyen sosyo-kültürel konular hakkında daha derin bir anlayış geliştirmek için sosyal bilimcilerle birlikte çalışmak. Sosyal bilim perspektiflerini entegre etmek, AI modellerinin analizini ve etkinliğini geliştirebilir.
- Kullanıcı gizliliği endişelerini gidermek için gelişmiş gizlilik koruma tekniklerini keşfetmek.
- Ayrıca, araştırmanın kapsamının farklı coğrafi bölgeleri ve dilleri içerecek şekilde genişletilmesi. Yanlış bilgilendirme kültürler arasında farklı şekillerde ortaya çıkmaktadır ve modellerin bu farklılıklara duyarlı olması gerekmektedir.

Etkili yanlış bilgi tespiti ile bireylerin mahremiyet hakkının korunması arasında bir denge kurmaya çalışılmalı. Bu, çeşitli yollar düşünülerek ve keşfedilerek başarılabilir. Bunu yaparak daha bilinçli ve dirençli bir dijital toplumu teşvik edilebilir.

KAYNAKLAR

Abdullah-All-Tanvir, Mahir, E. M., Akhter, S., & Huq, M. R. (2019). *2019 7th International Conference on Smart Computing & Communications (ICSCC)*.

Ahmad, I., Yousaf, M., Yousaf, S., & Ahmad, M. O. (2020). Fake News Detection Using Machine Learning Ensemble Methods. *Complexity*, *2020*. https://doi.org/10.1155/2020/8885861

Aïmeur, E., Amri, S., & Brassard, G. (2023). Fake news, disinformation and misinformation in social media: a review. *Social Network Analysis and Mining*, *13*(1). https://doi.org/10.1007/s13278-023-01028-5

Algabri, M., Abu Huliqah, E. N. A., Ghurab, M., Al-Khulaidi, A. A. G., & Al Gaphari, G. H. (2024). Fake News Detection On Social Media:Review of Literature. *مجلة جامعة صنعاء للعلوم التطبيقية والتكنولوجيا*, *2*(1), 7–15. https://doi.org/10.59628/jast.v2i1.369

Ali, A. M., Ghaleb, F. A., Al-Rimy, B. A. S., Alsolami, F. J., & Khan, A. I. (2022). Deep Ensemble Fake News Detection Model Using Sequential Deep Learning Technique. *Sensors*, *22*(18). https://doi.org/10.3390/s22186970

Al-Jarrah, O. Y., Yoo, P. D., Muhaidat, S., Karagiannidis, G. K., & Taha, K. (n.d.). *Efficient Machine Learning for Big Data: A Review*. http://cera.wdc-climate.de

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. In *Journal of Economic Perspectives* (Vol. 31, Issue 2, pp. 211–236). American Economic Association. https://doi.org/10.1257/jep.31.2.211

Alyoubi, S., Kalkatawi, M., & Abukhodair, F. (2023). The Detection of Fake News in Arabic Tweets Using Deep Learning. *Applied Sciences (Switzerland)*, *13*(14). https://doi.org/10.3390/app13148209

Alzubi, J., Nayyar, A., & Kumar, A. (2018). Machine Learning from Theory to Algorithms: An Overview. *Journal of Physics: Conference Series*, *1142*(1). https://doi.org/10.1088/1742-6596/1142/1/012012

Aslam, N., Ullah Khan, I., Alotaibi, F. S., Aldaej, L. A., & Aldubaikil, A. K. (2021). Fake Detect: A Deep Learning Ensemble Model for Fake News Detection. *Complexity*, *2021*. https://doi.org/10.1155/2021/5557784

Asta, R. S., & Budi Setiawan, E. (2023). Fake News (Hoax) Detection on Social Media Using Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) Methods.

International Conference on ICT Convergence, *2023-August*, 511–516. https://doi.org/10.1109/ICoICT58202.2023.10262617

Aziz, N., Akhir, E. A. P., Aziz, I. A., Jaafar, J., Hasan, M. H., & Abas, A. N. C. (2020). A Study on Gradient Boosting Algorithms for Development of AI Monitoring and Prediction Systems. *2020 International Conference on Computational Intelligence, ICCI 2020*, 11–16. https://doi.org/10.1109/ICCI51257.2020.9247843

Baptista, J., & Gradim, A. (2022). A Working Definition of Fake News. *Encyclopedia*, *2*(1), 632–645. https://doi.org/10.3390/encyclopedia2010043

Burgers, N., Imaad, T., & Aladeen, H. (2023). *Can Machine Learning Algorithms Really Stop Fake News in its Tracks?*

Cao, J., Qi, P., Sheng, Q., Yang, T., Guo, J., & Li, J. (2020). *Exploring the Role of Visual Content in Fake News Detection*. https://doi.org/10.1007/978-3-030-42699-6

Charbuty, B., & Abdulazeez, A. (2021). Classification Based on Decision Tree Algorithm for Machine Learning. *Journal of Applied Science and Technology Trends*, *2*(01), 20–28. https://doi.org/10.38094/jastt20165

Choudhary, A., & Arora, A. (2021). Linguistic feature based learning model for fake news detection and classification. *Expert Systems with Applications*, *169*. https://doi.org/10.1016/j.eswa.2020.114171

Collins, B., Hoang, D. T., Nguyen, N. T., & Hwang, D. (2021). Trends in combating fake news on social media–a survey. *Journal of Information and Telecommunication*, *5*(2), 247–266. https://doi.org/10.1080/24751839.2020.1847379

Cui, L., Shu, K., Wang, S., Lee, D., & Liu, H. (2019). dEFEND: A system for explainable fake news detection. *International Conference on Information and Knowledge Management, Proceedings*, 2961–2964. https://doi.org/10.1145/3357384.3357862

Deepa R. (2024). *An Ensemble Learning Frame Work for Robust Fake News Detection Sri Sairam Engineering college*. https://www.researchgate.net/publication/377956221

Dhaliwal, S. S., Nahid, A. Al, & Abbas, R. (2018). Effective intrusion detection system using XGBoost. *Information (Switzerland)*, *9*(7). https://doi.org/10.3390/info9070149

Fang, W., Chen, Y., & Xue, Q. (2021). Survey on Research of RNN-Based Spatio-Temporal Sequence Prediction Algorithms. *Journal on Big Data*, *3*(3), 97–110. https://doi.org/10.32604/jbd.2021.016993

Festus Ayetiran, E., & Özgöbek, Ö. (n.d.). *A review of deep learning techniques for multimodal fake news and harmful languages detection A R T I C L E I N F O Keywords: Fake news Abusive language Deep learning Hate speech Harmful languages Offensive language Toxic language Online trolling Cyberbullying Cyberaggression Extremism Multimedia data fusion*. https://ssrn.com/abstract=4691091

Fong, A. C. M. (2010). Welcome Message from the Editor-in-Chief. *Journal of Advances in Information Technology*, *1*(1). https://doi.org/10.4304/jait.1.1.1-1

Gelfert, A. (2018). Fake news: A definition. *Informal Logic*, *38*(1), 84–117. https://doi.org/10.22329/il.v38i1.5068

George, J., Gerhart, N., & Torres, R. (2021). Uncovering the Truth about Fake News: A Research Model Grounded in Multi-Disciplinary Literature. *Journal of Management Information Systems*, *38*(4), 1067–1094. https://doi.org/10.1080/07421222.2021.1990608

Gifu, D. (2023). An Intelligent System for Detecting Fake News. *Procedia Computer Science*, *221*, 1058–1065. https://doi.org/10.1016/j.procs.2023.08.088

Goldani, M. H., Momtazi, S., & Safabakhsh, R. (2021). Detecting fake news with capsule neural networks. *Applied Soft Computing*, *101*. https://doi.org/10.1016/j.asoc.2020.106991

Goldani, M. H., Safabakhsh, R., & Momtazi, S. (2021). Convolutional neural network with margin loss for fake news detection. *Information Processing and Management*, *58*(1). https://doi.org/10.1016/j.ipm.2020.102418

GÜLER, G., & GÜNDÜZ, S. (2023). Deep Learning Based Fake News Detection on Social Media. *International Journal of Information Security Science*, *12*(2), 1–21. https://doi.org/10.55859/ijiss.1231423

Hailemichael, E. N. (2021). *FAKE NEWS DETECTION FOR AMHARIC LANGUAGE USING DEEP LEARNING ADAMA, ETHIOPIA*.

Hakak, S., Alazab, M., Khan, S., Gadekallu, T. R., Maddikunta, P. K. R., & Khan, W. Z. (2021). An ensemble machine learning approach through effective feature extraction to classify fake news. *Future Generation Computer Systems*, *117*, 47–58. https://doi.org/10.1016/j.future.2020.11.022

Hamed, S. K., Ab Aziz, M. J., & Yaakub, M. R. (2023). A review of fake news detection approaches: A critical analysis of relevant studies and highlighting key challenges

associated with the dataset, feature representation, and data fusion. In *Heliyon* (Vol. 9, Issue 10). Elsevier Ltd. https://doi.org/10.1016/j.heliyon.2023.e20382

Hanshal, O. A., Ucan, O. N., & Sanjalawe, Y. K. (2023). Hybrid deep learning model for automatic fake news detection. *Applied Nanoscience (Switzerland)*, *13*(4), 2957–2967. https://doi.org/10.1007/s13204-021-02330-4

Hu, B., Mao, Z., & Zhang, Y. (2024). An overview of fake news detection: From a new perspective. In *Fundamental Research*. KeAi Communications Co. https://doi.org/10.1016/j.fmre.2024.01.017

Hu, L., Wei, S., Zhao, Z., & Wu, B. (2022). Deep learning for fake news detection: A comprehensive survey. *AI Open*, *3*, 133–155. https://doi.org/10.1016/j.aiopen.2022.09.001

Imran, A., Uddin Ahmed, M., Hussain, S., & Iqbal, J. (n.d.). *Social Engagement Analysis For Detection Of Fake News On Twitter Using Machine Learning* (Vol. 19, Issue 3). http://www.webology.org

Islam, M. R., Liu, S., Wang, X., & Xu, G. (2020). Deep learning for misinformation detection on online social networks: a survey and new perspectives. In *Social Network Analysis and Mining* (Vol. 10, Issue 1). Springer. https://doi.org/10.1007/s13278-020-00696-x

Jain, A., Shakya, A., Khatter, H., & Gupta, A. K. (2019, September 1). A smart System for Fake News Detection Using Machine Learning. *IEEE International Conference on Issues and Challenges in Intelligent Computing Techniques, ICICT 2019*. https://doi.org/10.1109/ICICT46931.2019.8977659

Jia, F. (2020). Misinformation Literature Review: Definitions, Taxonomy, and Models. *International Journal of Social Science and Education Research*, *3*, 2020. https://doi.org/10.6918/IJOSSER.202012_3(12).0011

John, A., & Journals, S. (2022). *Fake News Detection Using N-Gram Analysis and Machine Learning Algorithms*. https://doi.org/10.37591/JoMCCMN

Kaggle. (2022, April 5). *fake and real news dataset ISOT*. https://www.kaggle.com/datasets/clmentbisaillon/fake-and-real-news-dataset

Khan, A., Sohail, A., Zahoora, U., & Qureshi, A. S. (2020). A survey of the recent architectures of deep convolutional neural networks. *Artificial Intelligence Review*, *53*(8), 5455–5516. https://doi.org/10.1007/s10462-020-09825-6

Khan, J. Y., Khondaker, Md. T. I., Afroz, S., Uddin, G., & Iqbal, A. (2021). A benchmark study of machine learning models for online fake news detection. *Machine Learning with Applications*, *4*, 100032. https://doi.org/10.1016/j.mlwa.2021.100032

Khanam, Z., Alwasel, B. N., Sirafi, H., & Rashid, M. (2021). Fake News Detection Using Machine Learning Approaches. *IOP Conference Series: Materials Science and Engineering*, *1099*(1), 012040. https://doi.org/10.1088/1757-899x/1099/1/012040

Kondamudi, M. R., Sahoo, S. R., Chouhan, L., & Yadav, N. (2023). A comprehensive survey of fake news in social networks: Attributes, features, and detection approaches. In *Journal of King Saud University - Computer and Information Sciences* (Vol. 35, Issue 6). King Saud bin Abdulaziz University. https://doi.org/10.1016/j.jksuci.2023.101571

Lakshmikumar, P. C., Vinay, R., Setty, J., & Mishra, R. (2019). *Fake News Detection-A Deep Neural Network Vinay Jayarama Setty Rahul Mishra*.

Li, J., & Lei, M. (2022). A Brief Survey for Fake News Detection via Deep Learning Models. *Procedia Computer Science*, *214*(C), 1339–1344. https://doi.org/10.1016/j.procs.2022.11.314

Limon, F. A., & Jahan, N. (n.d.). *FAKE NEWS DETECTION USING DEEP LEARNING*.

Melvern, A., Sibaroni, Y., & Prasetiyowati, S. S. (2023). Fake News Detection: Hybrid Deep Supervised Learning Approach. *2023 International Conference on Data Science and Its Applications, ICoDSA 2023*, 414–419. https://doi.org/10.1109/ICoDSA58501.2023.10277274

Mjaaland, H. L., Vinay, R., Setty, J., & Anand, A. (2020). *Detecting Fake News and Rumors in Twitter Using Deep Neural Networks Vinay Jayarama Setty*.

Nawar, S., & Mouazen, A. M. (2017). Comparison between random forests, artificial neural networks and gradient boosted machines methods of on-line Vis-NIR spectroscopy measurements of soil total nitrogen and total carbon. *Sensors (Switzerland)*, *17*(10). https://doi.org/10.3390/s17102428

Pennycook, G., & Rand, D. G. (2021). The Psychology of Fake News. In *Trends in Cognitive Sciences* (Vol. 25, Issue 5, pp. 388–402). Elsevier Ltd. https://doi.org/10.1016/j.tics.2021.02.007

Prameela, M., Karnataka, U., Deekshith, I., Punith, I., & Balu, R. (2024). Advancements in Machine Learning for Combatting Misinformation: A Comprehensive Analysis of Fake

News Detection Strategies. In *International Journal of Innovative Science and Research Technology* (Vol. 9, Issue 1). www.ijisrt.com313

Reddy, H., Raj, N., Gala, M., & Basava, A. (2020). Text-mining-based Fake News Detection Using Ensemble Methods. *International Journal of Automation and Computing*, *17*(2), 210–221. https://doi.org/10.1007/s11633-019-1216-5

S, T. V. (2024). *Identifying Fake News On ISOT Data Using Stemming Method With A Subdomain Of AI Algorithms*. *21*(S6), 775–787. www.migrationletters.com

Samadi, M., Mousavian, M., & Momtazi, S. (2021). Deep contextualized text representation and learning for fake news detection. *Information Processing and Management*, *58*(6). https://doi.org/10.1016/j.ipm.2021.102723

Scheufele, D. A., & Krause, N. M. (2019). Science audiences, misinformation, and fake news. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(16), 7662–7669. https://doi.org/10.1073/pnas.1805871115

Schöpfel, J., & Kergosien, E. (n.d.). *Digital Transformation and the Changing Information Landscape Grey Literature in Open Repositories: New Insights and New Issues*. https://www.ccsd.cnrs.fr/2021/04/evolution-de-la-typologie-des-documents-dans-hal-les-resultats-du-groupe-de-travail/

Sharma, R. (2020). Study of Supervised Learning and Unsupervised Learning. *International Journal for Research in Applied Science and Engineering Technology*, *8*(6), 588–593. https://doi.org/10.22214/ijraset.2020.6095

Sharma, U., Saran, S., & Patil, S. M. (2020). *Fake News Detection using Machine Learning Algorithms*. www.ijert.org

Sheykhmousa, M., Mahdianpari, M., Ghanbari, H., Mohammadimanesh, F., Ghamisi, P., & Homayouni, S. (2020). Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review. In *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* (Vol. 13, pp. 6308–6325). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/JSTARS.2020.3026724

Shiri, F. M., PERUMAL, T., MUSTAPHA, N., & MOHAMED, R. (2023). *A Comprehensive Overview and Comparative Analysis on Deep Learning Models: CNN, RNN, LSTM, GRU*.

Shu, K., Cui, L., Wang, S., Lee, D., & Liu, H. (2019). Defend: Explainable fake news detection. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 395–405. https://doi.org/10.1145/3292500.3330935

Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). *Fake News Detection on Social Media: A Data Mining Perspective*. http://arxiv.org/abs/1708.01967

Smita, M. (2021). *94- LOGISTIC REGRESSION MODEL –A REVIEW*.

Song, C., Ning, N., Zhang, Y., & Wu, B. (2021). A multimodal fake news detection model based on crossmodal attention residual and multichannel convolutional neural networks. *Information Processing and Management*, *58*(1). https://doi.org/10.1016/j.ipm.2020.102437

Sudhakar, M., & Kaliyamurthie, K. P. (2024). Detection of fake news from social media using support vector machine learning algorithms. *Measurement: Sensors*, *32*, 101028. https://doi.org/10.1016/j.measen.2024.101028

Tandoc, E. C., Lim, Z. W., & Ling, R. (2018). Defining "Fake News": A typology of scholarly definitions. In *Digital Journalism* (Vol. 6, Issue 2, pp. 137–153). Routledge. https://doi.org/10.1080/21670811.2017.1360143

Trung Tin, P. (2018). *A Study on Deep Learning for Fake News Detection*.

Umer, M., Imtiaz, Z., Ullah, S., Mehmood, A., Choi, G. S., & On, B. W. (2020). Fake news stance detection using deep learning architecture (CNN-LSTM). *IEEE Access*, *8*, 156695–156706. https://doi.org/10.1109/ACCESS.2020.3019735

Verma, P. K., Agrawal, P., Amorim, I., & Prodan, R. (2021). WELFake: Word Embedding over Linguistic Features for Fake News Detection. *IEEE Transactions on Computational Social Systems*, *8*(4), 881–893. https://doi.org/10.1109/TCSS.2021.3068519

Wang, Y., Ma, F., Jin, Z., Yuan, Y., Xun, G., Jha, K., Su, L., & Gao, J. (2018). EANN: Event adversarial neural networks for multi-modal fake news detection. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 849–857. https://doi.org/10.1145/3219819.3219903

Wu, K., Yang, S., & Zhu, K. Q. (2015). *False Rumors Detection on Sina Weibo by Propagation Structures*.

Yang, Y., Zheng, L., Zhang, J., Cui, Q., Li, Z., & Yu, P. S. (2018). *TI-CNN: Convolutional Neural Networks for Fake News Detection*. http://arxiv.org/abs/1806.00749

Zervopoulos, A., Alvanou, A. G., Bezas, K., Papamichail, A., Maragoudakis, M., & Kermanidis, K. (2022). Deep learning for fake news detection on Twitter regarding the 2019 Hong Kong protests. *Neural Computing and Applications*, *34*(2), 969–982. https://doi.org/10.1007/s00521-021-06230-0

Zhou, X., Zafarani, R., Shu, K., & Liu, H. (2019). Fake News: Fundamental theories, detection strategies and challenges. *WSDM 2019 - Proceedings of the 12th ACM International Conference on Web Search and Data Mining*, 836–837. https://doi.org/10.1145/3289600.3291382