
GAZİ ÜNİVERSİTESİ
Gazi
1926

SOSYAL MEDYADAKİ TÜRKÇE SAHTE HABERLERİN DERİN ÖĞRENME YAKLAŞIMIYLA TESPİTİ VE SINIFLANDIRILMASI

Gülsüm KAYABAŞI KORU

DOKTORA TEZİ
ADLİ BİLİŞİM ANA BİLİM DALI

GAZİ ÜNİVERSİTESİ
BİLİŞİM ENSTİTÜSÜ

OCAK 2024

ETİK BEYAN

Gazi Üniversitesi Bilişim Enstitüsü Tez Yazım Kurallarına uygun olarak hazırladığım bu tez çalışmasında;

- Tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi,
- Tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu,
- Tez çalışmasında yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi,
- Kullanılan verilerde herhangi bir değişiklik yapmadığımı,
- Bu tezde sunduğum çalışmanın özgün olduğunu,

bildirir, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim.

Gülsüm KAYABAŞI KORU

22/01/2024

SOSYAL MEDYADAKİ TÜRKÇE SAHTE HABERLERİN DERİN ÖĞRENME YAKLAŞIMIYLA TESPİTİ VE SINIFLANDIRILMASI

(Doktora Tezi)

Gülsüm KAYABAŞI KORU

GAZİ ÜNİVERSİTESİ

BİLİŞİM ENSTİTÜSÜ

Ocak 2024

ÖZET

Sosyal ağları kullanan kişi sayısının artmasıyla daha fazla kişinin haber ihtiyacını karşılamak maksadıyla sosyal medya platformları kullanılmaktadır. Kullanıcılar gazete haber sayfaları yerine özellikle Twitter üzerinden haberlere ulaşarak gündemi takip etmenin daha kolay olduğunu düşünmektedir. Ancak sahte haberler sosyal medyada giderek daha fazla yer almakta ve insanların kısmi haber sayfalarından ya da kısa Twitter paylaşımlarından doğru haberlere ulaşabilmesi her zaman mümkün olmamaktadır. Twitter'da paylaşılan haberlerin doğru olup olmadığının anlaşılması önemli bir problemdir. Sahte tivitlerin tespit edilmesi her dilde olduğu gibi Türkçe'de de büyük önem taşımaktadır. Bu çalışmada, Twitter'daki doğrulama platformlarından elde edilen sahte haberler ve ana akım gazetelerin Twitter hesaplarından elde edilen gerçek haberler etiketlenerek, Türkçe için geliştirilen Zemberek doğal dil işleme aracı kullanılarak ön işleme tabi tutulmuş ve TR_FaRe_News isimli bir veri kümesi oluşturulmuştur. Ardından, TR_FaRe_News veri kümesi, sahte haber tespiti için topluluk yöntemleri ve BoW, TF-IDF ve Word2Vec vektörleştirme yöntemleri kullanılarak araştırılmıştır. Ayrıca, fastText ile kelime yerleştirme işlemi yapılarak fastText dil modelimiz içerisinde bulunan sk-learn kütüphaneleri kullanılarak haberlerin sınıflandırma başarısı farklı parametreler ile analiz edilmiştir. Ardından, önceden eğitilmiş bir BERT derin öğrenme modeline ince ayar yapılmış ve Bi-LSTM ve Evrişimsel Sinir Ağı katmanları ile genişletilen modelin dondurulmuş ve dondurulmamış parametreler yöntemleriyle varyasyonları araştırılmıştır. Performans değerlendirmesi BuzzFeedNews, GossipCop, ISOT, LIAR, Twitter15, Twitter16 ve LLM tarafından üretilen bir sahte haber veri kümesi olmak üzere yedi karşılaştırılabilir veri kümesi kullanılarak gerçekleştirilmiştir. Türkçe tweetlerin analiz edilmesi ve LLM tarafından üretilen sahte haber veri kümlerinin kullanılması önemli bir katkı olarak değerlendirilmektedir. BERT modelleriyle %90 ile %94 arasında, BERTurk+CNN modeliyle %94 oranında doğruluk elde edilmiştir.

Bilim Kodu : 92401

Anahtar Kelimeler : sahte haber, gerçek haber, Twitter, fastText, BERT

Sayfa Adedi : 135

Danışman : Prof. Dr. Çelebi ULUYOL

DETECTION AND CLASSIFICATION OF FAKE NEWS IN TURKISH ON SOCIAL MEDIA USING A DEEP LEARNING APPROACH (Ph. D. Thesis)

Gülsüm KAYABAŞI KORU

GAZİ UNIVERSITY

INSTITUTE OF INFORMATICS

January 2024

ABSTRACT

With the increasing number of people using social networks, social media platforms are being leveraged to address the news demands of more people. Users prefer accessing news on Twitter rather than traditional newspaper websites as it allows for a more streamlined and convenient way to follow the agenda. However, fake news is rapidly emerging on social media, and it is not always viable for individuals to acquire genuine news from fragmentary news pages or short Twitter tweets. Determining the veracity of news disseminated on Twitter is a crucial issue. Identifying fake tweets is crucial not only in Turkish but also in any other language. In this work, fake news received via verification services on Twitter and true news obtained from Twitter accounts of mainstream publications are tagged and preprocessed using the Zemberek natural language processing tool built for Turkish, and a dataset entitled TR_FaRe_News is constructed. Next, the TR_FaRe_News dataset is examined to identify false news using ensemble techniques and vectorization methods such as Bag-of-Words (BoW), Term Frequency-Inverse Document Frequency (TF-IDF), and Word2Vec. In addition, word embedding is conducted with fastText, and the classification effectiveness of the news is analyzed with different parameters using sk-learn libraries in our fastText language model. Subsequently, a pre-trained BERT deep learning model was further refined and expanded upon. The study examined the utilization of Bi-LSTM and Convolutional Neural Network layers, along with their different modifications, using both frozen and unfrozen parameter approaches. The performance evaluation is conducted using seven datasets that are comparable: BuzzFeedNews, GossipCop, ISOT, LIAR, Twitter15, Twitter16, and a synthetic fake news dataset created by LLM. Analyzing Turkish tweets and using fake news datasets created by LLM is considered a noteworthy contribution. The BERT models had accuracies ranging from 90% to 94%, while the BERTurk+CNN model achieved an accuracy of 94%.

Science Code : 92401
Key Words : fake news, real news, Twitter, fastText, BERT

Page Number : 135
Supervisor : Prof. Dr. Çelebi ULUYOL

TEŞEKKÜR

Doktora çalışmalarımın her aşamasında katkılarıyla motive eden ve eleştirileriyle yol gösteren, akademik çalışmalarımda her zaman bana güvenen, ilerleyeceğime olan inancımı teşvik etmekten asla vazgeçmeyen değerli danışmanım Prof. Dr. Çelebi ULUYOL'a, tezimin şekillenmesine katkı sağlayan alanım özelinde üst düzey bilgiler, değerlendirmeler ve tavsiyeler veren Doç. Dr. Ömer Özgür TANRIÖVER ve Dr. Öğr. Üyesi Uraz YAVANOĞLU'na, çalışmalarıma ilişkin fikir alışverişinde bulunduğum arkadaşlarıma, okul hayatım boyunca her türlü zorluğa rağmen desteklerini ve sevgilerini hissettiğim bugünlerime ulaşmamın mimarları annem, babam ve kız kardeşime, zor günlerimde hep yanımda olan sevgisi ve sabrı bitmeden manevi desteğini esirgemeyen sevgili eşime, oyun zamanlarını ertelemek zorunda kaldığım biricik oğluma teşekkür ve şükranlarımı sunarım.

İÇİNDEKİLER

Sayfa

Sayfa

ÇİZELGELERİN LİSTESİ

Çizelge **Sayfa**

Çizelge **Sayfa**

Çizelge **Sayfa**

ŞEKİLLERİN LİSTESİ

Şekil **Sayfa**

Şekil **Sayfa**

SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış simgeler ve kısaltmalar, açıklamaları ile birlikte aşağıda sunulmuştur.

Kısaltmalar	Açıklamalar
AA	Anadolu Ajansı
API	Application Programming Interface (Uygulama Programlama Arayüzü)
BERT	Bidirectional Encoder Respresentation from Transformers (Çift Yönlü Transformatör Kodlayıcı Temsilleri)
Bi-LSTM	Bidirectional Long Short Term Memory (Çift Yönlü Uzun Kısa Süreli Bellek
BOW	Bag of Words (Kelime Çantası)
CART	Classification and Regression Trees (Sınıflandırma ve Regresyon Ağaçları)
CBOW	Continuous Bag of Words (Sürekli Kelime Çantası)
CNN	Convolutional Neural Network (Erişimsel Sinir Ağı)
DDİ	Doğal Dil İşleme
DHA	Demirören Haber Ajansı
DL	Deep Learning (Derin Öğrenme)
GCN	Graph Convolutional Network (Çizge Evrişimsel Ağ)
GNN	Graph Neural Network (Çizge Sinir Ağı)
HBLC	Harmonic Boolean Label Crowdsourcing (Harmonik Mantıksal Etiket Kitle Kaynak Kullanımı)
IDF	Inverse Document Frequency (Ters Doküman Sıklığı)
İHA	İhlas Haber Ajansı
LIWC	Linguistic Inquiry and Word Count (Dil Sorgulama ve Kelime Sayımı)

LLM	Large Language Model (Büyük Dil Modeli)
LSVM	Linear Support Vector Machine (Doğrusal Destek Vektör Makinesi)
LR	Logistic Regression (Lojistik Regresyon)
ML	Machine Learning (Makine Öğrenmesi)
MLM	Masked Language Modelling (Maskeli Dil Modelleme)
MLP	Multilayer Perceptron (Çok Katmanlı Algılayıcı)
MNB	Multinomial Naive Bayes (Çok Terimli Naive Bayes)
NLP	Natural Language Processing (Doğal Dil İşleme)
NMF	Non-Negative Matrix Factorization (Negatif Olmayan Matris Çarpanlara Ayırma)
NSP	Next Sentence Prediction
RF	Random Forest (Rasgele Orman)
RNN	Recurremt Neural Network (Yineleyen Sinir Ağı)
SGD	Stochastic Gradient Descent (Stokastik Gradyan İnişi)
SVM	Support Vector Machine (Destek Vektör Makineleri)
TF	Term Frequency (Terim Sıklığı)
VC	Voting Classifier (Oylama Sınıflandırıcı)
YSA	Yapay Sinir Ağı

1. GİRİŞ

İnternet gazeteciliği çağımızda çok sık kullanılan bir kavramdır. Haber gruplarının abonelerine gönderdiği iletiler ile başlayan örnekler, internet üzerinden çeşitli yazılım ve donanım uygulamalarının da gelişimiyle internet gazeteciliği dediğimiz sanal gazeteciliği ortaya çıkarmıştır [1]. Sanal gazeteciliğin ortaya çıkması ile ise sosyal medya üzerinden haber alma yönelimi başlamıştır.

Sosyal medya, sadece medya olarak değerlendirilmemelidir. Web teknolojilerinin gelişimiyle birlikte geleneksel medyadan farklı olarak birçok boyutu içerisinde barındıran bütünleşik bir kavramdır [1]. Sosyal medya ile gazetecilik ilişkisini değerlendirecek olursak, geleneksel medyanın bugünün tüketicilerinin isteklerinin farkına varması gerektiğine dikkat çekmek gerekir. Bugünün yeni tüketicilerinin bilgiye daha eleştirel yaklaştıklarını düşünürsek, gazetecilik için de artık bilinen haberleşme araçlarının ötesinde siber alemdeki izleyicilere de ulaşmak gerektiğine dikkat çekilmelidir [2]. Bu nedenle yapılan sosyal medya istatistiklerinde, Ekim 2022 itibarıyla dünya çapında yaklaşık 4.74 milyar aktif sosyal medya kullanıcısı bulunduğu görülmüştür [3]. Söz konusu sosyal medya üzerinde paylaşılan haberler olduğundan ebizmba'ın yaptığı araştırmaya göre 2022 yılı ve 2023 yılının ilk çeyreğinde en popüler 15 web sayfası arasında haber siteleri yer almaktadır [4].

"We Are Social Dijital 2023 Global ve Türkiye Raporu" incelendiğinde; Türkiye'de nüfusun %83,4'ü internet kullanıcısıyken toplam nüfusun %73,1'i aktif olarak sosyal medya kullanmaktadır. Bu rapora göre Türkiye'de internet kullanımının 15 nedeni ve yüzdesel oranı verilmiştir. Bu nedenler arasında %67,6 oran ile 2. sırada yer alan haber ve etkinliklerden haberdar olmak amacıyla kişilerin sosyal medya kullandığı görülmüştür. Aynı rapora göre en çok tercih edilen sosyal medya sıralamasında %66.5 oran ile Instagram, Whatsapp ve Facebook'tan sonra Twitter 4. sırada yer almaktadır [5]. Ayrıca Türkiye'de Haberler ve Medya Yayıncıları kategorisinde en çok tıklanan gazete sayfaları hürriyet, habertürk, sözcü, sabah ve mynet olmuştur [6]. Yukarıdaki istatistikler de göz

önünde bulundurulduğunda insanlar, web sayfalarından haberlere ulaşmayı tercih ettikleri kadar Twitter üzerinden de haberlere erişmektedir.

Topluma ciddi zararlar veren sahte haberler de haber sayfalarının ve Twitter'ın haberlere ulaşımını kolaylaştırdıkça artış göstermektedir. Bu durum, sahte haber tespiti ve gerçek haber doğrulama tekniklerinin geliştirilmesi de dahil olmak üzere sahte haberle ilgili birçok çalışmayı tetiklemiştir. Pek çok insanın kullandığı sosyal ağlar üzerinden kullanıcılar tarafından paylaşılan haberlerin doğrulanması maksadıyla, uzmanlar Türkçe olarak oluşturulan sahte haberlerin tespit edildiği doğrulama platformları oluşturmuşlardır. Bu platformların web sayfalarında paylaştıkları haberler Twitter hesaplarında da paylaşılmaktadır [7]. Bu hesapların ana özelliği haber doğrulama yaparken doğruladıkları haberlerin sahte olup olmadığını kanıtlar nitelikte çalışıyor olmalarıdır. Bu hesaplar teyitorg, dogrulukpayicom, dogrulaorg, gununyalanlari ve malumatfurusorgtur. Bu çalışmada, sahte haberlerin tespit edilmesi maksadıyla bu hesaplar kullanılmıştır. Teyitorg ve doğrulukpayicom hesapları, dünya çapında doğrulama platformlarını bir araya getiren ve bu platformların denetimini ve hesap verebilirliğini artırmak maksadıyla kurulan Uluslararası Doğrulama Ağı'nın (IFCN) bir üyesidir [8]. Diğer hesaplar ise üye değil ancak bu ağın gerekliliklerini yerine getiren hesaplardır. Bu hesaplar sadece sahte haber yayınlamakla kalmamakta, bunun yanı sıra sahte haberlerin tespitini kolaylaştırmak amacıyla kişileri bilinçlendirmek için de yayınlar paylaşmaktadır. Çizelge 1.1'de söz konusu doğruluk kontrol hesaplarına ait bilgiler yer almaktadır. Çalışmada diğer bir odak noktası olan gerçek haberler ise en çok tıklanan ana akım gazetelerin Twitter hesaplarından çekilmiştir [8]. Çalışmanın veri seti bölümünde bu kısımla ilgili ayrıntılı bilgilere yer verilecektir.

Çizelge 1.1. Çalışma kapsamında kullanılan doğruluk kontrol platformları

Platform Twitter Hesabı	**Kuruluş Tarihi**	**Doğruluk kontrolü görevi**	**IFCN Üyesi**
malumatfurusorg	2009	Gazete köşe yazarları doğruluk kontrolü	Hayır
dogrulukpayicom	2014	Siyasi beyanat doğrulaması	Evet
teyitorg	2015	Sahte haber iddialarını çürütme	Evet
gununyalanlari	2015	Sahte haber iddialarını çürütme	Hayır
dogrulaorg	2017	Sahte haber iddialarını çürütme	Hayır

Problem durumu / Konunun tanımı

Basılı gazeteciliğin yerini sanal gazeteciliğe bıraktığı günümüzde ülkemizde en çok kullanılan sosyal platformlardan biri olan Twitter'da kişilerin yapılan paylaşımlar konusunda tiviti oluşturan kişiyi bilmeden bir kullanıcıdan diğerine geçirilen bir ifadenin arkasındaki motivasyonları anlaması zor olabilmektedir. Ayrıca sahte haber sosyal medya aracılığıyla yayılabilmekte, bu da utanç verici veya geri dönüşü olmayan zararlara yol açabilmektedir [9].

Son zamanlarda meydana gelen, dünya ve Türkiye gündemini oluşturan bazı sahte haber yayılımları sosyal medya üzerinden paylaşılarak toplumun ilgisini çekerek bu haberlere inanmalarına sebebiyet vermiştir. Sahte haber problemini toplumu nasıl etkilediği konusunda açıklamak maksadıyla örnek olacak birkaç haber makalesi tespit edilmiştir.

11 Mart 2020 tarihi itibariyle Dünya Sağlık Örgütü tarafından "pandemi" olarak adlandırılan COVID-19 hastalığı ortaya çıkmasıyla beraber çok şiddetli bir şekilde yayılım göstermiştir. Hastalığın can kayıplarını önlemek ve yayılma hızını yavaşlatmak için çeşitli aşı çalışmalara yapılmıştır. Bu çalışmalar yapılırken, hazırlanan aşıların içerikleri, üretim süreçleri, insanlara nasıl uygulanacağı konusunda çeşitli belirsizlikler de mevcuttur. Bu belirsizlik nedeniyle, bu konu hakkında haberler de ortaya çıkmış ve günden güne çoğalmıştır. Aşı konusunda yapılan haberlere olan alaka arttıkça, aşının niteliği ve insanlar üzerindeki yan etkileri hususunda da sahte haberler artmaya başlamıştır [10]. Komplo teorilerini de içerisinde barındıran sahte haberler şu şekildedir:

- Aşı olan insanların kollarına çip takılacağı,
- Türkiye'de yaşayan halka aşının diğer ülkede yaşayanlara göre yan etkilerinin daha çok olacağı,
- İnsanların aşı olduktan sonra robotlaşacakları,
- Aşının kısırlığa neden olacağı,

- Aşının felce (özellikle yüz) neden olacağı,
- Aşıyla kişilerin DNA'larının değiştirileceği,
- Aşı Avrupa'da geliştirildiği için yapılan aşıların daha etkili olacağı şeklinde COVID-19 hastalığı iyileştireceği düşünülen aşı konusunda birçok haber ortaya atılmıştır [10].

Rusya ile Ukrayna arasındaki savaşın yeniden alev almasına sebep olan "Ukrayna'da çarmıha gerilen çocuk" sahte haberidir. Söz konusu haber Rus medyasında yer almış ve zamanlama olarak Doğu Ukrayna krizinin yaşandığı zaman olması manidar olmuştur. Ukrayna'nın şehirlerinden biri olan ve Ukrayna'dan ayrılarak Rusya'ya bağlanmak isteyen grupların kontrol noktası olan Sloviansk şehrinde, bir militanın annesinin göz önünde bir bebeği öldürdüğünü iddia ettiği haberi kendini Rus mülteci olarak nitelendiren Galina Pişniyak'tır. Sahte haberin içeriğinde Lenin Meydanı diye sahte olan bir yerde Ukraynalı askerlerin Sloviansk'ta yaşayan halkını köşeye sıkıştırdığı iddia edilmiştir[11]. Kriz döneminde bu sahte haber her zamankinden daha da hızla yayılmıştır ve modern kitle iletişim araçlarında en büyük sahte haber yayılımı örneklerinden biri olarak görülmüştür. Bu sahte haberin hem Ukrayna hem Rusya'da insanları aldattığı ve bununla beraber silahlanmaya teşvik ettiği söylenmektedir [12].

Örnek verilen haberlerin doğruluk kontrolü yapan web siteleri tarafından sahte oldukları teyit edilmiştir. İnternet ortamında doğruluk kontrolü yapan web sitelerinden sahte haberlerin teyit edilmesi mümkündür. Son zamanlarda, Twitter'daki doğruluk kontrol hesapları sayesinde de sahte haberlerin erişimi olağan hale gelmiştir. Ayrıca Türk dili kullanılarak oluşturulmuş sahte ve gerçek haberden oluşan veri setleri literatürde erişilebilir durumda değildir.

Anlatılan tüm hususlar göz önünde bulundurulduğunda sahte haber sorununa yönelik çözülmesi gerekli olan 3 temel problem tespit edilmiştir. Bunlar;

- İnsanların sosyal medya ortamlarında bilinçsiz olarak yapılan sahte haber paylaşımlarından etkilenerek geri dönüşü olamayacak sonuçlar doğurması,
- Türkiye'de yaşayan insanların sahte haber konusunda farkındalıklarının artırılabilmesi için Türk dilinde yapılan çalışmaların yetersiz olması ve

- Türk dilinde sahte haber tespiti için hazırlanan verilerin araştırmacılar tarafından erişebilir durumda bulunmamasıdır.

Bu tez çalışmasında, Twitter platformunda paylaşılan Türkçe tivitlerde, doğal dil işleme yöntemleri kullanılarak sahte haber tespiti yapılmıştır. Sahte ve gerçek tivitlerden oluşan Türkçe dilinde oluşturulan veri seti Ocak 2020 ve Aralık 2022 tarihleri arasında çekilmiş tivitlerden oluşmaktadır. Toplanan tivitler el ile etiketlenmiş ve fastText, makine öğrenmesi ve derin öğrenme yaklaşımları kullanılarak sınıflandırılmıştır.

Tezin Amacı, Kapsamı ve Önemi

Sosyal medyanın açık ve güncel olması, söylentiler, spam ve sahte haberler gibi yanlış bilgilerin oluşturulmasını ve yayılmasını büyük ölçüde kolaylaştırmaktadır. Sahte haberlerin yayılımının artması nedeniyle, demokrasiye, adalete ve kamuya olan güvende bir erozyona sebep olduğundan sahte haber tespitine olan talep artmaktadır. Disiplinlerarası bir konu olarak, sahte haberlerin incelenmesi bilgisayar ve adli bilişim biliminde uzmanların birlikte çalışması, siyaset bilimi, gazatecilik, sosyal bilim, psikoloji ve ekonomiyi teşvik etmektedir. Sahte haber tanımı yeni bir kavram değildir. Özellikle, yayıncıların çıkarlarını gözetmek için yanlış ve yanıltıcı bilgi kullandığı için internetin ortaya çıkmasından önce bile mevcut olan bir tanımdır. Web'in ortaya çıkmasının ardından, giderek daha fazla tüketici çevrimiçi platformlar için bilgiyi yaymak maksadıyla kullanılan geleneksel medya kanallarını terk etmeye başlamıştır [13]. Ayrıca, kullanıcıların bir oturuşta çeşitli yayınlara erişmesine izin vermekle kalmaz, aynı zamanda daha kolay ve daha hızlıdır. Bir yandan, düşük maliyeti, kolay erişimi ve bilginin hızlı dağıtımı, insanları sosyal medyadan haber aramaya ve tüketmeye yönlendirmektedir. Öte yandan, sahte haberlerin geniş bir şekilde yayılmasını mümkün kılmakta, yani kasıtlı olarak yanlış bilgiler içeren düşük kalitede haberler bulundurmaktadır [14]. Sahte haberlerin yayılımının artması, bireyler ve toplum üzerinde son derece olumsuz etki bırakmaktadır. Bu nedenle, sosyal medyada sahte haber tespiti son zamanlarda büyük ilgi gören bir araştırma haline gelmiştir. Sahte haberlerin sosyal medyada yaygınlaşması, yakın zamanda akademik olarak da büyük ilgi görmektedir [15]. Sosyal medyanın açık ve güncel olması, söylentiler, spam

ve sahte haberler gibi yanlış bilgilerin oluşturulmasını ve yayılmasını büyük ölçüde kolaylaştırmaktadır. Bizim sahte haberden kastımız yanlış bilgilendirmenin bir koludur. Son zamanlarda yaşanan yanlış bilgilendirme olaylarında görüldüğü gibi, bunun nasıl tespit edileceği de önemli bir sorun haline gelmektedir.

Tez çalışması kapsamında, kullanıcıların Twitter ortamını kullanarak paylaştıkları metinlerin incelenmesi amaçlanmıştır. Twitter'dan elde edilen metin verisinin ön işleme yöntemleri ile analizi yapılarak sahte paylaşımların tespit edilmesi ve tez kapsamında toplu verinin gerçek mi sahte mi olduğunun belirlenerek sınıflandırılması amaçlanmaktadır.
Literatürde yapılan çalışmalar incelendiğinde (Bkz. Çizelge 2.2.) sosyal medya ortamlarını Türkiye'de kullanan pek çok tez bulunmasına rağmen sahte haber tespiti ve sınıflandırılması üzerine yönelik çok az sayıda çalışma bulunmaktadır.

Yapılacak çalışmanın;

- Twitter'daki doğrulama platformlarından elde edilen sahte haberler ve ana akım gazetelerin Twitter hesaplarından elde edilen gerçek haberlerin etiketlenmesi, etiketlenen haberlerin Türkçe için Zemberek doğal dil işleme aracı ile ön işleme tabi tutulması,
- Ön işleme sonrasında TR_FaRe_News isimli veri kümesinin oluşturulması,
- TR_FaRe_News veri kümesinde, sahte haber tespiti için topluluk yöntemleri ve BoW, TF-IDF ve word2vec bektörleştirme yöntemlerinin kullanılarak incelenmesi,
- Ön eğitilmiş bir BERT derin öğrenme modeline ince ayar yapılması, dondurulmuş ve dondurulmamış parametre yöntemleriyle Bi-LSTM ve CNN katmanlarıyla genişletilen modelin değişkenlerinin test edilmesi,
- Performans değerlendirmesinin yedi adet karşılaştırılabilir veri kümesi (BuzzFeedNews, GossipCop, ISOT, LIAR, Twitter15, Twitter16 ve GPT-2 tarafından üretilen sahte haber veri kümesi dahil) ile yapılması,
- TR_FaRe_News veri kümesinde yer alan verilerin özgün Türkçe verilerinden oluşması ve büyük dil modelleri tarafından üretilen sahte haberlerin sınıflandırma için kullanılması,
- BERT modelleriyle %90 ile %94 arasında ve BERTurk+CNN modeliyle %94 oranında doğruluk değeri elde edilmiş olmasıdır.

Sınırlılıklar

Yapılan tez çalışmasında karşılaştığımız zorluk ve sınırlılıklar aşağıda listelenmiştir;

- Twitter API'sinin ücretsiz sürümünde, yalnızca son yedi güne ait tivit cümlelerinin kullanılması ve bu durumun veri elde etmeyi zorlaştırması,
- Verileri Twitter kullanıcı adları kullanılarak yeniden düzenlenmiştir, böylece yedi gün içinde bu kullanıcı adlarından gelen tüm tivitler elde edilmiştir. İki yıllık bir süre boyunca (Ocak 2020'den Aralık 2022'ye kadar) tivitleri elde ettiğimiz göz önünde bulundurursak, uygulama her çalıştırıldığında her seferinde maksimum 3250 adet tivit elde edilmesi,
- Twitter'ın karakter sayısı açısından kullanım şekli göz önünde bulundurulduğunda, çok az kelimeden oluşan paylaşımların da olması,
- Kelime sayısının azlığının metin sınıflandırma işlemlerinin başarısını düşürmesi,
- Veri ön işleme adımları sonrasında hiç kelime içermeyen tivitlerle karşılaşılması,
- Türkçe, İngilizce'ye kıyasla belirgin bir şekilde farklı bir yapı sergilemektedir. Bu nedenle, kelime ilişkilerinin incelemek için Zemberek doğal dil işleme aracı kullanılmış olup, bu konuda kullanılabilecek kaynak sayısının oldukça az olması,
- GPT-2 gibi büyük bir dil modeli tarafından oluşturulan veri kümelerinin büyüklüğü, bihai veri kümesinin kullanılmasında zamansal sorunlara yol açmasıdır.

2. LİTERATÜRDEKİ ÇALIŞMALARIN İNCELENMESİ

Çalışmanın bu bölümünde sahte haber kavramı ele alınacak ve sosyal ağ platformları ile bağlantısı hususunda bilgilendirme yapılacaktır. Daha sonra sahte haber tespitine yönelik çalışmalardan ve literatürde sahte haber tespiti yapmak için kullanılan veri setlerinden bahsedilecektir. Tez kapsamında yaptığımız çalışmaya ışık tutması için çeşitli arama sözcükleri kullanılarak 2011 yılından günümüze gerçekleştirilmiş araştırmalara yer verilecektir. 2011-2023 yılları arasındaki, çalışmalardan özellikle son 5 yıla ait olanlar güncelliğini koruduğu düşünülerek daha çok tercih edilmiştir. Son 5 yıl içerisinde yayınlanan çalışmalardan da elemeler yapılmıştır.

2.1. Sahte Haber Kavramı

Sahte haber, kasıtlı ve gerçekte yanlış olan bir haberdir. Bu dar tanım, aşağıdaki nedenlerden dolayı yaygın olarak benimsenmiştir. İlk olarak, sahte haberlerin altında yatan amaç, sahte haberlerin daha derin bir şekilde anlaşılmasını ve analiz edilmesini sağlayan hem teorik hem de pratik değerinin olmasıdır. İkincisi, sahte haberlerin dar tanımına uygulanan her türlü gerçek doğrulama tekniği, daha geniş tanım kapsamında da uygulanabilir. Üçüncüsü, bu tanım, sahte haberler ile ilgili kavramlar arasındaki belirsizliği de ortadan kaldırabilir [9].

Tanıma göre şu kavramlar sahte haber kapsamındadır:

- İnsanları yanıltma veya aldatma niyeti olan ve olgusal olarak yanlış algılanması muhtemel hiciv haberleri;
- Haber olaylarından kaynaklanan söylentiler;
- Doğru veya yanlış olarak doğrulanması zor olan komplo teorileri;
- Kasıtsız olarak yaratılan yanlış bilgiler ve
- Yalnızca eğlenceyle motive edilen ya da hedeflenen kişileri dolandıran aldatmacalardır.

Sahte haberin içerik özellikleri, bir haberle ilgili meta bilgileri tanımlamaktadır. Temsili haber içeriği özelliklerinin şu şekildedir:
Kaynak: Haberin yazarı ve yayıncısı.
Başlık: Okuyucuların dikkatini çekmeyi amaçlayan ve haberin konusunu açıklayan kısa başlık metni.
Gövde Metni: Haberin ayrıntılarını anlatan ana metin; genellikle özellikle vurgulanan ve yayıncının açısını şekillendiren büyük bir iddia bulundurur.
Resim/Vidyo: Bir haberin gövde içeriğinin bir kısmı hikayeyi çerçevelemek için sağlayan görsel ipuçlarıdır.

Bu ham içerik özelliklerine dayalı olarak, sahte haberlerin ayırt edici niteliklerini ortaya çıkarmak için farklı türde özellik sunumları oluşturulabilir. Tipik olarak, bakılan haber içeriği çoğunlukla metinsel özellikler, görsel özellikler ve stil özellikleri olacaktır. Çalışma kapsamında bu özelliklerden sadece metinsel özellikler ile ilgilenilmiştir.

Metinsel Özellikler

Metinsel özellikler, doğal dil işleme teknikleriyle haber içeriğinden çıkarılmaktadır [9]. Dilbilimsel, düşük dereceli ve sinirse metin özelliklerinin nasıl çıkarılacağı da tanımlanmaktadır.

Dilbilimsel Özellikler

Dilbilimsel özellikler, metin içeriğinden karakterler, kelimeler, cümleler ve belgeler olarak çıkarılır. Sahte haberler ile gerçek haberlerin farklı yönlerini yakalamak maksadıyla hem ortak dil özellikleri hem de alana özgü dil özellikleri kullanılır. Doğal dil işlemede çeşitli görevler için belgeleri temsil etmek amacıyla sıklıkla kullanılan ortak dil özellikleri mevcuttur. Bu ortak dil özellikleri şunlardır:

- Sözcüksel özelliklere, toplam kelime, kelime başına karakter, büyük kelimelrin sıklığı ve benzersiz kelimeler gibi karakter seviyesi ve kelime seviyesi özellikleri dahildir.
- Sözdizimsel özelliklere, işlev bakımından kelimelerinin ve ifadelerinin sıklığı (yani “n-gram” ve BoW yaklaşımları [16]) veya noktalama ve kelime öbeği (POS) etiketleme [17] gibi cümle düzeyinde özellikler dahildir. Alıntı yapılan

kelimeler, harici bağlantılar, grafik sayısı ve ortalama grafik uzunluğu gibi haber alanına özel olarak hizalanmış alana özgü dil özellikleri [18].

Düşük Dereceli Metinsel Özellikler

Metin verileri için düşük dereceli modelleme, farklı alanlarda geniş çapta araştırılan bir konudur [19]. Düşük dereceli yaklaşım, yüksek boyutlu ve gürültülü ham özellik matrisinden düşük boyutlu bir metin gösteriminin öğrenmektir. Mevcut düşük dereceli modeller çoğunlukla matris çarpanlarına ayırma veya tensör çarpanlara ayırma tekniklerine dayalıdır ve bunlar terim-haber matrisini boyutlu bir gizli uzaya yansıtmaktadır.

Metin modelleme için matris ayrıştırması, kümeleme gibi belge temsillerini öğrenmek için yaygın olarak kullanılmaktadır [20]. Her boyutun, ham özellik matrisinden tutarlı gizli konuyu temsil ettiği düşük boyutlu bir temel öğrenebilmektedir. Negatif olmayan matris çarpanlarına ayırma (NMF) yöntemleri, haber kelimeleri matrisini çarpanlara ayırırken negatif olmayan kısıtlamalar getirmektedir [21]. NMF'yi kullanarak, belge-kelime matrisini, belge-kelime ilişkilerinin uzaydaki iç çarpım olarak modelleneceği şekilde, düşük boyutluluğa sahip ortak bir gizli anlamsal faktör uzayına yansıtmaya çalışılır. Tensör çarpanlarına ayırmanın amacı ise, belgedeki kelimelerin uzaysal ilişkilerini dikkate alarak haber makalelerinin temsilini öğrenmektir.

Sinirsel Metinsel Özellikler

DDİ'deki derin sinir ağlarının son gelişmeleriyle, gizli metinsel özellik temsillerini öğrenmek için evrişimli sinir ağları (CNN) ve tekrarlayan sinir ağları (RNN) gibi farklı sinir ağı yapıları geliştirilmiştir. Sinirsel metinsel özellikler, yüksek boyutlu ve seyrek özellikler yerine yoğun vektör temsillerine dayanmaktadır ve çeşitli DDİ görevlerinde üstün sonuçlar elde etmiştir [22]. CNN'ler, RNN'ler ve bunların bazı varyantları dahil olmak üzere büyük temsili derin sinirsel metinsel özellikler mevcuttur.

2.2. Sahte Haber Tespitine Yönelik Yapılan Çalışmalar

Doğal dil işleme (DDİ) alanında biz gibi çalışmalar yapan araştırmacılar, çeşitli makine öğrenmesi ve derin öğrenme yaklaşımları kullanarak sahte haberleri tespit etmeye ve bunlarla mücadele etmeye yönelik uğraşılarda bulunmuşlardır. Çalışmanın bu kısmında sahte haberlerin yayılmasını anlamak, azaltmak ve tespit etmek maksadıyla yapılmış mevcut çalışmalar gözden geçirilecek ve özetlenecektir.

Sahte haber tespitine yönelik çalışmaları gözden geçirirken yapılan araştırma süreci, [23,24] çalışmalarında da kullanılan Google Scholar, Science Direct, IEEE Explore, Wiley, Web of Science, Scopus ve ACM Digital Library gibi dijital kütüphanelerden alınan yaptığımız çalışmanın konusuna uygun olan yayınların bir araya getirilmesiyle oluşmaktadır. İngilizce dilinde yazılmış orijinal araştırma ve tarama/derleme makaleleri incelenmiştir. Makaleler, sahte haberlerle ilgili anahtar kelimeler ve "social media", "twitter", "fake news", "measurement", "score" gibi ilişkili terimler kullanılarak yukarıda bahsi geçen dijital kütüphanelerden elde edilmiştir. Tablo 1'de tüm araştırma öğesi ve zaman aralıklarına karşılık kullanılan anahtar sözcükler sunulmuştur. Burada zaman aralığı olarak 2011 ile 2023 yılları arası seçilmiş ancak son 5 yıla ait yayınlar güncelliğini halen koruyabildiği düşünülerek tercih sebebi olmuştur. Kullanılan anahtar sözcükler birlikte kullanılma (aralarında "ve" bağlacı bulunan) ve tamamı kullanılan (aralarında "veya" bağlacı bulunan) şeklinde Çizelge 2.1'de açıklanmıştır.

Çizelge 2.1. Arama dizesi içerisinde kullanılan anahtar sözcükler

Arama Öğesi ve Zaman Aralığı	**Kullanılan Anahtar Sözcükler**
2011 ile 2023 yılları arasında tüm arama anahtarları için	'fake news' AND 'social media' OR 'fake news' AND 'measurement' OR 'fake news' AND 'twitter' or 'fake news' and 'score'

Sahte haber tespitine yönelik model ve yöntemlerin daha net anlaşılabilmesi için yapılan incelemede çalışmalar özetlenmiş ve daha sonra yazar, yıl, platform, dil, özellik temsili, en iyi sınıflandırıcı ve sınıflandırma metrikleri doğrultusunda karşılaştırmalı olarak gösterilmiştir. Yapılan karşılaştırma sonucu hazırlanan özet tablo çizelge şeklinde sunulmuştur.

Sahte haberlerin çoğalmasının sosyal medyada yayılmasının yıkıcı etkisi nedeniyle endişe verici boyut gelmiştir. Bu nedenle sahte haber tespiti için önceden eğitilmiş bir dizi gelişmiş dil modelinin yanı sıra geleneksel ve derin öğrenme modellerini inceleyen ve karşılaştıran çalışmalar yapılmıştır. Üç farklı veri seti üzerinde 19 farklı makine öğrenmesi yaklaşımı kullanan çalışmada bir performans analizi sunulmaktadır. 19 modelden sekizi geleneksel öğrenme modelleri, altısı geleneksel derin öğrenme modelleri ve kalan beşi BERT gibi önceden eğitilmiş gelişmiş dil modelleridir [25].

Günümüzde haberin güvenilirliğinin herhangi bir garantisi bulunmamaktadır. Bu nedenle sahte haberleri ve özelliklerini derinlemesine bilmek gerekmektedir. İçerik tabanlı yaklaşım ve denetimsiz öğrenmeyi analiz eden çalışmada %85 doğruluk ile evrişimli ağ kullanılmıştır [26].

Sahte haberlerin siyasi, ekonomik, dini ve sosyal istikrar bakımından tehdit oluşturduğu bir dönemde İngilizce dilinde yapılan sahte haber tespit çalışmaları muazzam miktardadır. Ancak Bangla gibi morfolojik açıdan karmaşık ve kaynak bakımından kısıtlılık bulunan bir dilde sahte haberleri tespit eden çalışmalarda bir açık mevcuttur. Özellik çıkarımı için 1D CNN ve sınıflandırma için standart makine öğrenmesi teknikleri kullanan Bangla sahte haberlerini tespit etmek için hibrit bir model öneren çalışmada BanFakeNews veri seti oluşturularak %99 ve %82'lik bir performans elde edilmiştir [27].

Covid-19 salgınının internet ve sosyal medya kullanımını artırdığına yönelik küresel istatistikler bulunmaktadır. Kullanıcıların içgüdüsel olarak karşılaştıkları şeylere inanma ve bunlara dayanarak sonuçlara varma eğilimi vardır. Tüketicilerin bir sonuca varmadan önce haber ve haber kaynağı hakkında bir anlayış ve ön bilgiye sahip olması esastır. Bir haber makalesinin doğruluğunu tahmin etmek için hibrit bir model öneren çalışmada Pasif Agresif Sınıflandırıcı içeren hibrit modelde %88 ile en yüksek doğruluğu elde etmişlerdir. Ayrıca derin sinir ağları içeren modelleri %75-80 civarında doğruluk göstermiştir. Sinhala dilinde gerçekleştirilen çalışmada verimli sonuçlar elde ettiklerini düşünmektedirler [28].

Twitter'daki sahte haber kaynaklarının tespitini zenginleştirmek maksadıyla düğüm yerleştirmeleri ve kullanıcı tabanlı özellikleri birleştiren hibrit bir yaklaşım önerilmiştir. Sınıflandırıcı performansına göre f1-puanının %98 olduğu görülmüştür [29].

Dilbilimsel özellikleri (kelime sayısı, okuma kolaylığı vb.) ve bilgiye dayalı yaklaşımları birleştiren hibrit bir çalışmada, önerdikleri sistemin her iki özellik türünü de kullanarak %94,4'lük bir doğruluğa ulaşabildiği görülmüştür. Dilbilimsel özelliklerin ayrı ayrı kullanılmasıyla elde edilen %89,4 doğruluk oranından daha iyi bir doğruluğa ulaşabilmişlerdir [30].

Sahte haber tespiti, kullanıcılar tarafından sahte ve onur kırıcı bilgiler içeren bir yazıyı görmek için kullanılabilir. Bu çalışmada Naive Bayes sınıflandırıcı kullanılarak sahte haber tespiti için kolay bir teknik kullanmıştır. Test setinde, %80'lik bir doğruluk etmişlerdir [31].

Sahte haber tespitindeki en büyük zorluğun çok fazla bilgi olması ve benzerlik yayılımından kaynaklandığını iddia eden çalışmalarında, mevcut modellerin sınırlamalarının üstesinden gelerek haberlerdeki benzerlikler arasındaki düzeltmelerin öğrenilmesinde zorluğundan üstesinden geldiklerini ifade etmektedirler. Bunun için her bir farklı mod arasındaki ilişkiyi öğrenebilen karşılıklı dikkat sinir ağır (MANN) önermişlerdir. Weibo veri setini kullandıkları çalışmalarında %77 oranında doğruluğa sahip olmuşlardır [32].

Etki alanları arası metin sınıflandırması ve kaynak alan bilgisini kullanmayı kendilerine hedef olarak belirleyen bir model oluşturan çalışmada etki alanları arası modellerde açıklanabilirliğin etkisini incelemek için dört düzeyli, bir alanlar arası strateji önermişlerdir. BERT'de ince ayar yapmak için yeni ve açıklanabilir ELI5 ve SHAP modellerini geliştirmişlerdir [33].

Derin öğrenme modelini kullanan başka bir çalışmada önerilen sistem, bilgi mesajında bulunan herhangi bir kelimeyi ideal bir ölçüm vektörüne dönüştürmektir. Kelime vektörleştirme, verilerin yüksek boyutluluğunun değişimini yönetmektedir. LSTM modelinin kesinliği, CNN ve RNN modelleriyle karşılaştırıldığında yüksektir ve kabaca %91,73'tür [34].

LIAR veri setini kullanan çalışmada SVM ve RF en yüksek doğruluk olarak bulunmuş ve %97 değerindedir [35]. Çok dilli sahte haber algılama sorunlarını çözmek için sınırlı verilerden yararlanmanın en iyi yolunu bulmak için aktif öğrenme tekniklerini uygulayan çalışmada Çok dilli-BERT ile %97,1 f1 puanı performansı elde edilmiştir [36].

Sahte haber tespiti için özellik çıkarma modülü kullanan çalışmada, haber öyküsündeki ilk ön işleme aşamasından sonra önceden işlenmiş çıktılar alınır. Çıkarılan özellikler daha sonra sahte puanları keşfetmek amacıyla üretken çekişmeli ağ (GAN) ve sinir ağı (NN) kullanılarak analiz edilmektedir. Sinir ağı sınıflandırıcısını eğitmek için karışık çoban optimizasyon algoritması (RSSO) entegre edilmiştir. Veri etiketlemek meşakkatli bir iştir. GAN'ların denetlenmiyor oluşu nedeniyle onları eğitmek için verilerin etiketlenmesine gerek kalmamıştır. RSSO tabanlı NN+GAN sınıflandırıcı %91,2'lik doğruluk, %92'lik hassasiyet ve %88,8'lik özgüllük ile iyileştirilmiş performans sağlamıştır [37].

Haber makalelerini sınıflandırmak için önerilen bir başka yaklaşımda yararlı özelliklerin çıkarılıp analiz edildiği ve çıkarılan özellikleri kullanarak Balina Optimizasyon Algoritması (WOA) tarafından optimize edilen Extreme Gradient Boosting Tree (xgbTree) algoritması kullanıldığı iki aşama mevcuttur. Önerilen yaklaşımın iyi bir sınıflandırma doğruluğu ve f1 puanı oranı elde ettiği ve makalelerin yüzde 91'inden fazlasını başarıyla sınıflandırdığını göstermişlerdir [38].

Farklı dilde bir sahte haber tespiti çalışmasında ise, Arapça bir haber veya iddia verildiğinde otomatik olarak gerçeği tespit edebilen bir model geliştirilmiştir. Özellikle CNN'den yararlanarak sahte ve gerçek haber iddialarını sınıflandırabilen derin bir sinir ağı yaklaşımı önermişlerdir. Arapça veri setine model uygulandığında %91'lik doğruluk sonucu elde edilmiştir [39]. Önerilen başka bir Arapça söylenti tespit modelinde, transformatör tabanlı derin öğrenme mimarisi kullanılmıştır. Yakın zamanda geliştirilen iki transformatör tabanlı model, AraBERT ve MARBERT, üç Arapça veri kullanılarak test edilmiş ve değerlendirilmiştir. Örnekleme tekniği kullanarak dengesiz eğitim veri

kümelerinin getirdiği zorluklar da hafifletilmiştir. Deneysel sonuçlara göre %97 doğruluk oranı elde edilmiştir [40].

Az sınıf problemini aşmak için sentetik sahte veriler üreten ve en gelişmiş dil modeli olan BERT kullanarak sınıflandırma yapan AugFake-BERT tekniği ile tanıştıran çalışma; metin verilerini geliştirmek için önceden eğitilmiş çok dilli BERT'ten oluşur ve artırılmış verileri, yerleştirmeler oluşturmak için ince ayarlı bir BERT'e besler. Sahte veri üreterek veri seti oluşturulmuş ve artırılmış bu veriler gömülü BERT modeline beslenmiştir. Önerilen modelin doğruluk puanı %92,45'tir [41].

Veri analizinin LIAR ve FakeHealth veri setinde yapıldığı çalışmada metin özellikleri, BoW, TF-IDF, word2Vec ve doc2Vec vektörleştirmeleri kullanılarak vektör uzayı gösteriminden ve belirli kelimelerin ve kalıpları oluşumuna dayalı olarak elle çıkarılan özelliklerden oluşur. FakeHealth veri setinde yapılan deneylerde en iyi özellik seti, alt doğrusal terim frekansının kullanıldığı TF-IDF gösteriminden ve Gradient Boosting Regresör tarafından tahmin edilen 10000 boyuttan oluşmaktadır. En iyi temsil yöntemi Word2Vec ve Doc2Vec olmuş ve çalışmanın f1 ölçüsü %81,2 olarak elde edilmiştir [42].

Daha önceki sahte haber tespiti yaklaşımları, metinsel ve görsel özellikleri birleştirmiş ancak kelimeler arasındaki anlamsal korelasyonlara değinilmemiş ve birçok bilgilendirici görsel özellik kaybolmuştur. Bu sorunu ele almak maksadıyla yüksek bilgi içeriğine sahip çok modlu bir özellik vektörü oluşturmak amacıyla metinsel ve görsel özellikleri birleştiren otomatik bir sahte haber tespiti önerilmiştir. Kelimeler arasındaki anlamsal ilişkileri korumak için BERT modelinden faydalanmışlardır. Önerilen model Politifact ve Gossipcop veri kümeleri için sırasıyla %93 ve %92 gibi önemli ölçüde bir sınıflandırma doğruluğu sağlamıştır [43].

Facebook ve Twitter'da otomatik haber makalesi sınıflandırması için makine öğrenmesi optimizasyonu öneren bir teknikte tekrarlanmayan kaynaklardan gelen haberleri çarpıtmak amacıyla sosyal forum haber bulguları için doğal dil işlemenin stratejik olarak uygulanmasıyla kolaylaştırılmıştır. Çalışmadan elde edilen sonuç, Hibrit Destek Vektör Makinesi tarafından %91,23 doğruluk edilmiş olmasıdır [44].

Üç hakem tarafından değerlendirilen Endonezya dilinde 250 sayfalık sahte ve gerçek haberlerden oluşan veri seti manuel olarak sınıflandırılmıştır. Gözden geçiren üç kişinin oylamasıyla elde edilmiştir. Naive Bayes algoritması kullanılarak en yüksek doğruluk %78,6 elde edilmiştir. Çalışmada kullanılan veri kümesinin, gelecekteki araştırmalar için çoğaltılabilmesi ve sonucun test edilerek karşılaştırılabilmesine uygun olduğunu değerlendirmişlerdir [45].

Hazırlandığı dönem için Türkçe çalışmanın mevcut olmaması motivasyonları olarak kabul edilmiş olan çalışmada terim frekansı, TF-IDF, n-gram, stil işaretçileri, argo/küfür kullanımı, url, haberdeki erişilebilir link özelliği, Başlık ve Haber içeriği uyumu, Abartılı Başlık kullanımı gibi özellikler kullanılarak SVM ve NB sınıflandırıcılar kullanmışlardır. SVM sınıflandırıcı ile f1-ölçütü %79 elde etmişlerdir [46].

Çalışmalarında n-gram analizini kullanarak spam görüşlerin tespiti yapan ve n-gram analiz yönteminin sahte içerikler üzerinde avantajlı olduğunu sunmaktadır. Çalışmada, SGD, SVM, KNN, LR, DT algoritmaları denenmiştir. Doğruluk hesabı olarak %90 oranında SVM algoritmasıyla başarım elde edilmiştir [47].

Bir yazılım sistemi olarak uygulanan ve Facebook haber gönderilerinden oluşan bir veri kümesine karşı test edilen çalışmada, basit gibi görünmesine rağmen aldığı iyi sonuç yaklaşık olarak %74'tür. Sınıflandırıcı olarak Naive Bayes kullanmışlardır [48].
Basılı(gazete) haberciliğin biçimini ve stilini taklit eden hiciv haberlerinin benzersiz özelliklerini detaylandırarak ve göstererek hiciv ve mizaha kavramsal bir bakış sunan bir çalışma mevcuttur. Bu çalışmada 4 alanda (yurttaşlık bilgisi, bilim, iş ve yumuşak haberler) 12 güncel haber konusunda hicivli haberler, meşru haber muadilleriyle eleştirilerek incelenmiştir. Hiciv tespitinde önceki çalışmalara dayanarak, 5 tahmini özellikle (saçma, mizah, dilbilgisi, olumsuz etki ve noktalama) zenginleştirilmiş SVM tabanlı algoritma geliştirmişler ve 360 adet haber makalesinde demişlerdir. Çalışmanın kesinliği %90, f1-puanı ise %87 olarak hesaplanmıştır [49].

Sahte haber tespiti görevi için yedi farklı haber alanını kapsayan iki yeni veri seti sunan çalışmada, veri toplama, ek açıklama ve doğrulama sürecini ayrıntılı olarak açıklamışlar ve sahte ve meşru haber içeriğindeki dilsel farklılıkların belirlenmesine yönelik analizler yapmışlardır. Sahte haber tespiti yapmak için bir dizi öğrenme deneyi yürütmüşlerdir. Ayrıca sahte haberlerin otomatik ve manuel olarak tanımlanmasına ilişkin karşılaştırmalı analizler sunmuşlardır. Çalışmanın f1-puanı %73 olarak elde edilmiştir [50].

Kütüphanelerde sunulan dijital içeriğin güvenilirliğini manuel olarak doğrulama olmadan kontrol edebilen otomatik mekanizmaların geliştirilmesi gerektiğini savunan çalışmalarında Türkçe dijital haber içeriğine dayalı otomatik bir sahte haber tespit sistemi geliştirmişlerdir. Kütüphaneler aracılığıyla kontrolsüz sahte haber yayılımını önlemeye çalışmaktadırlar. El ile derledikleri ve GDELT projesi ile ana akım haber sayfalarından elde ettikleri gerçek haberler ve teyit.org doğrulama platformu ve el ilse derledikleri sahte haberlerden bir veri seti oluşturmuşlar ve ismine TR_FN vermişlerdir. ExtraTrees sınıflandırıcı ile %96.81 oranında bir doğruluk elde etmişlerdir [51].

Çalışmasında Ağustos 2019'da teyit.org platformunda birçok konudaki sahte haberler arasında Twitter'da en çok paylaşılan iki konu belirlenmiştir. Twitter Scraper kullanılarak veriler toplanmış. Geliştirdikleri etiketleme platformu ile gönderilerin gerçek veya sahte olduklarını belirlemişlerdir. Oluşturulan veri seti sadece 3 olayla ilgili tweet içermektedir. 1287 adet tweet bulunduran veri kümesi test edilmiştir. Sosyal medya için ilk Türkçe sahte haber tespiti örneği olduğunu savunmaktadırlar. Makine öğrenmesi ile birlikte derin öğrenme ve sosyal ağ analizini kullanmışlardır [52].

Hem Hintçe hem de İngilizce dilleri için sahte haber tespiti yapan çalışmalarında Covid-19 ile ilgili sahte haberleri tespit etmek hedeflenmiştir. Çalışmanın, İngilizce dilindeki doğruluğu en yüksek %93,45, Hintçe de %97'dir [53].

Farklı dil kökenlerine ait birden çok dil (İngilizce, Portekizce, Bulgarca) için sahte haber tespiti yapan çalışmada hem sosyal medya hem de web sitelerinde bulunan haberler için değerlendirmeler yapılmıştır. Geleneksek makine öğrenmesi algoritmaları kullanılarak elde edilen en başarılı sonuçlar şu şekildedir. FakeBrCorpus için %91, TwitterBR için %81, Faknewsdata1 için %86, Fakeorrealnews için %94 be btvsyle için %95'tir. [54].

Makine öğrenmesi topluluğu yaklaşımını öneren ve haber makalelerinin otomatik olarak sınıflandırılmasını sağlayan çalışmada, sahte içerikleri gerçek içeriklerden ayırt etmek maksadıyla farklı metinsel özellikler kullanılmıştır. Bu özellikler kullanılarak, çeşitli topluluk yöntemleriyle farklı makine öğrenmesi algoritmalarının bir kombinasyonu eğitilmekte ve performansları 4 adet gerçek dünya veri kümesiyle değerlendirilmektedir Deneysel değerlendirme sonucu, bireysel öğrenicilere kıyasla önerdikleri toplu öğrenme yaklaşımının daha üstün performans gösterdiğini iddia etmişlerdir [55].

Sahte haberlere odaklanarak sosyal medyadaki sahte haberleri tespit etmek maksadıyla iki aşamalı yöntem öneren çalışmanın ilk adımında, yapılandırılmamış veri setlerini yapılandırılmış veri setine dönüştürmek için kullanılan veri setlerine bir takım ön işlemler uygulanmıştır. Haberleri içeren veri setindeki metinler, TF-IDF yöntemiyle vektör temsilleri elde edilmiştir. İkinci adımda, metin madenciliği yöntemleri ile yapılandırılmış formattaki veriler yirmi üç adet denetimli yapay zeka algoritması kullanılarak sınıflandırma başarıları elde edilerek değerlendirme metrikleri kıyaslanmıştır [56].

Literatürde mevcut veri setlerinin yeterli sayıda olmaması ve aynı zamanda var olan veri setlerinde de haber içeriği, sosyal bağlam ve uzay-zamansal bilgi gibi sıklıkla gerekli olan özellikler çalışmalarda yer almamaktadır. Bu nedenle, sahte haberlerle ilgili araştırmaları kolaylaştırmak maksadıyla, haber içeriği, sosyal bağlam ve uzay-zamansal bilgilerde farklı özelliklere sahip kapsamlı veri seti içeren bir sahte haber veri havuzu olan FakeNewsNet'i tanıtan çalışma mevcuttur. BuzzFeed ve GossipCop veri setlerinin farklı açılardan keşfedici bir analizini gerçekleştiren çalışma diğer yandan FakeNewsNet'in sosyal medyadaki sahte haber çalışmalarına ilişkin potansiyel uygulamalar için faydalarını da tartışmaktadır [57].

Otomatik aldatmaca sistemlerine duyulan ihtiyaca yönelik olarak, Facebook gönderilerinin onları "beğenen" kullanıcılara göre yüksek doğrulukla sahte veya sahte olmayan olarak sınıflandırabileceğini gösteren çalışmada lojistik resregyona dayalı ve bool kitle kaynak kullanımı algoritmalarının yeni bir uyarlamasına dayanan iki sınıflandırma tekniği önermişlerdir. 15.500 adet Facebook gönderisi ve 909.236 adet kullanıcıdan oluşan veri

setinde, eğitim seni gönderilerin %1'inden az olduğunda bile %99'u aşan bir sınıflandırma başarısı elde etmişlerdir [58].

Yalnızca içeriğine bağlı olarak bir haberin sahte olup olmadığını tahmin edebilen bir sınıflandırıcı oluşturmak ve böylece soruna tamamen NLP perspektifinden yaklaşan bir çalışmada ana hedef, birden çok farklı model uygulamasından elde edilen sonuçları karşılaştırmak, raporlamak ve bulgularını analiz etmektir. Çalışmalarında çeşitli mimarileri denemişler ve evrişimli sinir ağlarında dikkat benzeri bir mekanizma içeren yeni bir tasarım sunmuşlardır [59].

2016 başkanlık kampanyası zamanında yayınlanan kötü amaçlı içeriği tespit edebilmek amacıyla otomatik bir sistem kurmaya çalışan çalışmada Snopes tarafından yanlış olarak işaretlenen bir dizi makale ve önde gelen haber kuruluşlarından bir dizi makale kullanılarak, çeşitli makine öğrenmesi algoritmaları ile sadece her makalenin metinsel yapısı kullanılarak eğitim yapılmaktadır. Bir makalenin genel duyarlılığı ile gerçek doğruluğu arasındaki olası herhangi bir korelasyondan yararlanmak için her makalenin duyguyla ilgili özelliklerine de bakılmaktadır [60].

Yaklaşımlarında, haber başlıklarını ve makaleleri kelime kelime okuyan tekrarlayan sinir ağlarından yararlanan bir çalışmada uygulanan modeller son teknoloji sistemlerle karşılaştırılmakta ancak ciddi sorunun aşırı uydurma olduğu gözlemlenmektedir [61].

Özellik çıkarma ve duruş sınıflandırması yapan çalışmada derin sinir ağları kullanılmıştır. RNN modelleri ve uzantıları, sınıflandırmada önemli farklılıklar göstermiştir. Çift yönlü LSTM modeli, geniş ve ayrıntılı sınıflandırma için en iyi doğruluğu vermiştir [62].

Bir haber makalesinde duruş tespiti yaparken haber metninin alaka düzeyini ve iddiasını belirlemek gerekmektedir. Bu soruna etkili bir çözüm sunan çalışmada nöral, istatistiksel ve harici özellikleri birleştiren bir çalışma mevcuttur. Derin tekrarlayan modelden nöral yerleştirmeyi, ağırlıklı n-gram kelime çantası modelinden istatistiksel özellikleri ve özellik mühendisliği yardımıyla manuel harici özellikleri hesaplamışlardır. Kapsamlı deneyler yapılarak, önerdikleri modelin sahte haber tespitinde son teknoloji tekniklerden daha iyi performans gösterdiğini iddia etmişlerdir [63].

Daha doğru ve otomatikleştirilmiş bir tahmin için üç özelliği birleştiren bir model öneren çalışma mevcuttur. Üç özellikten motive olarak, üç modülden oluşan CSI adlı bir model önerilmiştir. Yakalama, Puanlama ve Bütünleştirme. İlk modül yanıt ve metne dayalıdır; belirli bir makaledeki kullanıcı etkinliğinin geçici modelini yakalamak için tekrarlayan sinir ağı kullanmışlardır. İkinci modül, kullanıcıların davranışlarına dayalı olarak kaynak özelliklerini öğrenmekte ve iki modül üçüncü modülle entegre olarak haber makalesini sahte veya gerçek olarak sınıflandırmaktadır. CSI modeli mevcut modellerden daha yüksek doğruluk elde ettiği iddia edilmektedir [64].

Sahte haberlerin hızlı ve doğru tespiti için üç seviyeli hiyerarşik dikkat ağı (3HAN) aracılığıyla derin öğrenmeye dayalı otomatik bir algılayıcı kullanan bir çalışma mevcuttur. 3HAN'ın her biri kelimeler, cümleler ve başlık olmak üzere üç düzeye sahiptir. Haber başlığı, sahte haberlerin ayırt edici bir özelliği olarak bilinmektedir. 3HAN, üç dikkat katmanı nedeniyle bir makalenin bölümlerine farklı bir önem vermektedir. Büyük bir gerçek dünya veri seti üzerinde yapılan deneylerde 3HAN'ın doğruluğunun %96,77 olduğunu gözlemlemişlerdir [65].

130 bin haber gönderisini şüpheli veya doğrulanmış olarak sınıflandırmak için tahmin modelleri oluşturan ve şüpheli haberlerin dört alt türünü (hiciv, aldatmaca, tıklama tuzağı ve propaganda) tahmin eden bir çalışmada, tivit içeriği ve sosyal ağ etkileşimleri üzerine eğitilen sinir ağı modellerinin sözcüksel modellerden daha iyi performans gösterdiğini iddia etmişlerdir. Aldatma tespiti üzerine önceki çalışmalardan farklı olarak, modellerimize sözdizimi ve gramer özellikleri eklemenin performansı iyileştirmediğini göstermişlerdir. Dil özelliklerini birleştirmek, sınıflandırma sonuçlarını iyileştirdiğini iddia etmişlerdir [66]. Halka açık olarak sunulan LIAR veri setinin ilk defa sunulduğu çalışmada, ampirik olarak, yüzey seviyesindeki dil kalıplarına dayalı olarak otomatik sahte haber tespiti yapmaktadır. Meta verileri metinle bütünleştirmek için yeni hibrit bir evrişimsel sinir ağır tasarlamışlardır. Bu hibrit yaklaşımın sadece metin içeren bir derin öğrenme modelini geliştirebileceğini göstermişlerdir [67].

Bir haberin yalnızca içeriğine dayalı olarak sahte olup olmadığını tahmin edebilen bir sınıflandırıcı oluşturmak işteyen bir çalışmada, RNN modelleri ve LSTM'ler ile soruna tamamen derin öğrenme perspektifinden yaklaşmaktadırlar. LIAR veri setini uygulamışlardır [68].

Evrişimli sinir ağları ve kendinden çok başlı dikkat mekanizması kullanarak haberlerin gerçekliğini sadece içeriğe dayalı olarak yüksek doğrulukla elde edebilen SMHA-CNN (Self Multi-Head Attention-based Convolutional Neural Networks) adlı bir model mevcuttur. Geçerliliğini kanıtlamak maksadıyla deneyler yapmışlar ve 5 kat çapraz doğrulama ile %95,5 oranında bir doğruluk elde etmişlerdir [69].

Farklı çekirdek boyutlarına ve filtrelere sahip tek katmanlı derin CNN farklı paralel bloklarını BERT tabanlı derin öğrenme yaklaşımıyla sının çalışmada, doğal dil işlemedeki belirsizliğin üstesinden gelinmiş ve %98,9 doğruluk göstermiştir [70].

Echo-FakeD isimli yaklaşımda, hem haber içeriğinin hem de sosyal bağlamın gizli bir temsilini elde etmek için eşleştirilmiş bir matris-tensör çarpanlarına ayırma yöntemini izleyerek haber içeriği tensörle birleştirilmiştir. Modelin, her yoğun katmanında farklı sayıda filtre bulunacak şekilde tasarlanmıştır. Haber içeriğini ve sosyal bağlama dayalı bilgileri tek tek ve kombinasyon halinde sınıflandırma için, optimum hiper parametrelerle derin bir sinir ağır kullanılmış ve gerçek dünya veri setleri üzerinde doğrulanmıştır [71].

Çizge sinir ağlarını kullanan bir çalışmada, GNN'lerin, en son teknolojiye sahip yöntemlerle herhangi bir metin bilgisi olmadan karşılaştırılabileceği veya üstün performans elde edebileceği ve belirli bir veri kümesi üzerinde eğitilen GNN'ler yeni, görünmeyen veriler üzerinde düşük performans gösterebileceği ve doğrudan artımlı olarak eğitmek için sürekli öğrenme tekniklerinin kullanılabileceğini önermişlerdir [72].

Geometrik derin öğrenmeye dayalı yeni bir otomatik sahte haber algılama modelinin altında yatan çekirdek algoritmalar, içerik, kullanıcı profili ve etkinliği, sosyal grafik ve haber yayılımı gibi heterojen verilerin kaynaşmasına izin veren klasik evrişimli sinir ağlarının grafiklere genelleştirmektir [73].

Haberlerden çıkarılan özelliklere odaklanan çalışmada, sahte haberlerin otomatik tespiti için denetimli makine öğrenmesi algoritmaları kullanılmış ve karşılaştırılmıştır [74].

Derin öğrenmeyi geleneksel makine öğrenmesi algoritmalarıyla beraber kullanan çalışmada, beş makine öğrenmesi modeli ve üç derin öğrenme modelinin performansı, farklı boyutlardaki iki veri seti üzerinde çapraz doğrulama yaparak değerlendirmişlerdir. Yüksek test doğruluğuna erişme sebeplerinin yığınlama olduğunu iddia etmişlerdir [75].

Sahte haberleri tespit etmek için evrişimli sinir ağlarını marj kaybıyla önermiş ve modeli farklı kelime yerleştirme yöntemleriyle test etmiş olan çalışmada ISOT veri setine oranla LIAR veri setinde çok düşük performans sonuçları elde edilmiştir [76].

Sahte haber tespit görevinde kapsül sinir ağlarını kullanmayı amaçlayan bir çalışmada, farklı uzunluktaki haberler için farklı kelime yerleştirme modelleri kullanılmıştır. Statik sözcük yerleştirme, kısa haberler için kullanılırken, eğitim aşamasında artımlı eğitim ve güncellemeye izin veren statik olmayan sözcük yerleştirmeler, orta uzunlukta veya büyük haber ifadeleri için kullanılmıştır [77].

Sosyal medya için Türkçe sahte haber tespiti yaklaşımlarının ilk örnekleri arasında olan çalışmada, yeni bir kıyaslama veri seti olan SOSYalan Türkçe veri seti oluşturulmuştur. Derin öğrenme modelleri hem Türkçe hem de İngilizce dilleri için mevcut literatürden daha iyi performans göstermiştir. Derin öğrenme algoritmalarının eğitim ve test aşamaları makine öğrenmesi algoritmalarına kıyasla daha fazla zaman almaktadır [79].

Söylenti tespitine yönelik mevcut çalışmaların çoğu, mikroblog içerikleri, kullanıcılar ve yayılma modelleriyle ilgili modelleme özelliklerine odaklanmakta, ancak zaman içinde ileti yayılımı sırasında bu sosyal bağlam özelliklerinin değişiminin önemini göz ardı etmektedir. Yaptıkları çalışmada, çeşitli sosyal bağlam bilgilerini birleştirmek maksadıyla zaman serisi modelleme tekniğini uyguladıkları söylentinin yaşam döngüsünün zaman serisine dayalı olarak bu özelliklerin zamansak özelliklerini yakalayan bir yaklaşım sunmuşlardır. SVM-TS (SVM-Time Series) adını verdikleri yaklaşım, buluşsal kuralları

kullanan bir yöntemdir ve sahte haberleri tespit etmek için doğrusal bir SVM sınıflandırıcı kullanmıştır [80].

Söylentileri belirlemek için mikroblog olaylarının sürekli temsillerini öğrenen başka bir çalışmanın önerdiği modelde, ilgili gönderilerin bağlamsal bilgilerinin zaman içindeki değişimini yakalayan gizli temsilleri öğrenmek için RNN kullanılmıştır. RNN yönteminin, el yapımı özellikleri kullanan son teknoloji söylenti tespit modellerinden daha performans gösterdiğini, RNN tabanlı algoritmanın performansının, gelişmiş tekrarlayan birimler ve ekstra gizli katmanlar yoluyla daha iyileştirilebileceği ve RRN tabanlı yöntem, önde gelen çevrimiçi söylentileri çürütme hizmetleri de dahil olmak üzere mevcut tekniklerden daha hızlı ve doğru bir şekilde tespit edebileceği savunulmuştur [81].

CNN'e dayalı yeni bir yöntem olan CAMI (Convolutional Approach for Misinformation Identification- Yanlış Bilgi Tespiti için Konvolüsyonel Yaklaşım) adlı yaklaşımı sunan çalışmada, bir giriş dizisi arasına dağılmış temel özellikleri esnek bir şekilde çıkarabilen ve önemli özellikler arasındaki üst düzet etkileşimleri çıkarabilen ve önemli özellikler arasındaki üst düzey etkileşimleri şekillendirebilen bir yapı bulunmaktadır. Bu da yanlış bilgileri etkili bir şekilde tanımlamayı sağladığı ve pratik erken tespit sağlamaya yardım olduğu savunulmuştur [82].

Uzun haber makaleleri için ayrıntılı şekilde sahte haber sınıflandırması yapmak maksadıyla özellik mühendisliği ihtiyacını ortadan kaldıran grafik sinir ağı tabanlı bir model önerilmiştir. Önerdikleri yöntemin güçlü nöral temelleri aştığını ve mevcut veri kümelerinde son teknoloji doğruluk elde ettiklerini iddia etmişlerdir, GCN-text adı verilen yaklaşımda, haberlerdeki cümleler arasındaki ilişkileri grafik yapıları kullanarak modellemişler ve sahte haberlerin tespit edilmesi görevini bir grafik sınıflandırma problemine haline getirmişlerdir [83].

Çok modlu sahte haberleri tespit etmek için aynı anda modlar arası korelasyonu modelleyen, varlıkla güçlendirilmiş çok modlu füzyon çerçevesi sunan çalışmada BERT kullanılmıştır. Bert-text, haber makalesinin temsilini elde etmek için önceden eğitilmiş bir BERT ve sınıflandırma için tamamen bağlantılı bir katman kullanmaktadır [84].

Denetimli ve denetimsiz makine öğrenmesi algoritmalarını Türkçe sahte veri kümeleri üzerinde değerlendiren çalışmada, denetimli algoritmalar için %86 ve denetimsiz algoritmalar için %72 f1 puanı elde etmiştir [85].

Gelişmiş bir derin dil dönüştürücüsü öneren çalışmada, Türkçe dilinde COVID-19 için bir sahte haber tanımlama modeli geliştirmenin önemi vurgulanmaktadır. Çalışma, türkiye'de sosyal medya kaynaklı gerçek COVID-19 haberlerini tespit etmek için bir model geliştirmiştir. Beş geleneksel makine öğrenimi algoritması ve LSTM, Bi-LSTM, CNN ve GRU gibi derin öğrenme algoritmaları test edilmiş, BERT ve varyasyonları verimliliği artırmış ve %98,5 doğruluk elde etmiştir [86].

Sahte haberlerin tespiti için son teknolojite kıyasla daha iyi bir doğruluk elde ettiklerini idda ettikleri çalışmalarında bir topluluk öğrenme modeli önermişlerdir. Önerilen model, sahte haber veri kümelerinden önemli özellikleri çıkarmakta ve çıkarılan özellikler daha sonra karar ağacı, rastgele orman ve ekstra ağaç sınılandırıcı olmak üzere üç popüler makine öğrenmesi modelinden oluşan topluluk modeli kullanılarak sınıflandırılmaktadır. LIAR veri kümesibde sırasıyla %99,8 ve %44,15 eğitim ve test doğruluğu elde etmişlerdir. ISOT veri kümesi için %100 eğitim ve test doğruluğu elde ettiklerini iddia etmişlerdir [78].

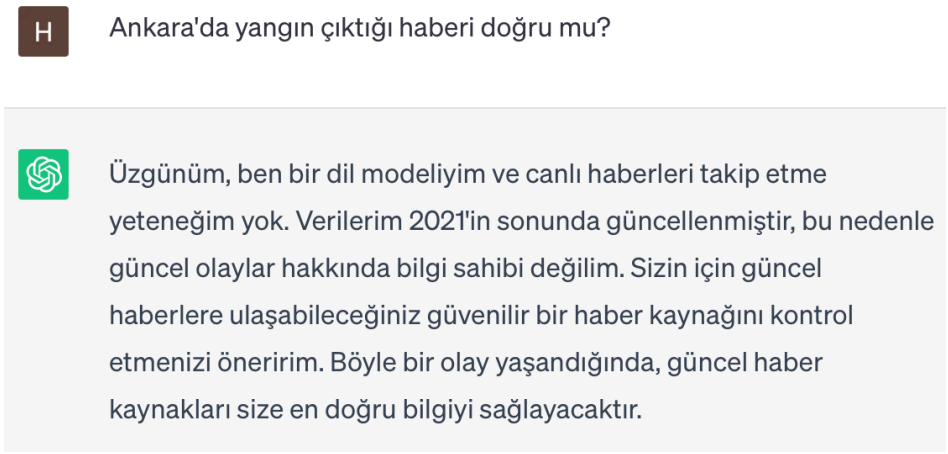
Endonezya sahte haber veri kümelerindeki dengesiz veri sorunuyla başa çıkmak için bir topluluk öğrenme tekniği öneren çalışmada, rastgele orman sınıflandırıcısının, 0.98 gibi etkileyici bir f1 skoru elde ederek topluluk sınıflandırmasında çok terimli sınıflandırıcıyı geride bıraktığını göstermişlerdir. Topluluk modeli olmayan ve destek vektör makinesi sınıflandırıcıları 660 belgeyi değerlendirmek için kullanılmıştır. Elde edilen f1 skorları sırasıyla 0.43 ve 0.74 olmuştur [87].

Gerçek ve sahte haberleri sınıflandırmak için 11 makine öğrenmesi algoritması ve Gradient Boosting ve Ada Boosting gibi tespit teknikleri kullanan çalışmada %94,5 doğruluk elde eden bir Topluluk oylama sınıflandırıcı kullanan akıllı bir sistem önermişlerdir [88].

GPT-2, ilk olarak Şubat 2019'da OpenAI tarafından duyurulan, tutarlı metin paragrafları oluşturan büyük ölçekli, denetimsiz bir dil modelidir [89]. Modelin küçük ile büyük arasında değişen dört çeşidi bulunmaktadır. Aynı yılın mayıs ayında GPT-2 tarafından üretilen ve WebText örneklerinden oluşan bir veri seti yayınlamışlardır [90]. Bu çalışmada sahte ve gerçek haber içerikleri yer almaktadır [91]. Bu tez çalışmasında, söz konusu sahte haber veri setinin belirli bir kısmı kullanılmıştır.

Araştırmalarına laboratuvar ortamında OpenAI adı altında 2015 yılında başlayan bir grup araştırma ekibi 30 Kasım 2022 tarihinde ChatGPT adında bir sistemin lansmanını yapmışlardır. Kişilerin dilini öğrenme ve çok amaçlı yapay zeka asistanları gibi etkileşimde bulunan ChatGPT bir dil modeli yapısı kullanan, ele aldığı cümleyi okuyup, cümlenin kullandığı dil yapısını anlayıp cevaplayan bir sistemdir [87]. Türkçe dil desteği de bulunan sistemde, sorulan sorulara insan gibi yanıt veren ve bu soruları çok hızlı bir sürede anlayarak anlamlı cümleler haline getiren sistemi çalışma kapsamında literatüre tezin çalışma süresi içerisinde kazandırıldığı için incelemeler yapılmıştır.

ChatGPT'nin sahte haber kapsamında değerlendirilmesinde, gelişmiş bütün makine öğrenmesi uygulamalarındaki gibi potansiyel güvenlik endişeleri taşıdığı görülmüştür. ChatGPT üzerinden sahte veya istenmeyen çıktılar üretebilen saldırganların sosyal medyada geniş bir erişim alanına sahip platformlarla bağlantı kurmasıyla, sahte haber yayma potansiyelinin de artabileceği öngörülmektedir [87]. Diğer yandan, çalışmada ChatGPT'nin soru yanıtlama sistemine girilen sahte haberin tahminlemesini yapmak için denemeler yapılmıştır. Yapılan denemeler içerisinde güncel olabilecek sorular da geçmişe yönelik sorular da mevcuttur. ChatGPT ara yüzünden sorulan sorulardan biri olan "Ankara'da yangın çıktığı haberi doğru mu?" sorusudur ve sistemin soru cümlesini tanıyarak ve dil çözümlemesini yapmasını müteakip bize verdiği yanıt Şekil 2.1.'de gösterilmiştir.



Şekil 2.1. ChatGPT soru-yanıt örneği

Sorulan soruya Şekil 2.1'de görüldüğü gibi verilen cevapta, ChatGPT bir dil modeli olduğu ancak güncel haberleri takip etme yeteneğinin bulunmadığını ve besleme verilerinin 2021 yılı sonunda güncellendiğini belirtmiştir.

ChatGPT'ye sorulan örnek soruya ve alınan yanıta bakıldığında, dil modeli olduğunu doğrulayan ve güncellemesini 2021 yılında yaptığı için güncel verilere erişemediğini dile getiren bir sistem görülebilmektedir.

Çalışmanın bu bölümünün devamında ise sahte haber tespitine yönelik çalışmalar ve kullanılan yaklaşımlar çeşitli başlıklar altında incelenmiş ve tabloya dökülmüştür. Söz konusu özet tablo yapılan çalışma, yayınladığı yıl, kullandığı veri seti, kullanılan dil, özellik çıkarım yöntemi, en performanslı yaklaşım ve performans kriteri göz önünde bulundurularak Çizelge 2.2.'de gösterilmiştir.

Çizelge 2.2. Sahte haber tespitine yönelik çalışmalar

Ref	Yıl	Veri seti	Dil	Özellik Çıkarımı	En iyi model	Performans
[25]	2021	LIAR,	İngilizce	BERT Kelime Yerleştirme	BERT, RoBERTa	0.62
		fake or real news,			RoBERTa	0.98

Ref	Yıl	Veri seti	Dil	Özellik Çıkarımı	En iyi model	Performans
		fake or real news,			RoBERTa	0.96
[26]	2022	Twitter veri seti	İngilizce	Kullanıcı özellikleri	CNN	%85
[27]	2021	BanFakeNews	Bangla	1D CNN	RF, DT, Adaboost, SVM, KNN	%99
[28]	2022	Covid-19 Sinhala Dataset	Sinhale	n-gram	Passive Agressive Sınıflandırıcı	%88
[29]	2019	ego-Twitter	İngilizce	Node2vec	Lienar Discriminant Analysis	%98
[30]	2022	FNC	İngilizce	Dil özellikleri	XGBoost	%89.4
[31]	2021	fake or real news	İngilizce	Vektör sayma, TF-IDF	Naive Bayes	%80
[32]	2023	Weibo	İngilizce	Metin özellikleri	MANN	%77
[33]	2023			yok	ELI5, SHAP	%75
[34]	2021	News	İngilizce	Kelime uzunluğu	LSTM	%91,73
[35]	2023	LIAR	İngilizce	N/A	SVM,RF	%97
[36]	2023	Covid-19 Fake News Dataset	İngilizce	N/A	Multilingual-BERT	%97,1
[37]	2022	FakeNewsNet	İngilizce	dill özellikleri	NN+GAN	%91,2
[38]	2021	ISOT	İngilizce	Özel semboller, argo, özel kelimeler, içerik uzunluğu, kelime sayısı	WOA, xgbTree	<%90
[39]	2022	Verify ve ve Reuters'dan elde edilen veri seti	Arapça	Faraşa segmentasyonu ile elde edilen kelime vektörleri	CNN	%91
[40]	2022	Covid-19, Arabic fake news ve Arabic rumor-non rumor tweetlerinden oluşan veri setleri	Arapça	Word2vec ve fastText kelime yerleştirme, kelime frekansı	AraBERT, MARBERT	%97
[41]	2022	BanFakeNews	Bengalce	fastText, word2vec, n-	AugFake-BERT	%92.45

Ref	Yıl	Veri seti	Dil	Özellik Çıkarımı	En iyi model	Performans
				gram		
[42]	2017	LIAR ve FakeHealth	İngilizce	BoW, TF-IDF, word2Vec ve doc2Vec	Gradient Boostin Regressor	F1 ölçütü: %81.2
[43]	2022	PolitiFact ve GosspCop	İngilizce	-	BERT	%93 ve %92
[44]	2021	Facebook verisi	İngilizce	TF-IDF, BoW	Hibrit SVM	%91.23
[45]	2017	Google'dan elde edilen haber makaleleri	Endonezca	-	NB	%78,6
[46]	2018	Özel veriseti	Türkçe	terim frekansı, TF-IDF, n-gram, stil işaretçileri, argo/küfür kullanımı, url, haberdeki erişilebilir link özelliği, Başlık ve Haber içeriği uyumu, Abartılı Başlık kullanımı	SVM,NB	F1 ölçütü %79
[47]	2017	Kaggle veri seti	İngilizce	n-gram, anlamsal benzerlik	SGD, SVM, KNN, LR, DT	0.90
[48]	2017	Facebook haber gönderileri veri seti	İngilizce	-	NB	0.74
[49]	2016	2 hicivli haber sitesi (The Onion ve The Beaverton) ve 2 meşru haber kaynağından (The Toronto Star ve The New York	İngilizce	Absürt kelime, dilbilgisi, noktalama	SVM	F1-ölçütü: 0.87

Ref	Yıl	Veri seti	Dil	Özellik Çıkarımı	En iyi model	Performans
		Times) elde edilen haberler				
[50]	2017	FakeNewsAMT ve Celebrity veri seti	İngilizce	n-gram, noktalama, psikodilbilimsel özellikler, söz dizimi, okunabilirlik	SVM	F1 ölçütü: 0.73
[51]	2020	TR_FN	Türkçe	Kök sayısı, ham sayı kelime başı hece hatası, haber okunabilirliği, kaynak, haber kategorisi, haber adresi	NB, KNN, ExtraTrees, SVM, LR, RF, DT	%96,81
[52]	2022	Özel veri seti		TF-IDF ve word2vec	SVM	F1 ölçüt: 0.90
[53]	2021	Covid-19 fake news	İngilizce Hintçe	TF-IDf ve BoW	SVM	%93,44
[54]	2020	FakeBrCorpus, TwitterBR, Fakenewsdta1, fakerorreal news, byvlifesytle	İngilizce Portekizce Bulgarca	BoW ve word2vec	Makine öğrenmesi algoritmaları	%91, %81, %86, %94, %95
[55]	2020	ISOT, DS2, DS3, DS4	İngilizce	LIWC	RF, LR, KNN, CART, LSVM, DT, AdaBoost, XGBoost	%99, %94, %96, %91
[56]	2020	buzzFeed, ISOT, Random political news	İngilizce	TF-IDF	BayesNet, JRip, OneR, ZeroR, SGD, J48, DT vb.	%65.5, %96.8, %68
[57]	2020	BuzzFeed, GossipCop	İngilizce	Haber içeriği, sosyal bağlam ve uzay zamansal özellikler	SVM, LR, NB, CN, Social Article Fusion	%69.1, %72.3
[58]	2017	Özel veri seti	İngilizce	Beğeni sayısı	LR, HBLC	%99.3
[59]	2017	Kaggle veri seti	İngilizce	Haber içeriği	LR, RNN, GRU, LSTM, BiLSTM,	F1 ölçütü: 0.84

Ref	Yıl	Veri seti	Dil	Özellik Çıkarımı	En iyi model	Performans
					CNN	
[60]	2017	Snopestan elde edilen sahte, haber kuruluşlarından elde edilen gerçek haberler	İngilizce	Metin içeriği, duygu	KNN, SVM, LSTM	F1 ölçütü: 0.90
[61]	2018	FNC-1 açık kaynak veri seti	İngilizce	TF-IDF	LSTM, GRU	69,8
[62]	2018	Özel veri seti	İngilizce	Word2vec	RNN, LSTM, BiLSTM, GRU, BiGRU	0.84
[63]	2017	FNC-1 açık kaynak veri seti	İngilizce	n-gram, BoW, harici özellikler	MLP	F1 ölçütü: 83.8
[64]	2017		İngilizce		CSI, DT, SVM, LSTM, GRU	0.95
[65]	2017	PolitiFact tarafından etiketlenmiş 19 sahte, 9 gerçek haber makalesi	İngilizce	TF-IDF	3HAN, GRU	96.77
[66]	2017	Özel veri seti	İngilizce	LIWC	LR, RNN, CNN	0.92
[67]	2017	LIAR	İngilizce	Metin içeriği	LR, SVM, BiLSTM, CNN	0.27
[68]	2018	LIAR	İngilizce	Metin içeriği	LR, SVM, RNN, GRU, LSTM, Bi-LSTM,	0.27

Ref	Yıl	Veri seti	Dil	Özellik Çıkarımı	En iyi model	Performans
					CNN	
[69]	2019	Fakenews.mit.edu.tr sayfasından elde edilen açık kaynak veri seti	İngilizce	Word2vec	GRU, LSTM, BiLSTM, SMHA-CNN	95.5
[70]	2021	Kaggle veri seti	İngilizce	-	FakeBERT	98.9
[71]	2021	BuzzFeed, PolitiFact	İngilizce	Kelime sayısı, POS etiketi	EchoFakeD	91.8, 92.3
[72]	2020	PolitiFact, GossipCop	İngilizce	Doğrulanmış kullanıcı, zaman, takipçi ve arkadaş sayısı, beğeniler, listeler, durumlar	GNN	81.1, 85.3
[73]	2019	BuzzFeed	İngilizce	Kullanıcı profili, kullanıcı aktiviteleri, ağ ve yayılma, içerik	Geometrik derin öğrenme	%92.7
[74]	2019	BuzzFeed	İngilizce	Sözdizimi, sözcüksel özellikler, LIWC, anlamsal özellikler, öznellik	KNN, NB, RF, SVM, XGB	%86
[75]	2021	ISOT, KDnugget	İngilizce	TF, TF-IDF	LR, DT, KNN, RF, SVM, CNN, LSTM, GRU	%99.87, %92.82
[76]	2021	ISOT, LIAR	İngilizce	-	CNN	%99.1, %41.6
[77]	2021	ISOT, LIAR	İngilizce	N-gram	Kapsül ağları	%99.8, %40.9
[78]	2021	ISOT, LIAR	İngilizce	bOW	DT, RF, ExtraTrees	%100, %44.15
[79]	2023	BuzzFeed, ISOT,	İngilizce,	Word2vec,	CNN, RNN-	%93.41

Ref	Yıl	Veri seti	Dil	Özellik Çıkarımı	En iyi model	Performans
		SOSYalan	Türkçe	kelime başına vektör temsili	LSTM	, %99.9, %92.48
[80]	2015	Media-Weibo, Media-Twitter, PHEME, PolitiFact, GossipCop, NewsBag	İngilizce	Mikriblog içeriği, kullanıcılar, yayılma modelleri	SVM	0.64, 0.53, 0.64, 0.60, 0.50, 0.51
[81]	2016	Media-Weibo, Media-Twitter, PHEME, WeChat, LIAR, Fakeddit, Ti-CNN	İngilizce, Çince	Kullanıcılar, gönderiler, olaylar, söylentiler, olay başı ortalama zaman, olay başı gönderi sayısı	GRU	0.70, 0.63, 0,74, 0.73, 0.72, 0.83, 0.86
[82]	2017	Media-Weibo, Media-Twitter, PHEME, PolitiFact, GossipCop, WeChat, Fakeddit, NewsBag	İngilice, Çince	Paragraf vektörleri	CNN	0.74, 0.55, 0.78, 0.65, 0.77, 0.75, 0.83, 0.5
[83]	2019	Media-Weibo, Media-Twitter, PHEME	İngilizce	Cümle matrisi	GCN	0.81, 0.70, 0.83
[84]	2021	Media-Weibo, Weibo-20, PolitiFact, GossipCop, Fakeddit, Ti-CNN	İngilizce	POS etiketleri, görsel özellikler, gömülü metin	BERT	0.83, 0.9, 0.71, 0.86, 0.87, 0.87
[85]	2021	Özel veri seti	Türkçe	TF-IDF, word2vec, doc2vec	SVM	%86
[86]	2022	Twitter, doğrulama platormu Twitter hesapları, COVID-19 veri seti	Türkçe	Bilgi Kazancı, Kazanç Oranı, Korelasyon Tabanlı	BERTurk	%98.5

Ref	Yıl	Veri seti	Dil	Özellik Çıkarımı	En iyi model	Performans
				Özellikler		

Çizelge 2.2.' de görüldüğü gibi, incelenen çalışmalar arasında sahte haber tespiti gerçekleşen çalışmalar arasında Türkçe dilini kullanan çalışmalar el ile sayılacak kadar azdır. Kullanılan modeller düzeyinde bakıldığında tez çalışmasında oluşturulan ana modelde kullanılan modelleri kullanan az sayıda çalışma bulunmaktadır. Türkçe dilinde ana modele ait çalışma mevcut değildir. Bu bakımdan çalışmanın, Türkçe sahte haber tespiti konusunda BERTurk ve Turkish-Bert-base kullanan ilk çalışma olduğunu söylemek mümkündür. Makine öğrenmesi, fastText ve ana model olan iyileştirilmiş BERT modellerinin veri setlerine uygulanması hususunda yapılan çalışmalar modellerin anlatıldığı bölümlerde ayrıntılı olarak değerlendirilecek ve değerlendirme metrikleri kullanılan veri setlerine göre karşılaştırılacaktır.

2.3. Literatürdeki Veri Setleri

Sahte haber tespiti için algoritmaların performansının değerlendirilmesine yardımcı olacak mevcut veri setlerine odaklanılmıştır. Farklı makaleler, farklı veri kümelerini kullanan deneysel sonuçları rapor ettiğinden, yöntemleri makul bir ölçekte adil bir şekilde karşılaştırmak için, son zamanlarda birkaç kıyaslama veri kümesi önerilmiştir. Aşağıdaki veri kümeleri, gerçek dünyada sosyal medya platformlarından toplandığı ve sahte haber tespitinde yaygın olarak kullanıldığı için gözden geçirilmiştir.

BuzzFeedNews

BuzzFeedNews veri seti [88], 2016 ABD seçimlerinden bir hafta önce 19-23 Eylül ve 26-27 Eylül tarihlerinde 9 haber ajansı tarafından Facebook'ta yayınlanan haberlerin tam (eksiksiz) bir örneğini içermektedir. Her yazı ve bağlantılı makale 5 BuzzFeed gazetecisi tarafından hak talebinde bulunulmuş ve tek tek doğrulanmıştır. Bu veri seti, 826 ana akım makale, 356 sol görüşlü makale ve 545 sağ görüşlü makale dahil olmak üzere toplam 1627 adet makale içermektedir [89]. Ayrıca bu makalelere bağlı olarak, medyalar ve çeşitli veriler de dahildir [90].

LIAR

LIAR veri seti API'si aracılığıyla bilgi kontrol web sitesi PolitiFact'ten toplanmıştır [67]. Haber bültenleri, TV veya radyo röportajları, kampanya konuşmaları vb. gibi çeşitli bağlamlardan örneklenmiş 12836 insan etiketli kısa ifadeleri içerir. Haberler için etiketler, altı sınıftan oluşmaktadır: pants-fire, yanlış, hemen hemen doğru, yarı doğru, çoğunlukla doğru ve doğru [91]. Ayrıca konular, taraf, bağlam ve konuşmacılar hakkındaki bilgileri de içermektedir.

BS Dedector

BS Dedector veri kümesi, haber doğruluğunu kontrol etmek için geliştirilen BS Dedector adı verilen bir tarayıcı uzantısından toplanır [92]. Manuel olarak uygun bir etki alanı listesine bakarak kontrol ederek belirli bir web sayfasındaki tüm bağlantıları güvenilmez kaynaklara yapılan başvuruları arar. BS dedektörün çıkış noktaları insan açıklamalarından değil, etiketlerden oluşmaktadır [93].

CREDBANK

CREDBANK veri seti [94], Ekim 2015'ten başlayarak 96 günü kapsayan yaklaşık 60 milyon tweetten oluşan kitle kaynaklı büyük bir veri kümesidir. Bu veri setindeki tweetlerin hepsi 100 ve üzeri sayıda haberle ilişkilendirilmiştir. İlişkilendirilen tweetlerin tümünün güvenilirliği Amazon MechanicalTurk'teki 30 yorumcu tarafından incelenmiş ve derecelendirilmiştir [94,95].

BuzzFace

BuzzFace veri kümesi [96], BuzzFeed veri kümesini Facebook'taki haber makaleleriyle ilgili yorumlarla genişleterek toplanmıştır. Veri kümesi, 2263 haber makalesi ve haber içeriğini tartışan 1,6 milyon yorum içermektedir.

FacebookHoax

FacebookHoax veri kümesi [58], Facebook API kullanılarak toplanan bilimsel haberler (aldatmaca olmayan) ve komplo sayfaları (aldatmaca) ile ilgili Facebook sayfalarından gelen gönderilere ilişkin bilgileri içermektedir. Veri seti, 2.300.000'den fazla beğeni ile 32 sayfadan (14'ü komplo ve 18'i bilimsel) 15.500 gönderi içermektedir.

Twitter15 ve Twitter16

Twittter15 ve Twitter16 veri setleri[97] sırasıyla 1381 ve 1181 yayılma ağacı içermektedir. Her veri kümesinde, ağaç yapısı, yanıtlar ve retweetler gibi yayılma dizilerinin yanı sıra kaynak tivitleri bir koleksiyonunu içermektedir. Her ağaç söylenti olmayan, yanlış söylenti, gerçek söylenti veya doğrulanmamış söylenti olarak kategorize edilmektedir. Bu veri setlerindeki her haber etiketi, söylentiyi yok eden web sitelerinden (Ör. Snopes.com, Emergent.info vb.) makalenin doğruluk etiketine göre etiketlenmektedir.

Ma-Twitter

Twitter'dan [7] toplanan Ma-Twitter veri setinde[76], gerçek zamanlı bir söylenti çürütme sitesi olan Snopes'tan[98] 498 söylenti toplanmaktadır. Ayrıca Snopes'tan 494 normal olay ve iki genel veri seti içermektedir. Veri kümesindeki her gnderi, Twitter içeriğini, yanıtları, yorumları ve yayıncının profilini içermektedir.

Media-Twitter

Media-Twitter veri seti [81], Çin'in sosyal medya sitesi Sina Weibo'dan [99] toplanmaktadır. Bilinen söylentiler, çeşitli yanlış bilgileri bildiren Sina topluluk yönetim merkezinden [100] toplanmıştır. Weibo API'si, orjinal mesajları ve etkinlik verilen tüm yeniden gönderme/yanıtlama mesajlarını yakalayabilmektedir. Gerçek haberler, söylenti olarak etiketlenmemiş normal başlıklardan gelen gönderiler toplanarak elde edilmiştir.

Media-Weibo

Media-Weibo veri seti [100], ilk olarak çok modlu sahte haber tespit görevi için sunulmuştur. Her gönderi, metin içeriği, kullanıcı profili ve ek resimlerden oluşan üç

bölümden meydana gelmektedir. Doğrulanan sahte haberler, Rwitter'a benzeyen Sina Weibo'nun resmi sahte haber ifşa sisteminden toplanmıştır. Verilerin dönemi Mayıs 2012'den Ocak 2016'ya kadardır. Gerçek haber, saygın bir Çin haber ajansı olan Xinhua Haber Ajansı'nndan alınmıştır [100]. Düşük kaliteli ve kopya görüntüler, önceki araştırmalardaki veri ön işleme yöntemlerine göre kaldırılmıştır. Veri kümesindeki her gönderi bir tweet kimliği, metin ve resim içermektedir.

Weibo-20

Weibo-20 veri seti[102], Media-Weibo veri setinin [101] genişletilmiş bir versiyonu olan Çince sahte haber tespit veri setidir. Haberleri sahte ve gerçek olmak üzere iki sınıfta tutmaktadır. Sahte haberler için Media Weibo'da 1355 sahte haber parçasını saklamaktadır ve Nisan 2014'ten Kasım 2018'e kadar Weibo Topluluk Yönetim Merkezi tarafından sahte olduğu resmi olarak doğrulanan haberleri toplamıştır. Gerçek haberler için Media-Weibo'nun 2351 gerçek haberini saklamaktadır ve sahte haberlerle aynı dönemde 850 benzersiz yeni parçayı bir araya getirmiştir. Yeni toplanan gerçek haberler, Weibo'daki şüpheli haber parçalarını keşfetmeye ve doğrulamaya odaklanan NewsVerift tarafından doğrulanan gerçek haberlerdir. Weibo-20 veri seti, 3161 sahte haber ve 3201 gerçek haber içermektedir.

Weibo-21

Weibo-21 veri seti [103], Çince'de açıklamalı alan etiketine sahip çok alanlı bir sahte haber veri setidir. Aralık 2014'ten Mart 2021'e kadar Sina Weibo'dan hem sahte hem de gerçek haberler toplanmıştır. Sahte veriler açısından Weibo21, resmi olarak Weibo Topluluk Yönetim Merkezi tarafından yanlış bilgi olarak değerlendirilmiştir. Weibo21, gerçek veriler için aynı dönemdeki gerçek haberleri toplamıştır. NewsVerify [104], Weibo'daki şüpheli haberleri keşfeden ve doğrulayan bir sahte haber doğrulama platformudur. Veri setinde her haber parçası, haber içeriğini, haberin görselini, zaman damgasını, kullanıcı yorumlarını ve analan adını içermektedir. Weibo-21, bilim askeri, eğitim, afetler, politika, sağlık, finans, eğlence ve toplum olmak üzere 9 alana ait haberleri içermektedir.

MCG-FNeWS

MCG-FNeWS veri seti [105], çok modlu bir sahte haber tespit veri setidir. Veri kümesindeki her haber öğesi, hem metin hem de beraberindeki bir görseli içermektedir. Veri kümesi, Mayıs 2012'den Kasım 2018'e kadar Sina Weibo haberlerini içermektedir.

NewsBag

NewsBag veri seti [106], sahte haberleri tespit etmek için çok modlu bir veri kümesidir. Eğitim setinde The Wall Street [107], kaynaklı gerçek haberler ve The Onion [108] kaynaklı sahte haberler olmak üzere 200.000 gerçek haber ve 15.000 sahte haber yer almaktadır. Test setinde 11.000 gerçek haber ve 18.000 sahte haber yer almakta olup, gerçek haberler The Real News[109] ve sahte haberler The Poke[110] kaynaklıdır.

Ti-CNN

Ti-CNN[111], sahte haberleri. Tespit etmek için kullanılan çok modlu bir veri kümeisidr. Veri setinde 11.941'i sahte, 8074'ü doğru olmak üzere toplam 20.015 adet haber yer almaktadır. Veri kümesindeki her haber makalesi bir başlık, metin, resim ve yazar bilgilerini içermektedir. Sahte haberler Kaggle'dan[112], gerçek haberler ise New York Times[113] ve Washington Post'tan[114] elde edilmiştir.

Politifact

Politifact [115], Amerika Birleşik Devletleri'nde siyasi beyanlar ve raporlar sunan, kar amacı gütmeyen ünlü bir bilgi doğrulama sitesidir. Politifact veri setindeki haberler Mayıs 2002'den Temmuz 2018'e kadar yayınlanmıştır. Etki alanı uzmanları, veri kümesindeki haberler için temel gerçeklik etiketleri (yanlış veya gerçek) sunmaktadır. PolitiFact'ten teyit edildikten sonra çıkan haberler, esas olarak politikacılar (Kongre üyeleri, Beyaz Saray çalışanları, lobiciler) ve siyasi gruplar tarafından yayınlanan açıklamalar veya haber makaleleridir. PolitiFact, bu haber makaleleri için web sitesinde orijinal içerikleri, doğrulama sonuçlarını ve kapsamlı doğrulama raporlarını sağlamaktadır. Platform, onları içeriğe ve konuya göre farklı konularda gruplamaktadır. Konunun kısa bir açıklaması

mevcuttur. Teyit sonuçları, doğru (2149 adet), çoğunlukla doğru (2676 adet), yarı doğru (2764 adet), çoğunlukla yanlış (2539 adet), yanlış(2601 adet) pants fire (1322 adet) şeklinde haber makalelerini etiketlemiştir. Pants fire, yanlış, çoğunlukla yanlış etiketlerini sahte haber olarak, doğru, çoğunlukla doğru, yarı doğruyu gerçek haber olarak algılamaktadır. Böylece 6465 adet sahte haber ve 7590 adet gerçek haber elde edilmiştir.

GossipCop

GossipCop [116] da bir bilgi kontrol sitesidir. GossipCop veri setindeki haberler Temmuz 2000'den Aralık 2018'e kadar yayınlanmıştır. Etki alanı uzmanları, veri kümesindeki haberler için temel gerçeklik etiketleri vererek haber etiketlerinin kalitesini sağlamaktadır.

FakeNewsNet

FakeNewsNet [116], bilgi doğrulama web siteleri BuzzFeed [117] ve PolitiFact'ten haberler içermektedir. Veri kümesi haber içeriğini, kullanıcı bilgilerini ve retweet'leri içermektedir. Veri setinde toplam 23.196 haber makalesi ve 69.733 retweet bulunmaktadır.

PHEME

PHEME [118] veri seti, Twitter platformundan atılan tweetlerden oluşmaktadır. Ayrıca, her biri bir tweet koleksiyonuna sahip beş ader son dakika haber kaynağından toplanmıştır. Her tweet haberi, metinlerin yanı sıra görüntüleri de içermektedir.

WeChat

WeChat [119], yarı denetimli bir sahte haber algılama veri kümesidir. Veri tabanı hem haber makalelerini hem de kullanıcı yorumlarını içermektedir. Veri seti, Mart. 2018 ile Ekim 2018 arasındaki sosyal medya platformu WeChat [119]'ten gelen haberleri içermektedir. Veri kümesindeki haberlerin yalnızca küçük bir kısmı WeChat ekibinin

uzmanları tarafından etiketlenirken, çoğu etiketlenmeden bırakılmıştır. Ancak haber etiketli olsun veya olmasın hepsinde yorum bilgisi bulunmaktadır.

Fakeddit

Reddit platformunun birden çok alt dizininden kaynaklanmaktadır. Fakeddit [120], veri kümesi, metin cümleleri, resimler ve sosyal bağlam bilgileri dahil olmak üzere zengin bilgiler toplamaktadır. Ayrıca Fakeddit veri seti, her örnek için 2'li 3'lü ve 6'lı etiketler sağlayarak araştırmacıların daha ayrıntılı bir sahte haber tespit modeli geliştirmesine olanak tanımaktadır. 2'li sınıflandırma, haberin gerçek mi yoksa sahte mi olduğunu belirlemektedir. 3'lü sınıflandırma, haberin tamamen gerçek olup olmadığını, sahte olup olmadığını ve doğru metin içerip içermediğini veya sahte olup ve sahte içerik içerip içermediğini değerlendirmektedir. Temel bir ikili veya üçlü sınıflandırma yerine, farklı sahte haber türlerini ayırt etmek maksadıyla 6'lı sınıflandırma geliştirilmiştir. Altı kategori etiketi ise şunlardır: doğru, hiciv/parodi, aldatıcı içerik, sahte içerik, yanlış bağlam, manipüle edilmiş içerik.

FakeHealth

FakeHealth veri seti [121], sağlıkla ilgili sahte haberleri tespit etmeye yönelik bir veri kümesidir. Veri kümesindeki veriler, Health News Review [122], web sitesinden alınmıştır. Veri seti, haber içeriğini, kullanıcıların yorumlarını ve insanlar için sosyal ağları içermektedir.

CoAID

CoAID veri seti [123], COVID-19 ile bağlantılı sahte haberleri tespit etmeye yönelik bir veri kümesidir. Veri kümesi, haber makalelerini, kullanıcı yorumlarını ve kullanıcı verilerini içermektedir. Veri seti, 4251 adet haber makalesi, 296.000 adet kullanıcı yorumu ve güncel haber etiketlerini içermektedir.

FakeCovid

FakeCovid veri seti [124], 92 teyitçiden 105 ülkede dolaşan 5182 makaleyi toplayan çok dilli bir etki alanları arası veri kümesidir. Makaleler, doğruluğu kontrol edilmiş 11 haber kategorisine manuel olarak eklenmektedir. Makalelerin %40,8'i İngilizce'dir.

ReCOVery

NewsGaurd [125], web sitesinden toplanan, COVID-19 ile ilişkili çok modlu bir sahte haber algılama veri kümesidir [126]. Veri seti, haberlerle ilgili görüntüleri, metin içeriğini ve sosyal bağlam bilgilerini içermektedir. Bu veri seti, 2029 adet haber makalesi ve 140820 adet tweet içermektedir.

MM-COVID

COVID-19 ile ilgili çok dilli bir sahte haber tespit veri kümesidir [127]. Veri seti, haber içeriğini, sosyal paylaşımları ve haberlerle ilgili mekansal bilgileri içermektedir. Bu veri kümesi altı farklı dilden haberler içermektedir. Bu diller: İngilizce, İspanyolca, Portekizce, Hintçe, Fransızca ve İtalyanca'dır.

CHECHED

Çince bir COVID-19 ile ilgili sahte haber tespit veri setidir [128]. Veri seti, sosyal medya platformu Sna Weibo'dan 2104 haber makalesinden oluşmaktadır. Veri kümesi, metinsel haber içeriği, resimler, yayılma bilgileri ve etiketler içermektedir.

Cross-lingual COVID-19

COVID-19 hakkında hem İngilizce hem de Çince haberleri içeren diller arası bir COVID-19 veri setidir [129]. Veri kümesi iki bölüme ayrılmıştır: eğitim ve test setleri. Eğitim

setinde 2840 adet İngilizce rapor bulunmaktadır ve haberlerin %49,43'ü sahte haberdir. Test seti, %43'ü sahte haber olmak üzere 200 adet Çince haber içermektedir.

SLN

SLN veri seti [130] toplu olarak 360 adet haber içermektedir. Amerika Birleşik Devletleri ve Kanada'daki ulusal gazetelerde yer alan konuları içermektedir. Veri seti iki ayrı set halinde oluşturulmuştur. İlk set, iki hiciv haber sitesi olan The Onion ve The Beaverton'dan [131] ve iki meşru haber kuruluşu olan The Toronto Star [132] ve The New York Times'tan toplanmıştır. Yine hiciv ve meşru haber makaleleri arasında bölünmüş olan ikinci set, 6 meşru ve 6 hiciv Kuzey Amerika çevrimiçi haber kaynağından alınmıştır.

LUN

LUN veri seti[133], çok kategorili bir sahte haber algılama veri kümesidir. Veri kümesi, PolitiFact ve kardeş sitelerinden (PunditFact, vb.) haberler içermektedir. Veri kümesindeki haberler altı kategoride sınıflandırılmıştır. Bu kategoriler: doğru, çoğunlukla doğru, yarı doğru, çoğunlukla yanlış, yanlış ve pants-firedır. Ek olarak, ilk üçn derecelendirmeti doğru kalan üç derecelendirmeyi yanlış olarak Kabul eden iki sınıflı bir değişken mevcuttur.

GermanFakeNC

GermanFakeNC [134], Almanca bir sahte haber tespit veri kümesidir. Veri seti 490 adet haber makalesinden oluşmaktadır. Temel gerçek etiketleri, Alman tanınmış haber medyası web sitelerinden alınmıştır. Bu veri kümesi, sosyal bilgiler olmadan yalnızca haberler için metin bilgilerini içermektedir.

ISOT

ISOT veri seti[135], gerçek ve sahte kategorilerine neredeyse eşit olarak dağıtılan 45000 haber makalesinden oluşmaktadır. Gerçek makaleler Reuters web sitesinden ve çeşitli kaynaklardan gelen sahte makaleler Wikipedia[136] ve PolitiFact tarafından sahte kaynaklar olarak etiketlenmiştir. Veri kümeleri, her makalenin tam gövdesini, başlığını

tarihini ve konusunu içermektedir. Ana makale konuları siyaset ve dünya haberleridir ve tarihler 2016 ile 2017 yılları arasındadır [137].

TR_FN

TR_FN, Türkçe sahte haberleri içermesinden ötürü böyle bir isim verilmiştir. Çevirimiçi haber kaynaklarının taranmasıyla belirli bir formatta hazırlanmış veri arşivi projesi olan GDELT Projesi[138] ile haber metinleri çekilmiştir. Haberlerin geçerli olmasını garanti altına almak maksadıyla 3 büyük haber ajansı olan AA, DHA ve İHA'ya ait haberler kullanılmıştır. Sahte haberler için teyit.org kullanılmış ve ayrıca hem sahte hem de gerçek haber veri setine elle derlenen veriler de eklenmiştir. Sonuç itibariyle eğitim setinde elle derlenen ve GDELT'ten elde edilen toplam 83546 geçerli haber, test setinde ise 210 adet geçerli haber bulunmaktadır. Test veri setinde ise elle derlenen ve teyit.org'tan elde edilen toplam 1188 adet sahte haber ve test setinde 210 adet sahte haber bulunmaktadır. Veri setindeki alan bilgileri şunlardır: haber kaynağı, tarih bilgisi, sınıf bilgisi, haber sitesi, kategori, metin tanımlayıcı, stilistik çıkarımlar.

SOSYalan

SOSYalan veri seti [79], sahte veya gerçek olarak etiketlenmiş 9361 haber tweeti içermektedir. Sahte haberler Twitter API aracılığıyla '@teyitorg', '@dogrulukpayicom', '@zaytung', '@dogrulaorg', '@DeminHaber' Twitter hesaplarından elde edilen 4196 adet haberdir. Bu veri setinin sütunları haber numarası, haber metni ve haber etiketini içermektedir.

GPT-2 Sahte haber veriseti

Şubat 2019'da, dünyanın önde gelen apay zeka labarotuvarlarından biri olan Open AI, GPT-2 olarak bilinen üretici bir metin dili modeli yayınlamıştır. GPT-2, bir makale başlığı gibi birkaç kelime veya kelime öbeği hazırlayarak, girdiye dayalı olarak yüksek sağlamlıkta

tutarlı metin paragrafları oluşturabilmektedir. Bu veri kümesi WebText test setinden 250 bin belge içermektedir.

Açıklamalı olarak anlatılan, verilerini Sina Weibo, Twitter ve diğer sosyal medya platformlarının yanı sıra gerçeklik doğrulama sitelerinden toplayan Çizelge 2.3'de gösterildiği gibi lietratürdeki çalışmalarda kullanılan veri kümeleri sosyal bağlam özellikleri, haber içerikleri ve kullandıkları etiket sayıları kapsamında değerlendirilerek özetlenmiştir.

Çizelge 2.3. Literatürde kullanılan veri setleri

Veri Seti	Sosyal Bağlam			Haber İçeriği			Etiket Sayısı
	Kullanıcı Detayı	Gönderi Analizi	Ağ Analizi	Çok Dilli	Metin	Görsel	
BuzzFeedNews					√		4
LIAR					√		6
BS Dedector					√		
CREDBANK	√				√		5
BuzzFace		√			√		4
FacebookHoax	√	√			√		2
Twitter15 ve Twitter16	√	√			√		4
Ma_Twitter	√	√			√		2
Media_Twitter	√	√	√		√	√	2
Media_Weibo	√	√	√		√	√	2
Weibo_20	√	√	√		√	√	2
Weibo_21		√			√	√	2
MCG_FNeWS					√	√	2
NewsBag					√	√	2
Ti_CNN	√				√	√	2
PolitiFact					√		2,6
GossipCop					√		2
FakeNewsNet	√	√	√		√	√	2
PHEME	√	√	√		√	√	2
WeChat		√			√		2
Fakeddit		√			√	√	2,3,6
FakeHealth	√	√	√		√	√	2
CoAID		√			√	√	2
FakeCovid				√	√		2
ReCOVery	√	√	√		√	√	2

Veri Seti	Sosyal Bağlam			Haber İçeriği			Etiket Sayısı
	Kullanıcı Detayı	Gönderi Analizi	Ağ Analizi	Çok Dilli	Metin	Görsel	
MM_COVID	√	√	√	√	√	√	2
CHECKED	√	√			√	√	2
Cross_lingual COVID-19	√			√	√		2
SLN					√		2
LUN					√		2,6
German FakeNC					√		2
ISOT					√		2
TR_FN					√		2
SOSYalan					√		2
GPT-2 Sahte haber veriseti					√		2

Çizelge 2.3'de literatürde sahte haber tespiti yapmak amacıyla oluşturulan modellerin performanslarını değerlendirmek için kullanılan veri setleri verilmiştir. Tez çalışması kapsamında bu veri setleri içerisinden BuzzFeed, GosspCop, ISOT, LIAR, Twitter15, Twitter16 ve GPT-2 sahte haber veri setleri kullanılmıştır. Türkçe sahte haberleri tespit etmek için oluşturulan kıyaslamalı Türkçe veri setinin adı tezin kalan kısımlarında TR_FaReNews olarak bahsedilecektir.

3. MATERYAL VE METOT

Mevcut çalışmaların incelenmesinin ardından sahte haber tespitine yönelik yapılan tez çalışmasına ait materyallere ve çalışma metodolojisine bu bölümde yer verilecektir.

3.1. Materyal

Twitter sosyal platformu kullanılarak sahte haber tespiti amacıyla oluşturulması planlanan modele ait yöntemin belirlenmesi için kullanılan materyaller veri kaynaklarının belirlenmesi, veri setinin oluşturulması, modellerin oluşturulması ve verilerin sınıflandırılması işlemleriyle belirlenmiştir. Yararlanılan materyal bilgileri Şekil 3.1'de özet olarak verilmiştir.

Şekil 3.1. Yararlanılan materyal bilgileri

Şekil 3.1.'de görüldüğü gibi çalışma kapsamında yararlanılan materyal bilgileri alt başlıklar halinde açıklamalı olarak anlatılmıştır. Yararlanılan materyallere mevcut çalışmaların incelenmesini müteakip karar verilmiştir. Çalışmaya yönelik kullanılan materyal metodoloji içerisinde yapılan tüm adımları en iyi şekilde yapabilecek olanlar

seçilmiştir.

Veri kaynaklarının belirlenmesi

Çalışma kapsamında bahsedilen problemin çözümü için gerekli olan mevcut çalışmaların incelenmesi işlemi gerçekleştirilmiştir (Bkz. Çizelge 2.2). Farklı çalışmalar veri kaynaklarını belirlememiz için odağımızı oluşturmamıza yardımcı olmuştur. İnceleme neticesinde doğruluk kontrolü yapan web sitelerinin Twitter hesapları ve ana akım gazetelerin Twitter hesapları veri kaynağı olarak belirlenmiştir.

Veri setinin oluşturulması

Twitter'dan elde edilen veriler ile Türkçe veri seti oluşturulmuştur. Sahte ve gerçek haberlerden oluşan veri setimiz için Twitter'dan python programlama dilinin Tweepy kütüphanesi kullanılmış ve Google Colab Pro+ alt yapısı ile verilerimiz çekilmiştir. El ile etiketlenmiştir.

Modellerin oluşturulması

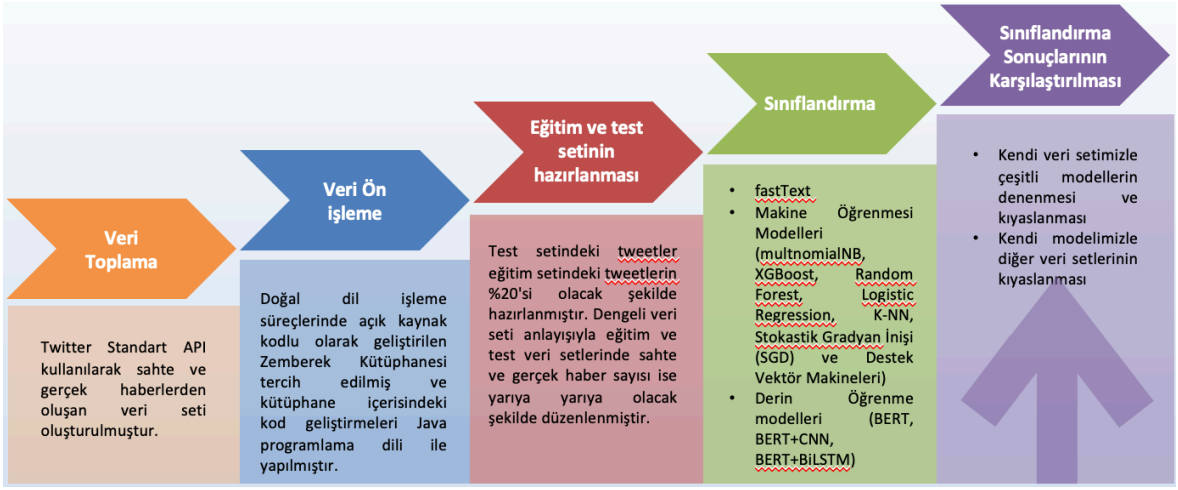
Mevcut çalışmalar incelendikten sonra model oluştururken Google Colab Pro+ altyapısı ile derin öğrenme çalışmalarında sıklıkça kullanılan Pytorch kütüphanesi [139] ve Zemberek kullanabilmek maksadıyla yazılan Java kodlarını çalıştırmak için Netbeans kullanılmıştır.

Sınıflandırma

Mevcut çalışmalar incelendikten sonra eğitim, test ve doğrulama veri setimiz hazırlanmış ve en uygun modeller uygulanarak en doğru sonuçları almak hedeflenmiştir.

3.2. Metot

Sahte haber tespiti için gerçekleştirilen tez çalışması kapsamında Çizelge 3.1'de gösterilen materyallerle yapılan değerlendirmeler sonrası Şekil 3.2'de gösterimi yapılan metodoloji ile çalışma sürecimiz yürütülmüştür.



Şekil 3.2. Çalışmanın metodolojisi

Şekil 3.2.'deki çalışma metodolojisi problem tanımının yapılmasıyla başlamıştır. Problem tanımı netleştikten sonra veri toplama, veri ön işleme eğitim ve test setinin hazırlanması, sınıflandırma ve sınıflandırma sonuçlarının karşılaştırması şeklinde devam etmiştir. Metodolojinin ayrıntılı sonraki sayfalarda anlatılmıştır.

4. VERİNİN HAZIRLANMASI VE ANALİZ EDİLMESİ

Tez metodolojisinin bu kısmında veri ile yaptığımız her türlü işlem ayrıntılı şekilde anlatılacaktır.

4.1. Veri Kaynakları

İnternet erişimi yaygınlaştıkça hem kullanıcı hem de üretilen içerik hacmi açısından sosyal medya kullanımı da önemli ölçüde artmıştır [140]. İnternete erişimde yaşanan bu hızlı artış, çevrimiçi forumların ve sosyal medya gruplarını patlamasına, dolayısıyla da Türk İnternet kullanıcılarının bilgi arama pratiklerinin dönüşmesine yol açmıştır [141] 2000'lerin sonuna doğru bu tür gruplar ve forumlar, şehir efsanelerinin ve yanlış bilgilerin başlıca kaynağı haline gelmiş, bu iddiaları doğrulayabilecek ve doğruluğunu kontrol edebilecek platformlara ihtiyaç duyulmasına neden olmuştur. Doğruluk kontrolü genellikle dezenformasyon ve sahte haber sorunlarına karşı başlıca panzehirlerden biridir. Uluslararası Doğrulama Ağı (IFCN), kuruluş amacı dünyanın her ülkesinde faaliyet gösteren doğrulama platformlarını bir araya getirmek, aynı zamanda bu platformların faaliyetlerine belli bir standardizasyon ve hesap verilebilirlik getirmektir. Dünya çapında IFCN üyesi olan 77 aktif doğrulama platformu bulunmaktadır. Türkiye'de ise sadece Doğruluk Payı ve Teyit.org IFCN üyesidir [8]. Çalışma kapsamında veri seti oluştururken kullanılan doğruluk kontrol platformlarının Twitter hesapları tercih edilmiştir. Türkiye'deki bu kapsamda kullanılan doğruluk platformlarını çalışmanın başında gösterimi yapılmıştır (Bkz Çizelge 1.1.)

Çizelge 1.1'de görüldüğü üzere sahte haber veri setimizi bu Twitter hesaplarından elde edilmiştir. Buradaki hesapların 2 tanesi IFCN üyesi geri kalan 3'ü ise IFCN kullarına riayet eden hesaplardır. Bu hesapların ortak özelliği Türkiye'de bulunmaları ve hesaplarında Türkçe tweetler paylaşan, çevrimiçi kanallar aracılığıyla yayılan sahte haberleri, sosyal medyada dolaşan doğrulanmamış haberleri, şehir efsanelerini paylaşan Twitter hesapları olmalarıdır. Bu sebeple 5 Twitter kullanıcısından toplamda 16250 adet tweet çektik. Bu hesaplar aynı anda doğru haber tweetlerinin teyidini de yaptığı için veri

ön işleme çalışmaları esnasında sadece sahte haber tweet paylaşımları aralarından manuel olarak etiketlenmiş ve çekilen tweet sayısında azalma meydana gelmiştir.

Çalışma kapsamında, Twitter platformundan farklı haber kaynakları kullanılarak tivitler çekilmiştir. Gerçek haberleri içeren veri seti, Twitter tarafından doğrulanmış haber hesaplarından çekilen tivitlerden oluşmaktadır. Yerel ve daha az bilinen haber kaynaklarının twitter hesapları göz ardı edilmiştir çünkü doğrulanması zor olacaktır. Bu sebeple 19 Twitter kullanıcısından toplamda 61750 adet gerçek haber tweeti çekilmiştir. Bu sayı veri ön işleme adımlarından sonra ana gereksinimlerde bahsedildiği gibi "Tweetler metinden oluşmalı ve homojen olmalı" maddesi gereğince biraz daha azalmıştır.

Veri derleminin oluşması için haber tweetlerinin toplandığı haber kaynakları, veriyi doğrulama ve gerçekleme hususu bu bölümde anlatılmıştır. Çalışmada, [142]'nin yayınlarında da kullandığı gibi, haberlerden oluşan derleme için gereken ana başlıkları literatürde yapılan araştırmalar neticesinde kılavuz olarak yararlanılan çalışmadaki gibi yapılmıştır [143] ve bu ana gereksinimler ve sosyal medya platformu olarak çalışmada Twitter platformunu tercih etme sebepleri aşağıdaki gibidir:

- Sahte ve gerçek haber maddelerini içeren tweetler olması,
- Tweetlerin sadece metinden oluşması,
- API yardımıyla biz gibi araştırmacıların Twitter'a erişiminin kolay olması,
- Doğrulanabilir bir gerçekliğe sahip olması,
- Uzunluğunun homojen olması,
- Sahte ve gerçek haberlerden oluşan tweetlerin aynı şekil ve amaçla sunulması,
- İstendiğinde zaman dilimine erişim sağlanması,
- Halka açık kaynak olarak sunulması,
- Dil ve kültür farkını gözetmesidir.

4.2. Veri Toplama ve Etiketleme

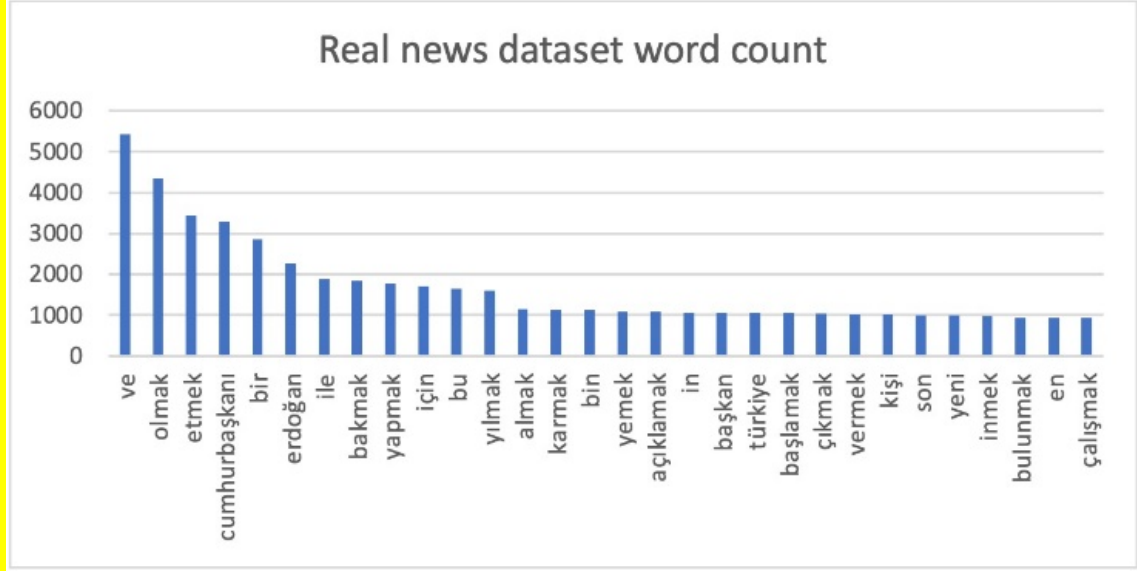
Veri seti oluşturma aşamasına başlamadan önce literatürde ihtiyaç duyduğumuz Türkçe tweetlerden oluşan veri setinin varlığı kontrol edilmiş ancak rastlanılmamıştır. Burada yanlış anlaşılmalara yer vermemek adına arayışta olduğumuz veri Twitter verisi ve hem sahte ham de gerçek haberlerden oluşmasıdır.

Söz konusu tweetleri Twitter Standart API kullanarak Twitter'dan sahte ve gerçek haberleri içeren Türkçe tweetleri Python'un tweepy modülünü kullanılarak çekilmiştir. Normal şartlarda kullanıcı bazlı olarak 200 adet tweet çekilebilen Twitter'dan Standart API kullanıcılarına son 7 gün içerisinde yayınlanan tweetleri indirmeye izin vermektedir. Ancak veriseti iyileştirme çalışmaları kapsamında Twitter üzerinden çekilen tweetleri ".json" uzantılı olacak şekilde çekildiğinde her kullanıcıdan bir anda çekilebilen tweet sayısı 3250 adet olmuştur. Böylelikle son 7 gün filtresiyle tivitler daha hızlı elde edilmiştir.

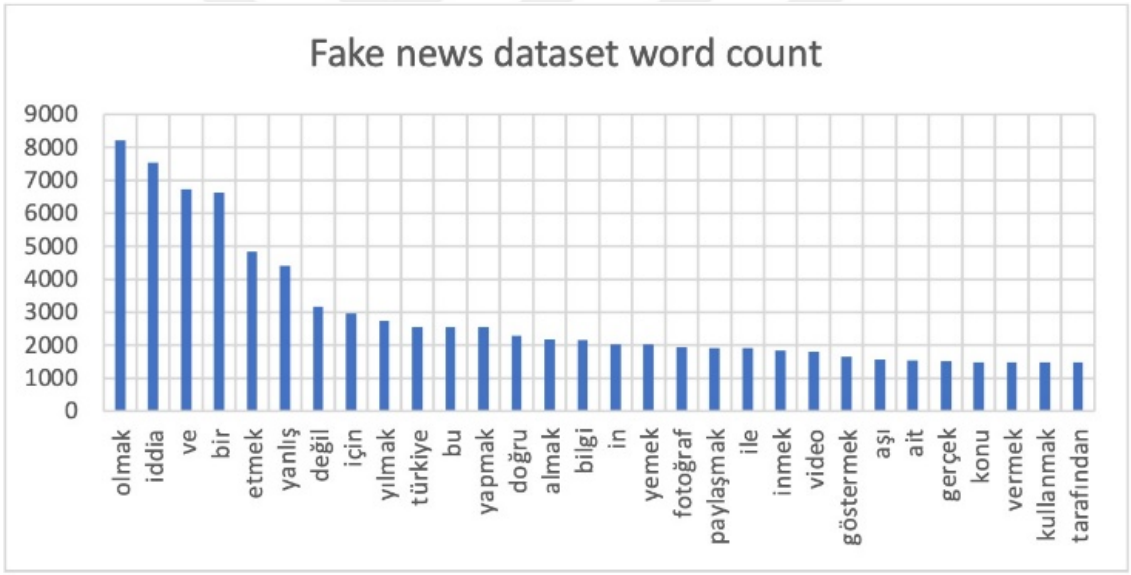
Bu çalışma için Türkçe sahte ve gerçek haberlerden oluşan TR_FaRe_News (Turkish Fake and Real News) adı verdiğimiz bir veri kümesi oluşturulmuştur. Twitter için geliştirilen Tweepy modülü kullanılarak atılan tivitler, ana akım gazetelerden alınan tivitler gerçek, doğruluk kontrol platformları tarafından paylaşılan sahte haberler ise sahte haber olarak etiketlenmiştir. Etiketleme işleminin ardından veri içerisinde benzer tivitler bulunmuş ve ayıklanmıştır. Ayıklama işleminden önce veri kümesindeki haberler için kosinüs benzerliği hesaplanmıştır. Çok benzer olan tivitler elenmiştir. Burada gerçek haber veri kümesindeki haber tiviti A, sahte haber veri kümesindeki haber tiviti B olarak kabul edilirse matematiksel işlem formülde görüldüğü gibi hesaplanmıştır.

$$\cos(\theta) = \frac{\mathbf{AxB}}{\|\mathbf{A}\|\boldsymbol{x}\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}}$$

Benzer tivitler ayrıştırıldıktan sonra, sahte ve gerçek haber veri kümelerinde sınıflandırmayı olumsuz etkileyecek yüksek kelime sayısına sahip kelimeleri içeren tivitler de düzenlenmiştir. Sahte ve gerçek veri kümeleri için hesaplanan kelime sayıları Şekil 4.1 ve Şekil 4.2'de gösterilmiştir.



Şekil 4.1. Gerçek haber veri kümesi kelime sayısı



Şekil 4.2. Sahte haber veri kümesi kelime sayısı

Şekil 4.2 ve Şekil 4.3'te görüldüğü gibi, her iki veri kümesinde de ortak kelimeler içeren tivitler tekrar incelenmiştir. Bu kelimeler 've', 'olmak', 'etmek', 'bir', 'cumhurbaşkanı' gibi en sık kullanılan kelimelerdir. Ayrıca, sahte haber veri setinde 'yanlış', 'doğru' ve 'değil' kelimelerini içeren tivitlerin de analizi yapılmıştır. Sonuç olarak, son mükerrer kayıtlar da temizlendikten sonra elimizde 18695 tivit cümlesi kalmıştır.

4.3. Veri ön işleme

Veri ön işleme, veri setini bilgisayarın anlayabileceği sayılara dönüştürmeden önce düzenlenmesi anlamına gelmektedir. Türkçe dili sondan eklemeli bir dil olduğundan, doğal dil işleme açısından zorluklar yaşatmaktadır [143]. Bu zorlukların üstesinden gelebilmek maksadıyla, metinleri çözümlemek ve kelime köklerine erişmek için Zemberek kullanılmıştır [144]. Java programlama dili kullanılarak Türkçe dili için geliştirilen ve açık kaynak kodlu doğal dil işleme kütüphanesi olan Zemberek'in StemmingAndLemmatization paketi kullanılmıştır. Burada yapılmak istenen işlem:

- Stemming işleminini bir kelimede yer alan son ekleri ve ön ekleri keserek kelime kökünü bulmaya çalışması,
- Lemmatization ile kelimelerin morfolojik analizlerinin yapılması esas alınmıştır. Böylelikle algoritmanın kelime kökünü elde edebilmesi için detaylı bir sözlüğe ihtiyacı bulunmaktadır.

Metin sınıflandırmasının etkili bir şekilde gerçekleştirilmesi ve sınıflandırma modelinin oluşturulması için veri setinin ön işlemeden geçmesi gerekmektedir. Literatür incelendiğinde ön işleme yapılan veri setlerinde verilerin sınıflandırması yapılırken elde edilen sonuçlar daha iyidir [143]. Ön işleme adımlarının uygulanmasındaki amaç daha küçük vektör uzayına ve daha az boyutla daha iyi performans elde etmektir. Çalışma için hazırladığımız veri seti için yapılan doğal dil işleme süreçlerinde açık kaynak kodlu Zemberek Kütüphanesi [144] tercih edilmiş ve kütüphane içerisindeki kod geliştirmeleri Java programlama dili ile yapılmıştır.

Twitter'dan alınan haber tivitleri üzerinde yapılan işlemlere göre toplam haber tiviti sayısı hesaplanmıştır. TR_FaRe_News veri kümesi için haber tivit istatistikleri Çizelge 4.1'de listelenmiştir.

Çizelge 4.1. TR_FaRe_News very kümesi haber tivit istatistikleri

İşlem	Ana akım haber tivitleri (AACanli, anadoluajansi, dhainternet, Hurriyet ve ihacomtr)	Doğruluk konrrol platrofmu haber tivitleri (dogrulaorg, dogrulukpayicom, gununyalanlari, malumatfurus, teyitorg)	Toplam
İşlem görmemiş tivit sayısı	27,298	38,170	65,468
Tekrar eden kayıtların silinmesi	27,286	35,209	62,495
Karakter temizleme	20,927	23,819	44,746
Zemberek DDİ	26,302	23,507	49,809
Son işlem sonrası tekrar eden kayıtların silinmesi	25,624	21,194	46,818

Hazırlık işlemlerinin ardından toplam 46818 tivit cümlesine sahip olan veri kümemiz, bir önceki bölümde açıklanan benzerlik teoremlerinin uygulanması ve mükerrer kayıtların tekrar silinmesinin ardından 18695 tivit cümlesinin veri kümemeize dahil etmiştir. Ön işleme için gerçekleştirdiğimiz işlemler aşağıda sıralanmıştır.

- Tekrar eden tweetler silinmiştir.
- Özel karakterler, emojiler ve noktalama işaretleri sözcüklerle beraber yazıldığında doğal dil işleme süreçlerini de etkilediği için silinmiştir.
- Kullanıcıların girdiği kelime hataları veya eksik harfler, veri külliyatına uygulanarak normalizasyon işlemi yapılmıştır.
- Bahsetmeler ve tweet etiketleri silinmiştir.
- Cümleye herhangi bir anlam katmayan durak sözcükleri silinmiştir.
- Kelimeler küçük harfe dönüştürülmüştür.

Yapılan veri ön işleme çalışması sonucu veri külliyatında yer alan tweet cümlelerine ait çıktılar Çizelge 4.2'de gösterilmiştir.

Çizelge 4.2. Orjinal tivitin önişleme örneği

İşlem	Çıktı
Orijinal tivit	#@milliyet Rusya'da savaşa gitmemek için denek olark tecavüzcüler kullanıldı \n\n https://url "
Karakter temizleme	#@milliyet Rusya'da savaşa gitmemek için denek olark tecavüzcüler kullanıldı n n https://url
Hatalı kelimeleri düzeltme	#@milliyet Rusya da savaşa gitmemek için denek olarak tecavüzcüler kullanıldı n n https://url
Bahsetme ve tivit etiketlerinin silinmesi	Rusya da savaşa gitmemek için denek olarak tecavüzcüler kullanıldı n n https://url
URL'lerin temizlenmesi	Rusya da savaşa gitmemek için denek olarak tecavüzcüler kullanıldı n n
Anlamsal kök çıkarma ve kelimeleri küçük harfe dönüştürme	rusya da savaşa gitmemek için denek olarak tecavüzcüler kullanıldı
Durak kelimelerin silinmesi	rusya savaş gitmek denemek olmak tecavüz kullanmak
Tivitin son hali	rusya ilaç gitmek denemek olmak tecavüz kullanmak

4.4. Eğitim, Test ve Doğrulama veri setinin hazırlanması

Sahte ve gerçek haberlerin sınıflandırma işlemi metin sınıflandırmanın alt başlığıdır. Dil modelleri geliştirmek de tespiti kolaylaştıran ve eğitim ve test verilerini oluşturmanın önemli bir adımıdır. Eğitim ve test verisi oluştururken dikkat edilmesi gereken en önemli husus veri kümelerinin dengeli olmasıdır. Hazırladığımız veri seti göz önünde bulundurulduğunda farklı etikete sahip tweet cümleleri tüm eğitim ve test setlerinde

orantılı olarak bulunmalıdır. Test setindeki tweetler eğitim setindeki tweetlerin %20'si olacak şekilde hazırlanmıştır. Dengeli veri seti anlayışıyla eğitim ve test veri setlerinde sahte ve gerçek haber sayısı ise yarıya yarıya olacak şekilde düzenlenmiştir. Aksi takdirde farklı sonuçlar elde etmek mümkündür. Literatür incelemesi neticesinde tabakalı tesadüfi örnekleme modelinin tez çalışması için uygun olacağı değerlendirilmiştir. Bu örnekleme modelinin avantajı, diğer örnekleme modellerine göre sonuçlarının doğruluk oranının yüksek olması, öğretmesi kolay ve araştırmacılar tarafından kolay anlaşılır ve en küçük boyutlu veri setleri için iyi sonuçlar veriyor olmasıdır [145]. Scikit-learn kütüphanesi python programlama dili üzerinde, veri setini 10 kat çapraz doğrulama için tabakalı rastgele örnekleme ile bölmek için kullanılmıştır [146].

Türkçe sahte haber sınıflandırması için manuel olarak oluşturulan TR_FaRe_News veri kümesi kullanılmıştır. Ayrıca, İngilizce sahte haberleri tespit etmek için kullanılan BuzzFeedNews, GossipCop, ISOT, LIAR, Twitter15, Twitter16 ve GPT-2 sahte haber veri kümeleri kullanılmıştır. Deneyler için veri kümesi üç bölüme ayrılmıştır: eğitim, test ve doğrulama. Veri kümelerine ait istatistikler bölüm, sınıf ve toplam sayısı olacak şekilde Çizelge 4.3'te sunulmuştur.

Çizelge 4.3. Veri kümesi istatistikleri

Veriseti	Bölüm	Gerçek	Sahte	Toplam
BuzzFeedNews	train	73	73	146
	val	8	8	16
	test	19	19	38
GossipCop	train	4,326	4,271	8,597
	val	480	474	954
	test	1,081	1,068	2,149
ISOT	train	16,336	14,323	30,659
	val	1,815	1,591	3,406
	test	4,084	3,580	7,664
LIAR	train	1,995	1,676	3,671
	val	221	186	407
	test	498	419	917

Veriseti	Bölüm	Gerçek	Sahte	Toplam
Twitter15	train	244	268	512
	val	27	29	56
	test	61	67	128
Twitter16	train	138	139	277
	val	15	15	30
	test	34	34	68
GPT-2 Fake news dataset	train	5,000	5,000	10,000
	val	2,500	2,500	5,000
	test	1,250	1,250	2,500
TR_FaReNews	train	6,869	6,867	13,736
	val	763	763	1,526
	test	1,717	1,716	3,433

Çizelge 4.3'te görüldüğü gibi veri setinin dengeli olması önemsenmiştir. Eğitim, test ve doğrulama veri setlerindeki ikili sınıflandırma kapsamında sahte ve gerçek haberlerin sayısının birbirine eşit olması sağlanmıştır.

5. MODEL OLUŞTURULMASI

Bu bölümde tez çalışmasında kullanılan 7 makine öğrenmesi algoritması, fastText ve derin öğrenme modellerinin kullanımını nasıl yapıldığı ve ana modelin nasıl oluşturulduğu ve model sonuçları yer alacaktır.

5.1. Makine Öğrenmesi

Makine öğrenmesi [147], tartışmasız bugün mevcut olan en önemli ve etkili teknolojilerden biridir. Yapay zeka, ML adı verilen bir alt kümeye sahiptir. Başka bir deyişle, 1959'da Amerikan bilgisayar oyunları ve yapay zeka öncüsü Arthur Samuel tarafından geliştirilen "ML" kavramı, "bilgisayarların açıkça talimat verilmeden öğrenmesine izin vermek" anlamına gelmektedir.

Makine öğrenmesi, geleneksel yaklaşımlardan daha farklı bir bilgisayar bilimidir. Algoritmalar, bilgisayarların sorunları analiz etmesine çözmesine yardımcı olmak için tasarlanmış bir dizi talimattır. Öte yandan makine öğrenmesi teknikleri, bilgisayarların, tanımlanmış bir aralıktaki üretim değerlerinin yanı sıra tanımlayıcı istatistikler yoluyla girdi verilerinden öğrenmesine olanak tanımaktadır. Bu nedenle makine öğrenmesi, bilgisayarların örnek verilerden model verileri oluşturmasının kolaylaştırmakta ve ayrıca gerçek dünya girdilerine dayalı karar verme süreçlerini hızlandırmaktadır [147].

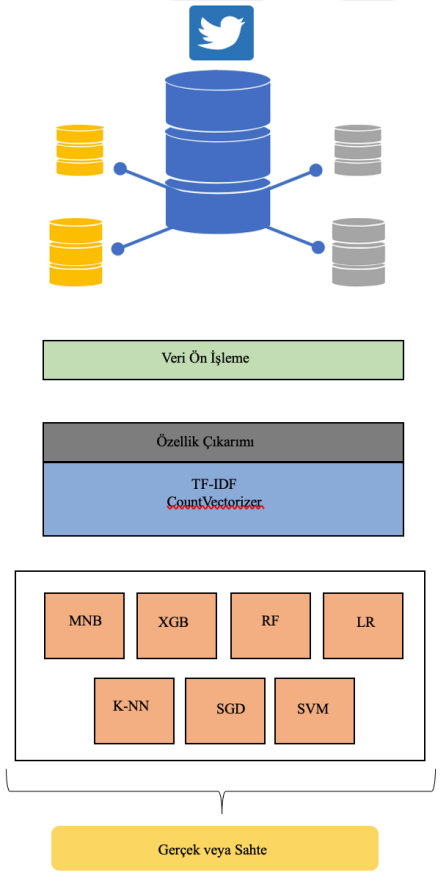
Makine öğrenmesi, yazılım programları için, sonuçları açıkça programlamadan gelen veri trendlerini ve ilişkilerini daha başarılı bir şekilde tahmin etmek ve onları buna göre değiştirmek için birkaç yöntem içermektedir [148].

Aşağıdakiler de dahil olmak üzere çok çeşitli alanlarda makine öğrenmesi için birçok uygulamak mevcuttur:

- Tahmin: sınıflandırma, tahmin sistemlerinde, hava tahminlerinde ve yanlış-sonuç olasılık analizlerinde makine öğrenmesi tekniklerini kullanmak mümkündür,

- Bilgisayar görüşü: görüntü işleme yapmak için makine öğrenmesi teknikleri kullanılabilmektedir.
- Ses tanıma: metni sese çevirmek, sesli talimatlar ve ses tanıma yoluyla otonom araçları ve robotları çalıştırmak için makine öğrenmesi teknikleri kullanılabilmektedir.
- Sağlık teşhisleri: Makine öğrenmesi algoritmaları kullanılarak tıbbi durumlar tahmin edilebilmektedir.

Makine öğrenmesinde, analizler genellikle referans kolaylığı için geniş kategoriler halinde gruplandırılmaktadır. Nesneleri kategorilere ayırmak için makinelerin geri bildirimi nasıl öğrendiğine ve kullandığına bakılmaktadır. Algoritmaları eğitmek için insan etiketli veri ve bilgileri kullanan denetimli öğrenme ve bir tekniğin net örneklerini sunmayan ve girdi verilerinde anlamın keşfedilmesini talep eden denetimsiz öğrenme, makine öğrenmesinde en sık kullanılan yöntemlerdir [148].



Şekil 5.1. Makine öğrenmesi sınıflandırma modeli mimarisi

Çalışmada sınıflandırma yapmak amacıyla 7 adet sınıflandırma algoritmasından faydalanılmıştır. Bu algoritmalar sırasıyla multnomialNB, XGBoost, Random Forest, Logistic Regression, K-NN, Stokastik Gradyan İnişi (SGD) ve Destek Vektör Makineleridir. Sınıflandırma modelinin mimarisi Şekil 5.1.'de gösterilmiştir.

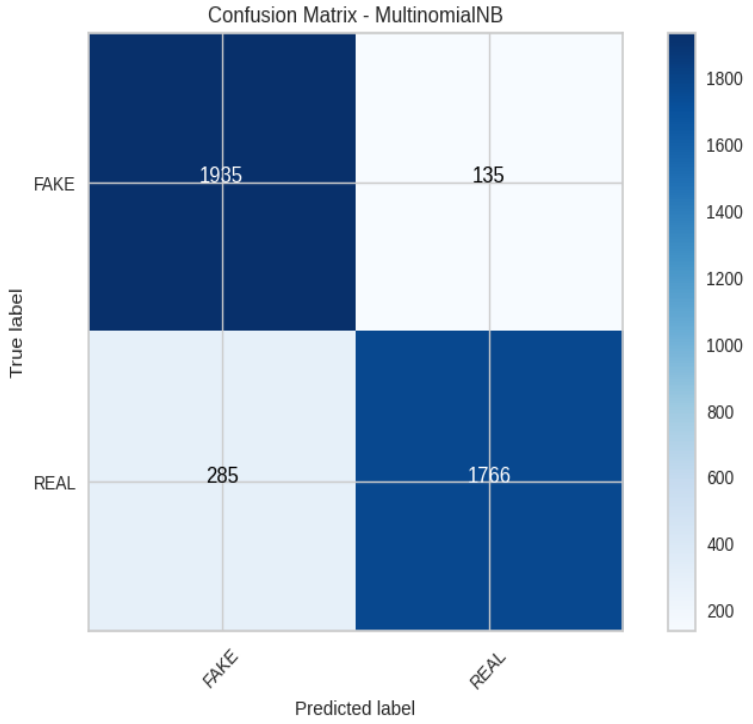
Şekil 5.1.'deki makine öğrenmesi model mimarisine uygun olarak kullanılan algoritmaların açıklamalı olarak sonuçları elde edilmiştir.

5.1.1. Multinomial Naive Bayes (MNB)

Makine öğrenmesinde Bayes sınıflandırıcıları, basit makine öğreniminin bir parçasıdır. MNB, ardışık düzen kavramlarının kullanarak sahte veya gerçek haberlerin doğruluğunu bulmak için kullanılan popüler algoritmalardan biridir [149].

Çalışmamızda scikit-learn kütüphanesinin çok terimli Naive Bayes algoritması kullanılmıştır. Düzeltme (yumuşatma) parametresi alfa, varsayılan değer olarak 1'e ayarlanmıştır. Ayrıca fit-prior parametresi doğrudur, yani model önceki sınıf olasılıklarına göre öğrenmektedir. MNB, özellikle metinsel veriler için NB'nin basitleştirilmiş bir çeşididir. TF-IDF vektörünün çıktısı, bir kelimenin bir sınıfa verilen frekanslarına dayalı olarak koşullu olasılığını tahmin eden çok terimli sınıflandırıcıya (MNB) beslenmektedir. Yüksek boyutlu bir veri kümesi için iyi çalışmaktadır ve çok az ayarlanabilir parametresi bulunduğundan son derece hızlı çalışmaktadır [150].

Çalışmada kullanılan MNB algoritmasının sınıflandırma aşamasında oluşan karmaşıklık matrisi Şekil 5.2.'de sunulmuştur.



Şekil 5.2. MNB Algoritması karmaşıklık matrisi

Şekil 5.2.'deki MNB algoritması sonucu ortaya çıkan karmaşıklık matrisi ışığında elde edilen sonuçların yer aldığı doğruluk tablosu Çizelge 5.1.'de sunulmuştur.

Çizelge 5.1. MNB Algoritması doğruluk tablosu

	Kesinlik	Duyarlılık	F1-skor
Sahte	0.87	0.93	0.90
Gerçek	0.93	0.86	0.89
Doğruluk			0.90
Macro avg	0.90	0.90	0.90
Weighted avg	0.90	0.90	0.90

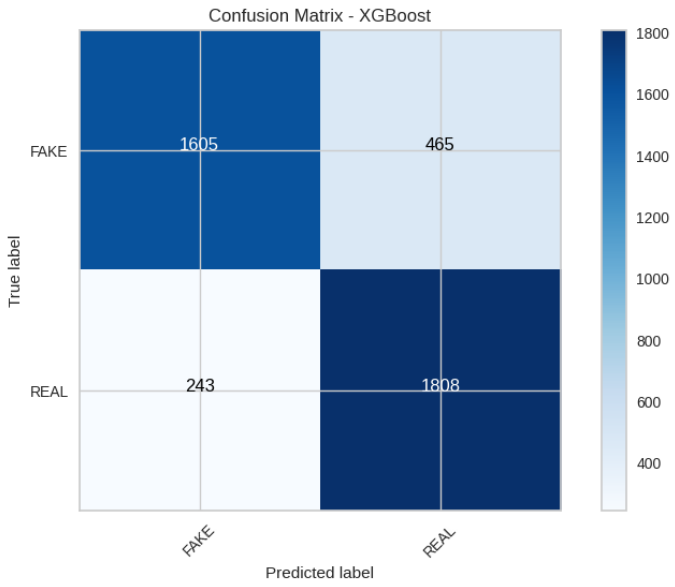
Çizelge 5.1.'de yer alan değerlendirme metriklerine göre MNB algoritması doğruluk değeri %90 elde edilmiştir.

5.1.2. XGBoost

XGBoost, makine öğrenmesi uygulayıcıları tarafından oldukça itibar gören ve rakip sınıflandırıcılar arasında popüler olan bir algoritmadır. XGBoost, hız ve performans için tasarlanmış gradyan destekli karar ağaçlarının bir uygulamasıdır. Bu algoritma, tüm

popüler veri analizi dilleri için bulunabilmekte ve ikili veya çok sınıflı sınıflandırma problemlerinde oldukça iyi performans göstermektedir [151].

XGBoost'un başarısının arkasındaki en önemli faktör, tüm senaryolarda ölçeklenebilir olmasıdır. Sistem, tek bir makinede mevcut popüler çözümlerden on kat daha hızlı çalışır ve dağıtılmış veya sınırlı bellekli ayarlarda milyarlarca örneğe ölçeklenir. XGBoost'un ölçeklenebilirliği, birkaç önemli sistem ve algoritmik optimizasyondan kaynaklanmaktadır. Bu yenilikler şunları içerir: seyrek verileri işlemek için yeni bir ağaç öğrenme algoritması; teorik olarak doğrulanmış bir ağırlıklı kuantil çizim prosedürü, yaklaşık ağaç öğrenmede örnek ağırlıklarının işlenmesini sağlar. Paralel ve dağıtılmış bilgi işlem, öğrenmeyi hızlandırır ve bu da daha hızlı model keşfi sağlar. Daha da önemlisi, XGBoost çekirdek dışı hesaplamadan yararlanır ve veri bilimcilerin bir masaüstünde yüz milyonlarca örneği işlemesine olanak tanır [152].
Son olarak, en az miktarda küme kaynağıyla daha da büyük verilere ölçeklenen uçtan uca bir sistem oluşturmak için bu teknikleri birleştirmek muazzamdır [152]. Çalışmada kullanılan XGBoost algoritmasının sınıflandırma aşamasında oluşan karmaşıklık matrisi Şekil 5.3.'te sunulmuştur.



Şekil 5.3. XGBoost algoritması karmaşıklık matrisi

Şekil 5.3.'teki XGBoost algoritması sonucu ortaya çıkan karmaşıklık matrisi ışığında elde edilen sonuçların yer aldığı doğruluk tablosu Çizelge 5.2.'de sunulmuştur.

Çizelge 5.2. XGBoost algoritması doğruluk tablosu

	Kesinlik	Duyarlılık	F1-skor
Sahte	0.87	0.78	0.82
Gerçek	0.80	0.88	0.84
Doğruluk			0.83
Macro avg	0.83	0.83	0.83
Weighted avg	0.83	0.83	0.83

Çizelge 5.2.'de yer alan değerlendirme metriklerine göre XGBoost algoritması doğruluk değeri %83 elde edilmiştir.

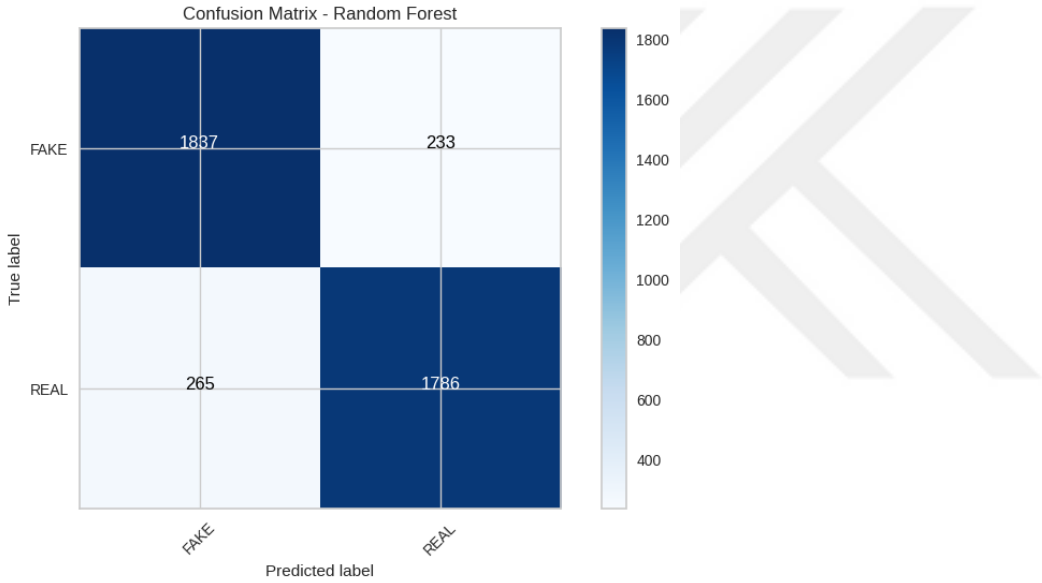
5.1.3. Random Forest

Genellikle Rastgele Karar Ormanları olarak anılan Rastgele Ormanlar, sınıflandırma ve regresyon sorunlarını çözmek için kullanılabilmektedir. Denetimsiz yaklaşımlarda da kullanılabilir. Birkaç ağacın tahminleri, rastgele orman algoritması tarafından oluşturulmaktadır. Bir özellik alt kümesi kullanılarak her bir karar ağacı oluşturulur. Her karar ağacı bir sınıf üretmekte ve sonunda Rastgele Orman tekniğinden daha iyi bir doğruluk elde etmek için kararlar önyüklenmektedir. Karar ağacındaki eylen planını açıklamak için ağaç biçimli bir model kullanılmaktadır. Herhangi bir düğümde bir karar verilecektir [153].

Sckit-learn'ün önceden tanımlanmış bir RF sınıflandırıcı kütüphanesi kullanılarak sahte haber tespiti için bir RF sınıflandırıcı uygulanmıştır. Farklı veri örnekleri üzerinde birden çok karar ağacına uyan bir sınıflandırıcıdır ve doğruluğu artırmak ve aşırı uydurmayı kontrol etmek için ortalamayı kullanmaktadır. Ormanda yaklaşım 200 karar ağacı tahmin edici olarak kullanılmıştır ve değerleri grid aramasına göre belirlenmiş ve paralel çalışacak iş sayısı 3 olarak ayarlanmıştır. 'Gini' kriteri varsayılan olarak kullanılmıştır, diğer 'Bilgi kazanımı' ise biraz daha yavaş hale getiren logaritmik bir fonksiyonun hesaplanmasını içermektedir. Bölme kalitesini ölçmek için 'Gini' kriteri kullanılmaktadır. Her sınıfın olasılık karelerinin toplamının 1'den çıkarılmasıyla ölçülmekte ve daha büyük bölümler için uygun hale getirilmektedir [153]. Ayrıca aşağıdaki adımlar izlenmiştir:

- Veriler, veri kazancı maksimum olacak şekilde bölünmüştür. (Kazanç, bölme işlemi sırasında entropi azalmasının hesaplanmasıdır.)
- Bölme için maksimum kazancı sunan bir düğümdeki durum seçilmiştir.
- Bölme, entropi sıfırı geçene kadar yapılmıştır.
- Birkaç karar ağacı oluşturmak için, yukarıdaki süreçler tekrarlanır ve sonunda, önemli karara bağlı olarak bir örneklem için bir sınıf (sahte veya gerçek) seçilmiştir.

Çalışmada kullanılan RF algoritmasının sınıflandırma aşamasında oluşan karmaşıklık matrisi Şekil 5.4.'te sunulmuştur.



Şekil 5.4. RF karmaşıklık matrisi

Şekil 5.4'te RF algoritması sonucu ortaya çıkan karmaşıklık matrisi ışığında elde edilen sonuçların yer aldığı doğruluk tablosu Çizelge 5.3.'te sunulmuştur.

Çizelge 5.3. RF Algoritması doğruluk tablosu

	Kesinlik	Duyarlılık	F1-skor
Sahte	0.87	0.89	0.88
Gerçek	0.88	0.87	0.88
Doğruluk			0.88

Macro avg	0.88	0.88	0.88
Weighted avg	0.88	0.88	0.88

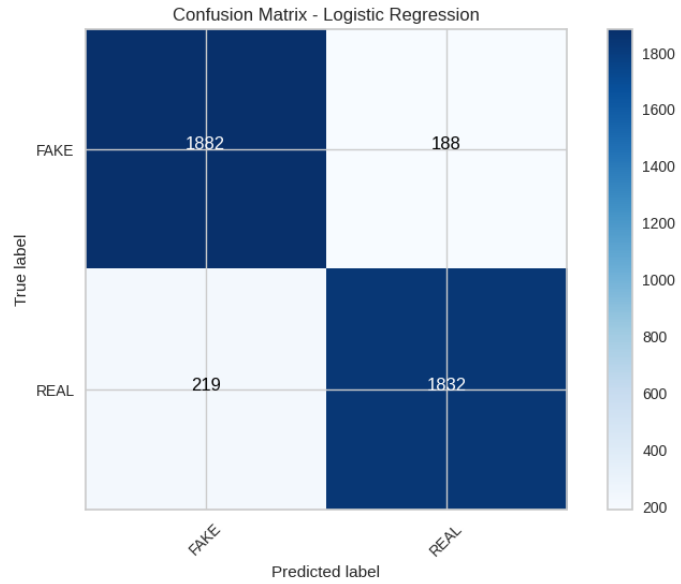
Çizelge 5.3.'te yer alan değerlendirme metriklerine göre RF algoritması doğruluk değeri %88 elde edilmiştir.

5.1.4. LR

Lojistik regresyon, denetimli öğrenme tekniği kapsamında yer alan popüler makine öğrenmesi algoritmalarından biridir. Belirli bir bağımsız değişken kümesini kullanarak kategorik bağımlı değişkeni tahmin etmek için kullanılmaktadır. Lojistik regresyon, kategorik bir bağımlı değişkenin çıktısını tahmin etmektedir. Lojistik regresyon, sürekli ve ayrık veri kümeleri kullanarak olasılıklar sağlama ve yeni verileri sınıflandırma yeteneğine sahip olduğu çin önemli bir makine öğrenmesi algoritmasıdır. Farklı veri türlerini kullanarak gözlemleri sınıflandırmak için kullanılabilmekte ve sınıflandırma için kullanılan en etkili değişkenleri kolayca belirleyebilmektedir. Regresyon olarak tahmini modelleme kavranımı kullanmaktadır; bu nedenle lojistik regresyon olarak adlandırılmıştır ancak verileri sınıflandırmak için kullanılmaktadır. Bu nedenle sınıflandırma algoritmaları kapsamına girmektedir [154].

Çalışmada, sahte haberleri tespit tahmin etmek için sk-learn kütüphanesinden lojistik regresyon kullanılmıştır. Çıkarılan özniteliklerin her biri sınıflandırıcıda kullanılmıştır. Model oluşturulduktan sonra değerlendirme metrikleri karşılaştırılmış ve karmaşıklık metrisi oluşturulmuştur.

Çalışmada kullanılan LR algoritmasının sınıflandırma aşamasında oluşan karmaşıklık matrisi Şekil 5.5.'te sunulmuştur.



Şekil 5.5 LR karmaşıklık matrisi

Şekil 5.5.'te verilen LR algoritması sonucu ortaya çıkan karmaşıklık matrisi ışığında elde edilen sonuçların yer aldığı doğruluk tablosu Çizelge 5.4.'te sunulmuştur.

Çizelge 5.4. LR Algoritması doğruluk tablosu

	Kesinlik	Duyarlılık	F1-skor
Sahte	0.90	0.91	0.90
Gerçek	0.91	0.89	0.90
Doğruluk			0.90
Macro avg	0.90	0.90	0.90
Weighted avg	0.90	0.90	0.90

Çizelge 5.4.'te yer alan değerlendirme metriklerine göre LR algoritması doğruluk değeri %90 elde edilmiştir.

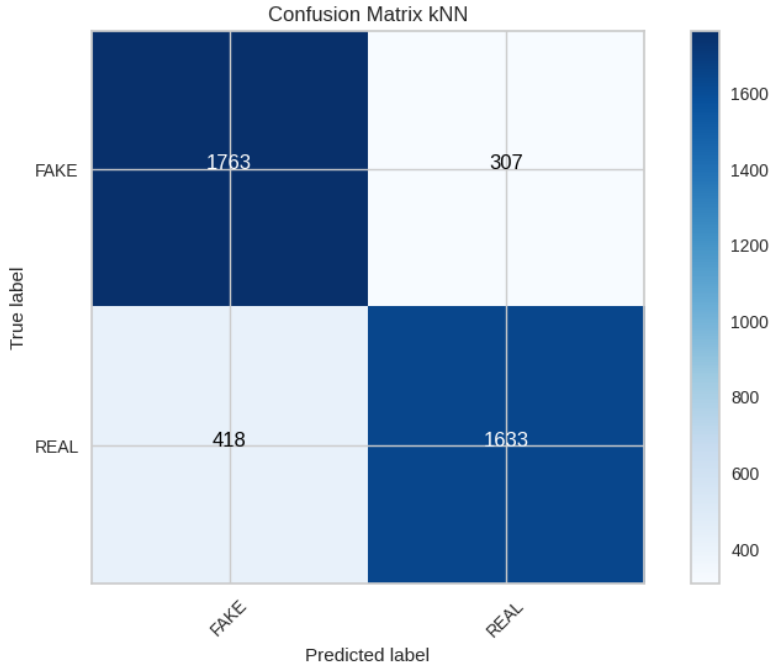
5.1.5. K-NN

K-NN, K En Yakın Komşu, makine öğrenmesindeki temel algoritmalardan biridir. Makine öğrenmesi modelleri, çıktı değerlerini tahmin etmek için bir dizi girdi değeri kullanır. K-NN, çoğunlukla sınıflandırma için kullanılan makine öğrenmesi algoritmalarının en basit

biçimlerinden biridir. Veri noktasını, komşusunun nasıl sınıflandırıldığına göre sınıflandırmaktadır. K-NN, daha önce saklanan veri noktalarının benzerlik ölçüsüne dayalı olarak yeni veri noktalarını sınıflandırır. Örneğin, bir domates ve muz veri kümemiz varsa K-NN, şekil ve renk gibi benzer ölçüleri saklayacaktır. Yeni bir nesne geldiğinde rengi sarı veya kırmızı ise, benzerliğini kontrol edecektir [155].

K En Yakın Komşu, sınıflandırma ve regresyon problemleri için kullanılan denetimli bir makine öğrenmesi algoritmasıdır. Verilen sınıflandırılmamış girdi için etiketi belirleyecek ve verecek bir fonksiyon geliştirilebilmesi için bir grup etiketli girdi verisini beslemektedir. "En yAkın Komşular" prensibi ile çalışmaktadır. K, bir grubun sahip olabileceği en yakın komşunun değeridir. Örneğin, her örnek uzayında koordinat (x,y) değerleri ile bir grafik üzerinde çizilen bir veri noktaları grubunu ele aldığımızda; yeni bir girdi alındığında, algoritma en yakın komşuya göre etkiketini belirlemiştir. "K" değeri sınıflandırmada büyük rol oynadığından, bunun için en uygun değere çeşitli deneme yanılma yöntemleri uygulanarak karar verilmiştir [155].

Çalışmada kullanılan K-NN algoritmasının sınıflandırma aşamasında oluşan karmaşıklık matrisi Şekil 5.6.'da sunulmuştur.



Şekil 5.6. KNN karmaşıklık matrisi

Şekil 5.6'da K-NN algoritması sonucu ortaya çıkan karmaşıklık matrisi ışığında elde edilen sonuçların yer aldığı doğruluk tablosu Çizelge 5.5.'te sunulmuştur.

Çizelge 5.5. K-NN Algoritması doğruluk tablosu

	Kesinlik	Duyarlılık	F1-skor
Sahte	0.81	0.85	0.83
Gerçek	0.84	0.80	0.82
Doğruluk			0.82
Macro avg	0.83	0.82	0.82
Weighted avg	0.82	0.82	0.82

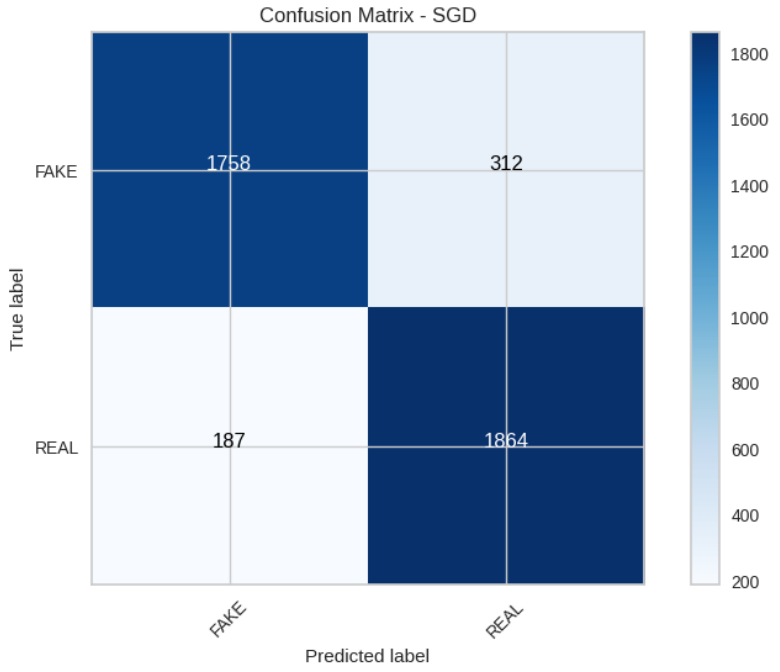
Çizelge 5.5.'te yer alan değerlendirme metriklerine göre K-NN algoritması doğruluk değeri %82 elde edilmiştir.

5.1.6. SGD

Gradyan iniş, optimizasyonu gerçekleştirmek için kullanılan bir algoritmadır ve sinir ağlarını optimize etmenin açık ara en yaygın yoludur [156]. Toplu gradyan iniş, her parametre güncellemesinden önce ilgili örnekler için gradyanları yeniden hesapladığından büyük veri kümeleri için gereksiz hesaplamalar gerçekleştirmektedir. SGD, her seferinde bir güncelleme gerçekleştirerek bu fazlalığı ortadan kaldırmaktadır. Bu nedenle, SGD çok daha hızlıdır ve çevrimiçi öğrenme için de kullanılabilmektedir. SGD, amaç fonksiyonunun yoğun bir şekilde salınmasına neden olan yüksek bir sapma ile sık güncellemeler gerçekleştirmektedir [156].

SGD, makine öğrenmesi topluluğunda uzun zamandır var olmasına rağmen, yakın zamanda büyük ölçekli öğrenim bağlamında önemli miktarda ilgi görmüştür [159]. SGD, metin sınıflandırmasında ve doğal dil işlemede sıklıkla karşılaşılan büyük ölçekli ve seyrek makine öğrenmesi problemlerine başarıyla uygulanmıştır. SGD sınıflandırıcı, temel olarak çeşitli kayıp fonksiyonlarını destekleyen ve sınıflandırmaya yardım olan basit bir SGD rutini uygulamaktadır Scikit-learn, SGD sınıflandırmasını uygulamak maksadıyla SGDClassifier modülü sağlamaktadır [157]. Çalışmamızda bu modülden faydalanarak sınıflandırma başarıları elde ettik.

Çalışmada kullanılan SGD algoritmasının sınıflandırma aşamasında oluşan karmaşıklık matrisi Şekil 5.7'de sunulmuştur



Şekil 5.7. SGD karmaşıklık matrisi

Şekil 5.7.'de gösterilen SGD algoritması sonucu ortaya çıkan karmaşıklık matrisi ışığında elde edilen sonuçların yer aldığı doğruluk tablosu Çizelge 5.6.'da sunulmuştur.

Çizelge 5.6. SGD Algoritması doğruluk tablosu

	Kesinlik	Duyarlılık	F1-skor
Sahte	0.90	0.85	0.88
Gerçek	0.86	0.91	0.88
Doğruluk			0.88
Macro avg	0.88	0.88	0.88
Weighted avg	0.88	0.88	0.88

Çizelge 5.6.'da yer alan değerlendirme metriklerine göre SGD algoritması doğruluk değeri %88 elde edilmiştir.

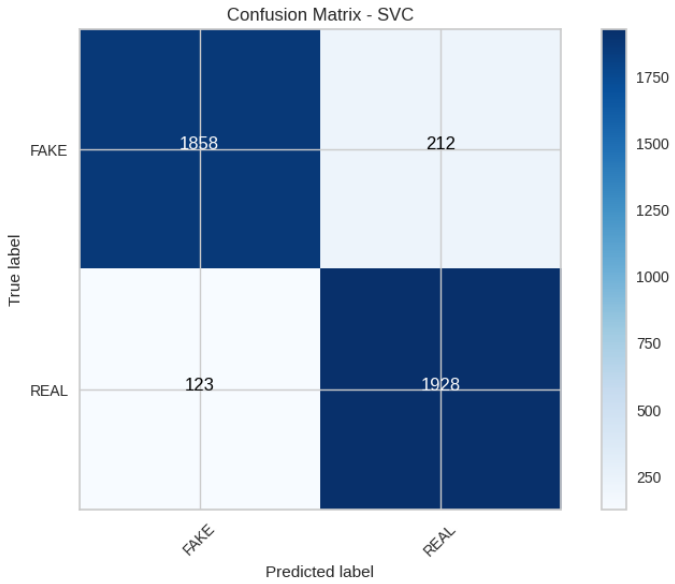
5.1.7. SVM

SVM algoritması, sınıflandırma ve regresyon problemlerinde en çok kullanılan algoritmadır. Verilerimizi dönüştürmek için çekirdek üzerinde kullanılan teknikler kullanmaktadır ve bu dönüşümlere dayanarak olası çıktılar arasında en uygun sınırı bulmaktadır. SVM'ler, tüm sınıfların en yakın veri noktalarına olan mesafeyi maksimize eden karar sınırını seçme biçimleri nedeniyle diğer sınıflandırma algoritmalarından farklıdır [158].

SVM algoritması, sınıflandırmada yaşanan güçlükleri çözmek için kullanılmaktadır. Her veri öğesi, n- boyutlu uzayda bir nokta olarak temsil edilmektedir ve her bir özelliğin değeri, SVM algoritmasında belirli bir koordinatın değeridir. İki sınıfı en iyi ayıran hiper düzlemi seçerek sınıflandırma gerçekleştirilmektedir. Destek vektörlerini hesaplamak için bireysel gözlem koordinatları kullanılmaktadır. SVM sınıflandırıcısı, iki sınıfı birbirinden en iyi şekilde ayıran bir sınırdır [158].

Scikit-learn, SVM sınıflandırmasını uygulamak maksadıyla SVClassifier modülü sağlamaktadır [159]. Çalışmada bu modülden faydalanarak sınıflandırma başarıları elde edilmiştir.

Çalışmada kullanılan SVM algoritmasının sınıflandırma aşamasında oluşan karmaşıklık matrisi Şekil 5.8.'de sunulmuştur.



Şekil 5.8. SVM karmaşıklık matrisi

Şekil 5.8.'de gösterilen SVM algoritması sonucu ortaya çıkan karmaşıklık matrisi ışığında elde edilen sonuçların yer aldığı doğruluk tablosu Çizelge 5.7.'de sunulmuştur.

Çizelge 5.7. SVM Algoritması doğruluk tablosu

	Kesinlik	Duyarlılık	F1-skor
Sahte	0.94	0.90	0.92
Gerçek	0.90	0.94	0.92
Doğruluk			0.92
Macro avg	0.92	0.92	0.92
Weighted avg	0.92	0.92	0.92

Çizelge 5.7.'de yer alan değerlendirme metriklerine göre SVM algoritması doğruluk değeri %92 elde edilmiştir.

Makine öğrenmesi algoritmalarıyla sınıflandırmaya başlamadan önce veri setimizde bulunan metinleri sayısal hale getirmemiz gerekmektedir. Bunun için de vektör uzaylarından faydalanılmıştır. En çok kabul gören yöntemlerden biri olan vektör uzayıyla metinlerim sayısallaştırılması, her tivit cümlesinin vektör olarak temsilinin yapıldığı BoW sürümüdür. Her boyut ayrı birer terime denk gelmektedir. Yani her bir tivit cümlesi, mevcuttaki kelimelerden oluşan bir AxB büyüklüğünde bir vektördür. Özetle vektörleri üst üste ekleyerek doküman-terim matrisi oluşturduk. AxB matrisinde A tane haber ve B tane

terim bulunmaktadır. Eğer ki ilgili terimler haber içeriğinde bulunuyorsa o terimin ağırlığı sıfırdan farklı demektir. Terimin ağırlık değeri denilen şey ilgili terimin metin üzerindeki etkisi anlamına gelmektedir.

Terim ağırlığını hesaplamak için iki yöntem kullanılmıştır. Count Vectorizer ve TF-IDF yaklaşımı. Terim saymak, ilgili terimin tivit içerisinde ne kadar geçtiğini hesaplamaktır. TD-IDF yönteminde hem iligili terimin haber tivitindeki sıklığına (TF) hem de tüm haber veri seti içerisindeki önemine (IDF) bakılmaktadır.

Kullanıdığımız yöntem TF-IDF yöntemidir. Aşağıdaki gibi hesaplanır.

$$\text{TF-IDF}\ (t,d) = \text{TF}(t,d) \times \log(N/\text{DF}(t))$$

Bu denklemde t terim sayısını, d haber tivitini, TF terim sıklığını, DF t'inci terimi içeren haber twiti sayısını ve N toplam haber tiviti sayısını göstermektedir.

Çalışma kapsamında kullanılan makine öğrenmesi modellerinin değerlendirmesi ayrıntılı bir şekilde gerçekleştirilmiştir. Makine öğrenmesi modellerimizi denerken veri setimizin kökleri alınmış hali kullanılmıştır.

Sınıflandırma modellerinin performansını değerlendirmek için yaygın olarak kullanılan metrikler doğruluk, kesinlik, duyarlılık ve f1 ölçümüdür. Bu metrikler hesaplanırken 4 parametre gereklidir; Gerçek Pozitif(TP), Gerçek Negatif (TN), Yanlış Pozitif (FP) ve Yanlış Negatif(FN).

• TP: Doğru tahmin edilen pozitif sınıf.

• TN: Doğru tahmin edilen negatif sınıf.

• FP: Yanlış tahmin edilen pozitif sınıf.

• FN: Yanlış tahmin edilen negatif sınıf.

Bu çalışmada model performansını değerlendirmek için kullanılan performans metriklerini hesaplayan formüller (1), (2), (3) ve (4) denklemlerinde sunulmuştur.

$$Accuracy := \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Precision := \frac{TP}{TP+FP} \quad (2)$$

$$Recall := \frac{TP}{TP+FN} \quad (3)$$

$$F1 - Measure := \frac{2*TP}{2*TP+FP+FN} \quad (4)$$

Değerlendirme metriklerine göre tez kapsamında oluşturulan TR_FaReNews veri seti için hesaplanan doğruluk, kesinlik, geri çağırma, f1-skor hesaplama sonuçları Çizelge 5.8.'de gösterilmiştir.

Çizelge 5.8. TR_FaReNews veri seti kullanarak makine öğrenmesi algoritmalarıyla elde edilen sonuçlar

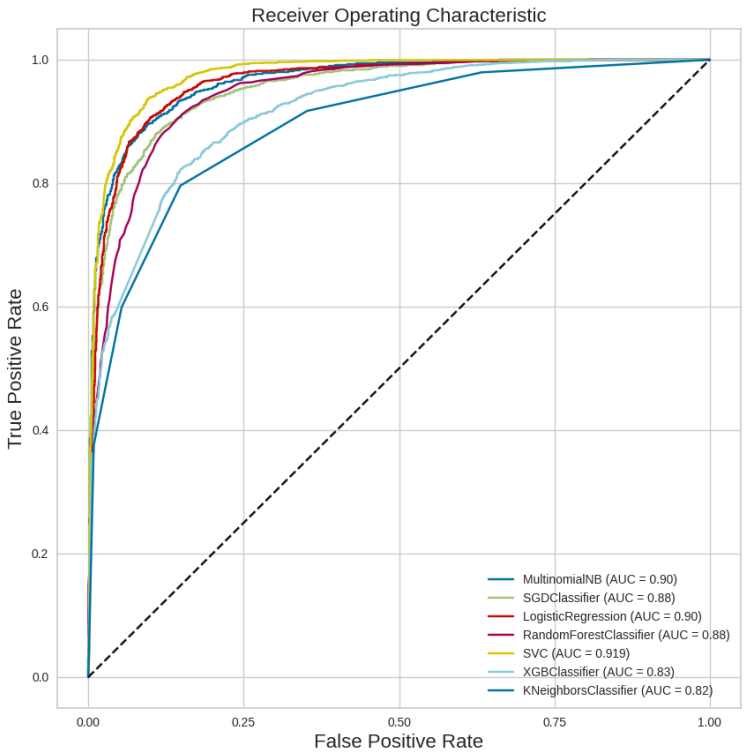
Algoritma	Doğruluk	Etiket	Kesinlik	Geri Çağırma	F-1 skor
MultinomialNB	0.90	Gerçek	0.87	0.93	0.90
		Sahte	0.93	0.86	0.89
SGDClassifier	0.88	Gerçek	0.90	0.85	0.88
		Sahte	0.86	0.91	0.88
Logistic Regression	0.90	Gerçek	0.90	0.91	0.90
		Sahte	0.91	0.89	0.90
RandomForest	0.88	Gerçek	0.87	0.89	0.88
		Sahte	0.88	0.87	0.88
SVM	0.92	Gerçek	0.94	0.90	0.92
		Sahte	0.90	0.94	0.92
XGBoost	0.83	Gerçek	0.87	0.78	0.82
		Sahte	0.80	0.88	0.84
K-NN	0.82	Gerçek	0.81	0.85	0.83
		Sahte	0.84	0.80	0.82

Hesaplamalar sonucunda en iyi doğruluk değerini SVM sınıflandırıcısı 0.92 doğruluk ile vermiştir. Ancak "eğitim" ve "test" verisini seçerken herhangi bir yanlışlık yapmadığımızdan ve veri seti içerisinde tekrarlama veya ezberleme olmadığından emin olmak için k-katmanlı çapraz sorgulama yapılmıştır. Burada k değeri fastText modelinde olduğu gibi 10 olarak belirlenmiştir. Yani veri setini 10 eş parçaya ayırarak 10 kez farklı verilerle doğruluk oranı hesaplanmıştır. Bu hesaplama sonucunda elde edilen sonuçlar Çizelge 5.9.'de sunulmuştur.

Çizelge 5.9. Çapraz doğrulama sonrası makine öğrenmesi algoritmalarıyla elde edilen sonuçlar

	Algoritma	Doğruluk
10-folds çapraz doğrulama	SVC	0.91 (+/- 0.02)
	Multinomial Naive Bayes	0.88 (+/- 0.02)
	SGD	0.87 (+/- 0.01)
	Logistic Regression	0.89 (+/- 0.02)
	Random forest	0.86 (+/- 0.02)
	KNN	0.81 (+/- 0.02)
	XGBoost	0.82 (+/- 0.03)

Sınıflandırma sonucunda 7 farklı modeli karşılaştırılarak en performanslı sonucu SVM algoritmasının verdiği görülmüştür. Ayrıca her bir model için ROC grafiği üretildi. ROC eğrisi çizdirmek tez çalışmasında olduğu gibi ikili sınıflandırma çok sık kullanılmaktadır. İyi boyutlu bir grafik olan ROC eğrisinde önemsediğimiz metrik AUC'dir. AUC ROC eğrisinin altında bulunan alan anlamına gelmektedir ve modelin gerçek kalitesini ve performansını temsil etmektedir. 0 ile 1 arasında olmalıdır. AUC değer 1'e ne kadar yakınsa model o kadar başarılır demektir. Çalışma kapsamında çizdirdiğimiz ROC eğrisi Şekil 5.9.'deki gibidir.



Şekil 5.9. Çalışma kapsamında çizilen ROC eğrisi

Şekil 5.9.'daki ROC eğrisinde de görüldüğü gibi 1'e en yakın değer SVM algoritmasına aittir. Modelimiz doğru çalışmaktığı tahmin edilmektedir.

Ayrıca kullanılan 7 adet makine öğrenmesi algoritması modeli, BuzzFeed, GosspCop, ISOT, LIAR, Twitter15 ve Twitter16 veri setlerinde de denenmiştir. Söz konusu veri setlerinde kullanılan alanlar modelimize uygun olacak şekilde güncellenmiş ve deneylerimizde kullanılmıştır.

Makine öğrenmesi modelimizde BuzzFeed veri setinin sadece haber başlıkları kullanılmış ve Şekil 5.1'deki model mimarisine göre ön işleme ve özellik çıkarımları yapıldıktan sonra 7 makine öğrenmesi algoritmasıyla çeşitli değerlendirme metrikleri elde edilmiştir. BuzzFeed veri setine ait metrik sonuçları Çizelge 5.10.'da sunulmuştur.

Çizelge 5.10. BuzzFeed veri seti değerlendirme metrikleri

Algoritma	Doğruluk	Etiket	Kesinlik	Geri Çağırma	F-1 skor
MultinomialNB	0.69	Gerçek	0.67	0.69	0.68
		Sahte	0.71	0.69	0.70
SGDClassifier	0.73	Gerçek	0.72	0.69	0.71
		Sahte	0.73	0.76	0.75
Logistic Regression	0.71	Gerçek	0.69	0.69	0.69
		Sahte	0.72	0.72	0.72
RandomForest	0.76	Gerçek	0.81	0.65	0.72
		Sahte	0.74	0.86	0.79
SVM	0.62	Gerçek	0.58	0.73	0.64
		Sahte	0.68	0.52	0.59
XGBoost	0.64	Gerçek	0.62	0.62	0.62
		Sahte	0.66	0.66	0.66
K-NN	0.65	Gerçek	0.64	0.62	0.63
		Sahte	0.67	0.69	0.68

Çizelge 5.10'da gösterildiği gibi, BuzzFeed veri setinin değerlendirme metrikleri hesaplanmış ve en yüksek doğruluk Random Forest algoritması ile %76 olarak elde edilmiştir.

Makine öğrenmesi modelimizde GossipCop veri setinin sadece haber metinleri kullanılmış ve Şekil 5.1'deki model mimarisine göre ön işleme ve özellik çıkarımları yapıldıktan sonra 7 makine öğrenmesi algoritmasıyla çeşitli değerlendirme metrikleri elde edilmiştir. GossipCop veri setine ait metrik sonuçları Çizelge 5.11.'de sunulmuştur.

Çizelge 5.11. GossipCop veri seti değerlendirme metrikleri

Algoritma	Doğruluk	Etiket	Kesinlik	Geri Çağırma	F-1 skor
MultinomialNB	0.82	Gerçek	0.82	0.99	0.90
		Sahte	0.88	0.30	0.45
SGDClassifier	0.77	Gerçek	0.77	1.00	0.87
		Sahte	0.83	0.04	0.08
Logistic Regression	0.81	Gerçek	0.87	0.87	0.87
		Sahte	0.59	0.59	0.59

Algoritma	Doğruluk	Etiket	Kesinlik	Geri Çağırma	F-1 skor
RandomForest	0.84	Gerçek	0.85	0.95	0.90
		Sahte	0.76	0.47	0.58
SVM	0.85	Gerçek	0.85	0.97	0.91
		Sahte	0.83	0.46	0.59
XGBoost	0.84	Gerçek	0.85	0.96	0.90
		Sahte	0.80	0.45	0.57
K-NN	0.80	Gerçek	0.85	0.90	0.87
		Sahte	0.60	0.50	0.55

Çizelge 5.11'de gösterildiği gibi, GosspCop veri setinin değerlendirme metrikleri hesaplanmış ve en yüksek doğruluk SVM algoritması ile %85 olarak elde edilmiştir.

Makine öğrenmesi modelimizde ISOT veri setinin sadece haber metinleri kullanılmış ve Şekil 5.1'deki model mimarisine göre ön işleme ve özellik çıkarımları yapıldıktan sonra 7 makine öğrenmesi algoritmasıyla çeşitli değerlendirme metrikleri elde edilmiştir. ISOT veri setine ait metrik sonuçları Çizelge 5.12'de sunulmuştur.

Çizelge 5.12. ISOT veri seti değerlendirme metrikleri

Algoritma	Doğruluk	Etiket	Kesinlik	Geri Çağırma	F-1 skor
MultinomialNB	0.94	Gerçek	0.95	0.91	0.93
		Sahte	0.92	0.96	0.94
SGDClassifier	0.94	Gerçek	0.95	0.91	0.93
		Sahte	0.92	0.96	0.94
Logistic Regression	0.95	Gerçek	0.95	0.95	0.95
		Sahte	0.95	0.96	0.96
RandomForest	0.95	Gerçek	0.94	0.95	0.95
		Sahte	0.96	0.95	0.95
SVM	0.96	Gerçek	0.96	0.97	0.96
		Sahte	0.97	0.96	0.97
XGBoost	0.93	Gerçek	0.89	0.96	0.93
		Sahte	0.96	0.90	0.93
K-NN	0.85	Gerçek	0.81	0.89	0.85
		Sahte	0.89	0.81	0.85

Çizelge 5.12'de gösterildiği gibi, ISOT veri setinin değerlendirme metrikleri hesaplanmış

ve en yüksek doğruluk SVM algoritması ile %96 olarak elde edilmiştir.

Makine öğrenmesi modelimizde LIAR veri setinin sadece iki etiketli hali (sahte ve gerçek) kullanılmış ve Şekil 5.1'deki model mimarisine göre ön işleme ve özellik çıkarımları yapıldıktan sonra 7 makine öğrenmesi algoritmasıyla çeşitli değerlendirme metrikleri elde edilmiştir. LIAR veri setine ait metrik sonuçları Çizelge 5.13'de sunulmuştur.

Çizelge 5.13. LIAR veri seti değerlendirme metrikleri

Algoritma	Doğruluk	Etiket	Kesinlik	Geri Çağırma	F-1 skor
MultinomialNB	0.59	Gerçek	0.56	0.64	0.60
		Sahte	0.61	0.54	0.57
SGDClassifier	0.60	Gerçek	0.58	0.65	0.61
		Sahte	0.63	0.56	0.59
Logistic Regression	0.54	Gerçek	0.53	0.57	0.55
		Sahte	0.56	0.52	0.54
RandomForest	0.61	Gerçek	0.60	0.57	0.59
		Sahte	0.62	0.65	0.63
SVM	0.60	Gerçek	0.58	0.62	0.60
		Sahte	0.62	0.59	0.60
XGBoost	0.60	Gerçek	0.55	0.52	0.57
		Sahte	0.58	0.61	0.59
K-NN	0.57	Gerçek	0.51	0.54	0.53
		Sahte	0.55	0.52	0.53

Çizelge 5.13'te gösterildiği gibi, LIAR veri setinin değerlendirme metrikleri hesaplanmış ve en yüksek doğruluk Random Forest algoritması ile %61 olarak elde edilmiştir.

Makine öğrenmesi modelimizde Twitter15 veri setinin haber metinleri kullanılmış ve Şekil 5.1'deki model mimarisine göre ön işleme ve özellik çıkarımları yapıldıktan sonra 7 makine öğrenmesi algoritmasıyla çeşitli değerlendirme metrikleri elde edilmiştir. Twitter15 veri setine ait metrik sonuçları Çizelge 5.14'te sunulmuştur.

Çizelge 5.14. Twitter15 veri seti değerlendirme metrikleri

Algoritma	Doğruluk	Etiket	Kesinlik	Geri Çağırma	F-1 skor
MultinomialNB	0.93	Gerçek	0.92	0.93	0.92
		Sahte	0.93	0.93	0.92
SGDClassifier	0.94	Gerçek	0.95	0.93	0.94
		Sahte	0.93	0.95	0.94
Logistic Regression	0.95	Gerçek	0.95	0.94	0.94
		Sahte	0.94	0.95	0.95
RandomForest	0.93	Gerçek	0.96	0.89	0.92
		Sahte	0.90	0.96	0.93
SVM	0.94	Gerçek	0.95	0.92	0.93
		Sahte	0.93	0.95	0.94
XGBoost	0.81	Gerçek	0.81	0.78	0.82
		Sahte	0.80	0.83	0.82
K-NN	0.92	Gerçek	0.88	0.95	0.92
		Sahte	0.95	0.88	0.92

Çizelge 5.14'te gösterildiği gibi, Twitter15 veri setinin değerlendirme metrikleri hesaplanmış ve en yüksek doğruluk Logistic Regression algoritması ile %95 olarak elde edilmiştir.

Makine öğrenmesi modelimizde Twitter16 veri setinin haber metinleri kullanılmış ve Şekil 5.1'deki model mimarisine göre ön işleme ve özellik çıkarımları yapıldıktan sonra 7 makine öğrenmesi algoritmasıyla çeşitli değerlendirme metrikleri elde edilmiştir. Twitter16 veri setine ait metrik sonuçları Çizelge 5.15'de sunulmuştur.

Çizelge 5.15. Twitter16 veri seti değerlendirme metrikleri

Algoritma	Doğruluk	Etiket	Kesinlik	Geri Çağırma	F-1 skor
MultinomialNB	0.92	Gerçek	0.91	0.93	0.92
		Sahte	0.93	0.91	0.92
SGDClassifier	0.95	Gerçek	0.96	0.95	0.95
		Sahte	0.95	0.96	0.96
Logistic Regression	0.95	Gerçek	0.96	0.95	0.95
		Sahte	0.95	0.96	0.96
RandomForest	0.93	Gerçek	0.98	0.87	0.92

		Sahte	0.89	0.98	0.93
SVM	0.94	Gerçek	0.93	0.95	0.94
		Sahte	0.95	0.93	0.94
XGBoost	0.86	Gerçek	0.83	0.89	0.86
		Sahte	0.88	0.82	0.85
K-NN	0.88	Gerçek	0.83	0.89	0.86
		Sahte	0.88	0.82	0.85

Çizelge 5.15'de gösterildiği gibi, Twitter16 veri setinin değerlendirme metrikleri hesaplanmış ve en yüksek doğruluk LR ve SVM algoritması ile %95 olarak elde edilmiştir. Tez kapsamında oluşturulan TR_FaReNews veri seti ve diğer veri setlerini toplu halde görebileceğimiz karşılaştırma tablosu Çizelge 5.16'da sunulmuştur.

Çizelge 5.16. Tüm veri setleriyle makine öğrenmesi modelleri karşılaştırma tablosu

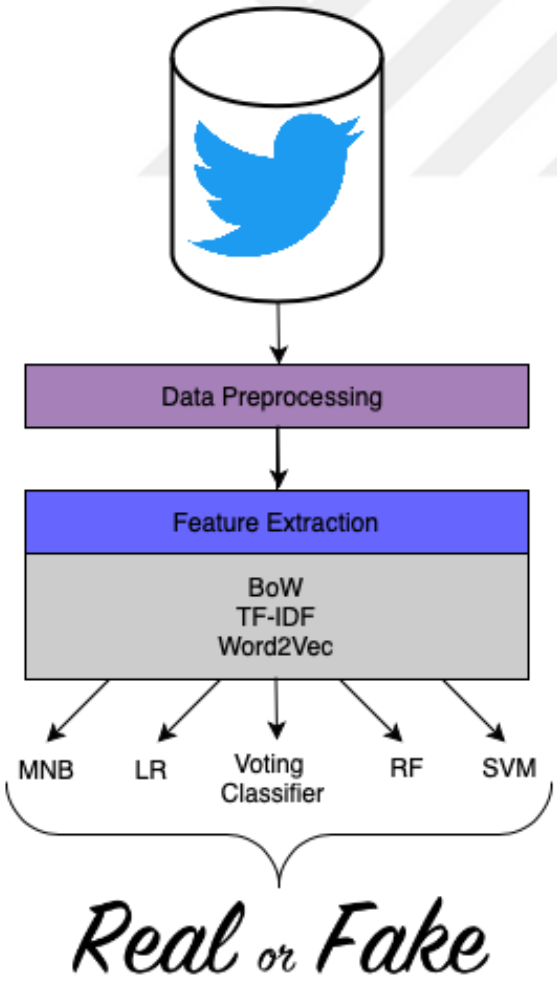
Algoritma	Veri seti						
	BuzzFeed	GossipCop	ISOT	LIAR	Twitter15	Twitter16	TR_FaRe_News
MNB	0.69	0.82	0.94	0.59	0.93	0.92	0.90
SGD	0.73	0.77	0.94	0.60	0.94	0.95	0.88
LR	0.71	0.81	0.95	0.54	**0.95**	**0.95**	0.90
RF	**0.76**	0.84	0.95	**0.61**	0.93	0.93	0.88
SVM	0.62	**0.85**	**0.96**	0.60	0.94	0.94	**0.92**
XGB	0.64	0.84	0.93	0.60	0.81	0.86	0.83
K-NN	0.65	0.80	0.85	0.57	0.92	0.88	0.82

Çizelge 5.16'da görüldüğü gibi, makine öğrenmesi modelimizin tez kapsamında oluşturulan veri seti TR_FaRe_News ile kıyaslamasını yapacak olursak, ISOT, Twitter15 ve Twitter16 veri kümelerinde TR_FaRe_News veri kümesinden daha iyi doğruluk oranı elde ettiği görülmüştür. Buradaki ana sebebin veri setinin Twitter'daki kısa tivitlerden oluşması ve Türk diler ait olan dil çözümleme zorlukları olduğunu tahmin edilmektedir.

5.1.8. Oylama sınıflandırıcı

Topluluk öğrenme, tahmine dayalı bir modele ulaşmak için iki veya daha fazla algoritma kullanmaktadır. Bunun sebebi, birden fazla algoritma kullanmanın daha iyi sonuçlar veya daha genelleştirilmiş sonuçlar verebileceğidir [160].

Topluluk öğrenme, tek bir modele güvenmenin aşırı uyum gibi sorunlara neden olabileceği fikrine dayanmaktadır. Örneğin bu durumda sahte ve gerçek olmak üzere iki sınıflı bir sınıflandırma problemimiz olduğundan bunların ortalaması (en çok oyu alanı) çoğunluğun kararı olarak değerlendirilmektedir. Yani bir sınıflandırma probleminde, problemi çözmek için oyları saymak yerine farklı çıktıların ortalamasını almak daha doğru bir yaklaşım olacaktır [160]. Bu kapsamda, MNB, LR, SVM, RF ve VC kullanarak topluluk öğrenme yaklaşımını kullanarak özgün bir yaklaşım sunulmuştur. Şekil XXX'te topluluk öğrenme modelinin mimarisi sunulmuştur.



Şekil 5.10. Topluluk öğrenme mimarisi

Şekil 5.10'da görüldüğü üzere TR_FaRe_News ve GPT-2 sahte haber veri kümelerindeki tivit metinleri öncelikle veri ön işleme adımlarından geçirilmiştir. BoW, TF-IDF ve word2vec ile özellik çıkarımı yapılarak ardından sınıflandırılmıştır. Sınıflandırma sonuçları Çizelge 5.17'de gösterilmektedir.

Çizelge 5.17. Topluluk öğrenmesi sonuçları

Metot	TR_FaRe_News Veri kümesi		GPT-2 Sahte haber veri kümesi	
	Doğruluk	f1-skor	Doğruluk	f1-skor
MNB+BoW	%71	%66.2	%63.2	%63
MNB+TF-IDF	%71	%66.2	%62.7	%62.7
MNB+Word2Vec	%85.2	%85.1	%69.6	%69.5
RF+BoW	%73.5	%79.1	%65.3	%66
RF+TF-IDF	%73.4	%79.4	%65.3	%65.2
RF+Word2Vec	%90	%90.9	%76.7	%76.6
LR+BoW	%73.6	%71.3	% 65.7	%65.5
LR+TF-IDF	%72.9	%71.2	%64.7	%64.6
LR+Word2Vec	%88.5	%88.4	%77.1	%77.1
SVM+BoW	%73	%70.4	%65.6	%65
SVM+TF-IDF	%71.9	%69.7	%64.1	%64.1
SVM+Word2Vec	%88	%88	%77	%77
VC+BoW	%73.5	%71.2	%65.8	%65.5
VC+TF-IDF	%72.8	%71	%64.9	%64.8
VC+Word2Vec	**%89**	%88.9	**%77.6**	%77.6

Çizelge 5.17. incelendiğinde, oluşturduğumuz TR_FaRe_News veri kümesi ile büyük bir dil modeli olan GPT-2 tarafından sahte haberlerden oluşan veri kümesinin karşılaştırdığımızda, TR_FaRe_News veri kümesi için en performanslı algoritma olan word2vec geliştirme süreci ile topluluk sistemlerinin oylama sınıflandırıcısının başarı oranı %89'dur. GPT-2 veri kümesi için en iyi performans gösteren algoritmanın oylama sınıflandırıcısı olduğu ve %77,6 başarı elde ettiği görülmüştür.

5.2. fastText

Makine öğrenmesi modelleriyle karşılaşılan zorlukların üstesinden gelebilmek maksadıyla kelime gömme ve kelime temsili adı verilen yöntemler geliştirilmiştir [161]. 2013 yılında geliştirilen word2Vec adı verilen kelime temsili yöntemi anlamları birbirine benzeyen

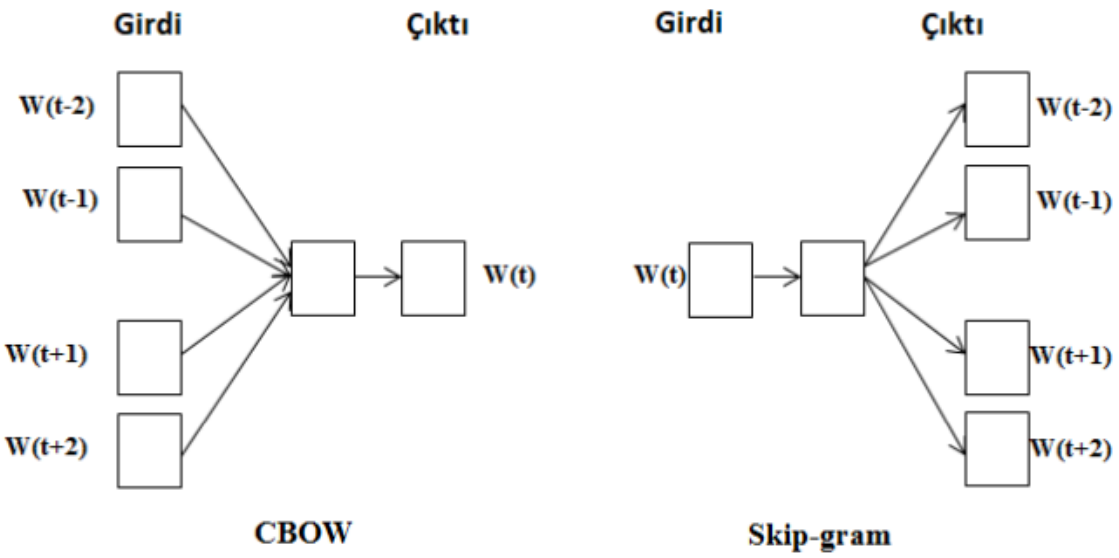
kelimelerin birbirlerine yakınlık göstermesi yöntemidir. Tahmine dayalı bu yöntem, CBOW ve Skip-gram adı verilen büyük miktarda yapılandırılmamış metin verisinin vektör temsillerinin verimliliğini yükseltmek maksadıyla geliştirilmiştir.

CBOW

CBOW modeli, dil çevirisi ve metin sınıflandırması gibi doğal dilleri işleme görevleri için bir sinir ağı yapısıdır. Çevreleyen kelimelerin bağlamına dayalı olarak bir hedef kelimeyi tahmin etmekte ve stokastik gradyan inişe benzer bir optimizasyon algoritması kullanılarak büyük bir metin veri kümesi üzerinde eğitilmektedir. CBOW modeli, bir kere eğitildikten sonra, sürekli bir vektör uzayında sözcüklerin anlamlarını yakalayan ve çeşitli DDİ görevlerinde kullanılabilen, kelime yerleştirmeleri olarak bilinen sayısal vektörler üretmektedir. Genellikle skip-gram modeli gibi diğer teknikler ve modellerle birleştirilebilkmekte ve Python'daki gensin gibi kütüphaneler kullanılarak uygulanabilmektedir [162].

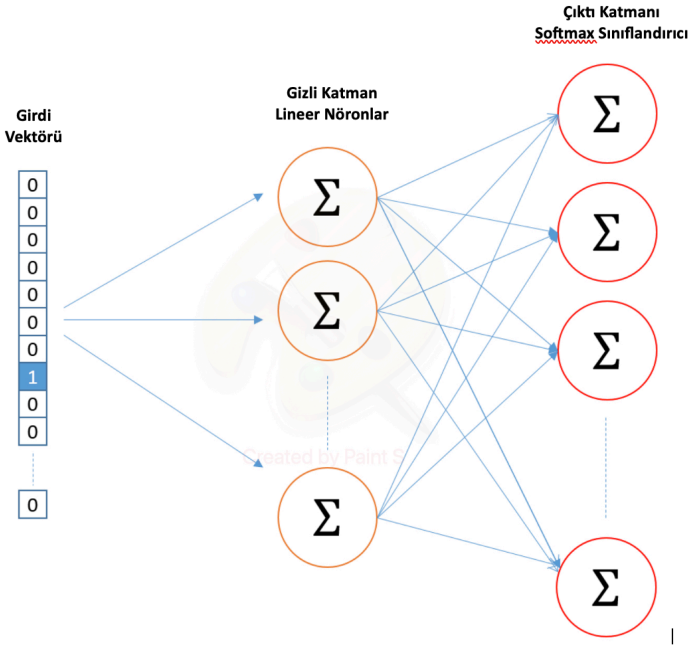
Skip-gram modeli ise CBOW'a benzemektedir, ancak bağlama göre mevcut kelimeyi tahmin etmek yerine, aynı cümledeki başka bir kelimeye dayalı olarak bir kelimenin sınıflandırmasını en üst düzeye çıkarmaya çalışmaktadır. Daha kesin olarak, mevcut her kelimeyi, sürekli izdüşüm katmanına sahip bir sınıflandırıcıya girdi olarak kullanırız ve mevcut kelimeden önce ve sonra belirli bir aralıktaki kelimeleri tahmin ederiz. Aralığı artırmanın, ortaya çıkan kelime vektörlerinin kalitesini iyileştirdiğini, ancak aynı zamanda hesaplama karmaşıklığını da artığını görülmüştür. Daha uzak kelimeler genellikle mevcut gelimeye yakın olanlardan daha az ilişkili olduğundan, eğitim örneklerimizde bu kelimelerden daha az örnek alarak uzaktaki kelimelere daha az ağırlık veririz [163].

CBOW ve Skip-gram modellerinin mimarisi Şekil 5.11.'de gösterilmiştir.



Şekil 5.11. CBOW ve Skip-gram model mimarisi (Türkçeleştirilmiştir) [166]

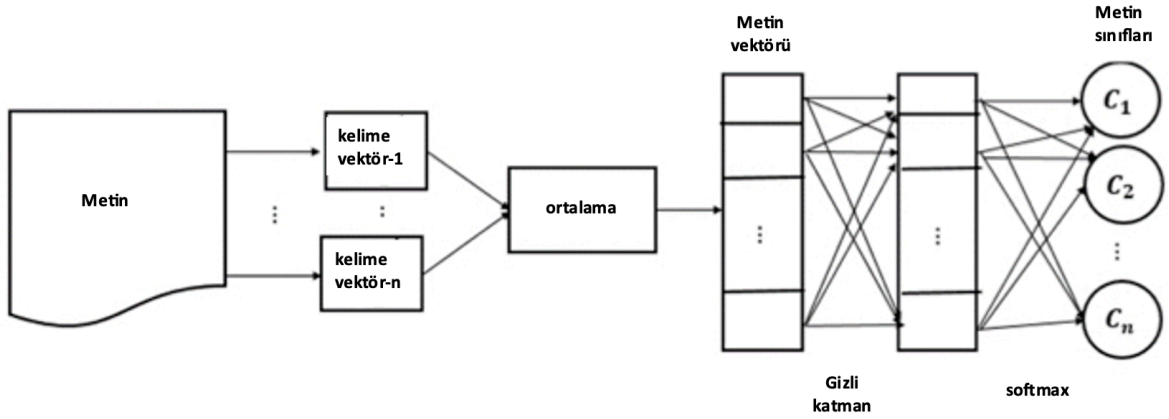
CBOW ve Skip-gram modellerini karşılaştıracak olursak; CBOW modelinde merkez noktadaki kelimeyi bulmak için etrafındaki (çevreleyen) kelimeler kullanılırken, skip-gram modelinde cümle ya da dokümanın etrafındaki kelimeleri tahmin edebilmek için merkez kelime kullanılmaktadır. Yukarıda da bahsedilen gibi word2vec yapay sinir ağı temelli bir modeldir [164]. Word2Vec için çizilmiş örnek gösterim Şekil 5.12'de sunulmuştur.



Şekil 5.12. word2vec mimarisi (Türkçeleştirilmiştir.) [164]

fastText, kelime temsillerinin ve metin sınıflandırmanın etkili bir şekilde öğrenilmesi için Facebook Araştırma Ekibi tarafından geliştirilen bir araç takımıdır [165]. fastText kelime yerleştirme modelinin ana katkılarından biri, algoritmanın özellikle Türkçe gibi morfolojik olarak zengin diller için faydalı olan kelime temsillerini öğrenirken iç yapısını dikkate almasıdır [166].

fastText kelime yerleştirme modelinin kelime temsili yaklaşımı, word2vec gibi diğer kelime yerleştirme modellerinden farklıdır. Word2vec her bir kelimeyi en küçük birim olarak kullanırken fastText bir kelimenin n-gram karakterden oluştuğunu varsayar, burada n'nin uzunluğu bir kelimede kelimenin uzunluğuna kadar değişebilir. Bu yaklaşımın yararı, yöntemin kelime vektörlerini karakter n-gramları olarak tuttuğu için, sözlükte doğrudan bulunmayan kelimeler için vektör temsillerini bulabilmesidir [167,168,169]. fastText sınıflandırmasının genel mimarisi Şekil 5.13'de sunulmuştur.



Şekil 5.13. fastText metodolojisinin genel mimarisi [168] (Türkçeleştirilmiştir.)

Çizelge 5.18'de fastText kelime yerleştirme modelinde kullanılan parametreler açıklamalarıyla birlikte sunulmuştur.

Çizelge 5.18. fastText modelinde kullanılan parametreler

Parametreler	**Açıklama**	**Kullanılan Değer**
Lr	Öğrenme oranı	0.1
Epochs	Epoch sayısı	75
Dim	Kelime vektörlerinin boyutu	300
Neg	Örneklenen negatiflerin sayısı	5
ws	İçerik penceresinin boyutu	2
wordNgrams	Maksimum kelime n gram uzunluğu	3

Öğrenme hızı, derin sinir ağlarının eğitim ayrıntılarını belirtmek için bir hiper parametredir. Epoch parametreleri, eğitim setlerinin ileri ve geri geçişini temsil eder. Sözcük vektörlerinin boyutu fastText sınıflandırma modelinde bir diğer kritik parametredir. Boyut değerinin maksimize edilmesi, kelime benzerlikleri belirlenirken gereksiz kelimelere yer verildiğinden sınıflandırma doğruluğunu azaltabilir. İçerik penceresinin boyutu, her kelimenin bağlamını belirlemek için kullanılan kelimelerin sayısını gösterir [168].

Çalışmanın farklı yanlarından biri fastText'in twitter veri setleri üzerinde sahte haber sınıflandırma amaçlı olarak uygulanmamış olmasıdır.

Çalışma kapsamında veri setimizde bulunan haberlerin sahte veya gerçek haber olup olmadıklarının tespit edilebilmesi için ikili bir sınıflandırma yapılması hedeflenmiştir. Bu kapsamda "gerçek" haber sayısı 6868 ve "sahte" haber sayısı 6867 olacak şekilde ikili sınıflandırmaya uygun olacak şekilde veri seti düzenlenmiştir. FastText modelini çalıştırabilmek için verilerin başına __label__sahte veya __label__gerçek şeklinde eklemelerimiz yapılmıştır. Bu çalışma için hazırlanan haber külliyatı ile fastText kullanarak bir model eğitmek için iki yöntem uygulanmıştır. İlk olarak, en iyi sonuçları elde etmek amacıyla farklı kombinasyonlarla eğitim için Çizelge 5.10'da sunulan hiper parametre değerleri kullanılmıştır. İkinci olarak, fastText'in belirli bir zaman diliminde en iyi hiper parametreleri otomatik olarak bulmaya çalışan "autotune özelliği" kullanılmıştır. Autotune özelliği ile fastText modelimiz 3 gün/72 saat/4320 dakika/259200 saniye çalıştırılmıştır. Modeller tabakalı örnekleme ile 10 kat çapraz doğrulama için önceden hazırlanmış test verileriyle test edilmiştir.

fastText, ikili sınıflandırma yaparken yukarıda anlatılan değerlendirme metriklerini otomatik olarak hesaplamamaktadır. Bu nedenle, bu metrikleri hesaplamak için Scikit-learn kütüphanesi işlevleri kullanılmıştır. 10 kat çapraz doğrulama için eğitim ve test işlemleri yapılmış ve yapılan testler doğrultusunda elde edilen en iyi performans sonuçları Çizelge 5.19'da sunulmuştur.

Çizelge 5.19. fastText ile elde edilen en iyi performans sonuçları

Set Durumu	Doğruluk	Etiket	Kesinlik	Geri çağırma	F-1 skor
Ham hali	0.83	Gerçek	0.85	0.81	0.83
		Sahte	0.81	0.86	0.84
Kök alınmış hali	0.84	Gerçek	0.84	0.84	0.84
		Sahte	0.84	0.84	0.84
Autotune edilmiş hali	0.93	Gerçek	0.95	0.91	0.93
		Sahte	0.91	0.96	0.94

fastText modelimizi sisteme yükledikten sonra çeşitli tivit tahmin denemeleri de yapılmıştır. Bu denemelerden birkaçı örnek olarak gösterilmiştir. fastText modelimiz ile yapılan tahminlerin gösterimleri Şekil 5.14'de sunulmuştur.

```
# Make predictions on new data
text = "ankara belediye başkanı mansur yavaş oldu"
labels, prob = model.predict(text)
print("Label:", labels[0])
print("Probability:", prob[0])
```

```
Label: __label__real
Probability: 0.9459689855575562
```

```
# Make predictions on new data
text = "ankara belediye başkanı melih gökçek oldu"
labels, prob = model.predict(text)
print("Label:", labels[0])
print("Probability:", prob[0])
```

```
Label: __label__fake
Probability: 0.9699885249137878
```

```
# Make predictions on new data
text = "uğur şahin aşı olmadı"
labels, prob = model.predict(text)
print("Label:", labels[0])
print("Probability:", prob[0])
```

```
Label: __label__fake
Probability: 0.9999842643737793
```

```
# Make predictions on new data
text = "fahrettin koca aşı olma tavsiyesi verdi"
labels, prob = model.predict(text)
print("Label:", labels[0])
print("Probability:", prob[0])
```

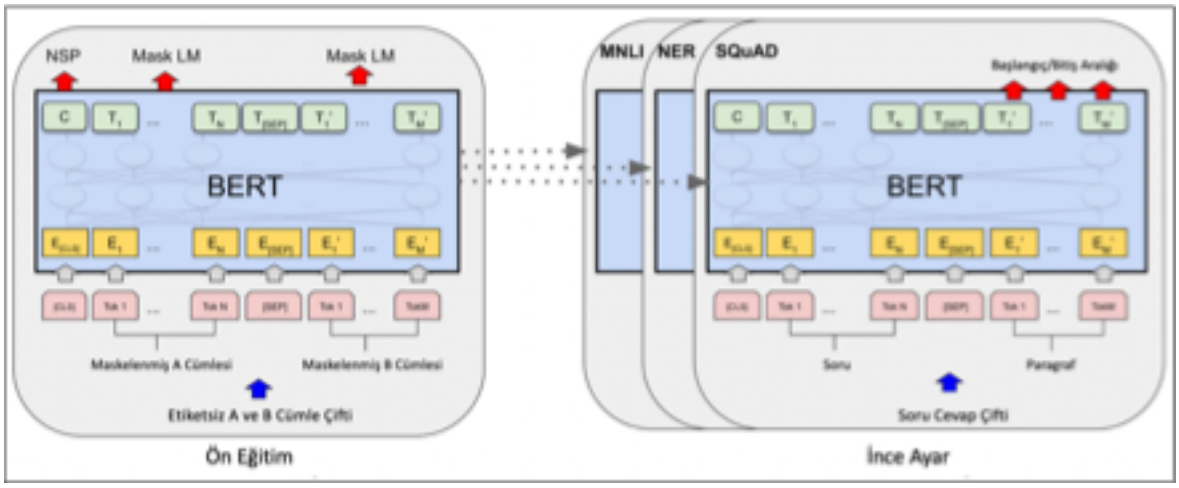
```
Label: __label__real
Probability: 0.9633703827857971
```

Şekil 5.14. fastText tahminleme örnekleri

5.3. BERT

BERT dil temsil modelinin tasarımı oldukça basittir ve dil modelleri arasında en güçlü olanlardan biridir. İlk olarak 2018 yılında tanıtımı yapılmıştır. Etiketlenmemiş verilerle işlem yapabilen çift yönlü bir şekilde ön eğitimleri tüm katmanlarında gerçekleştirecek şekilde tasarlanmıştır. Çift yönlü olması soldan sağa sağdan sola olacak şekilde çalışmasıdır. BERT modeli ön eğitimli bir modeldir. Sadece bir çıktı katmanı ekleyerek birçok farklı görev için state of art modellerin elde edilmesini sağlar. Bu görevlere örnek verecek olursak soru cevaplama, dil tanıma vb. olabilir. 11 doğal dil işleme görevinde BERT modeli kullanılarak state of art sonuçlar elde edilmiştir [170].

Ön eğitim ve ince ayar olmak üzere iki adıma sahiptir. BERT modelinin ön eğitiminde model etiketsiz veriler ile farklı ön eğitimli görevler için eğitilmektedir [170]. Özellikler ön eğitim gerçekleştirilen modelden çıkarılmaktadır [171]. İnce ayar işleminin ön eğitim işlemine göre maliyeti daha düşüktür. Bunun sebebi ön eğitimli parametreler ile ince ayar işleminin başlatılmasıdır. Çalışmanın türüne göre hazırlanan etiketli veri ile parametreler güncellenmektedir. Çıkış katmanları dışında ince ayar ve ön eğitim için kullanılan mimari aynıdır [170]. Türkçeleştirdiğimiz ön eğitim ve ince ayar mimarisi Şekil 5.15'teki gibidir.



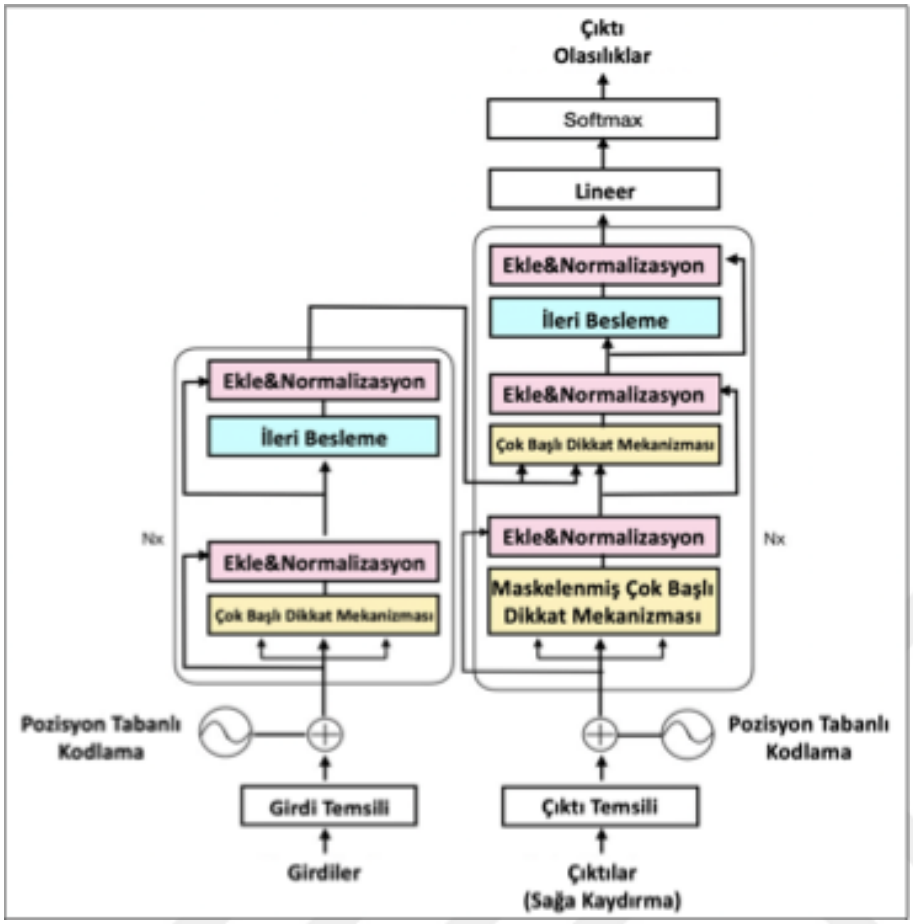
Şekil 5.15. BERT modeli için ön eğitim ve ince ayar mimarisi (Türkçeleştirilmiştir) [170]

BERT modelinin genel ön eğitim ve ince ayar işlemleri MLM ve NSP yöntemlerinin denetimsiz olarak uygulanmasıyla eğitilmektedir.

BERT modelinde giriş tokenı belirtecine [MASK] adı verilir ve derin çift yönlü kelime temsillerini eğitebilmek maksadıyla belirli bir oranda rasgele olacak şekilde maskelenir. Maskelenmiş tokenlar cümleden elde ettikleri bağlam doğrultusunda tahmin edilir. Bu yönteme MLM denir.

Ön eğitim sürecinde cümleler arası ilişkileri anlayabilecek bir model eğitmek maksadıyla model cümle çiftlerini girdi olarak alır. X ve Y cümleleri her ön eğitim örneğinde seçilir ve örneklerin %50'sinde Y cümlesi X cümlesinden sonra gelen cümle, kalan %50'sinde ise veri seti içerisinden rastgele bir cümle olarak belirlenir. Model eğitiminde bu öncelik sonralık durumunun olup olmadığı tahmin edilir.

BERT'nin model mimarisi, çok katmanlı çift yönlü bir Transformer kodlayıcı tabanlı olup orijinal uygulamaya dayandırılmıştır [170]. Transformer ile ortaya konulan basit ağ mimarisi yalnızca dikkat mekanizmalarına dayanmakta olup tekrarlama ve kıvrımlardan tamamen vazgeçilmiştir [170]. Bir dikkat işlevi, bir sorguyu ve bir dizi anahtar-değer çiftini, sorgunun, anahtarların, değerlerin ve çıktının tümünün vektör olduğu bir çıktıya eşlemek olarak tanımlanabilir. Çıktı, değerlerin ağırlıklı bir toplamı olarak hesaplanır; burada her bir değere atanan ağırlık, sorgunun karşılık gelen anahtarla bir uyumluluk fonksiyonu tarafından hesaplanır [170]. Transformer mimarisi Şekil 5.16'da sunulmuştur.



Şekil 5.16. Transformer model mimarisi (Türkçeleştirilmiştir) [170]

BERT modeli Şekil 5.16'da sunulan mimari doğrultusunda girdi metninin sözcükleri arasındaki bağlamsal ilişkileri öğrenen bir dikkat bölümünden oluşur. Girdi metnini okumak ve kelime gösterimlerini oluşturmak için bir kodlayıcı ve sonucu tahmin etmek için bir kod çözücü kullanır [170]. Bu sayede vektörler kelimelerin anlamsal bağlamına uygun üretildiği için dil daha iyi anlaşılmakta ve farklı amaçlarla kullanılmış sesteş kelimeler de ayırt edilebilmektedir [143].

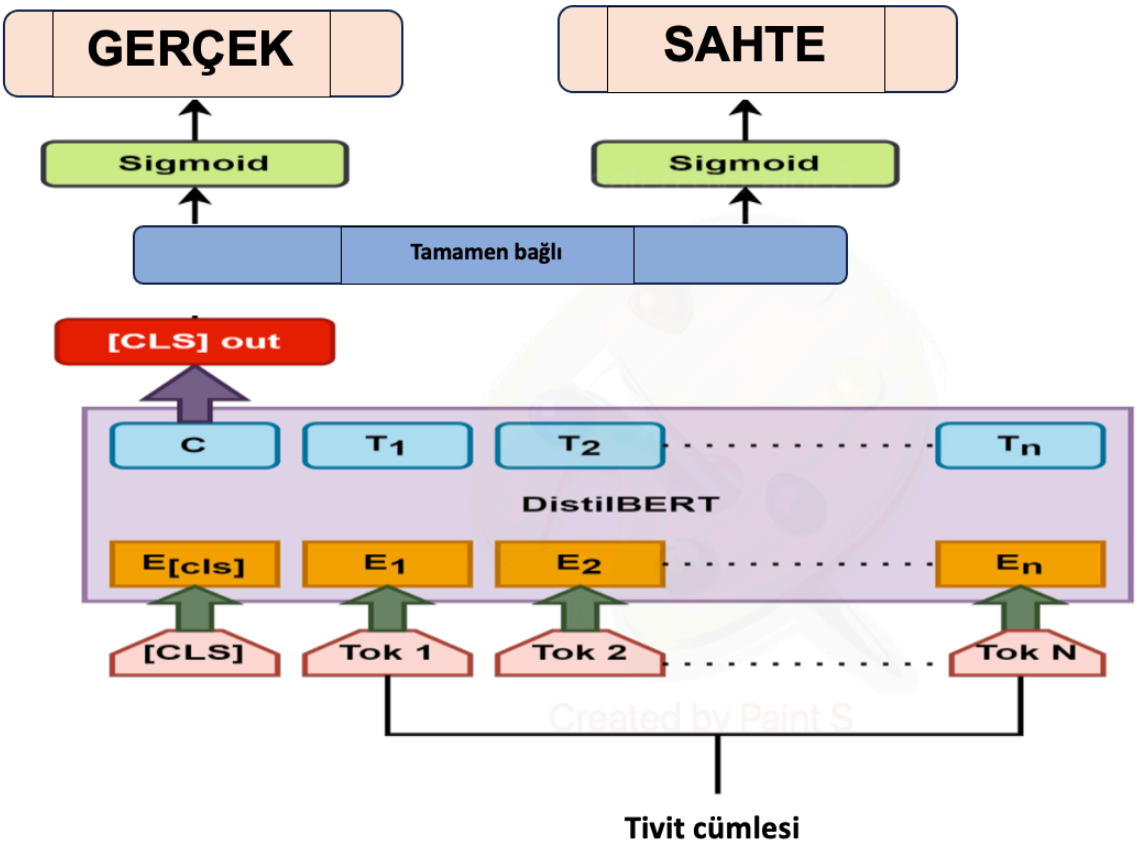
Bu tez kapsamında, eğitim ve test prosedürümüzü anlatmaya başlarken öncelikle BertForSequenceClassification modeli uygulanmıştır. Bu aşamada kullanılan BERT modelleri aşağıdaki gibidir:

- DistilBERT/DistilBERTurk ve
- BERT/BERTurk ön eğitimli modelleri kullanılmıştır.

5.3.1. DistilBERT/DistilBERTurk

DistilBERTurk [172], BERT'inkine benzer temel transformatör mimarisine sahip daha basit, daha ucuz ve daha hafif bir çözüm sunmaktadır. Göreve özgü ince ayar aşamasında damıtma yerine, burada damıtma eğitim öncesi aşamada yapılmaktadır. Katman sayısı yarıya indirilmiştir ve cebirsel işlemler optimize edilmiştir. Bu tür birkaç değişikliği kullanan DistilBERT, BERT'den %40 daha küçük olmasına rağmen rekabetçi sonuçlar sağlamaktadır [172].

Veriler, DistilBERTurk tokenizer kullanılarak tokenize edilmektedir ve ardından Şekil 5.17'de gösterildiği gibi bir dizi token haline getirilir, tensörlere dönüştürülür ve modele sağlanır. DistilBERTClass, bir sinir ağı oluşturmak için kullanılmaktadır. Bu ağ, nihai çıktıları, bir düşüşü ve bir doğrusal katmanı elde etmek için DistilBERT dil modelini kullanmaktadır. Veriler, DistilBERT dil modeline dahil edilmiştir. Modelin performansı için bir taban çizgisi elde etmek üzere [CLS] belirtecinin cümle düzeyinde gömülmesinin üstünde bir sigmoid aktivasyon işlevine sahip tek, yoğun bir çıktı katmanı oluşturulmaktadır. Model performansta birleşene kadar sınıflandırma katmanının rastgele başlatılan ağırlıklarını eğitmeye başlar. Ek sınıflandırma katmanlarını eğittikten sonra, DistilBERT'in gömme katmanını çözülmüş ve modelden daha da fazla performans elde etmek için tüm ağırlıklara daha düşük bir öğrenme oranıyla ince ayar yapılmıştır.



Şekil 5.17. DistilBERT model mimarisi

DistilBERT, BERT'in iyi performans sunan versiyonlarından biridir; bu nedenle performansını diğer BERT modelleri ile analiz etmek için uyguladık. DistilBERT konfigürasyonunu kullanmak için kullanılan parametreler Çizelge 5.20'de gösterilmiştir.

Çizelge 5.20. DistillBERT parametreleri

Parametreler	**Açıklama**	**Kullanılan Değer**
activation	Aktivasyon fonksiyonu	sigmoid
Epochs	Tur sayısı	5
Dim	Kelime vektörlerinin boyutu	768
Hidden_dim	Saklı boyut	3072
Vocab_size	Sözlük boyutu	32000
N_layers	Katman sayısı	6

Kullanılan parametreler doğrultusunda elde edilen sonuçlar doğruluk, kesinlik, geri çağırma ve f1-skor şeklinde hesaplanmıştır. DistilBERTurk doğruluk tablosu Çizelge 5.21.'de sunulmuştur.

Çizelge 5.21. DistilBERTurk doğruluk tablosu

	Kesinlik	Duyarlılık	F1-skor
Sahte	0.89	0.92	0.90
Gerçek	0.91	0.89	0.90
Doğruluk			0.90
Macro avg	0.90	0.90	0.90
Weighted avg	0.90	0.90	0.90

Ayrıca kullanılan DistillBERT modeli, BuzzFeed, GosspCop, ISOT, LIAR, Twitter15, Twitter16 ve GPT-2 veri setlerinde de denenmiştir. Söz konusu veri setlerinde kullanılan alanlar modelimize uygun olacak şekilde güncellenmiş ve deneylerimizde kullanılmıştır. Kendi veri setimizde kullanılan parametreler bahse konu veri setlerinde de kullanılmıştır.

DistillBERT modelimizde BuzzFeed veri setinin sadece haber başlıkları kullanılmış, distillBERTurk kullanım yapısına uygun hale getirilerek çeşitli değerlendirme metrikleri elde edilmiştir. BuzzFeed veri setine ait metrik sonuçları Çizelge 5.22'de sunulmuştur.

Çizelge 5.22. BuzzFeed veri seti değerlendirme metrikleri

	Kesinlik	Geri çağırma	F1-skor
Sahte	0.86	0.72	0.78
Gerçek	0.56	0.75	0.64
Doğruluk			0.73
Macro avg	0.71	0.73	0.71
Weighted avg	0.76	0.73	0.74

Çizelge 5.22'de görüldüğü gibi, BuzzFeed veri setinin haber başlıklarını kullanılan için veri seti küçük bir veri seti haline gelmiştir. Bu durum da hem Türkçe modeli yabancı veri setine uygulamamızdan hem de veri küçüklüğü nedeniyle doğruluk oranını düşürmüştür. BuzzFeed veri setini DistillBERT modelimizin doğruluk oranı %73 olarak elde edilmiştir.

DistillBERT modelimizde gossipCop veri setinin sadece haber metinleri kullanılmış, distillBERT kullanım yapısına uygun hale getirilerek çeşitli değerlendirme metrikleri elde edilmiştir. gossipCop veri setine ait metrik sonuçları Çizelge 5.23.'te sunulmuştur.

Çizelge 5.23. GossipCop veri seti değerlendirme metrikleri

	Kesinlik	Geri çağırma	F1-skor
Sahte	0.75	0.59	0.66
Gerçek	0.88	0.94	0.91
Doğruluk			0.85
Macro avg	0.81	0.76	0.85
Weighted avg	0.85	0.85	0.85

Çizelge 5.23.'te görüldüğü gibi, GossipCop veri setini DistillBERTurk modelimizde denediğimizde doğruluk oranı %85 olarak elde edilmiştir.

DistillBERT modelimizde ISOT veri setinin sadece haber metinleri kullanılmış, distillBERT kullanım yapısına uygun hale getirilerek çeşitli değerlendirme metrikleri elde edilmiştir. ISOT veri setine ait metrik sonuçları Çizelge 5.24'de sunulmuştur.

Çizelge 5.24. ISOT veri seti değerlendirme metrikleri

	Kesinlik	Geri çağırma	F1-skor
Sahte	0.98	0.97	0.97
Gerçek	0.96	0.98	0.97
Doğruluk			0.97
Macro avg	0.97	0.97	0.97
Weighted avg	0.97	0.97	0.97

ISOT veri seti büyük bir veri setidir. Eğitim süreci diğer veri setlerine göre daha uzun sürmüştür. Çizelge 5.24'te görüldüğü gibi, ISOT veri setini DistillBERT modelimizde denediğimizde doğruluk oranı %97 olarak elde edilmiştir.

DistillBERT modelimizde LIAR veri setinin sadece iki etiketi kullanılarak denenmiş, distillBERT kullanım yapısına uygun hale getirilerek çeşitli değerlendirme metrikleri elde edilmiştir. LIAR veri setine ait metrik sonuçları Çizelge 5.25'te sunulmuştur.

Çizelge 5.25. LIAR veri seti değerlendirme metrikleri

	Kesinlik	Geri çağırma	F1-skor
Sahte	0.64	0.52	0.57
Gerçek	0.59	0.70	0.64
Doğruluk			0.61
Macro avg	0.61	0.61	0.61
Weighted avg	0.61	0.61	0.61

LIAR veri setinin tamamı altı etiketlidir. Biz çalışmamızda sadece iki etiketini kullanılan için veri seti küçük bir küme haline gelmiştir. Çizelge 5.25'te görüldüğü gibi, LIAR veri setini DistillBERT modelimizde denediğimizde doğruluk oranı %61 olarak elde edilmiştir. Bu oran küçük görünüyor olsa da literatür araştırmaları tablomuz incelendiğinde LIAR veri seti için iyi bir doğruluk değer elde edilen görülmüştür (Bkz Çizelge 2.2):

DistillBERTmodelimizde Twitter15 veri setinin haber emetinleri kullanarak denenmiş, distillBERT kullanım yapısına uygun hale getirilerek çeşitli değerlendirme metrikleri elde edilmiştir. Twitter15 veri setine ait metrik sonuçları Çizelge 5.26'da sunulmuştur.

Çizelge 5.26. Twitter15 veri seti değerlendirme metrikleri

	Kesinlik	Geri çağırma	F1-skor
Sahte	0.90	0.96	0.93
Gerçek	0.95	0.87	0.91
Doğruluk			0.92
Macro avg	0.92	0.91	0.92
Weighted avg	0.92	0.92	0.92

Çizelge 5.26'da görüldüğü gibi, Twitter15 veri setini DistillBERT modelimizde denediğimizde doğruluk oranı %92 olarak elde edilmiştir.

DistillBERT modelimizde Twitter16 veri setinin haber emetinleri kullanarak denenmiş, distillBERT kullanım yapısına uygun hale getirilerek çeşitli değerlendirme metrikleri elde edilmiştir. Twitter16 veri setine ait metrik sonuçları Çizelge 5.27'de sunulmuştur.

Çizelge 5.27. Twitter16 veri seti değerlendirme metrikleri

	Kesinlik	Geri çağırma	F1-skor
Sahte	0.91	0.88	0.89
Gerçek	0.90	0.93	0.92

Doğruluk			0.91
Macro avg	0.91	0.90	0.90
Weighted avg	0.91	0.91	0.91

Çizelge 5.27'de görüldüğü gibi, Twitter16 veri setini DistillBERTurk modelimizde denediğimizde doğruluk oranı %91 olarak elde edilmiştir.

DistillBERT modelimizde GPT-2 sahte veri setinin haber metinleri kullanarak denenmiş, distillBERT kullanım yapısına uygun hale getirilerek çeşitli değerlendirme metrikleri elde edilmiştir. GPT-2 sahte haber veri setine ait metrik sonuçları Çizelge 5.28'de sunulmuştur.

Çizelge 5.28. GPT-2 sahte haber veri seti değerlendirme metrikleri

	Kesinlik	Geri çağırma	F1-skor
Sahte	0.97	0.82	0.89
Gerçek	0.85	0.98	0.91
Doğruluk			0.90
Macro avg	0.91	0.90	0.90
Weighted avg	0.91	0.90	0.90

Çizelge 5.28'de görüldüğü gibi, GPT-2 veri setini DistillBERTurk modelimizde denediğimizde doğruluk oranı %90 olarak elde edilmiştir.

DistilBERT/DistillBERTurk modelimiz ile elde edilen sonuçların tez kapsamında oluşturulan TR_FaRe_News veri setiyle kıyaslama tablosu Çizelge 5.29.'da sunulmuştur.

Çizelge 5.29. Tüm veri setleriyle DistillBERT karşılaştırma tablosu

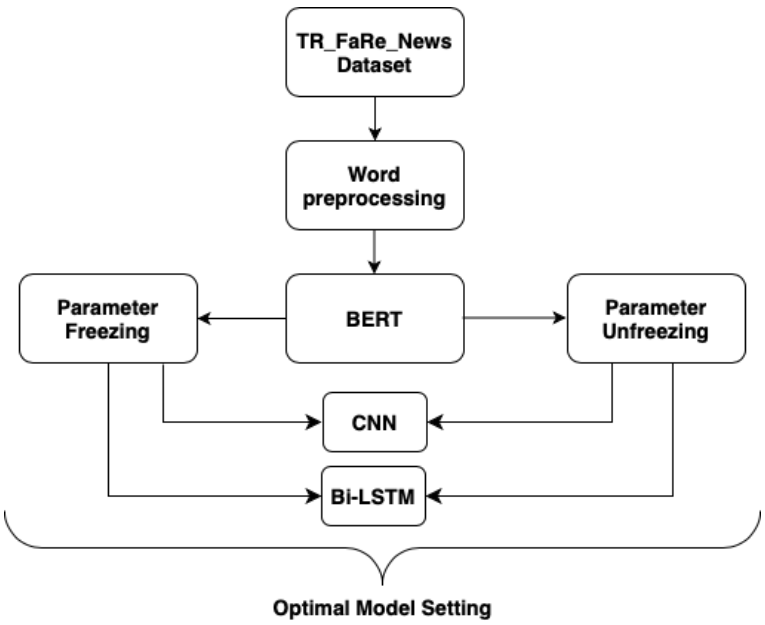
Veri seti	**Doğruluk**
BuzzFeedNews	0.73
GossipCop	0.85
ISOT	**0.97**
LIAR	0.61
Twitter15	**0.92**
Twitter16	**0.91**
GPT-2 veri seti	**0.90**
TR_FaReNews	0.90

Çizgelge 5.29'da görüldüğü gibi, DistilBERT/DistillBERTurk modelimizin tez

kapsamında oluşturulan veri seti TR_FaRe_News ile kıyaslamasını yapacak olursak, ISOT, Twitter15 ve Twitter16 veri kümelerinde TR_FaRe_News veri kümesinden daha iyi doğruluk oranı elde ettiği görülmüştür. Buradaki ana sebebin bizim veri setimizin Twitter'daki kısa tivitlerden oluşması ve Türk diller ait olan dil çözümleme zorlukları olduğunu tahmin edilmektedir. Ancak literatür tablosuna bakacak olursak, TR_FaRe_News veri setiyle de güzel sonuçlar elde edilen görülmektedir (Bkz. Çizelge 2.2)

5.3.2. Derin öğrenme kapsamında BERT

Bu tez kapsamında, eğitim ve test prosedürümüzü anlatmaya başlarken öncelikle BertForSequenceClassification modeli uygulanmıştır [173]. İkinci olarak, Bert modeline ince ayar yapılmıştır. Üçüncüsü, ince ayarlı modeldeki parametreleri hem dondurarak hem de dondurmadan CNN ve Çift yönlü LSTM dahil olmak üzere ince ayarlı modelden sonra ek katmanlar eklenmiştir. Ardından, eğitim, hiper parametre ayarı ve test işlemini gerçekleştirdik. Ana modelimiz olan iyileştirilmiş BERT/BERTurk modelinin model prosedürü Şekil 5.18'deki sunulmuştur.



Şekil 5.18. Model prosedürü

Sahte haberlerin sınıflandırmasının performansını değerlendirmek ve karşılaştırmak için hem BERTurk hem de Turkish BERT-base için oluşturulan beş farklı model (toplam 10

model) oluşturulmuştur. Modelin ilki BERT ince ayarlı model olarak tanımlanmaktadır. İkinci model donmuş parametrelerle CNN katmanları ile BERT ince ayarlı modeli olarak tanımlanmaktadır. Üçüncü model donmuş parametreler olmadan CNN katmanları ile BERT ince ayarlı modeli olarak tanımlanmaktadır. Ayrıca, dördüncü model donmuş parametrelerle BiLSTM katmanları ile BERT ince ayarlı modeli olarak tanımlanmaktadır. Son olarak, beşinci model donmuş parametreler olmasan BiLSTM katmanları ile BERT ince ayarlı modeli olarak tanımlanmaktadır.

BERT ince ayarı (Model 1)

BERT ince ayarı diye tanımladığımız Model 1 için, öğrenme oranı olarak 2e-5 kullanılmıştır. Modelde ince ayar yapmak maksadıyla 4 tur kullanılmıştır. BERT'in ince ayar mimarisi Şekil 5.15'te sunulmuştur.

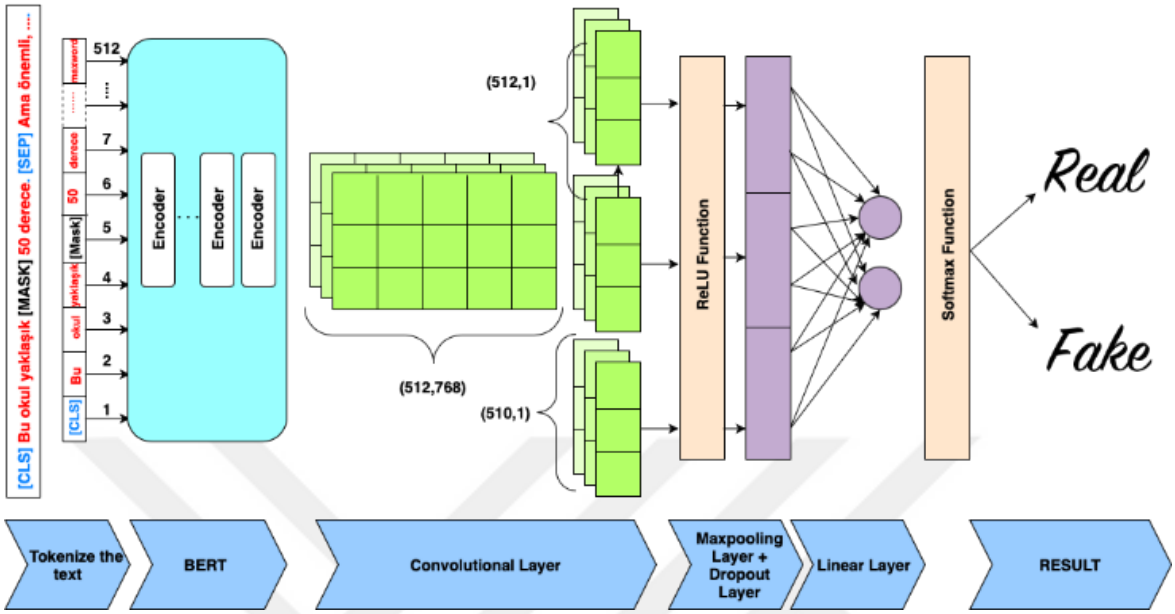
BERT ince ayarı + CNN (Model 2 ve Model 3)

Model 2 ve Model 3 için, BERT modelinden sonra, çekirdek boyutu (1,768) ve (2,768) olan ve ReLU aktivasyon fonksiyonuna sahip iki tane evrişimli katman uygulanmıştır. Daha sonra, çekirdek boyutu olarak önceki çıktı boyutuna ve adım olarak çıktının önceki yüksekliğine sahip bir maksimum havuzlama katmanı takip edilmektedir. Son olarak, softmax aktivasyon fonksiyonu ile lineer bir katman uygulanmıştır. Softmax aktivasyon fonksiyonu şu şekilde ifade edilmektedir:

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$$

Burada K=2'dir çünkü sahte ve gerçek haberlerden oluşan iki tane sınıf mevcuttur. Z değişkeni girişi göstermektedir. Sınıflandırma sonucu, softmax aktivasyon fonksiyonunun en yüksek çıktısına sahip sınıftır. Ayrıca bu iki model için öğrenme oranı BERT'in ince

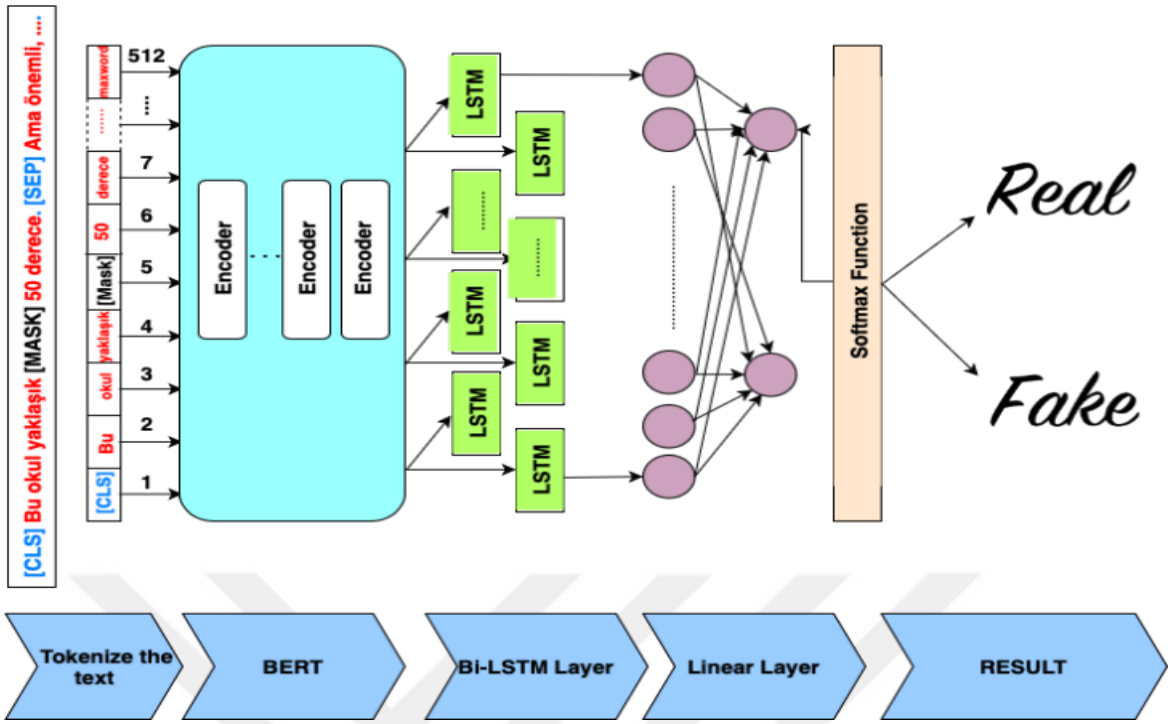
ayarında olduğu gibi 2e-5 ve tur sayısı 4'tür. BERT+CNN model mimarisi Şekil 5.19'da gösterilmiştir.



Şekil 5.19. BERT+CNN Model mimarisi

BERT ince ayarı + BiLSTM (Model 4 ve Model 5)

Model 4 ve Model 5 için BERT modelinden sonra Model 4'e 2 BiLSTM katmanı ve Model 5'e 1 BiLSTM katmanı uygulanmmıştır. Bundan sonra softmax aktivasyon fonksiyonu ile lineer bir katman uygulanmıştır. Ayrıca Model 4 ve Model 5 için öğrenme oranı 5e-5, Model 4 için kullanılan tur sayısı 10 ve model 5 için kullanılan tur sayısı ise 6'dır. BERT + BiLSTM model mimarisi Şekil 5.20.'de gösterilmiştir.



Şekil 5.20. BERT+BiLSTM model mimarisi

Kullanılan sınıflandırıcıların tahmin sonuçlarını sahte haber veri setinde test etmek maksadıyla çeşitli ölçümler kullanılmıştır. İlk olarak, birinci ölçüm olaran test doğruluğunu hesaplamaktır. Çalışmada test doğruluğu için (1)'de gösterilen hesap kullanılmaktadır.

$$\frac{doğru\ sınıflandırılmış\ haber\ sayısı}{toplam\ haber\ sayısı} \qquad (1)$$

Kullanılan ikinci metrik ise ROC AUC puanıdır. ROC AUC, ROC eğrisinin altındaki alanı ifade etmektedir. ROC AUC puanı 0 ile 1 arasında bir değerdir ve bir model için 1'e yakın bir AUC değeri, sınıflandırma tahminini iyi bir performans gösterdiğini anlatmaktadır. Kullanılan son metrik F1 puanıdır. F1 puanı da AUC değeri gibi 0 ile 1 arasında değişmektedir ve kesinlik sonuçlarının hatırlanmasıyla hesaplanmaktadır. Kesinlik, gerçek pozitif sonuçların sayısının tüm pozitif sonuçların sayısına bölümüdür ve geri çağırma, gerçek pozitif sonuçların sayısının pozitif olarak tanımlanması gereken tüm örneklerin sayısına bölünmesiyle elde edilmektedir. Puan ne kadar

yüksek olursa, o kadar iyi performans göstermektedir. F1 puanının formülü (2)'de gösterildiği gibidir.

$$F1\ puanı = 2 \cdot \frac{kesinlik \cdot geri\ çağırma}{kesinlik + geri\ çağırma} \quad (2)$$

5.3.3. BERT/BERTurk

Türk BERT modeli (BERTurk)[174], Oscar Corpus, Opus Corpora ve Wikipedia üzerinde önceden eğitilmiştir. Model 12 dödüştürücü katmandan oluşmaktadır. BERTurk modelleri 32K ve 128K kelime boyutunda farklılık göstermektedir ve her ikisinin de kasalı ve kasasız versiyonları mevcuttur.

Sahte haber sınıflandırmasının performans değerlendirmesini yapmak ve karşılaştırmak için BERTurk kullanılan 5 model oluşturulmuştur. Bu modellerin hiper parametre tablosu Çizelge 5.30'da gösterilmiştir.

Çizelge 5.30. BERTurk modeli için kullanılan hiper parametreler

Model	Öğrenme Oranı	Tur	Eklenen Katman Sayısı	Optimizasyon Algoritması
İnce ayarlı BERT	2e-5	4	1	AdamW
İnce ayarlı BERT+CNN (Donmuş parametreli)	5e-5	4	2	AdamW
İnce ayarlı BERT+CNN (Donmamuş parametreli)	5e-5	4	2	AdamW
İnce ayarlı BERT+BiLSTM(Donmuş parametreli)	5e-5	10	2	AdamW
İnce ayarlı BERT+BiLSTM(Donmamuş parametreli)	5e-5	6	1	AdamW

Oluşturulan modelleri Türkçe haber tweetlerinden oluşan TR_FaReNews veri setimizde eğittikten sonra canlı veri olarak adlandırdığımız veri setimiz içerisinde olmayan tweet cümleleriyle oluşturulan validation (doğrulama) veri seti ile denediğimizde çeşitli sonuçlar elde ettik. Validation veri seti Twitter'da yer alan parodi hesapları ve ana akım twitter hesaplarının güncel tweetlerinden oluşmuştur. BERTurk sonuçları sırasıyla Çizelge 5.31'de sunulmuştur.

Çizelge 5.31. BERTurk kullanılarak elde edilen sonuçlar

Model	Eğitim doğruluğu	Validation doğruluğu	ROC AUC	F1-ölçütü
İnce ayarlı BERTurk	0.98	0.93	**0.94**	0.93
İnce ayarlı BERTurk+CNN (Donmamış parametreli)	0.97	0.91	**0.94**	0.91
İnce ayarlı BERTurk+CNN (Donmuş parametreli)	0.81	0.78	0.91	0.81
İnce ayarlı BERTurk+BiLSTM(Donmamış parametreli)	0.84	0.78	0.91	0.82
İnce ayarlı BERTurk+BiLSTM(Donmuş parametreli)	0.84	0.79	0.91	0.82

Ayrıca ana modelimiz olarak adlandırdığımız BERT modeli, BuzzFeed, GosspCop, ISOT, LIAR, Twitter15, Twitter16 ve GPT-2 veri setlerinde de denenmiştir. Söz konusu veri setlerinde kullanılan alanlar modelimize uygun olacak şekilde güncellenmiş ve deneylerimizde kullanılmıştır. Kendi veri setimizde kullanılan parametreler bahse konu veri setlerinde de kullanılmıştır.

BERT modelimizde BuzzFeed veri setinin sadece haber başlıkları kullanılmış, BERT kullanım yapısına uygun hale getirilerek çeşitli değerlendirme metrikleri elde edilmiştir. BuzzFeed veri setine ait metrik sonuçları Çizelge 5.32'de sunulmuştur.

Çizelge 5.32. BuzzFeed veri seti ile elde edilen sonuçlar

Model	Eğitim doğruluğu	Validation doğruluğu	ROC AUC	F1-ölçütü
İnce ayarlı BERT	64	84	70	90
İnce ayarlı BERT+CNN (Donmamış parametreli)	99	79	86	83
İnce ayarlı BERT+CNN (Donmuş parametreli)	95	79	86	83
İnce ayarlı BERT+BiLSTM(Donmamış parametreli)	99	79	86	83
İnce ayarlı BERT+BiLSTM(Donmuş	98	79	86	83

Model	Eğitim doğruluğu	Validation doğruluğu	ROC AUC	F1-ölçütü
parametreli)				

Çizelge 5.32'de görüldüğü gibi, BuzzFeed veri setini BERT modeline CNN katmanı ekleyerek donmuş/donmamış parametreli ve BiLSTM katmanı ekleyerek donmuş/donmamış parametreli denediğimiz modelimizin doğruluk oranı %86 olarak elde edilmiştir.

BERT modelimizde gossipCop veri setinin sadece haber metinleri kullanılmış, BERT kullanım yapısına uygun hale getirilerek çeşitli değerlendirme metrikleri elde edilmiştir. GossipCop veri setine ait metrik sonuçları Çizelge 5.33'da sunulmuştur.

Çizelge 5.33. GossipCop veri seti ile elde edilen sonuçlar

Model	Eğitim doğruluğu	Validation doğruluğu	ROC AUC	F1-ölçütü
İnce ayarlı BERT	74	75	75	77
İnce ayarlı BERT+CNN (Donmamış parametreli)	70	74	74	75
İnce ayarlı BERT+CNN (Donmuş parametreli)	51	51	50	68
İnce ayarlı BERT+BiLSTM(Donmamış parametreli)	51	51	50	68
İnce ayarlı BERT+BiLSTM(Donmuş parametreli)	52	50	51	89

Çizelge 5.33'de görüldüğü gibi, GossipCop veri setini ince ayarlı BERT denediğimiz modelimizin doğruluk oranı %75 olarak elde edilmiştir.

BERT modelimizde ISOT veri setinin sadece haber metinleri kullanılmış, BERT kullanım yapısına uygun hale getirilerek çeşitli değerlendirme metrikleri elde edilmiştir. ISOT veri setine ait metrik sonuçları Çizelge 5.34'ta sunulmuştur.

Çizelge 5.34. ISOT veri seti ile elde edilen sonuçlar

Model	Eğitim	Validation	ROC AUC	F1-ölçütü

	doğruluğu	doğruluğu		
İnce ayarlı BERT	95	94	94	94
İnce ayarlı BERT+CNN (Donmamış parametreli)	84	90	91	91
İnce ayarlı BERT+CNN (Donmuş parametreli)	54	55	50	71
İnce ayarlı BERT+BiLSTM(Donmamış parametreli)	55	55	50	71
İnce ayarlı BERT+BiLSTM(Donmuş parametreli)	55	55	50	71

Çizelge 5.34'te görüldüğü gibi, GossipCop veri setini BERT modeline ince ayar yapılarak denediğimiz modelimizin doğruluk oranı %94 olarak elde edilmiştir.

BERT modelimizde LIAR veri setinin sadece haber metinleri kullanılmış, BERT kullanım yapısına uygun hale getirilerek çeşitli değerlendirme metrikleri elde edilmiştir. LIAR veri setine ait metrik sonuçları Çizelge 5.35'de sunulmuştur.

Çizelge 5.35. LIAR veri seti ile elde edilen sonuçlar

Model	Eğitim doğruluğu	Validation doğruluğu	ROC AUC	F1-ölçütü
İnce ayarlı BERT	61	60	58	42
İnce ayarlı BERT+CNN (Donmamış parametreli)	92	60	58	47
İnce ayarlı BERT+CNN (Donmuş parametreli)	94	59	59	57
İnce ayarlı BERT+BiLSTM (Donmamış parametreli)	92	60	60	58
İnce ayarlı BERT+BiLSTM (Donmuş parametreli)	93	59	60	58

Çizelge 5.35'de görüldüğü gibi, LIAR veri setini BERT modeline BiLSTM katmanı ekleyerek donmamış ve donmuş parametreli denediğimiz modelimizin doğruluk oranı %60 olarak elde edilmiştir.

BERT modelimizde Twitter15 veri setinin sadece haber metinleri kullanılmış, BERT kullanım yapısına uygun hale getirilerek çeşitli değerlendirme metrikleri elde edilmiştir. Twitter15 veri setine ait metrik sonuçları Çizelge 5.36'da sunulmuştur.

Çizelge 5.36. Twitter15 veri seti ile elde edilen sonuçlar

Model	Eğitim doğruluğu	Validation doğruluğu	ROC AUC	F1-ölçütü
İnce ayarlı BERT	95	89	89	90
İnce ayarlı BERT+CNN (Donmamış parametreli)	97	88	88	89
İnce ayarlı BERT+CNN (Donmuş parametreli)	98	85	84	87
İnce ayarlı BERT+BiLSTM(Donmamış parametreli)	98	88	88	89
İnce ayarlı BERT+BiLSTM(Donmuş parametreli)	98	88	88	89

Çizelge 5.36'da görüldüğü gibi, Twitter15 veri setini BERT modeline her türlü denediğimizde küsüratlı dahi de olsa modelimizin doğruluk oranı %89 olarak elde edilmiştir.

BERT modelimizde Twitter16 veri setinin sadece haber metinleri kullanılmış, BERT kullanım yapısına uygun hale getirilerek çeşitli değerlendirme metrikleri elde edilmiştir. Twitter16 veri setine ait metrik sonuçları Çizelge 5.37'de sunulmuştur.

Çizelge 5.37. Twitter16 veri seti ile elde edilen sonuçlar

Model	Eğitim doğruluğu	Validation doğruluğu	ROC AUC	F1-ölçütü
İnce ayarlı BERT	80	95	95	94
İnce ayarlı BERT+CNN (Donmamış parametreli)	99	97	97	97
İnce ayarlı BERT+CNN (Donmuş parametreli)	100	97	97	97
İnce ayarlı BERT+BiLSTM(Donmamış parametreli)	99	97	97	97
İnce ayarlı	99	97	97	97

BERT+BiLSTM(Donmuş parametreli)				

Çizelge 5.37'te görüldüğü gibi, Twitter16 veri setini BERT modeline her türlü denediğimizde küsüratlı dahi de olsa modelimizin doğruluk oranı %97 olarak elde edilmiştir.

BERT modelimizde GPT-2 veri setinin sadece haber metinleri kullanılmış, BERT kullanım yapısına uygun hale getirilerek çeşitli değerlendirme metrikleri elde edilmiştir. Twitter16 veri setine ait metrik sonuçları Çizelge 5.38'de sunulmuştur.

Çizelge 5.38. GPT-2 veri seti ile elde edilen sonuçlar

Model	Eğitim doğruluğu	Validation doğruluğu	ROC AUC	F1-ölçütü
İnce ayarlı BERT	0.99	0.94	0.93	0.93
İnce ayarlı BERT+CNN (Donmamış parametreli)	0.97	0.94	0.94	0.94
İnce ayarlı BERT+CNN (Donmuş parametreli)	0.98	0.93	0.93	0.92
İnce ayarlı BERT+BiLSTM(Donmamış parametreli)	0.98	0.95	0.95	0.94
İnce ayarlı BERT+BiLSTM(Donmuş parametreli)	0.98	0.95	0.95	0.94

Çizelge 5.38'de görüldüğü gibi, GPT-2 veri setini BERT modeline her türlü denediğimizde küsüratlı dahi de olsa modelimizin doğruluk oranı %95 olarak elde edilmiştir.

BERT modelimiz ile elde edilen sonuçların tez kapsamında oluşturulan TR_FaRe_News veri setiyle kıyaslama tablosu Çizelge 5.39'da sunulmuştur.

Çizelge 5.39. Tüm veri setleriyle BERT karşılaştırma tablosu

Veri seti	**Doğruluk**
BuzzFeedNews	0.86
GossipCop	0.75
ISOT	**0.94**

LIAR	0.60
Twitter15	0.89
Twitter16	**0.97**
GPT-2 veri seti	**0.95**
TR_FaReNews	**0.94**

Çizelge 5.39'da görüldüğü gibi, BERTurk modelimizin tez kapsamında oluşturulan veri seti TR_FaRe_News ile kıyaslamasını yapacak olursak, ISOT veri kümesinde ufak bir farkla TR_FaRe_News veri kümesinden daha iyi doğruluk oranı elde ettiği görülmüştür. Buradaki ana sebebin bizim veri setimizin Twitter'daki kısa tivitlerden oluşması ve Türk diler ait olan dil çözümleme zorlukları olduğunu söyleyebiliriz. Ancak literatür tablosuna bakacak olursak, TR_FaRe_News veri setiyle de güzel sonuçlar elde edilen görülmektedir (Bkz. Çizelge 2.2)

Ayrıca tez kapsamında kullanılan 2 adet BERTurk modelinin kendi oluşturulan TR_FaRe_News veri seti ile kıyaslama tablosu Çizelge 5.40'da sunulmuştur.

Çizelge 5.40. TR_FaRe_News verisetiyle çalışan BERTurk modellerinin kıyaslanması

Model	**Doğruluk**
distillBERTurk	0.91
İnce ayarlı BERTurk	0.94
İnce ayarlı BERTurk +CNN (Donmamış parametreli)	**0.94**
İnce ayarlı BERTurk +CNN (Donmuş parametreli)	0.91
İnce ayarlı BERTurk +BiLSTM(Donmamış parametreli)	0.91
İnce ayarlı BERTurk +BiLSTM(Donmuş parametreli)	0.91

Yapılan tez çalışması neticesinde kullanılan BERT modellerinin kıyaslama tablosu olan Çizelge 5.40'a bakıldığından en iyi sonuç veren modelin ince ayarlı BERTurk modeline CNN katmanı eklenmiş hali olup elde edilen sonuç %94'tür.

Çizelge 5.40'taki sonuçlara ulaşmamızı kolaylaştırmak için bir dizi işlem gerçekleştirilmiştir. ReLU aktivasyon fonksiyonu düşük hesaplama maliyeti nedeniyle kullanılırken, softmax aktivasyonu 2 sınıf ve 2 düğüm oluşturdu. Büyük boyutları azaltmak ve öğrenme oranını iyileştirmek için maksimum havuzlama ve bırakma kullanılmıştır. Maksimum havuzlama, boyutunu azaltmak için bir matristeki en büyük değeri seçmenin zaman alıcı sürecini en aza indirmek için kullanılmıştır. Dropout, verilerin ezberlenmesini

önlemek için kullanılırken, AdamW optimizasyon fonksiyonu optimizasyon, öğrenme oranını artırmak ve bir zamanlayıcı oluşturmak için kullanılmıştır.

6. TARTIŞMA

Tez çalışması kapsamında oluşturulan TR_FaRe_News veri kümesinde yer alan sahte ve gerçek haberlerin kelime bulutları çıkarılmıştır. Veri kümesindeki en yaygın ilk 50 kelime için kelime bulutları oluşturulmuştur. Kelime bulutları Şekil 6.1 ve Şekil 6.2'de gösterilmiştir.

Şekil 6.1. Gerçek haber veri seti kelime bulutu

Şekil 6.2. Sahte haber veri seti kelime bulutu

Derin öğrenme kullanılarak Türkçe sahte haber tespiti ve sınıflandırması için geliştirilen BERT modellerinin etkinliğini değerlendirmek amacıyla, sahte haber tespitine odaklanan Türkçe çalışmalar arasında bir karşılaştırma yapılmıştır. Karşılaştırma tablosu Çizelge 5.41'de sunulmuştur.

Çizelge 5.41. Sahte haber tespiti yapan Türkçe çalışmaların karşılaştırılması

Modeller	Veri seti	Görev	Performans
SVM, NB[46]	Özel veri seti	TF-IDF, n-gram, stil işaretleyici, argo kullanımı, link, haberin erişilebilir link özelliği, haber içeriği ve başlığı kullanarak sahte haber tespiti.	f1-score %71-79
NB, KNN, ExtraTrees, SVM, LR, RF, DT [51]	TR_FN	Kök sayımı, kök sayısı, kelime hatası, haber okunabilirliği, kaynak, haber kategorisi be haber adresi özellikleriyle sahte haber tespiti	%89-96
SVM [52]	Özel veri seti	TF-IDF and word2vec temsil yöntemiyle sahte haber tespiti	F1-score %57-89
CNN, RNN-LSTM [79]	BuzzFeed, ISOT, SOSYalan	word2vec ve her kelime için vektör temsili ile sahte haber tespiti	%87.14-92.48
LSTM, BiLSTM, GRU, Bi-GRU, BERT, RoBERTa, DistilBERT, BERTurk [86]	Twitter, doğruluk kontrol platformları Twitter hesapları, COVID-19 veriseti	Bilgi kazancı, Kazanma oranı ve korelasyon tabanlı özellikler ile sahte haber tespiti	%89-98.5
K-NN, SVM, RF, K-means, NMF, LDA [85]	Özel veri seti	TF-IDF, Word2Vec, Doc2Vec temsil yöntemleriyle sahte haber tespiti	F1-score %72-86

Modeller	Veri seti	Görev	Performans
MNB, RF, LR, SVM, Voting Classifier, DistillBERTurk, BERTurk	GPT-2 dataset	BoW, TF-IDF, word2vec temsil yöntemleri ve ince ayarlı BERT ile sahte haber tespiti	%62.7-95
MNB, RF, LR, SVM, Voting Classifier, DistillBERTurk, BERTurk	TR_FaReNews	BoW, TF-IDF, word2vec temsil yöntemleri ve ince ayarlı BERT ile sahte haber tespiti	%71-94

Çizelge 5.41'i inceledğimizde, denetimli makine öğrenmesi algoritmaları arasında SVM modellerinin Türkçe sahte haber tespiti için daha iyi performans gösterdiğini gözlemlenmiştir. [46] ve [51] incelendiğinde word2vec kelime yerleştirme işlemi yapıldığında oluşturduğumuz modelin performansının daha yüksek olduğu gözlemlenmiştir. Ayrıca makine öğrenmesi modelimizde sunduğumuz topluluk öğrenme bakış açısı eklendiğinde oylama sınıflandırıcı ile bu performans daha da artmış ve TR_FaRe_News veri kümesi ile eğitildiğinde %89, GPT-2 sahte haber veri kümesi ile eğitildiğinde ise %77,6'lık bir performans elde edilmiştir.

Henüz literastürle paylaşılmamış ancak kendi çalışmalarında özel bir veri kümesi oluşturan bu çalışma üç konu başlığına dayalı tivitleri içermektedir [52]. Yazarlar performansı değerlendirirken f1-skor sonuçlarının verdikleri için doğruluk açısından bir karşılaştırma yapmak uygun olmayacaktır. Ancak f1-skor sonuçlarını karşılaştırdığımızda söz konusu çalışmada sunulan modellerin skor sonuçlarının %57 ile %89 arasında olduğu ve bu sonuçların bizim çalışmamızdaki BERT modellerinin performans değerlerinin altında kaldığı görülmektedir. Çalışmamızda beş BERT modelinin tamamının f1 skoru %90'ın üzerindedir.

Ayrıca, Türkçe sahte haberlerin tespitine yönelik literatürde dikkat çeken bir çalışma olan SOSYalan veri kümesinde başarılı sonuçlar elde edilmiş ancak veri kümesi büyüklüğü açısından değerlendirildiğinde küçük bir veri kümesi olduğu için sınıflandırma başarısının yüksek olabileceği tahmin edilmiştir [79]. BERT modelleri çalışmalarına dahil edilmemiştir.

Veri setindeki veriler açısından COVID-19 konulu tivitlerin seçilmesiyle oluşturulan çalışma bizim çalışmamıza benzer görünmektedir. Çalışmamız tek bir konu içermediği için performansta hafif bir düşüş olsa da literatür çalışmaları arasında iyi bir orana sahiptir [86]. Ayrıca yaptıkları çalışmada [85] literatürdeki önemli veri kümeleri ile bir karşılaştırma tablosuna yer verilmediği görülmüştür.

Tez çalışmamızda kullandığımız büyük bir dil modeli olan GPT-2 tarafından üretilen sahte haber veri kümesi için de deneyler yapılmıştır. Deneyler sonucunda %62,7 ile %95 arasında bir doğruluk elde edilmiştir. Bu kullanım Türkçe literatürde hiçbir çalışmada karşımıza çıkmamaktadır. Ayrıca TR_FaRe_News isimli veri kümemiz kullanılarak elde edilen performans %71 ile %94 arasında gerçekleşmiştir. Bu performans, hazırladığımız veri kümesi düşünüldüğünde başarılı olarak değerlendirilmektedir.

8. SONUÇ ve ÖNERİLER

Bu tez çalışmasında; sosyal ağ platformlarından biri olan Twitter'da yapılan paylaşımların haber doğruluk platformları yardımıyla tespiti, sınıflandırması ve sınıflandırmaya yönelik modeller oluşturularak gerçek zamanlı tivit verisi ile başarısı test edilmiştir.

Yapılan çalışmalarda; öncelikle, sahte haber kavramı ve sahte haber tespitine yönelik bir literatür taraması yapılmış Twitter'daki sahte haber kavramının dezenformasyon, aldatmaca ve yanlış bilgilendirme arasındaki bağlantı ayrıntılı olacak şekilde ele alınmıştır. Daha sonra, uygulamalarında, test ve araştırmalarında sahte haber tespiti kapsamında kullanılan veri setleri detaylı olarak incelenmiş ve tez çalışması kapsamında kullanılan yöntem, metot, veri kaynakları gibi hususlara yol gösterici olmuştur. Bu bağlamda;

- Tez çalışması için belirlediğimiz metodolojiye uygun olacak şekilde öncelikle Türkçe veri setlerinin oluşturulmasına yönelik işlemler gerçekleştirilmiş, Twitter ortamında elde edilen veriler toplam 13736 adet tivit eğitim, 3433 test, 1526 adet tivit doğrulama için olmak üzere toplam 18695 adet tivit ve 2 etiketten oluşmuştur. Sahte haber tespiti için Türkçe diline ait hazırlanan veri setleri arasında yerini almış ve literatüre kazandırılmıştır. Bundan sonraki çalışmalara kaynaklık edecek olan bu veri setinin farklı dillerde gerçekleştirilen çalışmalar da düşünüldüğünde sahte haber tespitine yönelik hazırlanmış kapsamlı düzeyde denilecek bir veri seti olduğu düşünülmektedir.
- Veri kaynakları üzerinden verilerin toplanması ve etiketlenmesi aşamasını müteakip DDİ süreçleri için Zemberek kütüphanesi kullanılarak veri ön işleme adımları yapılmış ve kökler anlamsal olacak düzeye getirilmiştir. Makine öğrenmesi ve fastText ile gerçekleştirdiğimiz model için uygun hale getirilmiştir.

Sahte ve gerçek haberlerden oluşan TR_FaRe_News adını verdiğimiz veri setimizi kullanarak, sahte haber içeren tivitlerin tespitine yönelik makine öğrenmesi ve derin öğrenme çalışmaları yapılmıştır. Derin öğrenme kapsamında BERT yaklaşımları kullanılarak 2 farklı sınıf için DistillBERTurk dahil olmak üzere toplam 6 model

oluşturulmuştur. Bu modellerde %90-94 arası doğruluk değerleri elde edilmiştir. Diğer veri setleri üzerinde denenen modellerde de çeşitli farklılıklar mevcuttur. Oluşturulan modelin Türk diline göre oluşturulmuş olması ve bu aşamada böyle bir sonuç alınması Türk dili için geliştirilen BERT modellerinin kullanım biçimlerindeki farklılıktır.

Twitter'dan elde edilen veriler kısa metin halindedir. Kısa metinlerin sınıflandırılmasının da çeşitli dezavantajları mevcuttur. Kelime sayısının az olması ve frekanslarının düşük olması dahil metindeki konuların hızlı bir şekilde değişmesi sınıflandırma performanslarını negatif yönde etkilemektedir. Belirtilen dezavantajlara ek olarak, metin içerisindeki özel ifadeler, kısaltmalar, imla hataları, doğal dil işlemedeki kayıplar göz önünde bulundurulduğunda, çalışma kapsamında elde edilen doğruluk oranlarının yüksek bir performans olarak düşünülmektedir.

Sosyal ağlardaki sahte haberleri tespit etmek maksadıyla gerçekleştirilen bu çalışmada, Twitter'dan toplanan veriler ön işlemeden geçirilerek (veri temizleme, normalleştirme, veri etiketleme, durak kelimelerinin silinmesi) Türkçe bir veri seti oluşturulmuştur. Veri seti doğruluk kontrol platformlarının Twitter hesaplarından ve güvenilir haber ajanslarının hesaplarından yayınlanan tweetlerden elde edilmiştir. Veriseti TR_FaRe_News adı verilmiştir. Oluşturulan set, testlerin güvenilir olması bakımından tabakalı örnekleme yöntemi ile 10 eşit parçaya bölünmüştür. Çalışmada makine öğrenmesi algoritmaları ve fastText kütüphanesi kullanılarak bir dil modeli geliştirilmiş ve modelde en iyi sonuç fastText ile yapılan bölümde alınmıştır. Yapılan çalışma sonucunda en iyi performans %93 olarak ortaya çıkmıştır. Ancak burada en iyi fastText modeline ulaşmak için yaklaşık 3 gün gibi bir sürede modelin çalıştırıldığı unutulmamalıdır. Derin öğrenme kapsamında yapılan çalışma neticesinde kullanılan BERT modelleri içerisinde ise en iyi sonuç BERTurk+CNN modeline ait ve doğruluğu %94'tür.

Literatürde Türkçe sahte haber tespiti için referans alınabilecek çalışmalar mevcuttur ancak veri seti açısından ulaşılabilir bir veri seti bulunmamaktadır. Türkçe sahte haber tespiti için yapılan sınıflandırma çalışmasında elde edilen değerlerden yüksek bir performans elde edilmiştir. Sahte haber tespiti üzerinde İngilizce dili için yapılan çalışmalardan bizim çalışmamızla aynı sonucu elde edilen çalışmalar mevcuttur. Bu bakımdan çalışmada oluşturulan dil modelinin başarılı olduğu düşünülmektedir. Çalışmamızda, bir model oluşturularak sahte haberlerin tespit edilmesi için literatürde yaygın olarak kullanılmayan

GPT-2 sahte haber veri kümesi kullanılmıştır. Bu çalışmada geliştirilen dil modellerinin bu veri kümesinde de deneylerinin yapılarak yüksek performanslı sonuçlar elde etmesi hem özgün hem başarılı olduğunu göstermektedir. Büyük bir dil modelinin ürettiği verileri kullanarak kendi oluşturudğumuz modellerde kullanmak herhangi bir Türkçe çalışmada daha önce yapılmamıştır.

Çalışmamız kapsamında geliştirilen TR_FaRe_News veri setinin ileride çeşitli modellerde kullanılması planlanmaktadır. Gelecek çalışmalarda ayrıca GRU modelleri ve diğer çeşitli üretken modeller kullanılarak deneyler yapılacaktır. Geliştiridğimiz modeller GPT-2'nun ötesinde daha büyük dil modelleri ile test edilecektir. Çalışmamızın Türkiye'de gelecekte yapılacak araştırmalar için değerli bir kaynak teşkil etmesi beklenmektedir. TR_FaRe_News veri kümesi, sınıflandırma modelleri ve bulguları ile Türkçe Sahte Haber Tespiti literatürüne önemli bir katkı sağlayacağı düşünülmektedir. Bu veri kümesi, akademik araştırmalar için bir temel olarak kullanılabilecektir.

KAYNAKLAR

1. Taşkıran, İ. A. (2018). Dijital Yerli Gazetelerin Sosyal Medya Stratejileri ve Sosyal Medyanın Haber Okunurluğuna Etkisi. *Akdeniz Üniversitesi İletişim Fakültesi Dergisi*, (30), 218-240.

2. Beckett, C. (2011). *SuperMedia: Saving journalism so it can save the world*. John Wiley & Sons.

3. İnternet: 25 İÇİN 2023+ SOSYAL MEDYA İSTATİSTİKLERİ, GERÇEKLER VE TRENDLER. URL: https://www.websiterating.com/tr/research/social-media-statistics-facts/#sources , Son Erişim Tarihi: 16.06.2023.

4. İnternet: Top 15 Most Popular News Websites | June 2023 URL: http://www.ebizmba.com/articles/news-websites, Son Erişim Tarihi: 16.06.2023.

5. İnternet: We Are Social Dijital 2023 Global ve Türkiye Raporu Yayınlandı! URL: https://omgiletisim.com/we-are-social-dijital-2023-global-ve-turkiye-raporu-yayinlandi/, Son Erişim Tarihi: 16.06.2023.

6. İnternet: En Popüler Web Siteleri Sıralaması. URL: https://www.similarweb.com/tr/top-websites/turkey/news-and-media/, Son Erişim Tarihi: 16.05.2023.

7. İnternet: Twitter. İşte olup bitenler. URL: https://twitter.com/?lang=tr, Son Erişim Tarihi: 16.06.2023.

8. Ünver, H. A., & EDAM, O. C. (2020). *TÜRKİYE'DE DOĞRULUK KONTROLÜ VE DOĞRULAMA KURULUŞLARI*. Centre for Economics and Foreign Policy Studies.

9. Oshikawa, R., Qian, J., & Wang, W. Y. (2018). A survey on natural language processing for fake news detection. *arXiv preprint arXiv:1811.00770*.

10. Hakan, I. R. A. K. (2022). Post-Truth Çağda Dijital Dezenformasyon: Covid-19 Aşı Karşıtlığı Üzerine Bir İnceleme. *Ahi Evran Akademi*, *3*(1), 115-129.

11. İnternet: Savaş ve çatışmalara neden olan 3 sahte haber. URL: https://www.bbc.com/turkce/haberler-dunya-43908746, Son Erişim Tarihi: 16.06.2023.

12. Akyüz, S. S., & ÖZKAN, M. (2022). Kriz Dönemlerinde Enformasyon Süreçleri: Ukrayna-Rusya Savaşında Dolaşıma Giren Sahte Haberlerin Analizi. *Uluslararası Kültürel ve Sosyal Araştırmalar Dergisi*, *8*(2), 66-82.

13. Chen, Y., Conroy, N. K., & Rubin, V. L. (2015). News in an online world: The need for an "automatic crap detector". *Proceedings of the Association for Information Science and Technology*, *52*(1), 1-4.

14. Cao, J., Guo, J., Li, X., Jin, Z., Guo, H., & Li, J. (2018). Automatic rumor detection on microblogs: A survey. *arXiv preprint arXiv:1807.03505*.

15. Hassan, N., Zhang, G., Arslan, F., Caraballo, J., Jimenez, D., Gawsane, S., ... & Tremayne, M. (2017). Claimbuster: The first-ever end-to-end fact-checking system. *Proceedings of the VLDB Endowment*, *10*(12), 1945-1948.

16. Fürnkranz, J. (1998). A study using n-gram features for text categorization. *Austrian Research Institute for Artifical Intelligence*, *3*(1998), 1-10.

17. Ahmed, H., Traore, I., & Saad, S. (2017). Detection of online fake news using n-gram analysis and machine learning techniques. In *Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments: First International Conference, ISDDC 2017, Vancouver, BC, Canada, October 26-28, 2017, Proceedings 1* (pp. 127-138). Springer International Publishing.

18. Potthast, M., Kiesel, J., Reinartz, K., Bevendorff, J., & Stein, B. (2017). A stylometric inquiry into hyperpartisan and fake news. *arXiv preprint arXiv:1702.05638*.

19. Udell, M., Horn, C., Zadeh, R., & Boyd, S. (2016). Generalized low rank models. *Foundations and Trends® in Machine Learning*, *9*(1), 1-118.

20. Xu, W., Liu, X., & Gong, Y. (2003, July). Document clustering based on non-negative matrix factorization. In *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval* (pp. 267-273)

21. Shu, K., Wang, S., & Liu, H. (2017). Exploiting tri-relationship for fake news detection. *arXiv preprint arXiv:1712.07709*, *8*.

22. Young, T., Hazarika, D., Poria, S., & Cambria, E. (2018). Recent trends in deep learning based natural language processing. *ieee Computational intelligenCe magazine*, *13*(3), 55-75.

23. Ansar, W., & Goswami, S. (2021). Combating the menace: A survey on characterization and detection of fake news from a data science perspective. International Journal of Information Management Data Insights, 1(2), 100052.

24. Kabudi, T., Pappas, I., & Olsen, D. H. (2021). Ai-enabled adaptive learning systems: A systematic mapping of the literature. Computers and Education: Artificial Intelligence, 2, 100017.

25. Khan, J. Y., Khondaker, M. 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.

26. Saha, A., Tawhid, K., Miah, M. B., Somudro, A. A., & Nandi, D. (2022, March). A Comparative Analysis on Fake News Detection Methods. In *Proceedings of the 2nd International Conference on Computing Advancements* (pp. 392-399).

27. Adib, Q. A. R., Mehedi, M. H. K., Sakib, M. S., Patwary, K. K., Hossain, M. S., & Rasel, A. A. (2021, October). A deep hybrid learning approach to detect bangla fake news. In *2021 5th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)* (pp. 442-447). IEEE

28. Udurawana, L., Weerasinghe, R., & Pushpananda, R. (2022, November). A hybrid approach for detection of fake news in sinhala text. In *2022 22nd International Conference on Advances in ICT for Emerging Regions (ICTer)* (pp. 039-044). IEEE.

29. Hamdi, T., Slimi, H., Bounhas, I., & Slimani, Y. (2020). A hybrid approach for fake news detection in twitter based on user features and graph embedding. In *Distributed Computing and Internet Technology: 16th International Conference, ICDCIT 2020, Bhubaneswar, India, January 9–12, 2020, Proceedings 16*(pp. 266-280). Springer International Publishing.

30. Seddari, N., Derhab, A., Belaoued, M., Halboob, W., Al-Muhtadi, J., & Bouras, A. (2022). A hybrid linguistic and knowledge-based analysis approach for fake news detection on social media. *IEEE Access*, *10*, 62097-62109.

31. Yerlekar, A., Mungale, N., & Wazalwar, S. (2021, November). A multinomial technique for detecting fake news using the Naive Bayes Classifier. In *2021 International Conference on Computational Intelligence and Computing Applications (ICCICA)* (pp. 1-5). IEEE.

32. Guo, Y. (2023). A mutual attention based multimodal fusion for fake news detection on social network. *Applied Intelligence*, *53*(12), 15311-15320.

33. Kanneganti, D. (2023). A New cross-domain strategy based XAI models for fake news detection. *arXiv preprint arXiv:2302.02122*.

34. Kumar, K. K., Rao, S. H., Srikar, G., & Chandra, M. B. (2021). A Novel Approach for Detection of Fake News using Long Short Term Memory (LSTM). *International Journal*, *10*(5).

35. Choudhury, D., & Acharjee, T. (2023). A novel approach to fake news detection in social networks using genetic algorithm applying machine learning classifiers. *Multimedia Tools and Applications*, *82*(6), 9029-9045.

36. Pu, T. (2023). *An Active Learning Framework for Trustworthy Classification over COVID-19 Multilingual Fake News Dataset*(Doctoral dissertation, Fordham University).

37. Kanagavalli, N., & Baghavathi Priya, S. (2022). An Approach for Fake News Detection in Social Media Using Hybrid Classifier. *Cybernetics and Systems*, 1-24.

38. Sheikhi, S. (2021). An effective fake news detection method using WOA-xgbTree algorithm and content-based features. *Applied Soft Computing*, *109*, 107559.

39. Harrag, F., & Djahli, M. K. (2022). Arabic fake news detection: A fact checking based deep learning approach. *Transactions on Asian and Low-Resource Language Information Processing*, *21*(4), 1-34.

40. Bahurmuz, N. O., Amoudi, G. A., Baothman, F. A., Jamal, A. T., Alghamdi, H. S., & Alhothali, A. M. (2022). Arabic Rumor Detection Using Contextual Deep Bidirectional Language Modeling. *IEEE Access*, *10*, 114907-114918.

41. Keya, A. J., Wadud, M. A. H., Mridha, M. F., Alatiyyah, M., & Hamid, M. A. (2022). AugFake-BERT: Handling Imbalance through Augmentation of Fake News Using BERT to Enhance the Performance of Fake News Classification. *Applied Sciences*, *12*(17), 8398.

42. Bažík, B. M. (2020). *Automatic Detection of Fake News*. Master Thesis.

43. Palani, B., Elango, S., & Viswanathan K, V. (2022). CB-Fake: A multimodal deep learning framework for automatic fake news detection using capsule neural network and BERT. *Multimedia Tools and Applications*, *81*(4), 5587-5620.

44. Setiawan, R., Ponnam, V. S., Sengan, S., Anam, M., Subbiah, C., Phasinam, K., ... & Ponnusamy, S. (2021). Certain investigation of fake news detection from facebook and twitter using artificial intelligence approach. *Wireless Personal Communications*, 1-26.

45. Pratiwi, I. Y. R., Asmara, R. A., & Rahutomo, F. (2017). Study of hoax news detection using naïve bayes classifier in Indonesian language. In *2017 11th International Conference on Information & Communication Technology and System (ICTS)*(pp. 73-78). IEEE.

46. Mertoğlu, U., Genç, B., & Sever, H. (2018). Savunmada Yenilikçi bir Dijital Dönüşüm Alanı: Sahte Haber Tespit Modeli.

47. Ahmed, H. (2017). *Detecting opinion spam and fake news using n-gram analysis and semantic similarity* (Doctoral dissertation).

48. Granik, M., & Mesyura, V. (2017). Fake news detection using naive Bayes classifier. In *2017 IEEE first Ukraine conference on electrical and computer engineering (UKRCON)*(pp. 900-903). IEEE.

49. Rubin, V. L., Conroy, N., Chen, Y., & Cornwell, S. (2016). Fake news or truth? using satirical cues to detect potentially misleading news. In *Proceedings of the second workshop on computational approaches to deception detection* (pp. 7-17).

50. Pérez-Rosas, V., Kleinberg, B., Lefevre, A., & Mihalcea, R. (2017). Automatic detection of fake news. *arXiv preprint arXiv:1708.07104*.

51. Mertoğlu, U., & Genç, B. (2020). Automated fake news detection in the age of digital libraries. *Information Technology and Libraries*, *39*(4).

52. Taskin, S. G., Kucuksille, E. U., & Topal, K. (2022). Detection of Turkish fake news in Twitter with machine learning algorithms. *Arabian Journal for Science and Engineering*, *47*(2), 2359-2379.

53. Gupta, A., Sukumaran, R., John, K., & Teki, S. (2021). Hostility detection and covid-19 fake news detection in social media. *arXiv preprint arXiv:2101.05953*.

54. Faustini, P. H. A., & Covoes, T. F. (2020). Fake news detection in multiple platforms and languages. *Expert Systems with Applications*, *158*, 113503.

55. Ahmad, I., Yousaf, M., Yousaf, S., & Ahmad, M. O. (2020). Fake news detection using machine learning ensemble methods. *Complexity*, *2020*, 1-11.

56. Ozbay, F. A., & Alatas, B. (2020). Fake news detection within online social media using supervised artificial intelligence algorithms. *Physica A: statistical mechanics and its applications*, *540*, 123174.

57. Shu, K., Mahudeswaran, D., Wang, S., Lee, D., & Liu, H. (2020). Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media. *Big data*, *8*(3), 171-188.

58. Tacchini, E., Ballarin, G., Della Vedova, M. L., Moret, S., & De Alfaro, L. (2017). Some like it hoax: Automated fake news detection in social networks. *arXiv preprint arXiv:1704.07506*.

59. Bajaj, S. (2017). The pope has a new baby!. *Fake news detection using deep learning*, 1-8.

60. Patel, M. (2017). *Detection of Maliciously Authored News Articles* (Doctoral dissertation, Yüksek Lisans Tezi, The Cooper Union For The Advancement of Science and Art, New York).

61. Ågren, C., & Ågren, A. (2018). Combating fake news with stance detection using recurrent neural networks.

62. Rajendran, G., Chitturi, B., & Poornachandran, P. (2018). Stance-in-depth deep neural approach to stance classification. *Procedia computer science*, *132*, 1646-1653.

63. Bhatt, G., Sharma, A., Sharma, S., Nagpal, A., Raman, B., & Mittal, A. (2017). On the benefit of combining neural, statistical and external features for fake news identification. *arXiv preprint arXiv:1712.03935*.

64. Ruchansky, N., Seo, S., & Liu, Y. (2017). Csi: A hybrid deep model for fake news detection. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management* (pp. 797-806).

65. Singhania, S., Fernandez, N., & Rao, S. (2017). 3han: A deep neural network for fake news detection. In *Neural Information Processing: 24th International Conference, ICONIP 2017, Guangzhou, China, November 14-18, 2017, Proceedings, Part II 24* (pp. 572-581). Springer International Publishing.

66. Volkova, S., Shaffer, K., Jang, J. Y., & Hodas, N. (2017, July). Separating facts from fiction: Linguistic models to classify suspicious and trusted news posts on twitter. In *Proceedings of the 55th annual meeting of the association for computational linguistics (volume 2: Short papers)* (pp. 647-653).

67. Wang, W. Y. (2017). " liar, liar pants on fire": A new benchmark dataset for fake news detection. *arXiv preprint arXiv:1705.00648*.

68. Girgis, S., Amer, E., & Gadallah, M. (2018, December). Deep learning algorithms for detecting fake news in online text. In *2018 13th international conference on computer engineering and systems (ICCES)* (pp. 93-97). IEEE.

69. Fang, Y., Gao, J., Huang, C., Peng, H., & Wu, R. (2019). Self multi-head attention-based convolutional neural networks for fake news detection. *PloS one*, *14*(9), e0222713.

70. Kaliyar, R. K., Goswami, A., & Narang, P. (2021). FakeBERT: Fake news detection in social media with a BERT-based deep learning approach. *Multimedia tools and applications*, *80*(8), 11765-11788.

71. Kaliyar, R. K., Goswami, A., & Narang, P. (2021). EchoFakeD: improving fake news detection in social media with an efficient deep neural network. *Neural computing and applications*, *33*, 8597-8613.

72. Han, Y., Karunasekera, S., & Leckie, C. (2020). Graph neural networks with continual learning for fake news detection from social media. *arXiv preprint arXiv:2007.03316*.

73. Monti, F., Frasca, F., Eynard, D., Mannion, D., & Bronstein, M. M. Fake news detection on social media using geometric deep learning. arXiv 2019. *arXiv preprint arXiv:1902.06673*.

74. Reis, J. C., Correia, A., Murai, F., Veloso, A., & Benevenuto, F. (2019). Supervised learning for fake news detection. *IEEE Intelligent Systems*, *34*(2), 76-81.

75. Jiang, T. A. O., Li, J. P., Haq, A. U., Saboor, A., & Ali, A. (2021). A novel stacking approach for accurate detection of fake news. *IEEE Access*, *9*, 22626-22639.

76. Goldani, M. H., Safabakhsh, R., & Momtazi, S. (2021). Convolutional neural network with margin loss for fake news detection. *Information Processing & Management*, *58*(1), 102418.

77. Goldani, M. H., Momtazi, S., & Safabakhsh, R. (2021). Detecting fake news with capsule neural networks. *Applied Soft Computing*, *101*, 106991.

78. 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.

79. Güler, G., & GÜNDÜZ, S. (2023). Deep Learning Based Fake News Detection on Social Media. *International Journal of Information Security Science*, *12*(2), 1-21.

80. Ma, J., Gao, W., Wei, Z., Lu, Y., & Wong, K. F. (2015). Detect rumors using time series of social context information on microblogging websites. In *Proceedings of the 24th ACM international on conference on information and knowledge management* (pp. 1751-1754).

81. Ma, J., Gao, W., Mitra, P., Kwon, S., Jansen, B. J., Wong, K. F., & Cha, M. (2016). Detecting rumors from microblogs with recurrent neural networks.

82. Yu, F., Liu, Q., Wu, S., Wang, L., & Tan, T. (2017). A Convolutional Approach for Misinformation Identification. In *IJCAI* (pp. 3901-3907).

83. Vaibhav, V., Annasamy, R. M., & Hovy, E. (2019). Do sentence interactions matter? leveraging sentence level representations for fake news classification. *arXiv preprint arXiv:1910.12203*.

84. Qi, P., Cao, J., Li, X., Liu, H., Sheng, Q., Mi, X., ... & Yu, Y. (2021). Improving fake news detection by using an entity-enhanced framework to fuse diverse multimodal clues. In *Proceedings of the 29th ACM International Conference on Multimedia* (pp. 1212-1220).

85. TAŞKIN, S. G., Küçüksille, E. U., & Topal, K. (2021). Twitter üzerinde Türkçe sahte haber tespiti. *Balıkesir Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, *23*(1), 151-172.

86. Bozuyla, M., & ÖZÇİFT, A. (2022). Developing a fake news identification model with advanced deep languagetransformers for Turkish COVID-19 misinformation data. *Turkish Journal of Electrical Engineering and Computer Sciences*, *30*(3), 908-926.

87. Deng, J., & Lin, Y. (2022). The Benefits and Challenges of ChatGPT: An Overview. *Frontiers in Computing and Intelligent Systems*, *2*(2), 81-83.

88. Silverman, C., Strapagiel, L., Shaban, H., Hall, E., Singer-Vine, J., 2016. Hyperpartisan facebook pages are publishing false and misleading information at an alarming rate. Buzzfeed News 20, 68.

89. Potthast, M., Kiesel, J., Reinartz, K., Bevendorff, J., & Stein, B. (2017). A stylometric inquiry into hyperpartisan and fake news. *arXiv preprint arXiv:1702.05638*.

90. İnternet: BuzzFeedNews. URL: https://github.com/BuzzFeedNews/2016-10-facebookfactcheck/tree/master/data. Son Erişim tarihi: 18.11.2020

91. İnternet: Liar dataset. URL: https://www.cs.ucsb.edu/ william/data/liar dataset.zip, Son Erişim tarihi: 18.11.2020.

92. İnternet: Bs-detector. URL: https://github.com/bs-detector/bs-detector, Son Erişim tarihi: 18.11.2020.

93. İnternet: Getting Real about Fake News. URL: https://www.kaggle.com/mrisdal/fake-news , Son Erişim tarihi: 18.11.2020.

94. Mitra, T., & Gilbert, E. (2015). Credbank: A large-scale social media corpus with associated credibility annotations. In *Proceedings of the international AAAI conference on web and social media* (Vol. 9, No. 1, pp. 258-267).

95. Santia, G., & Williams, J. (2018). Buzzface: A news veracity dataset with facebook user commentary and egos. In *Proceedings of the international AAAI conference on web and social media* (Vol. 12, No. 1, pp. 531-540).

96. Ma, J., Gao, W., & Wong, K. F. (2017). Detect rumors in microblog posts using propagation structure via kernel learning. Association for Computational Linguistics.

97. Santia, G., & Williams, J. (2018). Buzzface: A news veracity dataset with facebook user commentary and egos. In *Proceedings of the international AAAI conference on web and social media* (Vol. 12, No. 1, pp. 531-540).

98. İnternet: Snopes, URL: http://www.snopes.com, Son Erişim Tarihi: 16.06.2023.

99. İnternet: Weibo. URL: http://www.weibo.com, Son Erişim Tarihi: 16.06.2023.

100. İnternet: Weibo Service. URL: http://service.account.weibo.com, Son Erişim Tarihi: 16.06.2023.

101. Jin, Z., Cao, J., Guo, H., Zhang, Y., & Luo, J. (2017). Multimodal fusion with recurrent neural networks for rumor detection on microblogs. In *Proceedings of the 25th ACM international conference on Multimedia* (pp. 795-816).

102. Zhang, X., Cao, J., Li, X., Sheng, Q., Zhong, L., & Shu, K. (2021). Mining dual emotion for fake news detection. In *Proceedings of the web conference 2021* (pp. 3465-3476).

103. Nan, Q., Cao, J., Zhu, Y., Wang, Y., & Li, J. (2021, October). MDFEND: Multi-domain fake news detection. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management* (pp. 3343-3347).

104. İnternet: NewsVerify. URL: https://www.newsverify.com/, Son Erişim Tarihi: 16.06.2023.

105. Cao, J., Sheng, Q., Qi, P., Zhong, L., Wang, Y., & Zhang, X. (2019). False news detection on social media. *arXiv preprint arXiv:1908.10818*.

106. Jindal, S., Sood, R., Singh, R., Vatsa, M., & Chakraborty, T. (2020). Newsbag: A multimodal benchmark dataset for fake news detection. In *CEUR Workshop Proc* (Vol. 2560, pp. 138-145).

107. İnternet: URL: www.wsj.com, Son Erişim Tarihi: 16.06.2023.

108. İnternet: The Onion, URL: www.theonion.com, Son Erişim Tarihi: 16.06.2023.

109. İnternet: The Real News Network. URL:www.therealnews.com, Son Erişim Tarihi: 16.06.2023.

110. İnternet: the poke time well wasted. URL:www.thepoke.co.uk, Son Erişim Tarihi: 16.06.2023.

111. Yang, Y., Zheng, L., Zhang, J., Cui, Q., Li, Z., & Yu, P. S. (2018). TI-CNN: Convolutional neural networks for fake news detection. *arXiv preprint arXiv:1806.00749*.

112. İnternet: Kaggle, URL: www.kaggle.com/mrisdal/fake-news, Son Erişim Tarihi: 16.06.2023.

113. İnternet: NewYork Times. URL: www.nytimes.com, Son Erişim Tarihi: 16.06.2023.

114. İnternet: Washington Post. URL: www.washingtonpost.com, Son Erişim Tarihi: 16.06.2023.

115. İnternet: PolitiFact. URL: https://www.politifact.com, Son Erişim Tarihi: 16.06.2023.

116. Shu, K., Mahudeswaran, D., Wang, S., Lee, D., & Liu, H. (2020). Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media. *Big data*, *8*(3), 171-188.

117. İnternet: BuzzFeed. URL: www.buzzfeed.com, Son Erişim Tarihi: 16.06.2023.

118. Wang, Y., Yang, W., Ma, F., Xu, J., Zhong, B., Deng, Q., & Gao, J. (2020, April). Weak supervision for fake news detection via reinforcement learning. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 34, No. 01, pp. 516-523).

119. İnternet: WeChat. URL: https://www.wechat.com/, Son Erişim Tarihi: 16.06.2023.

120. Nakamura, K., Levy, S., & Wang, W. Y. (2019). r/fakeddit: A new multimodal benchmark dataset for fine-grained fake news detection. *arXiv preprint arXiv:1911.03854*.

121. Dai, E., Sun, Y., & Wang, S. (2020, May). Ginger cannot cure cancer: Battling fake health news with a comprehensive data repository. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 14, pp. 853-862).

122. İnternet: Health News Review. URL: www.HealthNewsReview.org, Son Erişim Tarihi: 16.06.2023.

123. Cui, L., & Lee, D. (2020). Coaid: Covid-19 healthcare misinformation dataset. *arXiv preprint arXiv:2006.00885*.

124. Shahi, G. K., & Nandini, D. (2020). FakeCovid--A multilingual cross-domain fact check news dataset for COVID-19. *arXiv preprint arXiv:2006.11343*.

125. İnternet: NewsGuard, URL: www.newsguardtech.com, Son Erişim Tarihi: 16.06.2023.

126. Zhou, X., Mulay, A., Ferrara, E., & Zafarani, R. (2020, October). Recovery: A multimodal repository for covid-19 news credibility research. In *Proceedings of the 29th ACM international conference on information & knowledge management* (pp. 3205-3212).

127. Li, Y., Jiang, B., Shu, K., & Liu, H. (2020). MM-COVID: A multilingual and multimodal data repository for combating COVID-19 disinformation. *arXiv preprint arXiv:2011.04088*.

128. Yang, C., Zhou, X., & Zafarani, R. (2021). CHECKED: Chinese COVID-19 fake news dataset. *Social Network Analysis and Mining*, *11*(1), 58.

129. Du, J., Dou, Y., Xia, C., Cui, L., Ma, J., & Philip, S. Y. (2021). Cross-lingual covid-19 fake news detection. In *2021 International Conference on Data Mining Workshops (ICDMW)*(pp. 859-862). IEEE.

130. Rubin, V. L., Conroy, N., Chen, Y., & Cornwell, S. (2016, June). Fake news or truth? using satirical cues to detect potentially misleading news. In *Proceedings of*

the second workshop on computational approaches to deception detection (pp. 7-17).

131. İnternet: Beaverton Oregon. URL: www.beavertonoregon.gov, Son Erişim Tarihi: 16.06.2023.

132. İnternet: Toronto Star. URL: www.thestar.com, Son Erişim Tarihi: 16.06.2023.

133. Rashkin, H., Choi, E., Jang, J. Y., Volkova, S., & Choi, Y. (2017). Truth of varying shades: Analyzing language in fake news and political fact-checking. In *Proceedings of the 2017 conference on empirical methods in natural language processing*(pp. 2931-2937).

134. Vogel, I., & Jiang, P. (2019). Fake news detection with the new German dataset "GermanFakeNC". In *International Conference on Theory and Practice of Digital Libraries* (pp. 288-295). Cham: Springer International Publishing.

135. İnternet: ISOT LAB. URL: https://www.uvic.ca/engineering/ece/isot/datasets/fake-news/index.php, Son Erişim Tarihi: 16.06.2023.

136. İnternet: List of fake news websites, URL: https://en.wikipedia.org/wiki/List_of_fake_news_websites, Son Erişim Tarihi: 16.06.2023.

137. Nasir, J. A., Khan, O. S., & Varlamis, I. (2021). Fake news detection: A hybrid CNN-RNN based deep learning approach. *International Journal of Information Management Data Insights*, *1*(1), 100007.

138. İnternet: The GDELT Project. URL: https://www.gdeltproject.org/, Son Erişim Tarihi: 16.06.2023.

139. İnternet: Pytorch. URL: https://pytorch.org, Son Erişim Tarihi: 16.06.2023.

140. Dogramaci, E., & Radcliffe, D. (2015). How Turkey uses social media. *Digital News Report*, *3*, 61-68.

141. Nordberg, P., Kävrestad, J., & Nohlberg, M. (2020). Automatic detection of fake news. In *6th International Workshop on Socio-Technical Perspective in IS Development, virtual conference in Grenoble, France, June 8-9, 2020* (pp. 168-179). CEUR-WS.

142. Rubin, V. L., Conroy, N., Chen, Y., & Cornwell, S. (2016, June). Fake news or truth? using satirical cues to detect potentially misleading news. In *Proceedings of the second workshop on computational approaches to deception detection* (pp. 7-17).

143. Oflazer, K., & Saraçlar, M. (2018). Turkish and its challenges for language and speech processing. *Turkish Natural Language Processing*, 1-19.

144. Akın, A. A., & Akın, M. D. (2007). Zemberek, an open source NLP framework for Turkic languages. *Structure*, *10*(2007), 1-5.

145. İnternet: Available Master's thesis topics in machne learning. URL: https://www.uib.no/en/rg/ml/128703/available-masters-thesis-topics-machine-learning, Son Erişim tarihi: 16.05.2023

146. İnternet: Sklearn.naive_bayes.MultinomialNB. URL: https://scikitlearn.org/stable/modules/generated/sklearn.naive_bayes.MultinomialNB.html, Son Erişim tarihi: 16.05.2023

147. Mishra, S., Shukla, P., & Agarwal, R. (2022). Analyzing machine learning enabled fake news detection techniques for diversified datasets. *Wireless Communications and Mobile Computing*, 1-18.

148. Singh, L. (2021). Fake news detection using machine learning," *International Journal for Research in Applied Science and Engineering Technology*, vol. 9, no. 5, pp. 980–983.

149. Poovaraghan, R. J., Priya, M. K., Vamsi, P. S. S., Mewara, M., & Loganathan, S. (2019). Fake news accuracy using naive bayes classifier. *International Journal of Recent Technology and Engineering (IJRTE)*, *8*(1C2), 2277-3878.

150. Rohera, D., Shethna, H., Patel, K., Thakker, U., Tanwar, S., Gupta, R., ... & Sharma, R. (2022). A taxonomy of fake news classification techniques: Survey and implementation aspects. *IEEE Access*, *10*, 30367-30394.

151. Utsha, R. S., Keya, M., Hasan, M. A., & Islam, M. S. (2021). Qword at CheckThat! 2021: An Extreme Gradient Boosting Approach for Multiclass Fake News Detection. In *CLEF (Working Notes)* (pp. 619-627)

152. Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).

153. Gowthami, K., Ramani, S., & Mohan Kumar, S. (2020). Identification of fake news through SVM and random forest. *IJESC*, *10*(10), 4.

154. Rajeswari, M., Sriniranjani, A. (2022). FAKE NEWS DETECTION USING LOGISTIC REGRESSION ALGORITHM WITH MACHINE LEARNING.

155. Pandey, S., Prabhakaran, S., Reddy, N. S., & Acharya, D. (2022). Fake News Detection from Online media using Machine learning Classifiers. In *Journal of Physics: Conference Series*(Vol. 2161, No. 1, p. 012027). IOP Publishing.

156. Patil, D. R. (2022). Fake news detection using majority voting technique. *arXiv preprint arXiv:2203.09936*.

157. İnternet: Stochastic Gradient Descent, available online, URL: https://scikit-learn.org/stable/modules/sgd.html, Son Erişim tarihi: 16.06 2023.

158. Vinothkumar, S., Varadhaganapathy, S., Ramalingam, M., Ramkishore, D., Rithik, S., & Tharanies, K. P. (2022). Fake News Detection Using SVM Algorithm in Machine Learning. In *2022 International Conference on Computer Communication and Informatics (ICCCI)* (pp. 1-7). IEEE.

159. İnternet: Understanding Support Vector Machine(SVM) algorithm, available online, URL: https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html , Son Erişim tarihi: 16.06.2023.

160. İnternet: sklearn.svm.SVC, available online, URL: https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html, Son Erişim tarihi: 16.06.2023.

161. Mikolov, T., Sutskever, I., Chen, K., Corrado, G., Dean, J. (2013). *Distributed representations of words and phrases and their compositionality*. Proceedings of the 26th International Conference on Neural Information Processing Systems, 2, Lake Tahoe, Nevada.

162. İnternet: Continuous Bag of Words (CBOW) Model in NLP. URL: https://www.scaler.com/topics/nlp/cbow/ , Son Erişim tarihi: 16.06.2023.

163. Mikolov, T., Chen, K., Corrado, G. S., Dean, J. (2013). Efficient estimation of word representations in vector space. *Proceedings of Workshop at ICLR*.

164. İnternet: **Word2Vec, FastText, GloVe. URL:** https://medium.com/codable/word2vec-fasttext-glove-d4402fa8cce0 , Son Erişim tarihi: 16.06.2023.

165. Joulin, A., Grave, E., Bojanowski, P., Douze, M., Jégou, H. and Mikolov, T. 2016. Fasttext. zip: Compressing text classification models. arXiv preprint arXiv:1612.03651.

166. Elik, Z., & Ko, B. C. (2021). Classification of Turkish news text by TF-IDF, Word2vec and fasttext vector model methods. *Deu Muhendislik Fakultesi Fen ve Muhendislik*, *23*(67), 121-127.

167. Nergız, G., Safali, Y., Avaroğlu, E., & Erdoğan, S. (2019). Classification of Turkish news content by deep learning based LSTM using Fasttext model. In *2019 International Artificial Intelligence and Data Processing Symposium (IDAP)*(pp. 1-6). IEEE.

168. Kuyumcu, B., Aksakalli, C., & Delil, S. (2019). An automated new approach in fast text classification (fastText) A case study for Turkish text classification without pre-processing. In *Proceedings of the 2019 3rd International Conference on Natural Language Processing and Information Retrieval* (pp. 1-4).

169. Erdinç, H. Y., & Güran, A. (2019). Semi-supervised turkish text categorization with word2vec, doc2vec and fasttext algorithms. In *2019 27th Signal Processing and Communications Applications Conference (SIU)* (pp. 1-4). IEEE.

170. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

171. Alammary, A. S. (2022). BERT models for Arabic text classification: a systematic review. *Applied Sciences*, *12*(11), 5720.

172. Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*.

173. Wani, A., Joshi, I., Khandve, S., Wagh, V., & Joshi, R. (2021). Evaluating deep learning approaches for covid19 fake news detection. In *Combating Online Hostile Posts in Regional Languages during Emergency Situation: First International Workshop, CONSTRAINT 2021, Collocated with AAAI 2021, Virtual Event, February 8, 2021, Revised Selected Papers 1* (pp. 153-163). Springer International Publishing.

174. İnternet: BERTurk. URL: https://huggingface.co/dbmdz/bert-base-turkish-128k-uncased, Son Erişim Tarihi: 16.06.2023.

175. KORU, G. & ULUYOL, Ç. (2024). Detection of Turkish Fake News from Tweets with BERT Models. *IEEE Access*.

GAZİLİ OLMAK AYRICALIKTIR.