

**T.C.
MANİSA CELAL BAYAR ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**YÜKSEK LİSANS TEZİ
YAZILIM MÜHENDİSLİĞİ ANABİLİM DALI
YAZILIM MÜHENDİSLİĞİ BİLİM DALI**

**COVID-19 PANDEMİ SÜRECİNİN EĞİTİM ÜZERİNDEKİ ETKİLERİNİN
MAKİNE ÖĞRENME TEKNİKLERİYLE TESPİT EDİLMESİ**

Kemal KARGA

**Danışman
Dr. Öğr. Üyesi Mansur Alp TOÇOĞLU**



MANİSA-2022

TAAHHÜTNAME

Bu tezin Manisa Celal Bayar Üniversitesi Hasan Ferdi Turgutlu Teknoloji Fakültesi Yazılım Mühendisliği Bölümü'nde, akademik ve etik kurallara uygun olarak yazıldığını ve kullanılan tüm literatür bilgilerinin referans gösterilerek tezde yer aldığını beyan ederim.

Kemal KARGA



İÇİNDEKİLER

	Sayfa
İÇİNDEKİLER	I
SİMGELER VE KISALTMALAR DİZİNİ.....	III
ŞEKİLLER DİZİNİ.....	IV
TABLO DİZİNİ	V
TEŞEKKÜR.....	VI
ÖZET	VII
ABSTRACT.....	VIII
1. GİRİŞ	1
2. LİTERATÜR TARAMASI.....	3
3. GENEL TANIMLAR.....	7
3.1. Metin Madenciliği	7
3.2. Makine Öğrenmesi	8
3.2.1. Makine Öğrenmesi Türleri.....	9
3.2.1.1. Denetimli Öğrenme.....	9
3.2.1.2. Denetimsiz Öğrenme.....	11
3.2.1.3. Takviyeli / Pekiştirmeli Öğrenme	11
3.2.2. Makine Öğrenmesi Algoritmaları.....	12
3.2.2.1. Destek Vektör Makinaları	12
3.2.2.2. Naive Bayes	14
3.2.2.3. Lojistik Regresyon	15
3.2.2.4. Rastgele Orman.....	17
3.3. Duygu Analizi	18
3.3.1. Duygu Analizi Yöntemleri.....	19
3.3.1.1. Sözlük Tabanlı Yöntemler	19
3.3.1.2. Makine Öğrenmesine Dayalı Duygu Analizi.....	20
3.3.2. Öznitelik Yapılandırma.....	20
3.4. Derin Öğrenme	21
3.4.1. Derin Öğrenme Mimarileri	21
3.4.1.1. Evrişimli Sinir Ağları (CNN).....	22
3.4.1.2. Tekrarlayan Sinir Ağları (RNN)	23
3.4.1.3. Uzun Kısa Süreli Bellek Ağları (LSTM)	25
3.4.1.4. Geçitli Tekrarlayan Birim (GRU)	26

3.4.1.5. Çift Yönlü Kodlayıcı Temsilleri (BERT)	27
3.4.2. Derin Öğrenmeye Dayalı Duygu Analizi	29
3.4.2.1. Kelime Gömme Yöntemleri	29
3.4.2.2. Word2vec	30
3.4.2.3. FastText	30
3.4.2.4. GloVe	31
4. MATERYAL VE YÖNTEMLER	32
4.1. Veri Seti	32
4.2. Verilerin Toplanması	33
4.3. Veri Ön İşleme	35
4.4. Verilerin Etiketlenmesi	35
4.5. Deneysel Prosedür	37
5. SONUÇLAR	45
KAYNAKLAR	46
ÖZGEÇMİŞ	54

SİMGELER VE KISALTMALAR DİZİNİ

BERT	Bidirectional Encoder Representations from Transformers
BOW	Bag of Words (Kelime Çantası)
CBOW	Continuous Bag of Words (Sürekli Kelime Çantası)
CNN	Convolutional Neural Network (Evrışimli Sinir Ağı)
DNN	Deep Neural Networks (Derin Sinir Ağı)
DVM	Destek Vektör Makinesi
GRU	Gated Recurrent Unit (Geçitli Tekrarlayan Birim)
LR	Lojistik Regresyon
LSTM	Long-short Term Memory (Uzun-kısa Süreli Bellek)
NB	Naive Bayes
RF	Random Forest (Rastgele Orman)
RO	Rastgele Orman
RNA	Ribonükleik Asit
RNN	Recurrent Neural Network (Tekrarlayan Sinir Ağları)
SPSS	Statistical Package for the Social Sciences
SS	Snowball Stemmer
SVM	Support Vektör Machine (Destek Vektör Makinesi)
OOB	Out Of Bag
TF	Term Frequency (Terim Frekansı)
TF-IDF	Term Frequency- Inverse Document Frequency
WHO	World Health Organization (Dünya Sağlık Örgütü)

ŞEKİLLER DİZİNİ

	Sayfa
Şekil 3.1. Metin Madenciliği İş Akış Süreçleri.....	7
Şekil 3.2. Yapay Zekânın Alt Dalı Olarak Makine Öğrenimi.....	8
Şekil 3.3. Makine Öğrenmesi Yöntemleri.....	9
Şekil 3.4. Regresyon Yöntemi	10
Şekil 3.5. Sınıflandırma Yöntemi	10
Şekil 3.6. Kümeleme Yöntemi	11
Şekil 3.7. Destek Vektörleri ve Hiper Düzlemler	13
Şekil 3.8. Doğrusal Ayrılan DVM Modeli.....	14
Şekil 3.9. Doğrusal Olmayan DVM Modeli	14
Şekil 3.10. LR Verinin S Eğrisi Şeklinde Dağılımı	16
Şekil 3.11. Rasgele Orman modelinin basitleştirilmiş gösterimi.....	18
Şekil 3.12. Duygu Analizi İş Akış Süreçleri.....	19
Şekil 3.13. Örnek Görüntü İşleme CNN Mimarisi	23
Şekil 3.14. RNN Döngü Yapısı.....	24
Şekil 3.15. Tekrarlayan RNN Modeli	25
Şekil 3.16. LSTM Mimari Yapısı	26
Şekil 3.17. GRU Mimarisi	27
Şekil 3.18. Google Aramalarında BERT Öncesi ve Sonrası Örnek Gösterim....	27
Şekil 3.19. BERT ön eğitim ve ince ayar gösterimi.....	28
Şekil 3.20. CBOW ve Skip-gram Modeli	30
Şekil 3.21. Fasttext Mimari Modeli	31
Şekil 4.1. Google Form Anket Arayüzü	34
Şekil 4.2. Bölümlere Göre Cevap Dağılımları	35
Şekil 4.3. Dört derin öğrenme mimarisinin ortalama doğruluk sonuçlarının iki gömülü katman oluşturma formları açısından karşılaştırılması	41
Şekil 4.4. Üç gömme yönteminin ortalama doğruluk sonuçlarının iki gömme katmanı yapı formları açısından karşılaştırılması	41
Şekil 4.5. Her üç makine öğrenimi yaklaşımının ortalama doğruluk performanslarının karşılaştırılması	43
Şekil 4.6. Makine öğrenimi mimarileri tarafından elde edilen doğruluk değerlerinin temel etkiler grafiği.....	44
Şekil 4.7. Derin öğrenme mimarileri tarafından elde edilen doğruluk değerlerinin temel etkiler grafiği.....	45

TABLO DİZİNİ

	Sayfa
Tablo 4.1. Örnek Olumlu Metin Belgeleri	32
Tablo 4.2. Örnek Olumsuz Metin Belgeleri	33
Tablo 4.3. İptal Edilen Metin Belgesi Örnekleri	36
Tablo 4.4. Ham ve Değerlendirilmiş Belge Dağılımları	37
Tablo 4.5. Dört makine öğrenme mimarisinin doğruluk sonuçlarının iki farklı stemmer yöntemi açısından karşılaştırılması	38
Tablo 4.6. Dört derin öğrenme mimarisinin doğruluk sonuçlarının iki farklı stemmer yöntemi açısından karşılaştırılması	38
Tablo 4.7. Farklı temsil yöntemleri kullanılarak dört geleneksel makine öğrenmesi algoritmasının sınıflandırma performanslarının karşılaştırılması	39
Tablo 4.8. Üç farklı gömme yöntemi kullanılarak oluşturulan kendi kendini eğiten gömme katmanlarını kullanan dört derin öğrenme mimarisinin doğruluk değerlerinin karşılaştırılması	40
Tablo 4.9. Üç farklı gömme yöntemi kullanılarak oluşturulmuş önceden eğitilmiş gömme katmanları kullanılarak dört derin öğrenme mimarisinin doğruluk değerlerinin karşılaştırılması	40
Tablo 4.10. BERTurk önceden eğitilmiş dil modelinin dört farklı formu kullanılarak BERT sınıflandırıcısının sınıflandırma performansının karşılaştırılması	42

TEŐEKKÜR

Çalıőmamın her aőamasında bana destek olan, bilgi ve deneyimleri ile yol gösteren danıőman hocam Sayın Doktor Öğretim Üyesi Mansur Alp TOÇOĞULU' na, desteęini hiçbir zaman esirgemeyen ve çalıőmamda büyük emeęi geçen Sayın Doç. Dr. Aytuę ONAN hocama, teknik desteęiyle çalıőmama büyük katkılar saęlayan mesai arkadaőım Rıza EMET' e, lisansüstü eęitimim boyunca bana her zaman destek olan sevgili eőim ve hayat arkadaőım Selda KARGA' ya ve aileme yürekten teőekkür ederim.

Kemal KARGA
Manisa, 2022



ÖZET

Yüksek Lisans Tezi

COVID-19 PANDEMİ SÜRECİNİN EĞİTİM ÜZERİNDEKİ ETKİLERİNİN MAKİNE ÖĞRENME TEKNİKLERİYLE TESPİT EDİLMESİ

Kemal KARGA

Manisa Celal Bayar Üniversitesi
Fen Bilimleri Enstitüsü
Yazılım Mühendisliği Anabilim Dalı

Danışman: Dr. Öğr. Üyesi Mansur Alp TOÇOĞLU

2020 küresel COVID-19 pandemisi, afet niteliğinde ekonomik ve sosyal bozulmalara yol açmıştır. Pandemi, sağlık, gıda, ticari işletmeler ve eğitim dahil olmak üzere neredeyse hayatımızın her alanını etkilemiştir.

Yükseköğretimde eğitim-öğretimin dijitalleşmesiyle birlikte büyük bir değişim yaşanmıştır. Pandemi ile mücadelede dünya genelindeki birçok yüksek öğretim kurumu, lisans ve lisansüstü dersleri eşzamanlı ya da farklı zamanlı olarak çevrimiçi vermeye başlamıştır. Bu süreçte insanlar, haber, bilgi, sosyal bağlantılar, destek ve yardım elde etmek amacıyla büyük ölçüde sosyal medyayı kullanmışlardır. Sonuç olarak, Web’ de COVID-19 ile ilgili çok büyük miktarda elektronik metin belgesi paylaşılmıştır.

Bu çalışmada, COVID-19 pandemi sürecinde eğitimlerine uzaktan devam eden üniversite öğrencilerinin, pandemi sürecinde almış oldukları uzaktan eğitimin olumlu ve olumsuz etkilerini analiz etmek üzere derin öğrenmeye dayalı bir duygu analizi yaklaşımı sunulmaktadır. Bu bağlamda, geleneksel makine öğrenmesi algoritmalarının (Destek Vektör Makineleri, Naive Bayes, Lojistik Regresyon ve Rastgele Orman) ve derin sinir ağlarının (Evrışimli Sinir Ağı, Tekrarlayan Sinir Ağı, Uzun Kısa Süreli Bellek ve Geçitli Tekrarlayan Birim) tahmin performansları karşılaştırılmıştır. Bunun yanı sıra, transformatörlerden (BERT) çift yönlü kodlayıcı temsilleri ile elde edilen deneysel sonuçlar değerlendirilmiştir. Farklı metin temsil modelleri ve sınıflandırma algoritmaları ile elde edilen kapsamlı deneysel sonuçlar, derin sinir ağlarının COVID-19 ile ilgili metin belgelerinin yüksek öğretim üzerindeki etkisini analiz etmede umut vadeden sonuçlar verebileceğini göstermektedir.

Anahtar Kelimeler: Derin Öğrenme, Duygu Analizi, Metin Madenciliği, COVID-19, Yüksek Öğretim

2022, 54 sayfa

ABSTRACT

M.Sc. Thesis

DETERMINING THE EFFECTS OF THE COVID-19 PANDEMIC PROCESS ON THE EDUCATION VIA MACHINE LEARNING METHODS

Kemal KARGA

**Manisa Celal Bayar University
Graduate School of Applied and Natural Sciences
Department of Software Engineering**

Supervisor: Asst. Prof. Mansur Alp TOÇOĞLU

The global COVID-19 pandemic in 2020 has led to catastrophic economic and social disruption. The pandemic has affected almost every aspect of our lives, including health, food, business organizations, and education.

An essential shift in the higher education field has been occurred with the digitalization of instruction. In attempt to combat the pandemic, several higher education institutions throughout the world have begun to offer undergraduate and graduate courses online, either asynchronously or synchronously. During this period, people make considerable use of social media to gain news, information, social connections, and support. As a result, the immense quantity of electronic text documents has been shared on the Web related to COVID-19.

In this study, we present a deep learning-based sentiment analysis approach to analyze the impact of COVID-19 pandemic on the higher education. In this regard, the predictive performance of conventional machine learning algorithms (Support Vector Machines, Naïve Bayes, Logistic Regression, and Random Forest) and deep neural networks (Convolutional Neural Network, Recurrent Neural Network, Long Short-Term Memory, and Gated Recurrent Unit) are compared. In addition, the empirical results obtained by the bidirectional encoder representations from transformers (BERT) has been evaluated. The comprehensive empirical results with different text representation models and classification algorithms indicate that deep neural networks can yield promising results for the task of analyzing the impact of COVID-19 related text documents on the higher education.

Keywords: Deep Learning, Sentiment Analysis, Text Mining, COVID-19, Higher Education

2022, 54 pages

1. GİRİŞ

2020 yılında başlayan COVID-19 salgını tüm dünya genelinde yıkıcı ekonomik ve sosyal çalkantılara neden olmuştur. Salgın, sağlık, gıda, kurumsal örgütler, eğitim kurumları gibi hayatımızın neredeyse tüm unsurlarını etkilemiştir [1]. COVID-19 pandemi sürecinde hükümetler, akademik ve sanayi kuruluşları salgını önlemek gibi ortak bir amaç doğrultusunda iş birliği yapmak üzere bir araya gelmişlerdir. Bu iş birliği sağlık kaynakları yönetimi, sosyal politika oluşumu, salgın önleme, tedavi ve aşı araştırmaları alanlarında çok çeşitli sonuçları da beraberinde getirmiştir [2]. Eş zamanlı olarak, COVID-19 salgını süresince dünya genelinde çeşitli ülkelerde uygulanan sosyal politikalar, salgını önleme ve tedavi uygulamaları, aşı geliştirme süreçleri hakkında birçok sosyal medya gönderisi ve haber makaleleri web üzerinden medya ve iletişim platformlarında paylaşılmıştır. Resmi olmayan internet paylaşım platformlarının bulaşıcı hastalıklar ve salgın hastalıklarla ilgili önemli sayıda gönderiyi oluşturduğu ve bu platformların dünyaya ilk önce ve en güncel bilgileri sağladığı keşfedilmiştir.

Dünya Sağlık Örgütü (WHO), tüm büyük salgınların ilk önce yetkisiz internet siteleri yoluyla paylaşıldığını tespit etmiştir [3]. Sosyal medyaya ve diğer çevrimiçi iletişim kanallarına yüklenen haberler bulaşıcı hastalık salgınlarını izlemeye ve takip etmeye yardımcı olmaktadır. COVID-19 salgınını zaman ve mekânsal olarak araştırmak için, dünya genelinde bazı ülkeler önemli sayıda gerçek zamanlı, etkileşimli mobil veya çevrimiçi coğrafi bilgi sistemleri, web siteleri ve uygulamalar oluşturmuştur. COVID-19 salgını hakkında doğru ve zamanında bilgi edinmek için, bilgi ve iletişim teknolojilerindeki ilerlemelerin yanı sıra çeşitli kaynaklardan elde edilen veriler de çok önemlidir.

Pandemiyle mücadele kapsamında tüm dünyada birçok yükseköğretim kurumu lisans ve lisansüstü derslerini eşzamanlı olarak (senkron) ya da farklı zamanlı (asenkron) olarak çevrimiçi vermeye başlamışlardır [4]. Eğitim-öğretimin dijitalleşmesiyle yükseköğretimde önemli bir değişim meydana gelmiştir. COVID-19 pandemisi eğitim sistemleri önünde ciddi bir engel teşkil etmektedir [5]. Bu bağlamda, sosyal eşitsizliğin de bir göstergesi olarak, internete erişim, nispeten kısa bir süre içinde öğrencilerin sabit bir bağlantıya erişimi olanlar ve olmayanlar olarak daha da

fazla sınıflara ayrılmalarına sebep olmuştur [6]. Ayrıca, uygulamalı dersler, çevrimiçi öğrenme ile düzgün bir şekilde öğretilmeyen doğa bilimi, mühendislik ve tıp gibi bazı bölümlerde okuyan üniversite öğrencilerinin öğrenme sürecinde hayati önem taşımaktadır [5, 6].

COVID-19 pandemisi, üniversitelerdeki çevrimiçi derslerin etkinliğini ölçmek ve ayrıca bu süreçteki olası güçlükleri ve olumlu deneyimleri belirlemek için deneysel ortam oluşturmuştur. İnsanlar bu süre zarfında haber, bilgi, destek ve yardım elde etmek üzere sosyal medyadan kapsamlı bir şekilde yararlanmışlardır. Sonuç olarak, COVID-19 ile bağlantılı devasa boyutta elektronik metin belgesi Web ortamında yayımlanmıştır. Salgınla eğitimde yaşanan dijitalleşme, öğrenciler ve eğitimciler tarafından, yeni süreçle ilgili görüşlerini dile getirdikleri birçok çevrimiçi içerik oluşturulmasına da neden olmuştur. Bu bağlamda, bu çalışmanın temel amacı, yükseköğretimde COVID-19 salgınının bir duygu sınıflandırma çerçevesini sunmaktır. Yapılan bu analiz, üniversite öğrencilerinin COVID-19 sürecinde duygu yönelimini belirlemek için metin madenciliği, makine öğrenimi ve derin öğrenme tekniklerine dayanan ilk kapsamlı analizdir.

Bu çalışmada, COVID-19 salgınının yükseköğretim üzerindeki etkisini analiz etmek üzere derin öğrenmeye dayanan bir duygu analizi yaklaşımı ele alınmaktadır. Bu bağlamda, geleneksel makine öğrenmesi algoritmalarının (Destek Vektör Makineleri, Naïve Bayes, Lojistik Regresyon ve Rastgele Orman) ve derin sinir ağlarının (Evrışimli Sinir Ağı, Tekrarlayan Sinir Ağı, Uzun Kısa Süreli Bellek ve Geçitli Tekrarlayan Birim) tahmin (kestirim) performansları incelenmiştir. Bunun yanı sıra, transformatörlerden (BERT) çift yönlü kodlayıcı temsilleri ile elde edilen ampirik sonuçlar değerlendirilmiştir. Farklı metin temsil modelleri ve sınıflandırma algoritmaları ile elde edilen kapsamlı deneysel sonuçlar, derin sinir ağlarının COVID-19 ile ilgili metin belgelerinin yüksek öğretim üzerindeki etkisini analiz etmede umut vadeden sonuçlar verebileceğini göstermektedir.

2. LİTERATÜR TARAMASI

Metin sınıflandırması, metin belgelerinin belirtilen bir veya daha fazla sınıf etiketine atandığı metin madenciliğinin önemli bir uygulama alanıdır [7]. Web sayfası sınıflandırması [8], duygu analizi [9-14], spam filtreleme [15] ve metin türü belirlenmesi [16] metin sınıflandırmasının başarılı olduğu alanlardan sadece birkaçıdır. COVID-19 gönderilerini analiz etmede metin madenciliğinin kullanımı hakkında çeşitli bilimsel çalışmalar olsa da bu bölümün geri kalanı COVID-19 salgını sırasında eğitime ilişkin duygu analizine özel bir vurgu yaparak konuyla ilgili temel araştırmalara odaklanmaktadır.

Rahmanian ve Jahanbin [17], COVID-19 sürecini düzenlemek için sosyal medya platformlarında yapılan sosyal ağ gönderilerini ve salgınla ilgili haberleri izlemek ve takip etmek için bulanık c-ortalama kümeleme algoritmasına dayanan bir yöntem sundular. Gizli Dirichlet tahsisi konu modelleme algoritması, Ordun ve arkadaşları [18] tarafından temalara, anahtar ifadeler ve özelliklere dayalı olarak COVID-19 ile ilgili Twitter gönderilerini analiz etmek için kullanılmıştır.

Peng ve arkadaşları tarafından yapılan başka bir çalışmada [19], popüler bir Çin sosyal medya sitesi olan Sina Weibo'da yapılan gönderilere bakarak, COVID-19 zatürre vakalarına mekânsal-zamansal bir şekilde fayda sağlayıp sağlayamayacakları incelenmiştir. Şubat 2020'de yapılan bu çalışmada amaç, on günlük bir süre boyunca coğrafi etiketleme kullanılarak incelenen verileri kullanarak COVID-19 salgınının kentsel ve mekânsal özelliklerini simüle etmektir.

Li ve arkadaşları [20] Amerika Birleşik Devletleri'nde zaman-mekân ölçeğinde COVID-19 ile ilişkili stres semptomlarını tespit etmek için korelasyon ek açıklamalı öğrenme algoritmasına ve klinik tabanlı kelime dağarcığına dayalı bir algoritma önermektedir. Sunulan bu algoritma gizli Dirichlet tahsisine nazaran, insan etkileşimini azaltarak performansı iyileştirmeyi amaçlamaktadır. Benzer şekilde, Chen ve arkadaşları [21] COVID-19 tweetlerinin zamansal ve mekânsal unsurlarını analiz etmek için makine öğrenimi ve konu modellemeye dayanan bir yaklaşım geliştirdiler. Çalışmanın hedefi, sosyal medya kullanıcılarının zaman içinde COVID-19'a nasıl tepki verdiklerini anlamaktır. Bu ampirik çalışmanın sonuçları, çalışılan ülkelerdeki

Twitter gönderilerinin miktarı ile COVID-19 örnekleri arasında bir bağlantı olduğunu göstermektedir.

Başka bir çalışmada, Boon-Itt ve Skunkan [22] Twitter kullanıcılarının COVID-19 salgını hakkındaki mesajlarını, duygu analizine ve gizli Dirichlet tahsis algoritmasına dayanan konu modelleme yaklaşımlarını kullanarak analiz ettiler.

Onan [20] üç temel n -gram modelinin (1-gram, 2-gram ve 3-gram), konuşma kısmı 2-gram ve konuşma kısmı 3-gram özelliklerinin, kelime/konuşma kısmı çiftleri, karakter n -gram ($n=2$ için) ve karakter n -gram ($n=3$ için) özelliklerinin tahmin performansını araştırdı. Bunun yanı sıra COVID-19 ile ilgili tweetlerin mekânsal-zamansal analizi için bu özelliklerin birleştirilmesiyle oluşturulan topluluk özellik setlerinin verimliliğini araştırdı. Özellik kümelerinin performansını değerlendirmek için Naïve Bayes algoritması, Lojistik Regresyon, Destek Vektör Makineleri, C4.5 Karar Ağaçları, K-en yakın komşu algoritması ve Rastgele Orman algoritması kullanıldı.

Eğitim verilerinin duygu analizi, öğrenme materyallerinin kalitesinin nasıl artırılacağı ve öğrencilerin öğrenme uygulamalarının nasıl belirleneceği hakkında değerli içgörüler sağlayabilecek öğrenme materyalleri ve kaynakları hakkında geri bildirim sağlamak için kullanılabilir [24, 25]. Örneğin, Bustillos ve arkadaşları [26] akıllı bir öğrenme ortamında fikir madenciliği için makine öğrenimi ve derin öğrenme stratejilerinin derinlemesine incelenmesini gerçekleştirmişlerdir. Çok sayıda makine öğrenimi algoritması (Bernoulli Naïve Bayes, Multinomial Naïve Bayes, Destek Vektör Makineleri, Lineer Destek Vektör Makineleri, Stokastik Gradyan İniş ve K-En Yakın Komşu algoritması dâhil) ve derin öğrenme mimarileri (Evrışimli Sinir Ağı ve Uzun-Kısa Süreli Bellek dahil) kullanılmıştır. Benzer şekilde, Cabada ve arkadaşları [27] iki derin öğrenme tabanlı mimari (Evrışimli Sinir Ağı ve Uzun Kısa Süreli Bellek) kullanarak eğitimsel incelemelerini yüzde 84,32 doğrulukla sınıflandırmışlardır.

Nguyen ve Nguyen [28], video yorumlarının duygu analizinde çift yönlü evrişimli n -gram LSTM (Uzun Kısa Süreli Bellek) kelime gömme mimarisinin kullanımına öncülük etmişlerdir. Önerilen yöntem, kısa ve uzun mesafelerde anlamsal ve bağlamsal bilgilere sahip bir kelimeyi temsil etmektedir. Lin ve arkadaşları [29]

bilgi tabanlı ve makine öğrenmesine dayalı duygu analizi tekniklerinin öğrencinin eğitim-öğretim değerlendirmeleri üzerine tahmin (kestirim) performansını araştırmışlardır. López ve arkadaşları [30] yaptıkları bir başka bir çalışmada bir eğitim kaynakları platformunda fikir madenciliği ve anlamsal profillemeye için bir çerçeve oluşturmuşlardır.

COVID-19 pandemi sürecinde eğitimde metin madenciliği, son zamanlarda araştırmacıların önemli ölçüde ilgisini çekmektedir çünkü virüsün kullanıcı deneyimi üzerindeki etkisini analiz etme ve kullanıcıların gereksinimlerini ayrıntılı şekilde belirleme yeteneğine sahiptir. Örneğin, Chen ve arkadaşları [31] pandeminin çevrimiçi eğitim platformlarının kullanıcı deneyimi üzerindeki etkilerini belirlemek için metin madenciliğine dayalı bir duygu analizi çerçevesi sunmuşlardır.

Komasawa ve arkadaşları [32] son sınıf öğrencileri arasında COVID-19 pandemisi ile ilgili algılarını belirlemek için tıp öğrencilerinin yazdıkları deneme makale belgeleri üzerinde veri madenciliğine dayanan bir yaklaşım geliştirmişlerdir. Benzer şekilde, Kim ve arkadaşları [33], öğrencileri entegre teknoloji kabul modeline dayalı çevrimiçi öğrenme sistemlerini kullanmaya daha istekli hale getiren karar verme unsurlarının daha net bir şekilde anlaşılabilmesi için COVID-19 pandemi sürecinde öğrenciler tarafından sağlanan geri bildirimleri analiz etmişlerdir.

COVID-19 pandemi sürecinde gerçekleştirilen uzaktan eğitime ilişkin analizlerin büyük çoğunluğu anket şeklinde veri toplamaya dayalıdır. Toplanan veriler istatistiksel analiz yöntemleriyle incelenmiş olup, makine öğrenmesi tekniklerine başvurulmamıştır. Keskin ve Kaya [34] çalışmalarında birden fazla üniversite öğrencileri üzerinde, uzaktan eğitime yönelik anket sonuçlarını değerlendirirken sayısal oranlama skalası kullanmıştır.

Daroedono ve arkadaşları [35] uzaktan eğitimin öğrenciler üzerindeki etkilerini ölçmek için, anket aracılığıyla veri toplayarak, toplanan verilerin analizlerini SPSS, Microsoft Excel gibi hazır yazılım araçları ile gerçekleştirmişlerdir.

Literatürde uzaktan eğitim hakkında elde edilen verileri makine öğrenme tekniklerini kullanarak analiz eden az sayıda çalışma mevcuttur. Bu çalışmalar pandemi öncesi yapılmış olup, ayrıca uzaktan eğitimin öğrenciler üzerindeki sosyolojik ve psikolojik etkileri analize dahil edilmemiştir.

Osmanoglu ve arkadaşları [36] çalışmalarında, pandemi öncesi uzaktan açık öğrenim gören öğrencilerin, eğitim amaçlı kullanmış oldukları dijital ders materyalleri (sesli kitap, video) üzerinden geri bildirim ile toplanmış oldukları verileri, makine öğrenme teknikleriyle analiz ederek, dijital materyallerin basılı materyallere kıyasla öğrenciler üzerinde daha olumlu sonuç bıraktığını tespit etmişlerdir.



3. GENEL TANIMLAR

3.1. Metin Madenciliği

Metin madenciliği, makine öğrenimi tekniklerinden yararlanılarak, yapısal olmayan metinleri yapısal biçime dönüştürme sürecidir. Bu süreçte, makine öğrenmesi ve derin öğrenme algoritmaları uygulanarak, yapısal olmayan veriler içerisinde bilgi keşfi yapılmış olur [37]. Keşfedilen değerli bilgiler amacına uygun olarak kullanılır. Metin madenciliğinde kullanılan yöntemler ise genel olarak matematiksel ve istatistiksel yöntemlerdir. Metin madenciliği, duygu analizi, fikir madenciliği, yazar tanıma gibi birçok farklı alanda uygulanmaktadır [38].

Metin madenciliği genel olarak aşağıdaki maddelere göre uygulanmaktadır.

- Analiz edilecek olan metinleri belirlemek, metin içerisindeki kavram ve ilişkilerin ayırt edilmesi ile istatistiksel, yapısal ve dilsel tekniklerin uygulanması.
- Metinlerin, makine öğrenmesi ve istatistiksel teknikler ile sınıflandırılarak, metinlerdeki ilişkilerin ve anlamların ortaya çıkarılması.
- Oluşturulan modelin performansının değerlendirilmesi.
- Elde edilen sonuçların kullanıcıya sunulması [39].



Şekil 3.1. Metin Madenciliği İş Akış Süreçleri [40]

3.2. Makine Öğrenmesi

Makine öğrenmesi, elde edilen veriler üzerinde bilgisayar bilimleri ve istatistiksel yöntemler kullanılarak çıkarım yapan, elde ettiği çıkarımlara göre tahminleme yapan yöntemler bütünüdür. Makine öğrenmesindeki temel amaç, çeşitli girdilerle eğitilen modelin doğru tahminleme yapabilmesidir [41]. Günümüzde başta sağlık alanı olmak üzere, eğlence, otomotiv, bankacılık sistemleri gibi büyük bir alanda yaygın olarak kullanılmaktadır.

Makine öğrenmesi, genel olarak öğrenmeye odaklıdır. Herhangi bir sürecin çalışma modelini tasarlayarak, pratik kullanıma hazır hale gelmesini sağlar. Karşılaştığı yeni örneklerle nasıl bir yol izleyeceğini belirlemek için önceden öğrendiği bilgilerden çıkarımlar yaparak, insan müdahalesine gerek duymadan dinamik kararlar verir [42]. Makine öğrenmesinde kullanılan algoritmalar programlama ihtiyacı duymadan görevlerini yerine getirebilir. İstenilen görevleri yerine getirmek için elde edilen verilerden öğrenme olayını gerçekleştirirler [43].

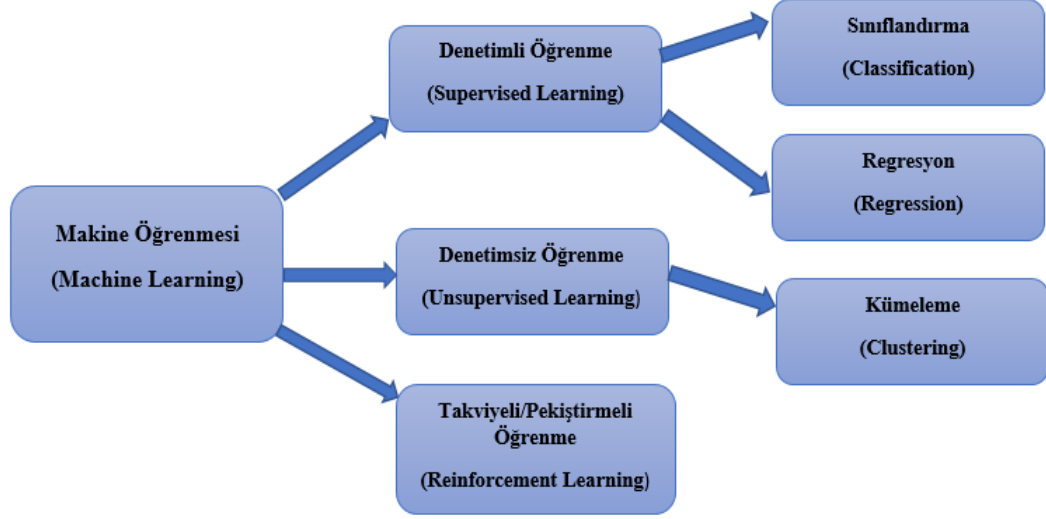
Yapay zekânın alt bir alanı olarak kabul edilen makine öğrenmesi, tahminleme yaparken pasif gözlemlere dayanır. Yapay zekâ ise öğrenme ve eylem isteklerini yerine getirmek için çevre ile etkileşime girmektedir [43]. Ayrıca günümüzde daha popüler kullanıma sahip olan derin öğrenme, makine öğrenmesinin alt dalı olarak kabul edilmektedir. Bu üç yaklaşım Şekil 3.2'deki gibi gösterilmektedir.



Şekil 3.2. Yapay Zekânın Alt Dalı Olarak Makine Öğrenimi [44]

3.2.1. Makine Öğrenmesi Türleri

Makine öğrenmesi ile kullanılan algoritma ve yöntemler, öğrenme süreçlerine göre üç kategoride incelenirler [43]. Bu yöntemler Şekil 3.3'te gösterildiği gibi kategorize edilmektedir.

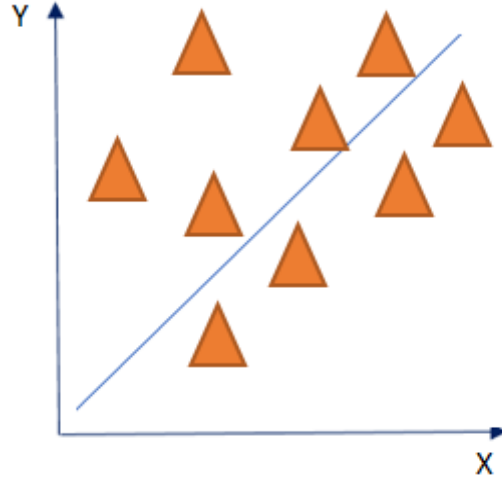


Şekil 3.3. Makine Öğrenmesi Yöntemleri

3.2.1.1. Denetimli Öğrenme

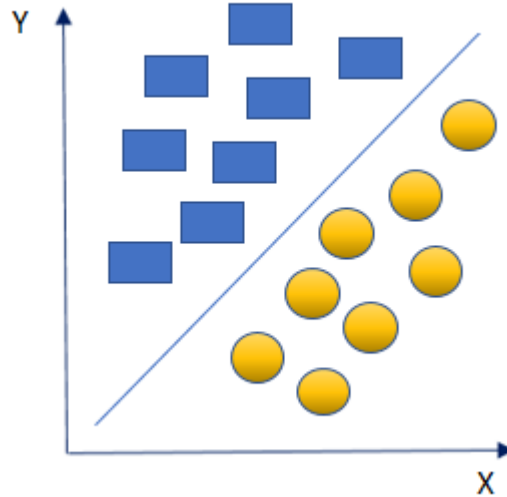
Etiketlenmiş veriler kullanılarak üretilen modelin, test verileri ile doğrulanmasını sağlayan öğrenme yöntemidir [45]. Veri kümesi içerisindeki data'nın büyük bir çoğunluğu, oluşturulacak olan modelin öğrenmesi ve öğrenim sürecinin test edilmesi ile ilgilidir. Denetimli öğrenme algoritmaları öncelik olarak tahmin problemlerine odaklanmaktadır. Veri seti içerisinde, eğitilmiş veri ile bağımlı ve birçok bağımsız değişken kullanılarak kestirim yapılır. Modele gönderilen test verileri ile bağımlı değişkenlerin tahminleme işlemi gerçekleştirilir [41].

Denetimli öğrenme yöntemleri Regresyon ve Sınıflandırma olarak iki kategoride ele alınmaktadır. Regresyon yöntemi, bağımlı olarak kullanılan bir değişkenin, bağımsız değişkenler üzerinde modellenmesidir. Sınıflandırma yöntemi ise, verinin hangi sınıfa ait olduğunun önceden bilindiği ve bu sınıfa göre eğitildiği, yeni gelen test verisinin ise hangi sınıfa dahil edileceğinin tahmin edilmesidir [46].



Şekil 3.4. Regresyon Yöntemi

Regresyon problemleri sayısal değerler üzerinde uygulanmaktadır. Mevcut rakamsal veri üzerinden (bağımlı değişken), daha sonra gelecek olan değerler (bağımsız değişken) tahmin edilir. Örneğin, bir araç markasının yıllık satış tutarlarının bulunduğu bir veri setinde, bir sonraki yılın satış tutarlarının tahmin edilmek istenmesi durumunda regresyon yöntemi uygun bir yaklaşım olacaktır.



Şekil 3.5. Sınıflandırma Yöntemi

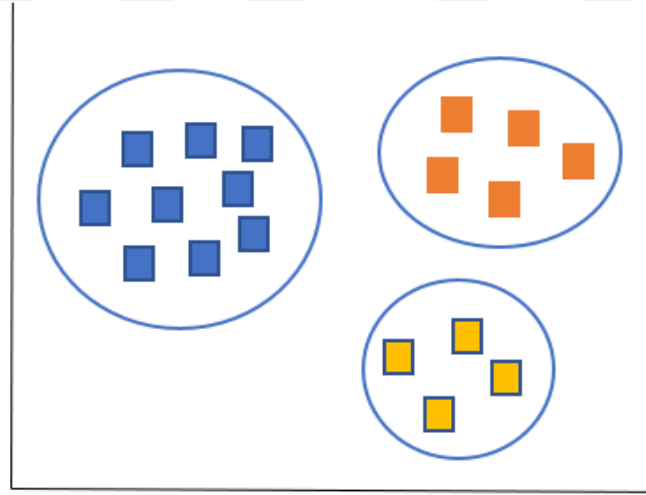
Sınıflandırma yönteminde elde edilen sonuçlar ayrı çıktılarla tahmin edilir. Bu yöntemi kullanan algoritmalar, veri seti içerisinde bulunan verilerin hangi sınıfa dahil

olacağını öğrenir ve test edilecek olan verileri doğru sınıflara yerleştirmeye çalışır. Örneğin, e-posta sunucularına gelen maillerin zararlı olup olmadığının tespit edilmesi probleminde, sınıflandırma yöntemi çözüm açısından uygun bir yaklaşım olacaktır.

3.2.1.2. Denetimsiz Öğrenme

Denetimsiz öğrenme, veri seti içerisinde bulunan örneklerin etiketlenmediği ve herhangi bir sınıfa dahil edilmediği öğrenme biçimidir. Denetimsiz öğrenme algoritmaları, verideki ortak yapıyı belirleyerek yeni gelecek olan test verisinde bu ortak özelliklerin bulunup bulunmadığına karar verir [47].

Denetimsiz öğrenmede en çok kullanılan yöntem, birbirine benzer yapıların olduğu verilerin aynı grup içerisine dahil edildiği kümeleme yöntemidir. Kümeleme yönteminde genel olarak veriler arasındaki mesafeler ölçülerek, ortak yapı bulma ve gruplama işlemi gerçekleştirilir. Basit uzaklık ölçütleri ile (Manhattan, Öklid) veriler arasındaki benzerlikler tespit edilir. Ayrıca komşuluk ve yoğunluk hesaplamalarında benzerlik tespiti için kullanılmaktadır [46].



Şekil 3.6. Kümeleme Yöntemi

3.2.1.3. Takviyeli / Pekiştirmeli Öğrenme

Takviyeli öğrenme, bulunduğu ortamı algılayarak kendi kendine kararlar alabilen ve değişen ortam koşullarına göre yeni kararlar oluşturabilen makine öğrenme yöntemidir. Takviyeli öğrenmede, model birden fazla karar içerisinden en doğru olanı nasıl bulacağını öğrenir. Ödül sistemine göre çalışan bu öğrenme yönteminde, yanlış

kararlarda cezalandırma veya daha az ödül verme yöntemi uygulanır. Modelin gerçekleştirmek istediği ana hedef, yeteri kadar deneme yaparak büyük ödülü alabileceği kararları almaktır [46].

Genellikle derin öğrenme mimarileriyle çalışan takviyeli öğrenme, sürü zekâsı, genetik algoritmalar, oyun teorisi gibi birçok alanda yöntem olarak kullanılmaktadır. Ayrıca pekiştirmeli öğrenme algoritmaları, matematiksel optimizasyon ve bilgisayar programlama yönteminin birlikte uygulandığı, dinamik programlama tekniklerini uygulamaktadır [48].

3.2.2. Makine Öğrenmesi Algoritmaları

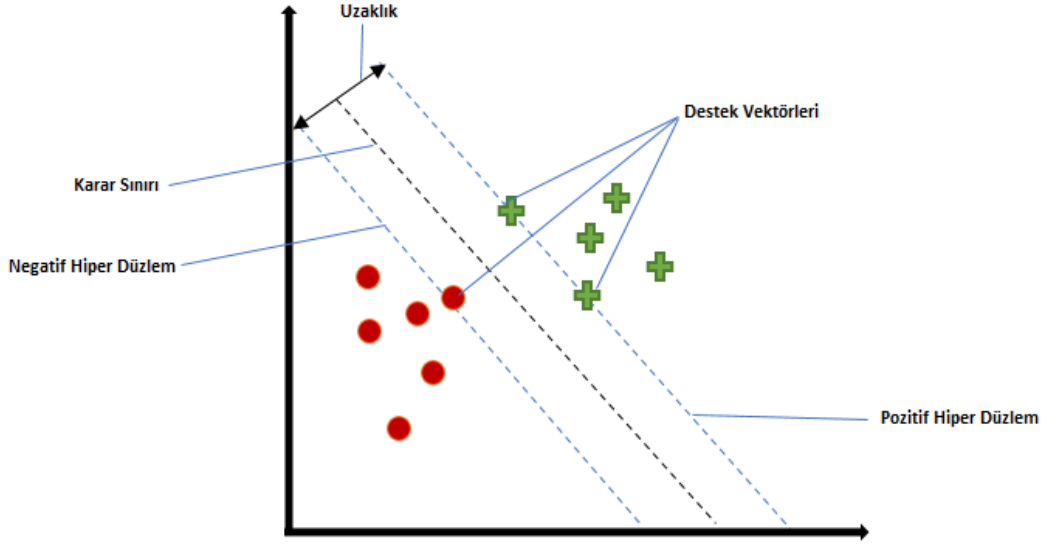
Bu çalışmada, özellik kümelerine dayalı öğrenme modelleri elde etmek üzere dört denetimli makine öğrenme algoritması kullanılmıştır. Bu algoritmalar aşağıda kısaca açıklanmıştır.

3.2.2.1. Destek Vektör Makinaları

Destek vektör makinaları (DVM), sınıflandırma ve regresyon görevlerine yönelik denetimli öğrenme algoritmalarıdır. DVM, farklı sınıfların örnekleri arasındaki sınırı belirlemek için daha yüksek boyutlu bir uzayda hiperdüzlem hesaplamaktadır [49].

DVM algoritmasında, doğrusal olmayan eşleme ile n boyutlu veri kümesi $m > n$ olacak biçimde m boyutlu veri setine dönüştürülür. Dönüştürülen veri kümesi üzerinde yüksek boyutta tahminleme ve doğrusal sınıflandırma işlemi yapılabilir. Veri, uygun dönüşüm yapılarak hiperdüzlem ile iki ayrı sınıfa ayrılabilir. Destek vektörleri, hiperdüzleme en yakın veriler olarak adlandırılır [50].

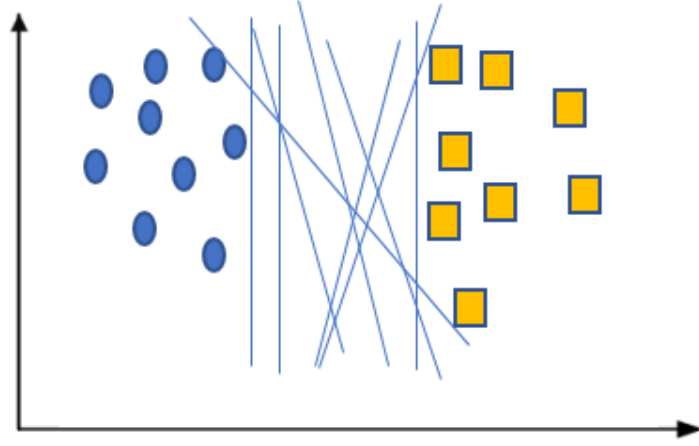
DVM ile sınıfların birbirinden ayrılması için, optimal ayırma hiperdüzleminin oluşturulması amaçlanmaktadır. Bu sayede, destek vektör çizgisine yakın olan sınıfların mesafeleri maksimize edilmektedir [51].



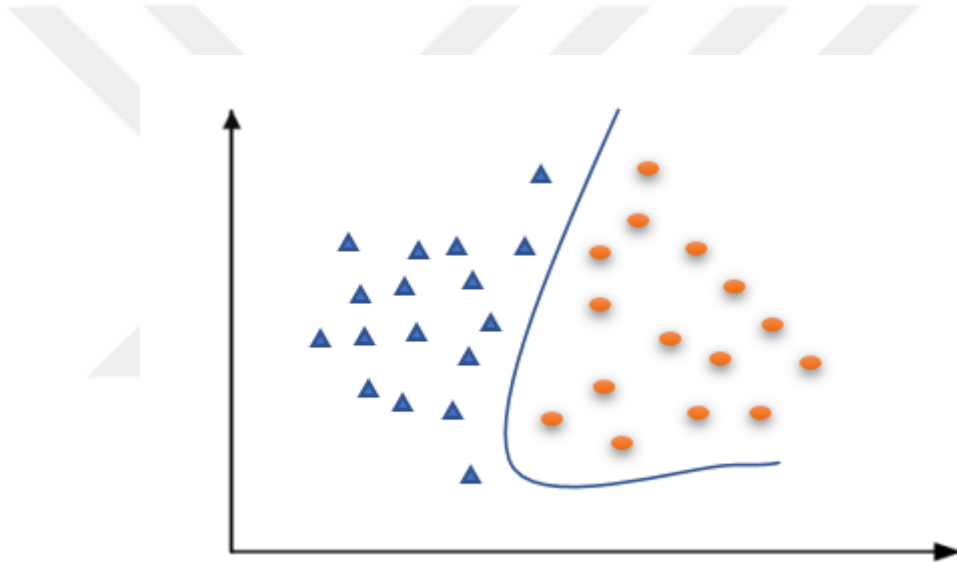
Şekil 3.7. Destek Vektörleri ve Hiper Düzlemler [51]

Şekil 3.7’de gösterilen DVM yönteminde, sınıfları birbirinden ayıran birden çok sayıda vektör bulunmaktadır. DVM ile, bu vektörler arasındaki en uzak mesafeye sahip düzlemin bulunmasını amaçlanır.

DVM modelinde, verinin doğrusal olarak sınıflandırıldığı veya sınıflandırılmadığı duruma göre iki farklı yaklaşım bulunmaktadır. Doğrusal olan DVM farklı iki sınıfı düzgün bir doğru ile ayırabilirken, doğrusal olmayan DVM yönteminde veri doğrusal olarak ayrılamaz. Bu durumda veri kümesinin sınıflandırma işlemi, çekirdek fonksiyonu aracılığıyla yüksek boyutlu uzaya taşınarak gerçekleştirilir [52].



Şekil 3.8. Doğrusal Ayrılan DVM Modeli



Şekil 3.9. Doğrusal Olmayan DVM

Şekil 3.9’da gösterildiği gibi doğrusal olmayan DVM yöntemi, verilerin doğrusal olarak ayıramadığı durumlarda kullanılmaktadır. Bu durumda veriler hatasız bir şekilde ayıramadığından bu yöntem tercih edilmektedir.

3.2.2.2. Naive Bayes

Bayes teoremine ve koşullu bağımsızlık varsayımına dayanan istatistiksel denetimli bir öğrenme algoritmasıdır [53]. Genel olarak algoritmanın çalışma biçimi, veri kümesi içerisinde bulunan her bir verinin bütün olasılık durumlarının

hesaplanması ve sınıflandırma işlemi gerçekleştirilirken olasılık değeri en yüksek olan verinin seçilme işlemidir.

Naive Bayes (NB), sınıflandırma işlemi denetimsiz öğrenme yöntemi ile gerçekleştirmektedir. Verilerin hangi sınıfa ait olduğu daha önceden bilinmektedir. Gelen veri kendi sınıfı içersine alınarak, bağlı olduğu sınıfı temsil etmektedir.

Denklem (3.1)' de NB modelinin çalışma prensibi olan Bayes formülü gösterilmektedir.

$$P(A/B) = \frac{P(B/A) \cdot P(A)}{P(B)} \quad (3.1)$$

$P(A/B)$ = B olayı gerçekleştiğinde A olayının gerçekleşme ihtimali

$P(A)$ = A olayının gerçekleşme ihtimali

$P(B/A)$ = A olayı gerçekleştiğinde B olayının gerçekleşme ihtimali

$P(B)$ = B olayının gerçekleşme ihtimali

Uygulanması nispeten kolay ve hızlı olan NB, birçok alanda kullanılmaktadır. Hastalıkların tedavi süreçleri hakkında bilgi verme ve hastalık teşhisleri, RNA dizilerinin sınıflandırılması, duygu analizi ve e-posta spam filtreleme gibi değişik disiplinlerde sıkça kullanılmaktadır.

3.2.2.3. Lojistik Regresyon

Doğrusal regresyon kullanılarak sınıflandırma problemlerini çözmeye bir çerçeve sağlayan doğrusal bir sınıflandırma algoritmasıdır. Lojistik regresyon (LR) modeli, doğrusal bir regresyon modeli kullanarak doğrusal bir sınıflandırma şeması oluşturur ve bu şemadaki hedef değişkenleri dönüştürür [54].

Lojistik regresyon genel olarak, sınıflandırma ve regresyon problemlerinde kullanılmaktadır. Algoritmanın çalışma prensibinde, girdi olarak alınan bağımsız değişkenlerle, bağımlı olan değişken arasındaki ilişkiyi modelleme bulunmaktadır. Tahmin edilecek olan bağımlı değişken, kategori sayısına göre ikili, sıralı ve çok kategorili regresyon olarak isimlendirilmektedir.

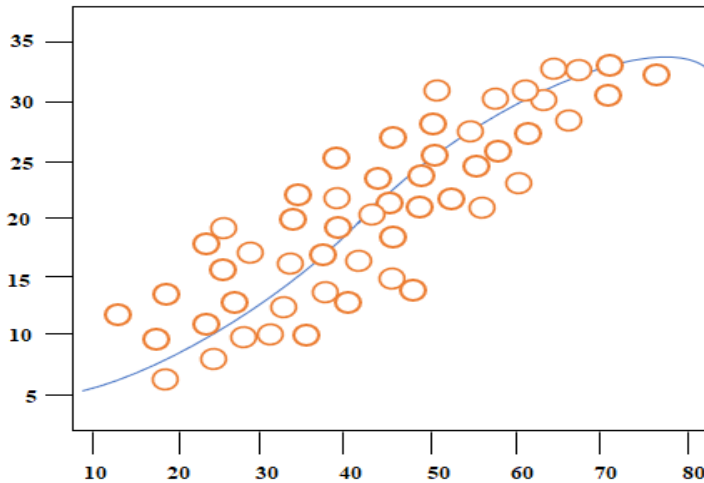
Lojistik regresyonda 1 olarak tanımlanan bağımlı değişkenin gerçekleşme ihtimali hesaplanır. Bu sebeple bağımlı değişkenin alacağı değer göz ardı edilir. LR modelinde üretilecek olan değerlerin 0 ile 1 arasında olması için, π tahmin değerleri ile bağımsız değişkenler arasındaki ilişki lojistik fonksiyon ile hesaplanır [55].

$$\pi(x) = P(Y = 1 | X = x) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p}}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)}} \quad (3.2)$$

(3.2) 'de belirtilen Denkleme göre $X=x$ olduğunda ve $Y=1$ olma olasılığında $\pi(x)$ değeri elde edilmektedir. Fonksiyonun doğrusallaştırılması için lojit dönüşüm uygulanmaktadır. Lojit dönüşüm için ise olayın gerçekleşme olasılığının, gerçekleşmeme ihtimaline oranıyla elde edilecek olan odds oranına gerek duyulmaktadır. Odds oranının doğal algoritması hesaplanılarak regresyon modeli elde edilmiş olur.

$$\frac{\pi(x)}{1 - \pi(x)} \quad (3.3)$$

LR modeli uygulanan veri kümesi üzerindeki dağılımın S-curve şeklinde davranış sergilemesi, veri kümesindeki tahminin başarı oranını arttırmaktadır.



Şekil 3.10. Lojistik Regresyon Modelinde veri dağılımının S eğrisi şeklinde dağılımı

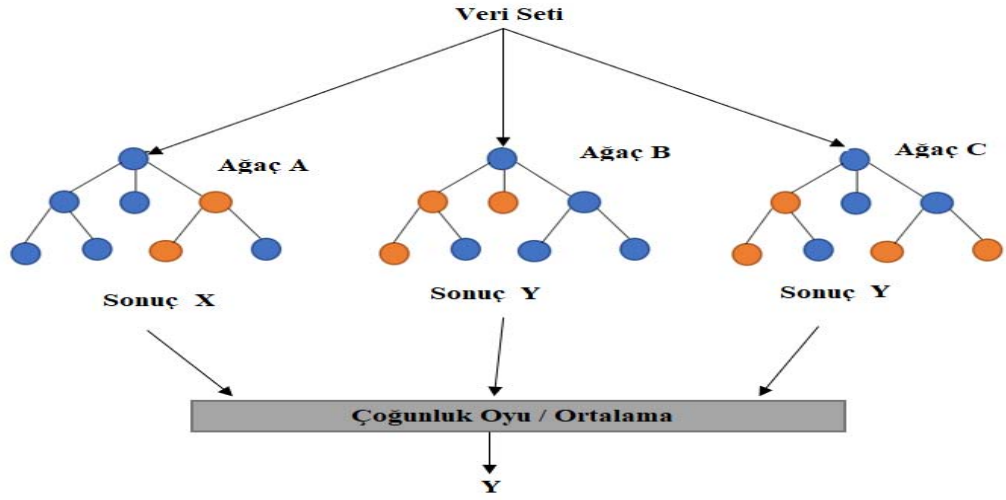
Şekil 3.10'da gösterilen lojistik regresyon modelinde veriler S-egrisi şeklinde dağılım göstermektedir. Bu eğrinin oluşması ile veri üzerinde tahmin yapmak daha kolay olacaktır. Ayrıca model tahminini yükseltmek için, modeldeki bağımlı değişkenlerle bağlantılı olacak biçimde koşullandırılmış değişkenler dahil edilmektedir.

3.2.2.4. Rastgele Orman

Rastgele orman algoritması, çantalama ve rastgele alt uzay algoritmalarının bir kombinasyonu olarak tanımlanmaktadır. Karar ağaçları bu algoritmada temel öğrenici olarak kullanılmaktadır. Her bir ağaç eğitim verilerinden alınan bootstrap örnekleri kullanılarak oluşturulmaktadır. Temel öğreniciler arasında çeşitlilik üretmek için rastgele bir özellik seçim süreci kullanılmaktadır. Böylece, gürültülü veya alakasız verilerin varlığında dahi, model tatmin edici sonuçlar üretmektedir [56].

RO modelinde veri kümesi test ve eğitim verisi olarak ayrılmadan bir bütün olarak ele alınabilir. Algoritmanın performansını ölçmek için, model iç hatası olarak adlandırılan Out-Of-Bag (OOB) hata oranı kullanılabilir. Ağaçlar oluşturulduktan sonra herbirine hata oranına göre bir ağırlık verilir. Verilen bu ağırlık, her bir ağaç için hesaplanan OOB hata oranına göre belirlenmektedir. Hata oranı en yüksek olan ağaç en düşük ağırlığı, hata oranı en düşük olan ağaç ise en yüksek ağırlığı almaktadır. Ağırlıklarına göre ağaçlar, sınıf tahmininde bulunmak için bir adet oy verme işlemi gerçekleştirirler. Ağaçlar tarafından oluşturulan ağırlıklı oylar RO algoritması tarafından toplanır [57].

Aşağıda Şekil 3.11'de gösterilen, üç adet karar ağacının her birinden elde edilen sınıflandırma örneği gösterilmektedir. Her bir ağaç veri kümesini ayrı olarak analiz ederek, tahmin sınıfını oluşturur. Sonuç tahmini ise çoğunluk oylamasıyla (Y olacak biçimde) belirlenir.



Şekil 3.11. Rasgele Orman modelinin basitleştirilmiş gösterimi [58]

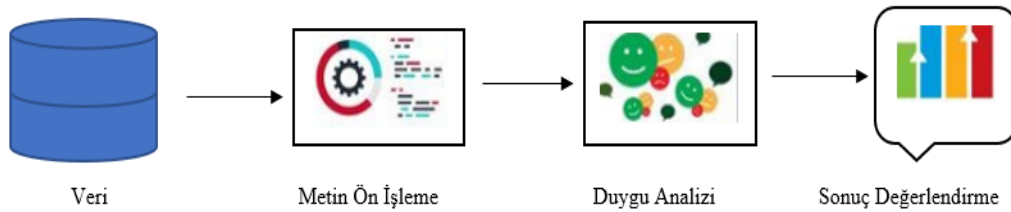
Rastgele orman algoritmasının doğruluk tahmini tek bir karar ağacının doğruluk tahmininden yüksektir. Ayrıca hata oranı tek bir karar ağacına göre daha başarılıdır. RO' da karar verme ve orman oluşturma olmak üzere iki adet aşama bulunmaktadır. Karar ağaçlarının ortalaması alınarak ve kararlar birleştirilerek RO sonucu elde edilir. Yapılan çalışmalarda bu algoritmanın karar ağacı algoritmasına göre daha başarılı sonuçlar ürettiği gözlemlenmiştir [59].

3.3. Duygu Analizi

Fikir madenciliği olarak da tanımlanan duygu analizi, makine öğrenmesi, doğal dil işleme gibi teknik ve yöntemlerin kullanılarak, belli bir metin içerisindeki görüş veya duyguların ortaya çıkarılmasını sağlayan analiz yaklaşımıdır. İnternet tabanlı teknolojilerinin artması ile birlikte insanlar duygu ve düşüncelerini çeşitli platformlarda paylaşma gereksinimi duymaktadır. Büyük bir hızla artan bu paylaşımlar, bireysel veya kurumsal karar vericiler tarafından ilgi odağı haline gelmektedir. Karar vericiler, bu veriler üzerinden duygu analizi yaklaşımını uygulayarak karar süreçlerini daha etkin hale getirmektedir. Duygu analizi yaklaşımı, pazarlama stratejileri, müşteri memnuniyeti belirleme, devletlerin kamusal fikirleri öngörmesi gibi ekonomik ve siyasi alanlarda sıkça uygulanmaktadır [60].

Duygu analizi iş süreci, birtakım yöntemler kullanılarak elde edilen verinin yapısal hale dönüştürülmesiyle başlar. Veri seti oluşturulduktan sonra veri ön işleme ve etiketleme süreçleri başlatılır. Verilerin ön işlem sürecinden geçmesinin nedeni, uygulanacak modelleme tekniklerinin (makine öğrenmesi, derin öğrenme algoritmaları) daha performanslı çalışması içindir. Veri ön işlemede, büyük küçük harf dönüşümleri, noktalama ve sayısal ifadelerin metin içerisinden çıkarılması, anlam ifade etmeyen kelimelerin (stopwords) silinmesi gibi adımlar uygulanmaktadır. Veri etiketleme süreci ise, metnin içerisinde geçen düşünce veya duyguların pozitif, negatif ve nötr duygu durumuna göre işaretlenmesidir.

Veri ön işleme süreçleri tamamlandıktan sonra makine öğrenmesi veya derin öğrenme algoritmaları kullanılarak model oluşturma aşamasına geçilir. Oluşturulan modelde, hangi algoritmaların kullanılacağı ile verilerin eğitim ve test süreçleri belirlenir. Modelin başarısı ölçüldükten sonra, sistem kullanıcıya sunulmak üzere hazır hale getirilir.



Şekil 3.12. Duygu Analizi İş Akış Süreçleri [61]

3.3.1. Duygu Analizi Yöntemleri

Duygu Analizi çalışmalarında uygulanan yöntemler temel olarak aşağıda belirtildiği gibi açıklanmıştır.

3.3.1.1. Sözlük Tabanlı Yöntemler

Veri kümesi içerisinde bulunan kelimelerin sözlük yardımıyla duygu durumlarının (pozitif, negatif, nötr) tespit edilip etiketlenmesi veya sayısal olarak ifade edilme işlemidir. Etiketlenen veriler toplam etiket sayısına bölünerek (olumlu ve olumsuz kelime / etiket sayısı) basit bir hesaplama işlemi gerçekleştirir. Elde edilen

sonuç, belirli bir eşik değerin üzerinde ise olumlu, altında ise olumsuz olarak değerlendirilir [46].

3.3.1.2. Makine Öğrenmesine Dayalı Duygu Analizi

Makine öğrenmesine dayalı duygu analizinde iki temel aşama bulunmaktadır: Verilerden özelliklerin (öznitelik) çıkarımını yaparak bunların özellik vektörleri olarak temsil edilmesi ve öğrenme modelini elde etmek için bu özellik vektörleri üzerinde denetimli öğrenme algoritmalarını eğitmek. Görünmeyen örneklerin sınıf etiketleri bu elde edilen öğrenme modeli kullanılarak belirlenmiştir [62].

Makine öğrenimi ve doğal dil işleme yaklaşımlarında uygulanacak olan algoritmaların daha etkin çalışabilmesi için, terim ağırlıklandırma işlemleri uygulanmaktadır. Terim ağırlıklandırma işlemi ile belgede geçen terimlerin ilgi düzeylerinin ölçümleri gerçekleştirilmektedir [63]. Duygu analizi yaklaşımlarında, belge içerisinden bilgi çıkarımı işlemlerinde genellikle terim ağırlıklandırma şemaları kullanılmaktadır.

Metin madenciliği çalışmalarında özellikle belge üzerinden bilgi çıkarımı işlemlerinde kullanılmak üzere birçok terim ağırlıklandırma yöntemi önerilmektedir. Bu ağırlıklandırma yöntemleri literatürde geleneksel yöntemler olarak nitelendirilmektedir. Bu yöntemler, ilgili terimin mevcut dokümanda bulunma veya bulunmama durumu (terim varlığı), terimin belge içerisindeki geçme sıklığı (Term Frequency- TF), ilgili terimin belge içerisinde yaygın olarak kullanılıp kullanılmadığı durumu (Inverse Document Frequency – IDF) ve bu iki yöntemin çeşitli yöntemlerle kombine edildiği TF-IDF yaklaşımlarıdır.

Bu çalışmada, makine öğrenmesine dayalı duygu analizi gerçekleştirmek için üç terim ağırlıklandırma şeması olan, terim varlığı, terim sıklığı (frekans) TF-IDF ve 3 N-gram (bigram, unigram ve trigram) modelleri kullanılmıştır.

3.3.2. Öznitelik Yapılandırma

Metin belgelerini denetimli öğrenme algoritmalarıyla işlerken, en kritik işlem metin belgelerinin özellik vektörlerine dönüştürülmesidir. Metin madenciliği ve bilgi geri alma çalışmalarında, sıklıkla kullanılan başarılı bir şema, kelime çantası (BOW)

çerçevesidir. Bir metin belgesi bu çerçevede bir kelime çantası olarak görülür ve sözdizimi, kelime sıralaması veya dilbilgisi dikkate alınmaksızın belgede karşılaşılan tüm kelimeleri içeren bir vektörle temsil edilir [64]. Her bir metin belgesi bu çerçevede, her bir kelimenin frekansı kullanılarak temsil edilmektedir.

Bu çalışmada öğrenme modelini elde etmek ve denetimli öğrenme algoritmasını eğitmek için bir dizi öznelik kullanılmıştır.

3.4. Derin Öğrenme

Makine öğrenmesinin alt dalı olarak kabul edilen derin öğrenme, birden fazla veri işleme katmanı bulunduran ve karar süreçlerinde yapay sinir ağlarını kullanan öğrenme türüdür. Derin öğrenme yalnızca kurallar oluşturmak yerine, örneklerden yararlanarak çıkarım elde eder ve bu çıkarımları kullanarak insan davranışına benzer tepkiler ortaya koyar. Makine öğrenmesi algoritmaları kısıtlı sayıda veri ile çalışırken, derin öğrenme algoritmaları büyük hacimli karmaşık veriler ile daha iyi performans göstermektedir [65].

Derin öğrenme mimarilerinde, klasik yapay sinir ağları kullanımından farklı olarak birçok gizli katman ve nöron kullanılmaktadır. Bu katman ve nöronlar sayesinde elde edilen ham verilerden yüksek seviyede soyutlama yapılarak, öznelik seçimleri otomatik olarak oluşturulmaktadır [66].

Yapay zekâ uygulamalarında derin öğrenme yöntemleri kullanılarak önemli ilerlemeler kaydedilmiştir. Derin öğrenme mimarileri, büyük ölçekteki karmaşık verilerin analizinin gerçekleştirilmesi sayesinde, birçok işletme ve bilim dalı alanında uygulama alanı bulmaktadır. Ses ve görüntü tanıma alanında klasik makine öğrenme yöntemlerine göre daha iyi performans göstermektedir. Ayrıca derin öğrenme yöntemleri ile, hastalıkların tanınması, doğal dil işleme, soru cevaplama ve duygu analizi gibi alanlarda umut verici çözümler üretilmektedir [67].

3.4.1. Derin Öğrenme Mimarileri

Derin öğrenme mimarileri özelliklerin çok düzeyli temsillerin elde edilmesini sağlamaktadır. Mimariler, doğrusal olmayan bilgilerin birden çok katmanının veya

aşamasının hiyerarşik olarak işlenmesi yoluyla öğrenme modellerini belirlemek için tasarlanmıştır [68].

Bu bölümün geri kalanında, ampirik analizde kullanılan dört adet derin öğrenme mimarisinin temel özelliklerinden bahsedilmiştir.

3.4.1.1. Evrişimli Sinir Ağları (CNN)

Evrişimli sinir ağları, verileri ızgara tabanlı bir topoloji kullanarak işleyen derin sinir ağlarına dayanan mimarilerdir. CNN, evrişim olarak bilinen belirli bir matematiksel işlem türü ile tanımlanmaktadır. Evrişim, bir veya daha fazla evrişim katmanı kullanılarak gerçekleştirilmiştir. Tipik bir evrişimli sinir ağı bir girdi katmanı ile bir çıktı katmanı ve gizli katmanlardan oluşmaktadır. Mimarinin gizli katmanları, evrişim katmanları, havuz katmanları, tam bağlantılı katmanlar ve normalleştirme katmanları dahil olmak üzere birkaç kategoriye ayrılmaktadır [68]. CNN mimarisinde ele alınan üç temel katmanın özellikleri kısaca açıklanmıştır.

Evrişim Katmanı: CNN mimarisinin temel katmanı olarak ele alınır ve hesaplama yoğunluğu yüksek olan katmandır. Giriş verisi ve filtreleme işlemleri bu katmanda gerçekleştirilir. Evrişim katmanının yaptığı temel ve en önemli iş, giriş verilerine filtre uygulayarak, verilerin özellik haritasının çıkarılmasıdır.

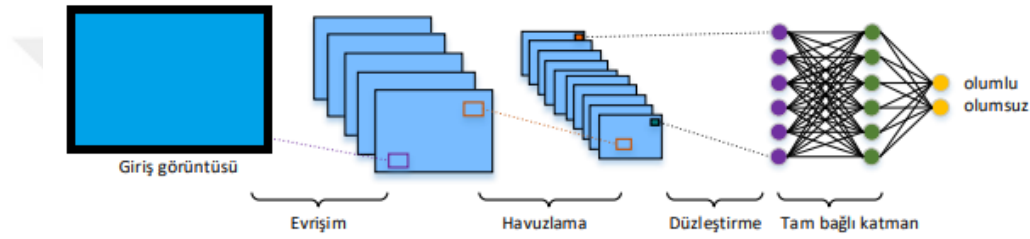
Havuz Katmanı: Havuz katmanının temel işlevi, evrişim katmanında oluşturulan özellik haritalarının boyutlarının indirgenmesidir. Boyut indirgeme işleminin temel amacı, sonraki katmanın hesaplama yükünü azaltmak ve verinin düşük boyutta temsil edilmesini sağlamaktır.

Tam Bağlantı Katmanı: CNN mimarisinin son katmanı olan bu katmanda klasik yapay sinir ağı işlemleri gerçekleştirilmektedir. Havuzlama katmanından gelen giriş verileri, düzleştirme işleminden sonra, bu katmanda tüm nöronlara bağlı olarak çalışırlar. Bu bağlılıktan dolayı tam bağlantı adını almaktadır. Ayrıca giriş verilerinin sınıflandırma işlemi bu katmanda yapılmaktadır.

CNN 'de giriş verilerinin özellik çıkarımları farklı filtreler kullanılarak evrişim katmanında yapılmaktadır. Havuzlama katmanı ile evrişim katmanı arasında ReLU

(Rectified Linear Units) aktivasyon fonksiyonu kullanılarak, giriş verisinden elde edilen özniteliklerin doğrusal olmaması sağlanır. Özellik haritalarının boyutları havuzlama katmanında düşürülerek, sonraki katmanların iş gücü hafifletilmektedir. CNN mimarisindeki son katman, tipik yapay sinir ağı hücrelerinden oluşan ve giriş verilerinin, nöronlara tam bağlı bir yapıda bulunduğu katmandır [69].

Giriş verisi olarak görüntü kümesi alan ve bu görüntü kümesini sınıflandıran bir CNN mimarisinin, katmanlar arasındaki bağlantılarını gösteren şema Şekil 3.13'te verilmiştir.



Şekil 3.13. Örnek Görüntü İşleme CNN Mimarisi [69]

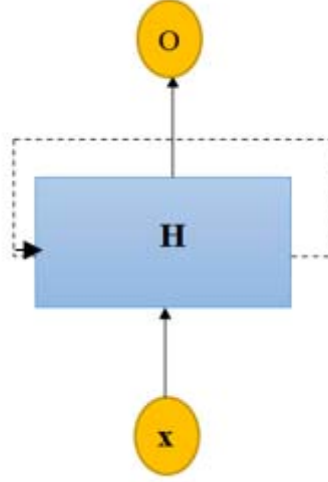
Evrşimli sinir ağları, genellikle görsel veriler için önerilmiş ve çok boyutlu girdiler içeren derin öğrenme mimarisidir. Daha az sayıda nöron bağlantısı kullanan bu mimaride, öğrenme sürecinde etiketlenmiş büyük veriye ihtiyaç duyulmaktadır. Görüntü işleme, ses tanıma ve doğal dil işleme problemlerinde yaygın olarak kullanılan bu mimaride başarılı sonuçlar elde edilmektedir [70].

3.4.1.2. Tekrarlayan Sinir Ağları (RNN)

Günümüzde doğal dil işleme problemlerinde sıkça kullanılan RNN, sıralı veri işlemede kullanılan bir sinir ağları modelidir. RNN mimarisi, sıralı bilgileri işlemek için ileri beslemeli sinir ağlarından farklı olarak dahili hafızasını kullanabilmektedir. Hafıza kullanımı ile önceden işlenmiş bilgilerin hatırlatılması amaçlanmaktadır. Bu mimarideki girdilerin sırasını işlemek için ise iç durum ağı oluşturulmaktadır [71].

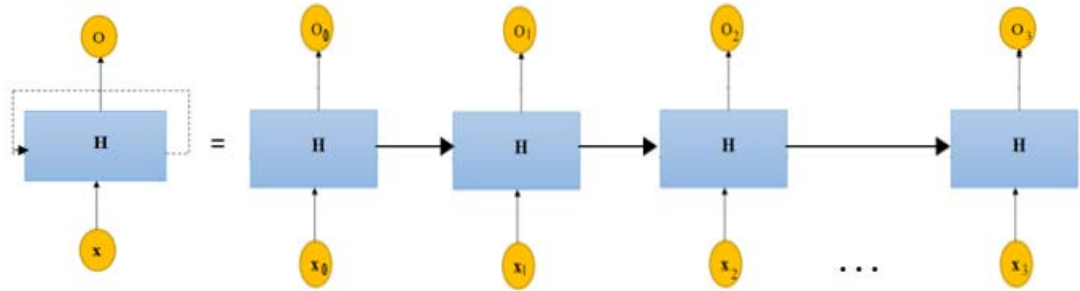
Şekil 3.14'te gösterilen RNN mimarisinde, H sinir ağı modelinde x girdi düğümü, o ise çıktı düğümü olarak tanımlanmaktadır. Örnek şekildeki RNN döngü

yapısı incelendiğinde, H sinir ağı t zamanda x girdisi ile o çıktısını oluşturmaktadır. Bu yapı sayesinde üretilen çıktı, bir sonraki adımda kullanılabilir.



Şekil 3.14. RNN Döngü Yapısı [72]

Birden fazla sinir ağının tekrarlanmasıyla oluşan RNN mimarisinde, her bir sinir ağı işlediği verinin çıktısını oluşturarak, bir sonraki sinir ağına aktarmaktadır. Sinir ağları, bir önceki ağın çıktı verisi ile kendi girdi verisini ilişkilendirerek sıralı bir yapıda ilerleme sağlamaktadır. Bu yapı ile birlikte, her bir ağın bir önceki giriş verilerini hatırlama imkânı olmaktadır. Hatırlama özelliği sayesinde RNN mimarisi daha çok seri verilerin işlendiği problemlerin çözümünde kullanılmaktadır.



Şekil 3.15. Tekrarlayan RNN Modeli [73]

RNN mimarisinde, patlayan veya kaybolan gradyan probleminden dolayı birkaç adım öncesine bakmak ve uzun süreli bağımlılıkların modellenmesi

zorlaşmaktadır. Bu problemin çözümü için, Uzun Kısa Süreli Bellek Ağı, Çift Yönlü ve Derin Çift Yönlü RNN mimarileri geliştirilmiştir. Çift yönlü RNN mimarisi iki adet RNN'nin üst üste eklenmesiyle oluşturulmaktadır. RNN'lerden biri ileri yönde bilgi işlerken, diğeri geriye dönük ters bir biçimde bilgiyi işler. Böylece yapı içerisindeki ileri ve geri yöndeki bilgi kolay bir şekilde elde edilmiş olur. Derin Çift Yönlü RNN mimarisi ise RNN yapısına benzemektedir. Bu mimari RNN yapısından farklı olarak, birden fazla katmana ve daha fazla eğitim verisine ihtiyaç duymaktadır [71].

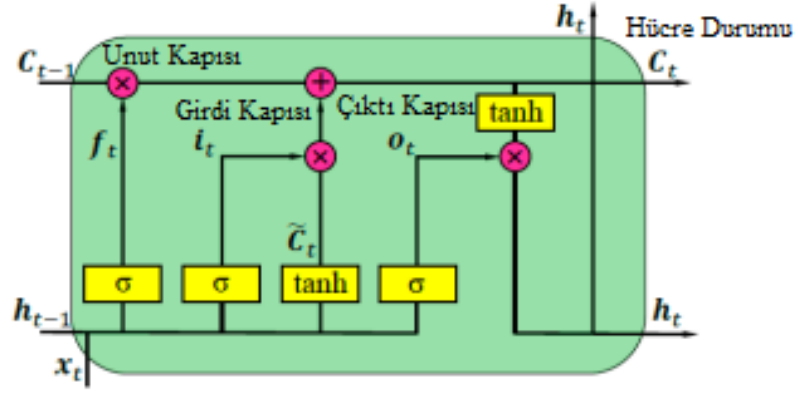
3.4.1.3. Uzun Kısa Süreli Bellek Ağları (LSTM)

Tekrarlayan sinir ağlarına dayalı derin öğrenme mimarisi olan Uzun Kısa Süreli Bellek Ağları, 1997 yılında Hochreiter ve Schmidhuber tarafından gündeme getirilmiştir. RNN mimarisindeki, gradyan patlaması veya kaybolması sorununa çözüm üretmek için geliştirilen LSTM, bu problemi çözmek için unutmaya kapıları kullanmaktadır [74].

LSTM mimarisinde sınırlı sayıda zaman adımına kadar hatanın geri yayılımına izin verilir. Tipik bir LSTM ünitesi bir hücre ve üç farklı kapı türünden oluşur. Bu kapılar giriş kapısı, çıkış kapısı ve unut kapısıdır. Kapıların açık ve kapalı işlemleri, hangi bilgilerin korunması ve ne zaman erişilmesi gerektiğini belirtmek için kullanılmıştır [75].

LSTM hücreleri genel olarak RNN mimarisindeki hücre yapısı ile benzerlik göstermektedir. LSTM algoritmasında, RNN' den farklı olarak girdi kapısı, çıktı kapısı ve unut kapısı bulunmaktadır.

- Girdi kapısı ile giriş verisinin aktarılması sağlanır.
- Çıktı kapısı ile çıkış verisinin aktarılması sağlanır.
- Unut kapısı ile, bir önceki hücreden aktarılan verinin ne kadarının unutulup unutulmadığına karar verilir [76].



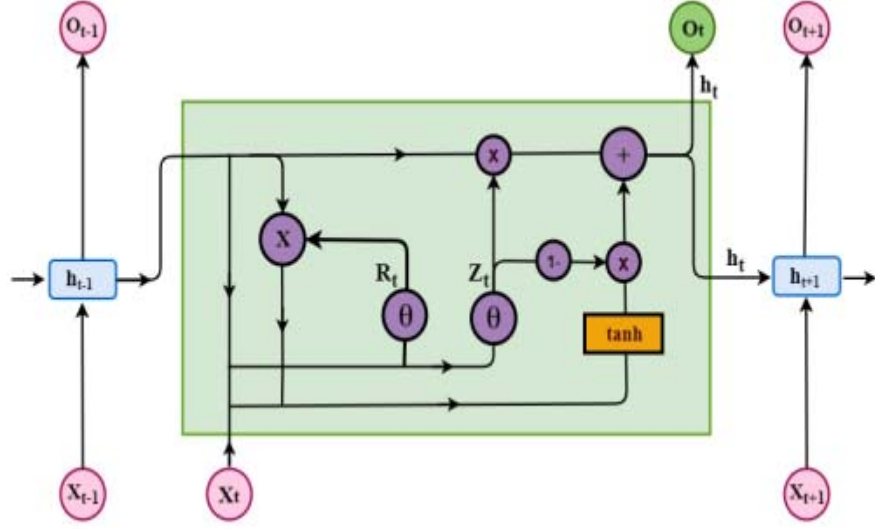
Şekil 3.16. LSTM Mimari Yapısı [77]

Şekil 3.16’da gösterilen LSTM mimarisinde, f_t unut kapısı, i_t girdi kapısı ve o_t çıkış kapısı olarak belirtilmiştir. Önceki ve sonraki bilgileri depolayan birim c_t hücre durumudur. Giriş ve çıkış bilgileri ise sırasıyla x_t ve h_t tarafından temsil edilmektedir [77].

3.4.1.4. Geçitli Tekrarlayan Birim (GRU)

LSTM mimarisine benzerlik gösteren GRU, unutma ve giriş kapısını tek bir güncelleme kapısında birleştirir. Hücre durumu ile gizli durumu ise tek bir gizli türde birleştirerek, güncelleme ve sıfırlama olarak adlandırılan iki adet kapı kullanır. Standart LSTM modelinden daha basit bir yapıya sahip olmasından dolayı yaygın olarak kullanılmaktadır [78].

Şekil 3.17’de gösterildiği gibi, geçitli tekrarlayan birim modeli, sıfırlama ve güncelleme kapılarına sahip denetleyiciler içermektedir. Sıfırlama kapısı, bir önceki bilginin ne kadarının unutulacağına karar verir. Güncelleme kapısı ise, bilgilerin nasıl saklanacağına ve yeni bilginin nasıl ekleneceğine karar verir.

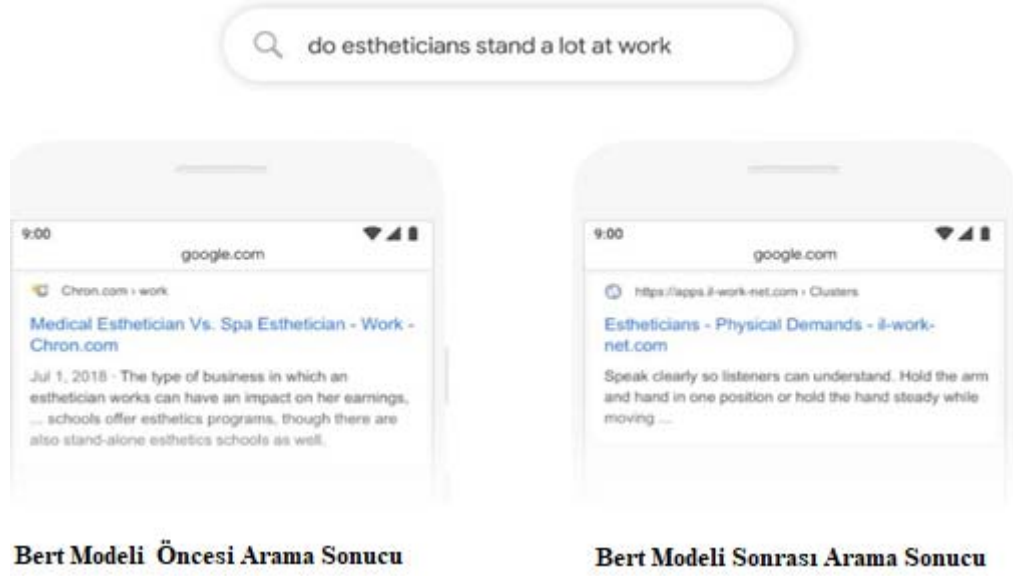


Şekil 3.17. GRU Mimarisi [79]

3.4.1.5. Çift Yönlü Kodlayıcı Temsilleri (BERT)

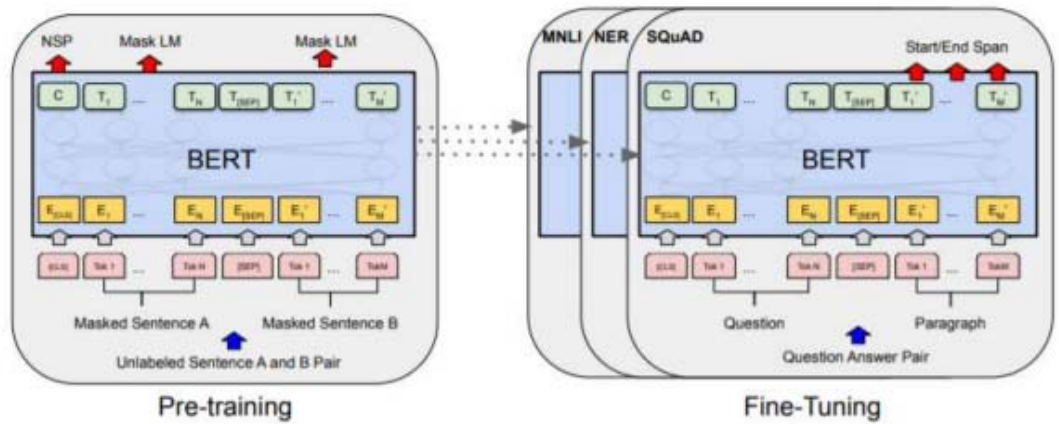
Yeni bir eğitim öncesi çift yönlü dil temsil modeli olan ve doğal dil işleme problemlerinde kullanılan BERT, 2018 yılında Google çalışanları tarafından tasarlanmıştır [52]. Google, arama sonuçlarında insan mantığına en uygun cevapları filtrelerken, sorgu kelimelerini ayrı ayrı işlemekteydi. BERT modeli ile sorgu kelimelerini mantıksal bir bütün olarak değerlendirip, daha mantıklı sonuçlara ulaşmayı sağlamıştır [80].

Google, 2020 yılında tüm İngilizce aramalarında bu modeli kullandığını duyurmuştur. Şekil 3.18’deki örnek Google sorgusunda “estetisyenler iş başında çok durur mu?” sorusuna karşılık sistem sonuç olarak “bağımsız” terimi ile sorgudaki “standart” kelimesini eşleştirerek, anahtar kelimeleri eşleştirme yaklaşımını uyguluyordu. BERT modeli kullanılarak ise, “stand” kavramının işin fiziksel talepleri ile ilgili olduğunu anlayarak daha mantıklı sonuçlar üretmektedir [81].



Şekil 3.18. Google Aramalarında BERT Öncesi ve Sonrası Örnek Gösterim [81]

Çift Yönlü Kodlayıcı Temsilleri, önceden eğitilmiş derin çift yönlü temsilleri ve cümleler arasındaki anlamsal ilişkileri keşfetmede sonraki cümle tahminini etkinleştirmek için eğitim öncesi aşamada maskeli dil modellemesini kullanmaktadır. Eğitim öncesi süreçte kullanılan yoğun mühendislik görevleri nedeniyle BERT, ilk ince ayar tabanlı temsil modeli olarak, önceden eğitilmiş modelin çok çeşitli sınıflandırma görevleri için ince ayarlı olmasını sağlamaktadır [82]. BERT algoritmasında kullanılan önceden eğitilmiş model ile ince ayar çalışma prensibi Şekil 3.19’ da gösterilmiştir.



Şekil 3.19. BERT ön eğitim ve ince ayar gösterimi [82]

BERT mimarisinde yapılan ilk aşama girdi değerlerinin parçalara ayrılarak vektörlere ayrılma işlemidir. Sonraki aşamada katmanlardan oluşan yapının içinde doğal dil işleme yöntemi gerçekleştirilir.

Günümüzde BERT algoritması kullanılarak, doğal dil işleme alanında birçok çözüm geliştirilmektedir. Metin belgelerinden özet çıkarma, cümle tahminleme, dil çevirileri ve soru cevaplama problemlerinde BERT modeli sıklıkla uygulanmaktadır [82].

3.4.2. Derin Öğrenmeye Dayalı Duygu Analizi

Son yıllarda yapılan çalışmalarda, metin madenciliği ve doğal dil işleme problemlerinde işlenecek olan metinlerin, derin öğrenme yaklaşımlarıyla analiz edilmesi, bu alandaki başarımları önemli ölçüde arttırmıştır. Özellikle evrimsel sinir ağları ve yinelenen sinir ağlarının metin çevirisi, dil modelleme ve duygu analizi gibi doğal dil işleme alanlarında başarılı sonuçlar ürettiği gözlemlenmiştir [70].

Makine öğrenmesine dayalı duygu analizi problemlerinde, metin belgesi içerisindeki kelimeler kelime çantası (BOW) yöntemiyle temsil edilmektedir. Derin öğrenmeye dayalı duygu analizi yaklaşımlarında ise, kelime gömme teknikleri uygulanarak (word2vec, LDA2vec, fastText, GloVe) temsil edilmektedir [83].

Bu çalışmadaki metin belgeleri, derin öğrenmeye dayalı duygu analizi için üç adet kelime gömme şeması (word2vec, fastText, GloVe) ile temsil edilmiştir. Temsil işlemi tamamlandıktan sonra, dört adet derin öğrenme mimarisi metni işlemek için kullanılmıştır. Bu bölümün geri kalanında çalışmada kullanılan kelime gömme yöntemleri kısaca ele alınmıştır.

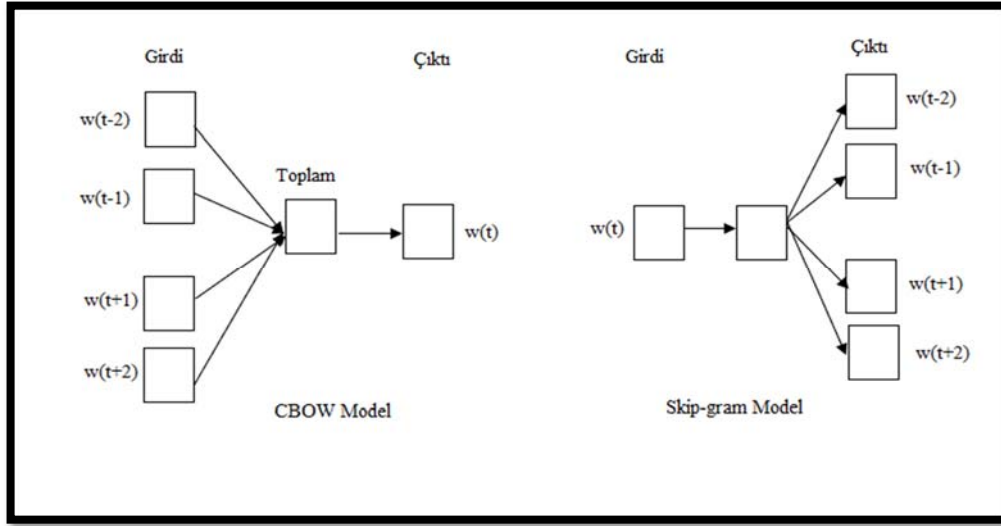
3.4.2.1. Kelime Gömme Yöntemleri

Literatürde kelime temsili olarak da bilinen ve doğal dil işleme problemlerinde yaygın olarak kullanılan kelime gömme yöntemi, metin belgesi içerisinde bulunan kelimelerin, biçimsel veya anlamsal benzerliklerinin ölçülmesi için kullanılmaktadır [84]. Kelime gömme tabanlı bir temsil, düşük boyutlu alandaki kelimeler için dağıtılmış ifadelerin öğrenirliğini sağlamaktadır [85].

3.4.2.2. Word2vec

Yapay sinir ağı tabanlı olan bu model, sinir ağlarına benzer katmanlardan oluşmaktadır. Metin belgelerinden kelime gömmeleri öğrenmek için denetimsiz, hesaplama açısından verimli bir modeldir.

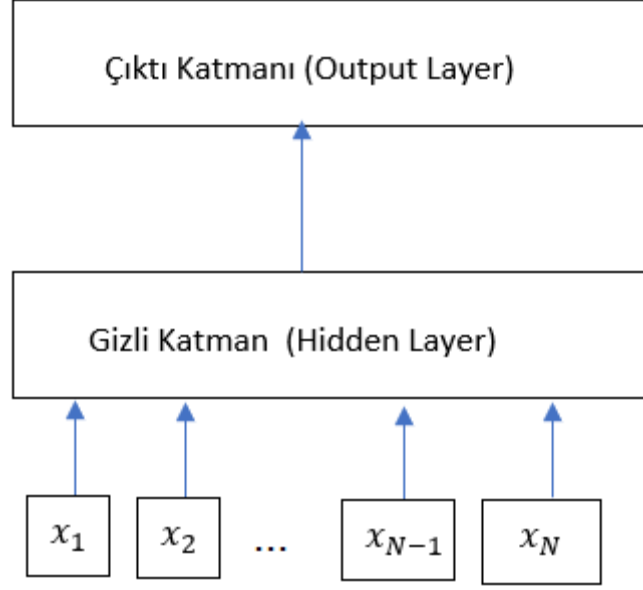
Word2vec modeli iki bileşenden oluşmaktadır. Bu bileşenler, sürekli bir kelime çantası (CBOW) mimarisi ile sürekli bir atlama mimarisi olan Skip-Gram'dır [86]. CBOW modeli, hedef kelimeyi bir k-word penceresi üzerinde çevreleyen bağlam kelimelerine göre tahmin eder. Skip-Gram modeli ise hedef kelimenin bağlam kelimelerini tahmin etmektedir.



Şekil 3.20. CBOW ve Skip-gram Modeli [87]

3.4.2.3. FastText

Metin belgelerinden kelime gömmelerini öğrenmek için kullanılan, hesaplama açısından verimli bir temsil şemasıdır. Her kelime bu modelde n-gram karakter koleksiyonu olarak kabul edilir [87]. Word2vec ile karşılaştırıldığında, fastText modeli morfolojik olarak zengin dillerde ve nadir sözcüklerde daha iyi performans gösterebilmektedir [88].



Şekil 3.21. fastText Mimari Modeli [89]

3.4.2.4. GloVe

GloVe modeli, kelimelerin vektör temsillerini oluşturmada kullanılan word2Vec tabanlı, denetimsiz bir tahmin algoritmasıdır. Bu model, küresel matris çarpanlarına ayırma şemasını, word2Vec modelinin yerel bağlama dayalı öğrenimini birleştirmek için kullanır. Model uygulanırken, metin belgesi içerisinde bulunan kelimelerin olasılık oranları dikkate alınır. Kelimelerin olasılık oranları, modelin hata fonksiyonunu hesaplamak için kullanılır. Metin belgesi içerisinde bulunan kelimelerin birlikte bulunma olasılıklarının yüksek olması, modelin eğitim sürecini olumlu yönde etkilemektedir [83].

4. MATERYAL VE YÖNTEMLER

4.1. Veri Seti

Bu çalışmada sınıflandırma aşamasında kullanılmak üzere “TURCOVIDEDU” isimli yeni bir veri seti oluşturulmuştur. Bunun için de Türkiye’de öğrenim gören önlisans, lisans ve yüksek lisans öğrencileri arasında online anket yapılarak ham veriler toplanmıştır. Her katılımcıya iki adet ucu açık soru yöneltilerek, COVID-19 pandemi sürecinde uzaktan eğitiminin avantaj ve dezavantajlarının yazılması istenmiştir. Sonuç olarak 12.018 belgeden oluşan ham bir veri seti elde edilmiştir.

Tablo 4.1. Örnek Olumlu Metin Belgeleri

Pandemi sürecinin meydana getirmiş olduğu durumlar göz önüne alındığında (psikolojik, ekonomik, sosyal) almış olduğunuz uzaktan eğitim sürecinde yaşamış olduğunuz <u>olumlu</u> bir olayı birkaç cümle ile anlatabilir misiniz?
Dersleri tekrar dinleme şansımız oluyor dersi kaçırma gibi sıkıntı olmuyor.
Temassız risksiz ders alıyoruz. En önemlisi sağlık diye düşünürsek bundan daha olumlu bir durum olamaz.
Sosyal yönden aile yanı çok önemli. İşte çalışmak ve üniversitede olduğu gibi masraf yapmamak ekonomik yönden. Psikolojik yönden de öğrenciyken tek mücadele etmek herşeyle zor ancak bu durumda uzaktan iyi
Derslerimiz gayet verimli anlaşılır. Hocalarımızla iletişim halindeyiz. Dersleri gayet iyi anlayabiliyorum. Geçen dönem corana geçirdiğim sırada sunumumu anlatacak sağlık durumum olmadığı için hocamız gayet anlayış gösterip daha sonra yapabilme imkânı tanıdı. Uygulamalı bilimler fakültesi hocalarımızın hepsi bu konularda çok destekçimiz oldular.
Dersleri durdurup ekstra kaynaklardan araştırma yapma fırsatı buldum. Ve tekrar tekrar ders kaydını izleyerek anlamadığım yerleri pekiştirme fırsatı buldum.
Ailem il dışında okuduğum için bana maddi destekte bulunuyordu. Pandemi döneminde sadece apartımın kirasını ödüyoruz tek artısı bu.
Ders çalışmak için daha fazla vakit bulabildim ve ekonomik anlamda olumlu katkılar edindim.

Tablo 4.2. Örnek Olumsuz Metin Belgeleri

Pandemi sürecinin meydana getirmiş olduğu durumlar göz önüne alındığında (psikolojik, ekonomik, sosyal) almış olduğunuz uzaktan eğitim sürecinde yaşamış olduğunuz <u>olumsuz</u> bir olayı birkaç cümle ile anlatabilir misiniz?
Sınavlarım sistem yüzünden iptal oldu, birçok ders aksiliği yaşadım vb
Okula gidemediğimiz için 1 sene boyunca ekran başında ders yapmaya çalıştık. Psikolojik olarak herkes çökmüş durumda çünkü çok uzun zamandır farklı hiçbir şey yapmıyoruz, arkadaş ortamlarından çok uzağız
Psikolojik olarak sıkılmış ve bezmiş durumdayım. Sosyal olarak yaşam enerjim kalmamıştır.
Yüz yüze eğitimde düşük ortalaması olan insanların online eğitimde ortalamalarını yükseltmeleri bana çok şüpheli geliyor.
Bağlantı sorunu yaşıyorum. Kesintisiz ders işleyemiyorum
Verim düştü. Dersleri anlamakta zorlanıyorum ki anlatımlar da güzel gelmiyor. Öğretmenlerin suçlu öğrenciymiş gibi davranması beni soğuttu ve daha çok streslendim. Olduğumdan daha gergin ve stresli hissediyorum. Sosyal hiçbir aktivitemiz yok. İmkanlar el verdiğince başka eğitimlere katılıyorum.
Psikolojik olarak sorun çektim sürekli odamda bilgisayarın başında olmak beni içe kapanıklığa itti. Konuşmamda sorun çekiyorum eskisi kadar akıcı konuşamıyorum. Ekonomik olarak zorluk çektim çünkü bilgisayarım ve internetim yoktu. Ödevler ve sorumluluklar arttı kendime vakit ayıramıyorum. Genel olarak olumsuz bir eğitim süreci.

4.2. Verilerin Toplanması

Çalışmada kullanılan veriler, Türkiye’de öğrenim gören üniversite öğrencilerine Google Form üzerinden gönderilen anket uygulaması ile toplanmıştır. Anket uygulanmadan önce Etik Kurul onayı alınmıştır. Etik kurul onayından sonra, anket linki Türkiye’de bulunan birçok üniversiteye gönderilmiştir. Yapılan ankette üniversite öğrencilerine “Covid-19 pandemi sürecinin meydan getirmiş olduğu durumlar göz önüne alındığında (psikolojik, ekonomik ve sosyal) almış olduğunuz uzaktan eğitim sürecinde yaşamış olduğunuz olumlu ve olumsuz olayları birer cümle ile anlatırmısınız?” şeklinde iki adet ucu açık soru yöneltilmiştir. Form doldurulurken gönüllülük ilkesi esas alınmıştır. Veriler yaklaşık üç ay içerisinde toplanarak, Google

Formun otomatik olarak aktardığı Microsoft Excel dosyasında incelenmek üzere biraraya getirilmiştir. Anket toplama işleminin bitiminde toplamda 12.018 adet ham veri elde edilmiştir.

Uzaktan Eğitim Değerlendirme Anketi

ARAŞTIRMA GÖNÜLLÜ KATILIM FORMU

Bu çalışma, "Covid-19 Pandemi Sürecinin Eğitim Üzerindeki Etkilerinin Makine Öğrenme Teknikleriyle Tespit Edilmesi" başlıklı bir Yüksek Lisans Tez çalışmasıdır. Çalışmanın amacı, COVID-19 pandemi sürecinde eğitimlerine uzaktan devam eden üniversite öğrencilerinin pandemi sürecinde almış oldukları uzaktan eğitimin psikolojik, ekonomik ve sosyal açılarından olumlu ve olumsuz etkilerini Türkçe metinlerde tespit edebilen bir duygu (sentiment) analiz modeli oluşturmaktır.

Gönüllü katılım formunu okumak ve değerlendirmek üzere ayırdığınız zaman için teşekkür ederiz. Çalışma hakkındaki sorularınızı kemalkarga@gmail.com adresine mail atarak iletebilirsiniz.

Bu araştırmaya katılımın gönüllü olduğu konusunda bilgilendirildim *

Evet , bu araştırmaya katılmayı kabul ediyorum

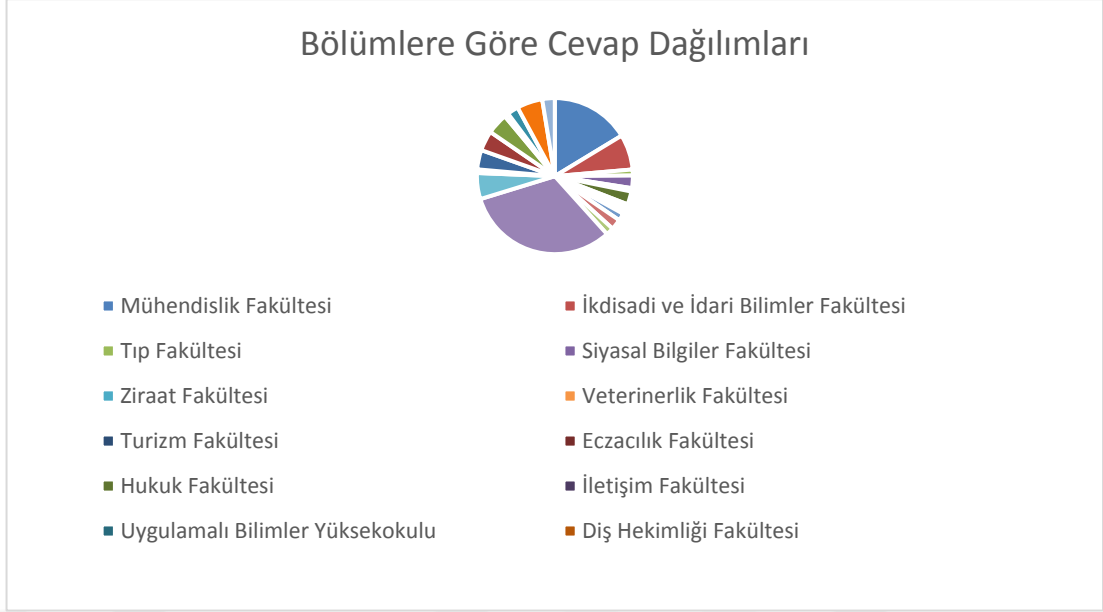
1-Pandemi sürecinin meydana getirmiş olduğu durumlar göz önüne alındığında (psikolojik, ekonomik, sosyal) almış olduğunuz uzaktan eğitim sürecinde yaşamış olduğunuz olumlu bir olayı birkaç cümle ile anlatabilir misiniz?

Uzun yanıt metni

2-Pandemi sürecinin meydana getirmiş olduğu durumlar göz önüne alındığında (psikolojik, ekonomik, sosyal) almış olduğunuz uzaktan eğitim sürecinde yaşamış olduğunuz olumsuz bir olayı birkaç cümle ile anlatabilir misiniz?

Uzun yanıt metni

Şekil 4.1. Google Form Anket Arayüzü



Şekil 4.2. Bölümlere Göre Cevap Dağılımları

4.3. Veri Ön İşleme

Veri ön işleme, metin madenciliği işlemlerinde gerekli modelin kurulmadan önce kullanılan veri seti üzerinde çeşitli işlemlerin gerçekleştirilmesidir. Genel olarak veri seti içerisindeki eksik verilerin tamamlanması, tekrar eden verilerin kaldırılması, veri temizleme ve normalleştirme adımları olarak ele alınmaktadır.

Online anket üzerinden toplanan TURCOVIDEDU veri seti, veri üzerindeki gürültünün azaltılması ve kullanılacak olan algoritmaların daha iyi performansta çalışabilmesi için bir takım veri ön işlemlerinden geçirilmiştir. Elde edilen belgeler tek bir dosya formatına getirilerek veri ön işleme adımları uygulanmıştır.

Veri seti içerisinde tekrar eden kelimeler tespit edilerek dosyadan çıkarılmıştır. Ardından sırasıyla sayısal karakterler, fazladan boşluklar ve noktalama işaretleri de kaldırılmıştır. Daha sonra, Snowball-stemmer (SS) kullanılarak her bir terim için kök çıkarma (stemming) işlemiyle veri seti normalize hale getirilmiştir.[34]

4.4. Verilerin Etiketlenmesi

Elde edilen belgelerden 5.882 tanesi olumlu (pozitif) ve kalan 6.136 tanesi ise olumsuz (negatif) olarak etiketlenmiştir. Veri setini doğrulamak adına, iki ek yorumcunun her bir belgeyi olumlu, olumsuz ya da nötr olarak etiketlediği ek bir

açıklama süreci gerçekleşmiştir. Değerlendiriciler arasında uyumsuzluk olması durumunda ilgili belge iptal edilmiştir. Onaylama süreci sonrası toplamda 606 adet belge iptal edilmiştir. İptal edilen belge örneklerinden bazıları Tablo 4.3'te gösterilmiştir.

Tablo 4.3. İptal Edilen Metin Belgesi Örnekleri

Uzaktan eğitim süreci benim için normal eğitim gibiydi. Fark göremedim.
Pandemi derdine düşmekten eğitim iyimi kötümü bilemedim açıkcası.
Olumlu bir olayda yok olumsuz bir durumda, söz konusu olmadı herşey normaldi.
Son sınıfta olduğum için derslere zaten gitmiyorum.
Uzaktan eğitim fark yok açıkcası okul bitti gibi bişey zaten.
Staj dönemindeyim belirgin bir durum yok nötr herşey.
İstedğim bölümde değilim sınava girip başka bir bölüm seçeceğim.
Sosyal etki.
Uzaktan eğitimin olumluda olumsuzda birçok etkisi var.
Hangi olumlu olay?
Pandemi döneminde uzaktan eğitimin olumlu veya olumsuz bir tarafıyla karşılaşmadım açıkcası.
Okul hayatım gayet normal bir şekilde devam ediyor. Beni etkileyen bir durum yok.
Olumlu da çok şey var olumsuzda.
Sektör eğitimindeyim derslerim bitti, uzaktan eğitim almıyorum.

Tablo 4.4'te veri setindeki ham ve değerlendirilmiş belgelerin dağılımı gösterilmektedir. Ayrıca, değerlendiriciler arasında uyum olduğunu gösteren Cohen' in kappa (K) katsayısı 0,82 olarak hesaplanmıştır.

Tablo 4.4. Ham ve Değerlendirilmiş Belge Dağılımları

	Olumlu	Olumsuz	Toplam
Ham	5,882	6,136	12,018
Değerlendirilmiş	5,588	5,824	11,412

4.5. Deneysel Prosedür

Deneysel analizde, dört adet tanınmış geleneksel makine öğrenme algoritmasının (SVM, NB, LR, RF), derin öğrenme mimarilerinin (LSTM, GRU, RNN, CNN) ve BERT sınıflandırıcısının sınıflandırma performansları karşılaştırılarak, COVID-19 pandemisinde uzaktan eğitimin etkinliğine odaklanılmıştır. Derin öğrenme mimarileri ve BERT sınıflandırıcısı üzerine deneyler için önceden eğitilmiş modeller kullanılarak, on kat çapraz doğrulama kullanılmıştır.

Çalışmadaki tüm durumlarda hiper parametre optimizasyonu kullanılmıştır. Girdi dizisi uzunluğunu her gömme katmanının 100' ü olarak ayarlanıp, tüm derin öğrenme mimarilerinde kayıp fonksiyonu için binary_crossentropy, optimize edici olarak adamx kullanılmıştır. Her bir mimarinin yoğun sinir ağının (DNN) katmanları sırasıyla girdi, gizli ve çıktı olarak oluşturulmuştur. Tüm ağların bırakma (dropout) oranı 0,3, gizli katman nöron sayısı 50 ve aktivasyon fonksiyonu relu olarak ayarlanmıştır.

CNN mimarisinin yapısı ile ilgili olarak, evrişimli katman parametrelerini, aktivasyon fonksiyonunu softmax, çekirdek boyutunu 16 ve filtre değeri 100 olarak ayarlanmıştır. Havuzlama boyutu 2 olan bir sonraki katman olarak havuzlama kullanılmıştır. Sonraki adımda 2B havuzlanmış özellik haritalarını tek boyutlu bir vektöre düzleştirilerek, LSTM, GRU ve RNN mimarilerinde birim boyutu 100 olarak belirlenmiştir. BERT sınıflandırıcısında, modelin parametrelerini küme büyüklüğü: 32, öğrenme oranı: 5×10^{-5} , eğitim dönemi sayısı: 4 ve optimize edici: Adam. olarak belirlenmiştir.

İlk olarak, veri setinin normalleşme (normalizasyon) sürecinin etkilerine odaklanılmıştır. Bunun için, veri setini ham ve SS olan iki kök formda kullanarak dört farklı geleneksel makine öğrenme algoritmasının (SVM, NB, LR ve RF) ve dört farklı

derin öğrenme mimarisinin (CNN, RNN, LSTM, GRU) tahmine dayalı performanslarını karşılaştırılmıştır. Tablo 4.5 ve 4.6’da gösterilen genel doğruluk değerlerine göre, ham form yerine kökü bulunmuş (stemmed) veri seti kullanarak tüm durumlar için daha yüksek performanslar elde edilmiştir. Aşağıdaki tüm deneylerde SS stemmer kullanılarak her terimin kökü belirlenmiştir.

Tablo 4.5. Dört makine öğrenme mimarisinin doğruluk sonuçlarının iki farklı stemmer yöntemi açısından karşılaştırılması

Stemmer	SVM	NB	LR	RF
Ham	0.8962	0.8925	0.8922	0.8570
SS	0.9026	0.8968	0.8941	0.8671

Tablo 4.6. Dört derin öğrenme mimarisinin doğruluk sonuçlarının iki farklı stemmer yöntemi açısından karşılaştırılması

Stemmer	CNN	RNN	LSTM	GRU
Ham	0.9081	0.8946	0.9047	0.9029
SS	0.9084	0.8977	0.9083	0.9102

Tablo 4.7’de dört geleneksel makine öğrenme algoritmalarının (SVM, NB, LR ve RF) üç farklı özellik çıkarma yöntemi (TF, TP ve TF-IDF) ve üç N-gram modeli (unigram, bigram ve trigram) kullanılarak sınıflandırma performansları karşılaştırılmıştır.

Algoritmaların performanslarına bakıldığında, SVM, NB ve LR, kısmen benzer sonuçlar vermiş ve RF, diğerleri arasında en düşük doğruluk sonucunu elde etmiştir. Öte yandan, verilerin unigram modeli ve TF-IDF ağırlıklandırma şeması kullanılarak temsili, 0,8901 olarak en yüksek ortalama doğruluk sonucunu elde etmiştir.

Tablo 4.7. Farklı temsil yöntemleri kullanılarak dört geleneksel makine öğrenmesi algoritmasının sınıflandırma performanslarının karşılaştırılması.

Temsil	SVM	NB	RF	LR	Ortalama
Unigram, TF	0.8840	0.8963	0.8684	0.8959	0.8862
Unigram, TP	0.8893	0.8974	0.8663	0.8955	0.8871
Unigram, TF-IDF	0.9026	0.8968	0.8671	0.8941	0.8901
Bigram, TF	0.8850	0.8919	0.8672	0.8919	0.8840
Bigram, TP	0.8886	0.8937	0.8636	0.8932	0.8848
Bigram, TF-IDF	0.8991	0.8915	0.8634	0.8912	0.8863
Trigram, TF	0.8847	0.8905	0.8642	0.8920	0.8828
Trigram, TP	0.8877	0.8929	0.8645	0.8917	0.8842
Trigram, TF-IDF	0.8983	0.8910	0.8631	0.8910	0.8858

Bir sonraki deneyin amacı; üç farklı gömme yöntemini (Glove, Word2Vec SG ve Word2Vec CBOW) kullanarak iki farklı gömme katmanı yapı formunun (kendi kendine eğitilmiş ve önceden eğitilmiş) tahmine dayalı performanslarını, dört derin öğrenme mimarisıyla (CNN, RNN, LSTM and GRU) ile birlikte karşılaştırmaktır.

Kendi kendine eğitilmiş formda, TURCOVIDEDU veri setini kullanarak gömme katmanı eğitilmiştir. Bununla birlikte, önceden eğitilmiş formda, gömme katmanlarının eğitim sürecinde COVID-19 ve eğitim haberleriyle ilgili 116K dokümandan oluşan yeni bir ham veri kümesi kullanılmıştır. Tablo 4.8 ve 4.9, sırasıyla kendi kendine eğitilmiş ve önceden eğitilmiş olarak gömme katmanlarını oluşturan üç farklı gömme yöntemiyle bağlantılı olarak dört derin öğrenme mimarisinin doğruluk değerleri açısından tahmin performanslarını göstermektedir.

Tablo 4.8. Üç farklı gömme yöntemi kullanılarak oluşturulan kendi kendini eğiten gömme katmanlarını kullanan dört derin öğrenme mimarisinin doğruluk değerlerinin karşılaştırılması

Gömme Yöntemi	CNN	RNN	LSTM	GRU
Glove	0.9041	0.8896	0.9064	0.9090
Word2Vec SG	0.9053	0.8924	0.9062	0.9069
Word2Vec CBOW	0.9047	0.8924	0.9010	0.9019

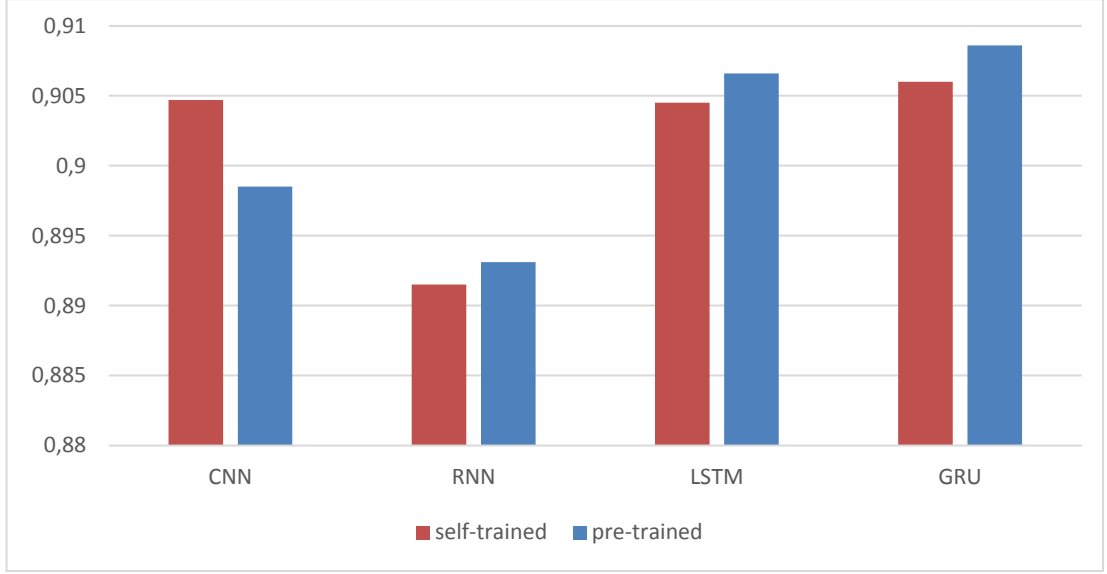
Tablo 4.9. Üç farklı gömme yöntemi kullanılarak oluşturulmuş önceden eğitilmiş gömme katmanları kullanılarak dört derin öğrenme mimarisinin doğruluk değerlerinin karşılaştırılması

Gömme Yöntemi	CNN	RNN	LSTM	GRU
Glove	0.8946	0.8885	0.9040	0.9055
Word2Vec SG	0.9016	0.8973	0.9117	0.9130
Word2Vec CBOW	0.8993	0.8934	0.9041	0.9074

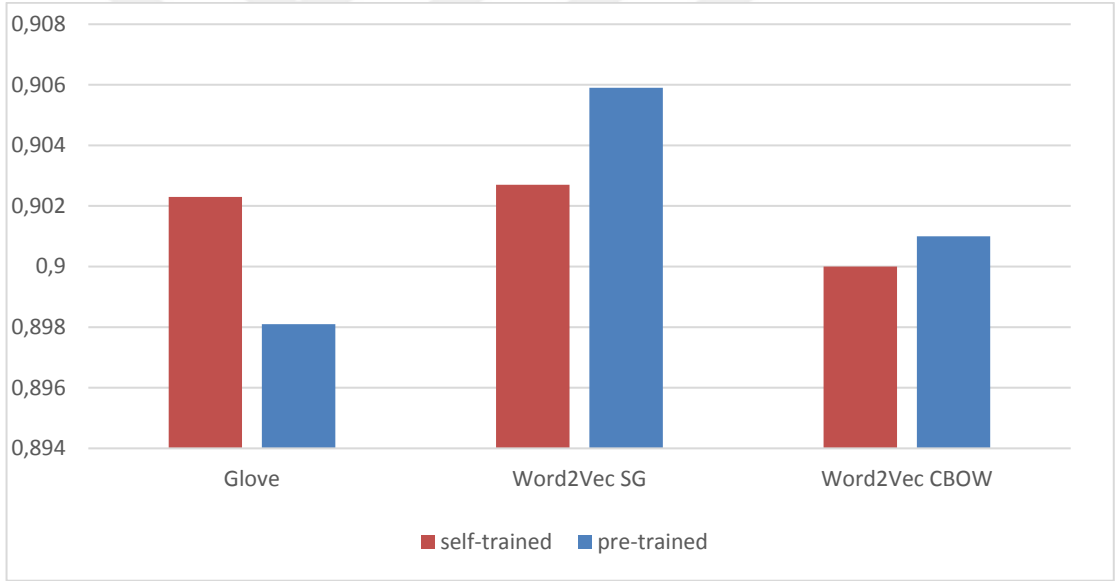
Dört derin öğrenme mimarisinin tahmine dayalı performanslarıyla ilgili olarak, GRU hemen her durumda diğer sınıflandırıcılardan daha iyi performans göstermiştir. LSTM ikinci en yüksek performansı; CNN en yüksek üçüncü performansı göstermiştir, en düşük tahmin performansları ise RNN sınıflandırıcısı tarafından elde edilmiştir.

Gömme yöntemlerinin tahmin edici performansları ile ilgili olarak, Word2Vec türü yöntemler, Glove'a kıyasla hemen her durumda daha yüksek doğruluk değerleri gerçekleştirmiştir. Kendinden eğitilmiş ve önceden eğitilmiş gömme katman yapı formlarının kullanım performansları göz önüne alındığında, farklı durumlarda her iki form için daha yüksek doğruluk değerleri elde edilmiştir. Tüm CNN mimarilerinde ve Glove ile ilgili sonuçlarda kendi kendine eğitilmiş gömme katmanlarının kullanılması daha yüksek performanslar sağlamıştır.

RNN, LSTM ve GRU mimarileri Word2Vec tipi gömme yöntemleri ile oluşturulmuş önceden eğitilmiş gömme katmanları kullanılarak daha yüksek doğruluk değerleri elde etmişlerdir. Bunun yanı sıra, GRU, Word2Vec SG gömme yöntemi ve önceden eğitilmiş yapı formunun birleşimi, diğerleri arasında 0,9130 ile en yüksek doğruluk değerini elde etmiştir. Şekil 4.3'te, iki gömülü katman oluşturma formu açısından dört derin öğrenme mimarisinin ortalama doğruluk sonuçlarının karşılaştırılması ve Şekil 4.4'te, iki gömülü katman oluşturma formu açısından üç gömme yönteminin ortalama doğruluk sonuçlarının karşılaştırılması gösterilmektedir.



Şekil 4.3. Dört derin öğrenme mimarisinin ortalama doğruluk sonuçlarının iki gömülü katman oluşturma formları açısından karşılaştırılması.



Şekil 4.4. Üç gömme yönteminin ortalama doğruluk sonuçlarının iki gömme katmanı yapı formları açısından karşılaştırılması.

Tablo 4.10' da BERTurk önceden eğitilmiş dil modelinin dört farklı formunu kullanarak BERT sınıflandırıcısının sınıflandırma performanslarını doğruluk, f-skoru, geri çağırma ve hassasiyet (kesinlik) değerleri açısından karşılaştırılmıştır. Sonuçlara göre BERTurk_cased (büyük küçük harf duyarlı) önceden eğitilmiş dil modelini kullanarak bu çalışmada 0.9666 olarak en yüksek doğruluk değeri elde edilmiştir. Ayrıca, büyük küçük harfe duyarlı ve duyarsız modellerin performanslarını

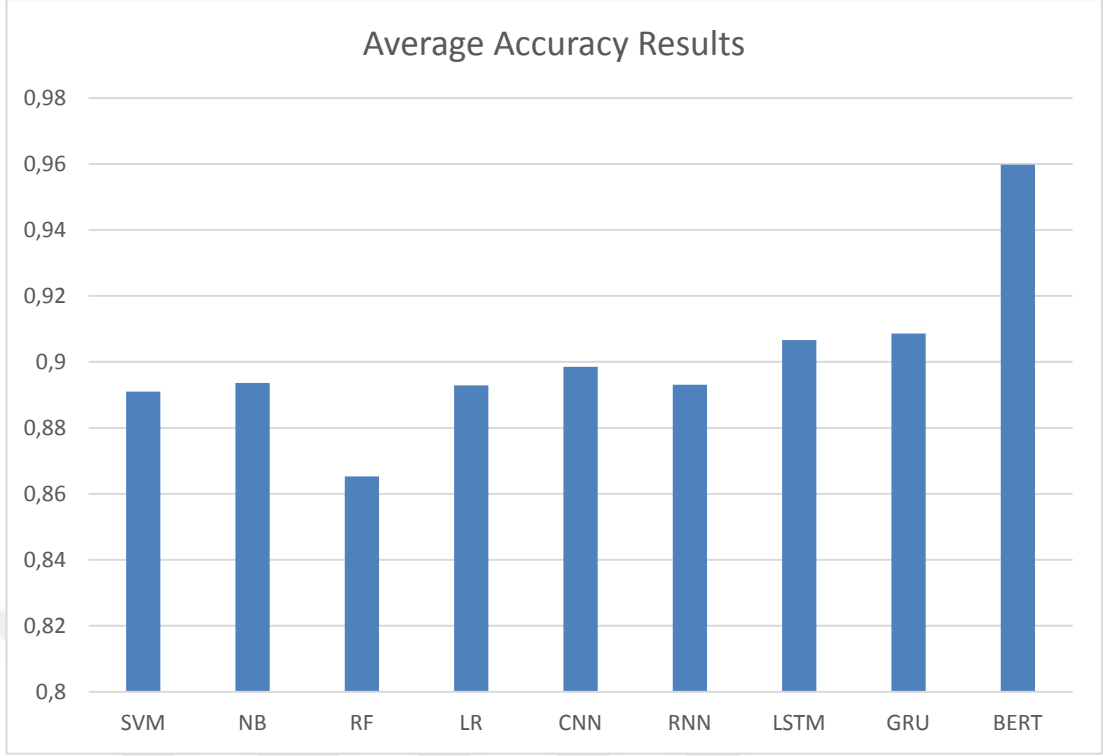
karşılaştırılarak, büyük küçük harfe duyarlı durumda model her durumda daha yüksek sonuçlar elde etmiştir. Öte yandan, önceden eğitilmiş bir modelin kelime boyutunun 128K'ya düşürülmesi, her durumda daha yüksek performans göstermemiştir.

Tablo 4.10. BERTurk önceden eğitilmiş dil modelinin dört farklı formu kullanılarak BERT sınıflandırıcısının sınıflandırma performansının karşılaştırılması.

Önceden Eğitilmiş Dil Modelleri	Doğruluk	f-score	Geri Çağırma	Kesinlik (Hassasiyet)
BERTurk_büyük-küçük harfe duyarlı	0.956117	0.956112	0.956259	0.956102
BERTurk_büyük küçük harfe duyarlı	0.966607	0.966595	0.966640	0.966560
BERTurk_büyük küçük harfe duyarlı 128K	0.957909	0.957896	0.957975	0.957846
BERTurk_büyük küçük harfe duyarlı 128K	0.958517	0.958510	0.958644	0.958472

Şekil 4.5'te, deneysel analizde kullanılan tüm makine öğrenimi yaklaşımlarının (geleneksel makine öğrenmesi algoritmaları, derin öğrenme algoritmaları ve BERT sınıflandırıcısı) ortalama doğruluk sonuçları karşılaştırılmıştır. BERT sınıflandırıcısı en yüksek ortalama doğruluk değerini 0.9598 ile gerçekleştirmiştir. LSTM ve GRU mimarileri benzer sonuçlar elde etmesine rağmen GRU, 0.9086 ile ikinci en yüksek ortalama doğruluk sonucunu elde etmiştir. Öte yandan, 0,8931 ile RNN mimarisi derin öğrenme mimarileri arasında en düşük performansı göstermiştir.

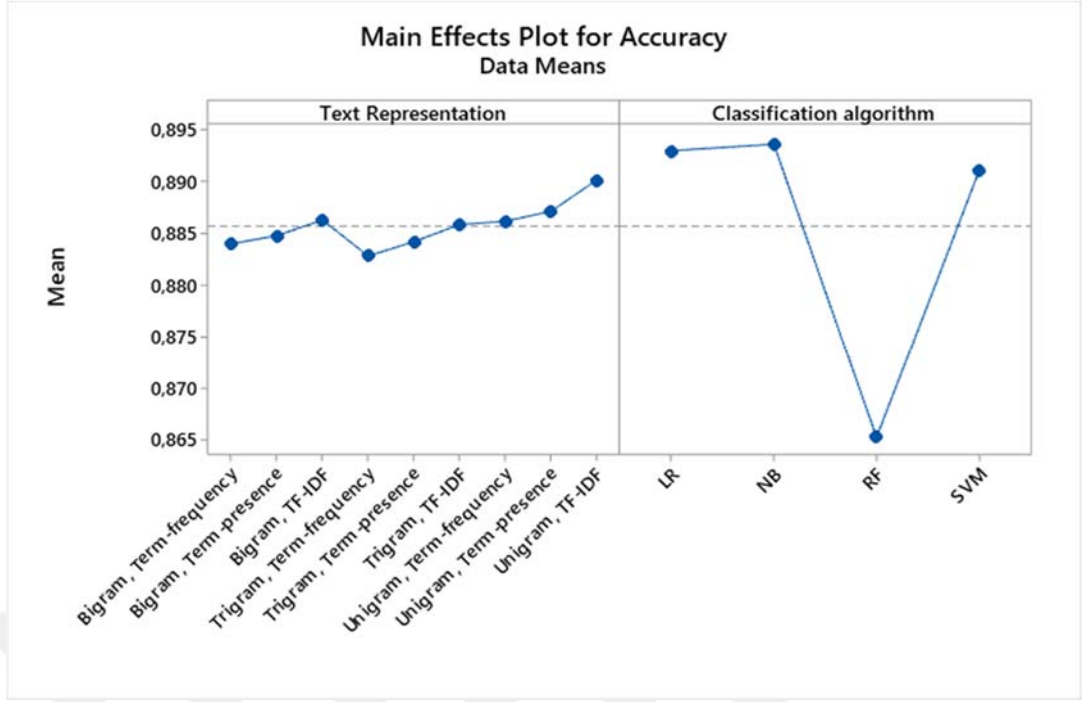
Geleneksel makine öğrenmesi algoritmalarının ortalama doğruluk sonuçlarına bakıldığında, NB ve LR algoritmaları, geleneksel algoritmalar arasında en yüksek değerleri gerçekleştirmiştir ve bunları SVM sınıflandırıcısı takip etmiştir. Son olarak RF ampirik analizin tüm sınıflandırıcıları arasında 0.8653 ile en düşük sonucu elde etmiştir.



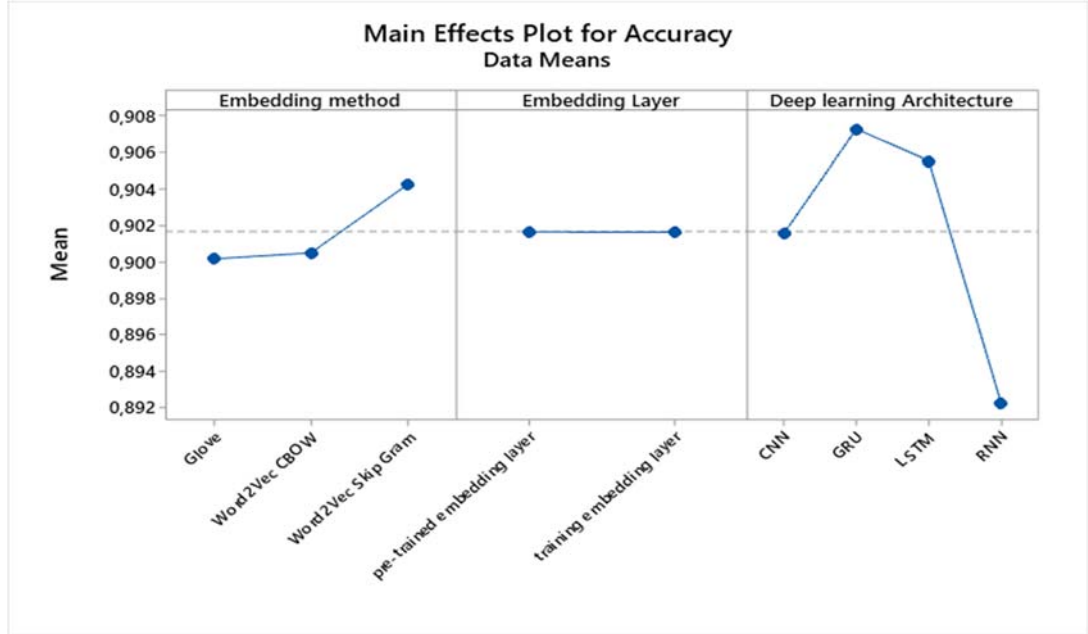
Şekil 4.5. Her üç makine öğrenimi yaklaşımının ortalama doğruluk performanslarının karşılaştırılması.

Deneysel sonuçların temel bulgularını özetlemek gerekirse, geleneksel makine öğrenimi algoritmaları ve derin öğrenme mimarileri için doğruluk değerlerine yönelik başlıca etkiler sırasıyla Şekil 4.6 ve Şekil 4.7’de gösterilmiştir. Şekil 4.6’da görülebileceği gibi, geleneksel makine öğrenimine dayalı duyu analizi ile ilgili olarak, TF-IDF terim ağırlıklandırması ile unigram metin temsil şeması, diğer metin temsil şemalarından daha iyi performans göstermektedir. Geleneksel denetimli öğrenme yöntemlerinin tahmine dayalı performansı ile ilgili olarak, Naïve Bayes algoritması diğer sınıflandırıcılardan daha iyi performans göstermiştir.

Şekil 4.7’de derin öğrenmeye dayalı duyu analizi ile elde edilen sonuçlar özetlenmiştir. Şekil 4.7’den de görülebileceği gibi, word2vec (Skip (atlama) Gram modeli) diğer sinir dili modellerinden daha iyi performans göstermiştir. Eğitim gömme katmanı, önceden eğitilmiş gömme katmanına nazaran daha yüksek tahmine dayalı performans sağlamıştır. Ayrıca, kapılı tekrarlayan ünite, görevlerde diğer geleneksel derin sinir ağlarından daha iyi performans göstermektedir.



Şekil 4.6. Makine öğrenimi mimarileri tarafından elde edilen doğruluk değerlerinin temel etkiler grafiği.



Şekil 4.7. Derin öğrenme mimarileri tarafından elde edilen doğruluk değerlerinin temel etkiler grafiği

5. SONUÇLAR

COVID-19 pandemisinde uzaktan eğitimin olumlu ve olumsuz etkilerinin otomatik olarak belirlenmesi, ilgili durumlara sosyal medya platformlarında sıklıkla rastlanabildiğinden, doğal dil işlemede zorlu bir iş olarak kabul edilebilir. Bu çalışma üç temel katkı sağlamıştır. Birincisi, Türkiye’ de öğrenim gören üniversite öğrencileri arasında bir anket yapılarak, 11.412 doğrulanmış belgeden oluşan yeni bir veri seti elde edilerek, doğrulama işlemi gerçekleştirilmiştir. İkincisi, üç farklı makine öğrenimi yaklaşımının kapsamlı bir ampirik analizi gerçekleştirilmiştir. Üçüncüsü ise, derin öğrenme mimarilerinde kullanılmak üzere 116.085 belgeden oluşan yeni bir ham veri kümesi kullanarak önceden eğitilmiş yeni bir gömme katmanı oluşturulmuştur.

Dört geleneksel makine öğrenimi algoritması (SVM, NB, LR, RF), dört derin öğrenme mimarisi (CNN, RNN, LSTM, GRU) ve BERT sınıflandırıcısı olmak üzere üç farklı makine öğrenimi yaklaşımı kullanarak bu çalışmanın ampirik analizinde umut verici tahmine dayalı sonuçlar elde edilmiştir. Genel sonuçlara göre BERT sınıflandırıcısı, 0,9598 olarak ortalama doğruluk sonucu elde ederek her durumda hem derin öğrenme yaklaşımlarından hem de geleneksel makine öğrenimi algoritmalarından daha iyi performans göstermiştir. İkinci en yüksek ortalama doğruluk değerleri sırasıyla derin öğrenme mimarileri GRU, LSTM ve CNN tarafından gerçekleştirilmiştir.

Derin öğrenme mimarilerine gömme katmanlarının oluşturulması aşamasında üç farklı gömme yönteminin (Glove, Word2Vec SG ve Word2Vec CBOW) performansları karşılaştırılmıştır. Genel sonuçlara bakıldığında, Word2Vec SG 0.9059 ile en yüksek ortalama doğruluk değerini gerçekleştirmiştir. Bunun yanı sıra, derin öğrenme mimarilerinde kullanılmak üzere yeni bir ham veri kümesi kullanarak önceden eğitilmiş yeni bir gömme katmanı oluşturulmuştur. Sonuçlara göre, genel olarak kendi kendine eğitilmiş gömme yapı formunu kullanarak daha yüksek performanslar elde edilmiştir.

Bu çalışma çeşitli boyutlarda genişletilebilir. İlk olarak, iki veri kümesindeki belgelerin sayısı artırılabilir. İkinci olarak ise, yeni bir derin öğrenme mimarisi önermek faydalı olabilecektir.

KAYNAKLAR

- [1] Chawla, S.K., Mittal, M., Chawla, M., Goyal, L.M. Corona Virus- SARS-CoV-2: An Insight to Another way of Natural Disaster. EAI Endorsed Transactions on Pervasive Health and Technology. 2020, 6(22), 1-9.
- [2] Wang, L.W., Lo, K. Text mining approaches for dealing with the rapidly expanding literature on COVID-19. Briefings in Bioinformatics. 2021, 22(2), 781–799.
- [3] Gajewski, K.N., Peterson, A.E., Chitale, R.A., Pavlin, J.A., Russell, K.L., Chretien, J.P. A Review of Evaluations of Electronic Event-Based Biosurveillance Systems. PLoS One. 2014, 9(10), e111222.
- [4] Bismala, L., Manurung, Y.H. Student satisfaction in e-learning along the COVID-19 pandemic with importance performance analysis. International Journal of Evaluation and Research in Education. 2021, 10(3), 753-759.
- [5] Daniel, S.J. Education and the COVID-19 pandemic. Prospects. 2020, 49, 91-96.
- [6] Bilecen, B. Commentary: COVID-19 Pandemic and Higher Education: International Mobility and Students' Social Protection. INTERNATIONAL MIGRATION. 2020, 58(4), 263–266.
- [7] Onan, A., Korukoğlu S., Bulut, H. Ensemble of keyword extraction methods and classifiers in text classification. Expert Systems with Applications. 2016, 57, 232–247.
- [8] Onan, A. Classifier and feature set ensembles for web page classification. Journal of Information. 2016, 42(2), 150-165.
- [9] Onan, A., Korukoğlu, S., Bulut, H. A multiobjective weighted voting ensemble classifier based on differential evolution algorithm for text sentiment classification. Expert Systems with Applications. 2016, 62, 1–16.
- [10] Onan, A., Korukoğlu, S. A feature selection model based on genetic rank aggregation for text sentiment classification. Journal of Information Science. 2015, 43(1), 25–38.
- [11] Onan, A. Hybrid supervised clustering based ensemble scheme for text classification. Kybernetes. 2017, 46(2), 330-348.

- [12] Onan, A., Tocoglu, M.A. A Term Weighted Neural Language Model and Stacked Bidirectional LSTM Based Framework for Sarcasm Identification. *IEEE Access*. 2021, 9, 7701-7722.
- [13] Onan, A., Korukoğlu, S., Bulut, H. A hybrid ensemble pruning approach based on consensus clustering and multi-objective evolutionary algorithm for sentiment classification. *Information Processing and Management*. 2017, 53(4), 814–833.
- [14] Toçoğlu, M.A., Onan, A. Satire detection in Turkish news articles: A machine learning approach,” in *Communications in Computer and Information Science*, Cham: Springer International Publishing, 2019, pp. 107–117.
- [15] Onan, A. Review spam detection based on psychological and linguistic features. 2018 26th Signal Processing and Communications Applications Conference (SIU), 2-5 Mayıs, 2018, İzmir (Bildiriler Kitabı, 1-4)
- [16] Onan, A. An ensemble scheme based on language function analysis and feature engineering for text genre classification. *Journal of Information Science*. 2018, 44(1), 28-47.
- [17] Rahmanian, V., Jahanbin, K. Using twitter and web news mining to predict COVID-19 outbreak. *Asian Pacific Journal of Tropical Medicine*. 2020, 13(8), 378-380.
- [18] Ordun, C., Purushotham, S., Raff, E. Exploratory analysis of covid-19 tweets using topic modeling, UMAP, and DiGraphs. *arXiv [cs.SI]*, 2020.
- [19] Peng, Z., Wang, R., Liu, L., Wu, H. Exploring urban spatial features of COVID-19 transmission in Wuhan based on social media data. *ISPRS International Journal of Geo-Information*. 2020, 9(6), 402.
- [20] Li, D., Chaudhary, H., Zhang, Z. Modeling spatiotemporal pattern of depressive symptoms caused by COVID-19 using social media data mining. *International Journal of Environmental Research and Puplic Health*. 2020, 17(14), 4988.
- [21] Chen, N., Zhong, Z., Pang, J. An exploratory study of COVID-19 information on Twitter in the Greater Region. *Big data and cognitive computing*. 2021, 5(1), 5.
- [22] Boon-Itt, S., Skunkan, Y. Public perception of the COVID-19 pandemic on Twitter: Sentiment analysis and topic modeling study. *JMIR Public Health and Surveilance*. 2020, 6(4), e21978.

- [23] Onan, A. COVID-19 ile İlgili Sosyal Medya Gönderilerinin Metin Madenciliği Yöntemlerine Dayalı Olarak Zaman-Mekansal Analizi. *European Journal of Science and Technology*, 2021, 26, 138-143.
- [24] Onan, A., Toçoğlu, M. A. Weighted word embeddings and clustering-based identification of question topics in MOOC discussion forum posts. *Computer Applications in Engineering Education*. 2021, 29(4), 675–689.
- [25] Onan, A. Sentiment analysis on massive open online course evaluations: A text mining and deep learning approach. *Computer Applications in Engineering Education*. 2021, 29(3), 572–589.
- [26] Bustillos, R.O, Cabada, R.Z., Barrón Estrada, M.L., and Hernández Pérez, Y. Opinion mining and emotion recognition in an intelligent learning environment. *Computer Applications in Engineering Education*. 2019, 27(1), 90-101.
- [27] Cabada, R. Z., Estrada, M. L. B., Bustillos, R. O. Mining of educational opinions with deep learning. *Journal of Universal Computer Science*. 2018, 24(11), 1604–1626.
- [28] Nguyen, H.T., Nguyen, M.L. Multilingual opinion mining on YouTube – A convolutional N-gram BiLSTM word embedding. *Information Processing and Management: an International Journal*. 2018, 54(3), 451–462.
- [29] Lin, Q., Zhu, Y., Zhang, S., Shi, P., Guo, Q., Niu, Z. Lexical based automated teaching evaluation via students’ short reviews. *Computer Applications in Engineering Education*. 2019, 27(1), 194–205.
- [30] López, M.B, Alor-Hernández, G., Sánchez-Cervantes, J.L., Pilar Salas-Zárate, M., Paredes-Valverde, M.A. EduRP: an Educational Resources Platform based on Opinion Mining and Semantic Web. 2018, *Journal of Universal Computer Science*. 24(11), 1515–1535.
- [31] Chen, T., Peng, L., Jing, B., Wu, C., Yang, J., Cong, G. The impact of the COVID-19 pandemic on user experience with online education platforms in China. *Sustainability*. 2020, 12(18), 7329.
- [32] Komasa, N., Terasaki, F., Nakano, T., Saura, R., Kawata, R. A text mining analysis of perceptions of the COVID-19 pandemic among final-year medical students. *Acute Medicine and Surgery*. 2020, 7(1), e576.

- [33] Kim, E.J., Kim, J. J., Han, S.H. Understanding student acceptance of online learning systems in higher education: Application of social psychology theories with consideration of user innovativeness. *Sustainability*. 2021, 13(2), 896.
- [34] Keskin, M., Kaya, D.Ö. COVID-19 Sürecinde Öğrencilerin Web Tabanlı Uzaktan Eğitime Yönelik Geri Bildirimlerinin Değerlendirilmesi. *İzmir Kâtip Çelebi Üniversitesi Sağlık Bilimleri Fakültesi Dergisi*, 2020, 5(2), 59-67.
- [35] Daroedono, E., Siagian, F. E., Alfarabi, M., Cing, J. M., Arodes, E. S., Sirait, R. H., Hutabarat, R. S. O. The impact of COVID-19 on medical education: our students perception on the practice of long distance learning. *International Journal of Community Medicine and Public Health*, 2020, 7(7), 2790-2796.
- [36] Osmanoglu, U.O., Atak, O.N., Caglar, K., Kayhan, H., Can, T.C. Sentiment Analysis for Distance Education Course Materials: A Machine Learning Approach, *Journal of Educational Technology & Online Learning*, 2020, 3, 31-48.
- [37] Çevrimiçi: IBM: <https://www.ibm.com/cloud/learn/text-mining> (Erişim Tarihi: 02.03.2022)
- [38] Kılınç, D., Borandağ, E., Yücalar, F., Tunalı, V., Şimşek, M., Özçift, A. KNN Algoritması ve R Dili ile Metin Madenciliği Kullanarak Bilimsel Makale Tasnifi. *Marmara Fen Bilimleri Dergisi*. 2016, 28(3), 89-94.
- [39] Çevrimiçi:AnalyticsVidyha:
<https://www.analyticsvidhya.com/blog/2015/06/blog/machine-learning-basics/> (Erişim Tarihi: 12.04.2022)
- [40] Çevrimiçi: AWS: <https://aws.amazon.com/blogs/big-data/applying-machine-learning-to-text-mining-with-amazon-s3-and-rapidminer/> (Erişim Tarihi: 19.04.2022)
- [41] Akay, E.Ç. *Ekonometride Büyük Veri ve Makine Öğrenmesi: Temel Kavramlar, Der Yayınları, İstanbul, 2020, 208 s.*
- [42] Yaraş, N. *Vehicle Type Classification With Deep Learning. İzmir Institute of Technology, Computer Engineering, İzmir, 2020, 77 s. (Yüksek Lisans Tezi)*
- [43] Alpaydın, E. *Introduction to Machine Learning, MIT Press, Cambridge, 2020, 537 s.*

- [44] Sindhu, V., Nivedha, S., Prakash, M. An Empirical Science Research on Bioinformatics in Machine Learning. *Journal of Mechanics of Continua and Mathematical Sciences*. 2020, 7, 86-94.
- [45] Russel, S., Norvig, P. *Artificial Intelligence: A Modern Approach*, Prentice Hall, New Jersey, 2003, 905 s.
- [46] Kılınç, D., Başeğmez, N. *Uygulamalarla Veri Bilimi*, Abaküs Yayınları, İstanbul, 2018, 158 s.
- [47] Jordan, M.I., Bishop, C.M. *Neural Networks, Second Edition Computer Science Handbook*, Ed: Allen B. Tucker, Florida, 2004.
- [48] Van Otterlo, M., Wiering M. *Reinforcement Learning and Markov Decision Processes*. 2012, 12, 3-42.
- [49] Vapnik, V. *The Nature of Statistical Learning Theory*, Springer, New York, 2014, 314 s.
- [50] Karagül K. İstanbul Menkul Kıymetler Borsasında İşlem Gören Firmaların Destek Vektör Makineleri Kullanılarak Sınıflandırılması. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*. 2014, 20(5), 174-178.
- [51] Yıldız, A., Zan, H. Segmantasyon Yapmadan Patolojik Kalp Sesi Kayıtlarının Tespiti İçin Bir Örüntü Sınıflandırma Algoritması. *Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi*. 2019, 10(1), 77-91.
- [52] Gunn, S.R. *Support Vector Machines for Classification and Regression*. Department of Science and Mathematics Engineering, University of Southampton, Southampton, UK, ISIS Technical Report, 14, 1998.
- [53] Li, X., Li, S., Li, J., Yao, J., Xiao, X. Detection of fake-video uploaders on social media using Naive Bayesian model with social cues. *Scientific Reports*. 2021, 11(1), 16068.
- [54] Hastie, T., Tibsharani, R., Friedman, J. *Springer Series in Statistics The Elements of. Mathematical Intelligencer*. 2009, 27(819), 83–85.
- [55] Keskin, M.V. *Büyük Veride Makine Öğrenmesi Uygulaması*. Yıldız Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Uygulamalı İstatistik Anabilim Dalı, İstanbul, 2018, 71 s. (Yüksek Lisans Tezi)
- [56] Breiman, L., Last, M., Rice, J. Random forests: Finding Quasars. *Statistical Challenges in Astronomy*. 2003, 243-254.

- [57] Akman, M., Genç, Y., Ankaralı, H. Random Forest Yöntemi ve Sağlık Alanında Bir Uygulama. *Türkiye Klinikleri Biyoistatistik*, 2011, 3(1), 36.
- [58] Subramaniam, P., Kaur, M.J. Review of security in mobile edge computing with deep learning. *Advances in Science and Engineering Technology International Conferences*. 2019, 1-5.
- [59] Diaz, P., Salas, J.C., Cipriano, A., Nunez, F. Random Forest Model Predictive Control for Paste Thickening. *Minerals Engineering*. 2021, 163, 106760.
- [60] Onan, A. Twitter Mesajları Üzerinde Makine Öğrenmesi Yöntemlerine Dayalı Duygu Analizi. *Yönetim Bilişim Dergisi*. 2017, 3(2), 1-14.
- [61] Aydoğan, M., Şener, A. Duygu Analizi Tabanlı Yeni Bir Hibrit Tavsiyeci Sistem. *Euroasia Journal of Applied Sciences*. 2020, 7(13), 48-62.
- [62] Lane, P. C. R., Clarke, D., Hender, P. On developing robust models for favourability analysis: Model choice, feature sets and imbalanced data. *Decision Support System*. 2012, 53(4), 712–718.
- [63] Aizawa, A. An information-theoretic perspective of tf-idf measures. *Information Processing and Management*. 2003, 39(1), 45-65.
- [64] G. Hackeling, *Mastering machine learning with scikit-learn-*, 2nd ed. Birmingham, England: Packt Publishing, 2017.
- [65] Çevrimiçi: Microsoft: <https://azure.microsoft.com/tr-tr/overview/what-is-deep-learning/> (Erişim Tarihi: 12.04.2022)
- [66] Ravi, D., Wong, C., Deligianni, F., Berhelot, M., Andreu-Perez, J., Lo, B., Yang, G.Z. Deep Learning for Health Informatics. *IEEE Journal of Biomedical and Health Informatics*. 2017, 21(1), 4-21.
- [67] LeCun, Y., Bengio, Y., Hinton G. Deep Learning. *Nature*. 2015, 521(7553), 436-444.
- [68] LeCun, Y. Generalization and Network Design Strategies. *Connectionism in Perspective*. 1989, 19, 143-155.
- [69] Salur, M.U. Derin Öğrenme Tabanlı Çok Modlu Duygu Analizi Yöntemlerinin Geliştirilmesi. Fırat Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Anabilim Dalı, Elâzığ, 2021, 101 s. (Doktora Tezi)

- [70] Küçük, D., Arıcı, N. Doğal Dil İşlemede Derin Öğrenme Uygulamaları Üzerine Bir Literatür Çalışması. Uluslararası Yönetim Bilişim Sistemleri ve Bilgisayar Bilimleri Dergisi. 2018, 2(2), 76-86.
- [71] L. Zhang, S. Wang, and B. Liu, "Deep learning for sentiment analysis: A survey. Wiley Interdisciplinary Reviews. 2018, 8(4), e1253.
- [72] Asefisaray, B. Uçtan-Uca Konuşma Tanıma Modeli: Türkçe' deki Deneyler. Hacettepe Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Anabilim Dalı, Ankara, 2018, 107 s. (Doktora Tezi)
- [73] Çevrimiçi: Medium: <https://aditi-mittal.medium.com/understanding-rnn-and-lstm-f7cdf6dfc14e> (Erişim Tarihi: 10.04.2022)
- [74] Petmezas, G., Haris, K., Stefanopoulos, L., Kilintzis, V., Tzavelis, A., Rogers, J. A., Maglaveras, N. Automated Atrial Fibrillation Detection using a Hybrid CNN-LSTM Network on Imbalanced ECG Datasets. Biomedical Signal Processing and Control. 2021, 63(8), 84-93
- [75] Rojas-Barahona, L.M. Deep learning for sentiment analysis: Language and Linguistics Compass. Language and Linguistics Compass. 2016, 10(12), 701-719.
- [76] Kayaalp, K., Süzen, A.A. Derin Öğrenme ve Türkiye' deki Uygulamaları, IKSAD Yayınları, Türkiye, 2014, 89 s.
- [77] Xiao, Y., Yin, Y. Hybrid LSTM Neural Network for Short-Term Traffic Flow Prediction. Information. 2019, 10(3), 105.
- [78] Chen, J., Jing, H., Chang, Y., Liu, Q. Gated recurrent unit based recurrent neural network for remaining useful life prediction of nonlinear deterioration process. Reliability Engineering & System Safety. 2019, 185, 372-382.
- [79] Bibi, I., Akhunzada, A., Malik, J., Iqbal, J., Musaddiq, A., Kim, S. A dynamic DL-driven architecture to combat sophisticated Android malware. IEEE Access. 2020, 8, 129600-129612.
- [80] Seveli, O., Kemaloğlu, N. Olağandışı Olaylar Hakkındaki Tweet'lerin Gerçek ve Gerçek Dışı Olarak Google BERT Modeli ile Sınıflandırılması. Veri Bilimi. 2021, 4(1), 31-37.
- [81] Çevrimiçi: Google: <https://blog.google/products/search/search-language-understanding-bert/> (Erişim Tarihi: 21.04.2022)

- [82] Devlin, J., Chang, M. W., Lee, K., Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018
- [83] Onan, A. Evrişimli Sinir Ağı Mimarilerine Dayalı Türkçe Duygu Analizi. *Avrupa Bilim ve Teknoloji Dergisi*. 2020, 374-380.
- [84] Liu, Y., Liu, Z., Chua, T. S., Sun, M. Topical word embeddings. In *Twenty-ninth AAAI conference on artificial intelligence*, 2015.
- [85] Bengio, Y., Senecal, J.S., Adaptive importance sampling to accelerate training of a neural probabilistic language model. *IEEE Transactions on Neural Networks*. 2008, 19(4), 713–722.
- [86] Rezaeinia, S.M., Rahmani, R., Ghodsi, A., Veisi, H. Sentiment analysis based on improved pre-trained word embeddings. *Expert Systems with Applications*. 2019, 117, 139–147.
- [87] Mikolov, T., Chen, K., Corrado, G., Dean, J. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781, 2013.
- [88] Joulin, A., Grave, E., Bojanowski, P., Douze, M., Jégou, H., Mikolov, T. Fasttext. zip: Compressing text classification models. arXiv preprint arXiv:1612.03651, 2016.
- [89] Joulin, A., Grave, E., Bojanowski, P., Mikolov, T. Bag of tricks for efficient text classification. arXiv preprint arXiv:1607.01759, 2016.

ÖZGEÇMİŞ

Adı Soyadı : Kemal KARGA

Doğum Yeri ve Yılı :

Medeni Hali

Yabancı Dili : İngilizce

E-posta :

Eğitim Durumu

Lise : Beylerbeyi Hacı Sabancı Lisesi, 2005

Lisans : Manisa Celal Bayar Üniversitesi, Yazılım Mühendisliği Bölümü, 2019

Yüksek Lisans : Manisa Celal Bayar Üniversitesi, Fen Bilimleri Enstitüsü, Yazılım Mühendisliği Anabilim Dalı, 2022

Mesleki Deneyim

Kurum bilgisi Trakya Üniversitesi 2011-2016

Kurum bilgisi Manisa Celal Bayar Üniversitesi 2016-(halen)