

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

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

**MAKİNE ÖĞRENME TEKNİKLERİYLE
KRİPTO PARA DUYGU ANALİZİ**

Gül Cihan HABEK

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

**II. Danışman
Doç. Dr. Aytuğ ONAN**



MANİSA-2022

**Gül Cihan
HABEK**

**MAKİNE ÖĞRENME Sİ TEKNİKLERİYLE
KRİPTO PARA DUYGU ANALİZİ**

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.

Gül Cihan HABEK



İÇİNDEKİLER

	Sayfa
İÇİNDEKİLER	I
SİMGELER VE KISALTMALAR DİZİNİ	III
ŞEKİLLER DİZİNİ.....	IV
TABLO DİZİNİ	V
TEŞEKKÜR.....	VI
ÖZET.....	VII
ABSTRACT.....	VIII
1. GİRİŞ	1
1.1. Kripto Para Nedir?.....	1
1.1.1.Kripto Para Tarihi	2
1.1.2.İlk Kripto Para: Bitcoin	2
1.1.3.Ethereum	3
1.1.4.Verit Seti İçerisinde Bulunan Diğer Altcoinler	3
1.1.5.NFT	4
1.1.6.DeFi	4
1.2. Twitter ile Duygu Analizi.....	6
2. LİTERATÜR ARAŞTIRMASI	7
2.1. Farklı Konularda Yapılan Duygu Analizi Çalışmaları.....	7
2.2. Kripto Paralar ile İlgili Duygu Analizi Çalışmaları	9
3. GENEL BİLGİLER	15
3.1. Metin Madenciliği	15
3.1.1.Metin Madenciliği Süreçleri	15
3.2. Duygu Analizi	17
3.3. Makine Öğrenmesi	18
3.4. Derin Öğrenme	20
4. MATERYAL VE YÖNTEMLER.....	22
4.1. Kelime Gömme Teknikleri.....	22
4.1.1. Terim Frekansı ve Ters Doküman Frekansı (TF-IDF)	22
4.1.2. Word2Vec.....	22
4.1.2.1. Sürekli Kelime Torbası (CBOW)	23
4.1.2.2. Skip-Gram.....	24
4.1.3. Google BERT	24
4.1.4. FastText	25
4.1.5. GloVe.....	25
4.2. Derin Öğrenme Algoritmaları	26
4.2.1. Evrişimli Sinir Ağları (CNN)	26
4.2.2. Tekrarlayan Sinir Ağları (RNN).....	27
4.2.3. Uzun Kısa Süreli Bellek (LSTM).....	29
4.2.4. İki Yönlü Uzun Kısa Süreli Bellek (BLSTM).....	33
4.2.5. Geçitli Tekrarlayan Birim (GRU).....	34
4.3. Hibrit Modeller.....	35
4.4. Verit Seti.....	43
4.4.1. Verilerin Toplanması.....	43
4.4.2. Verilerin Etiketlenmesi.....	45
5. ARAŞTIRMA BULGULARI VE TARTIŞMA	47
5.1. Kullanılan Yöntem	47
5.2. Deney Sonuçları	48
5.2.1. Derin Öğrenme Algoritmalarıyla Elde Edilen Sonuçlar.....	48

5.2.2. Hibrit Modeller Kullanılarak Elde Edilen Sonuçlar	54
6. SONUÇ VE ÖNERİLER	58
KAYNAKLAR	60
ÖZGEÇMİŞ	65



SİMGELER VE KISALTMALAR DİZİNİ

NFT	Non-Fungible Token (Değiştirilemez Para)
DEFI	Decentralized Finance (Merkeziyetsiz Finans)
ERC	Ethereum Request for Comments (Yorumlar için Ethereum Talebi)
API	Application Programming Interface (Uygulama Programlama Arayüzü)
KNIME	Konstanz Information Miner (Konstanz Bilgi Madencisi)
NLTK	Natural Language Toolkit (Doğal Dil Araç Seti)
NLP	Natural Language Processing (Doğal Dil İşleme)
PHP	Hypertext Preprocessor (Üstünyazı Önişlemcisi)
WEKA	Waikato Environment for Knowledge Analysis (Bilgi Analizi için Waikato Ortamı)
NB	Naive Bayes
LR	Lojistik Regresyon
DVM	Destek Vektör Makinesi
RF	Random Forest (Rastgele Orman)
KNN	K-Nearest Neighbors (K En Yakın Komşuluk)
YSA	Yapay Sinir Ağları
TF-IDF	Term Frequency-Inverse Document Frequency (Terim Frekansı-Ters Doküman Frekansı)
CBOW	Continuous Bag of Words (Sürekli Kelime Torbası)
BERT	Bidirectional Encoder Representations from Transformers (Transformatörlerden Çift Yönlü Kodlayıcı Temsilleri)
CNN	Convolutional Neural Network (Evrişimsel Sinir Ağı)
RNN	Recurrent Neural Network (Tekrarlayan Sinir Ağı)
LSTM	Long-Short Term Memory (Uzun-Kısa Süreli Bellek)
BLSTM	Bidirectional Long Short-Term Memory (İki Yönlü Uzun-Kısa Süreli Bellek)
GRU	Gated Recurrent Unit (Geçitli Tekrarlayan Birim)

ŞEKİLLER DİZİNİ

Sayfa

Şekil 1.1. Son 12 ay içerisinde Bitcoin, Ethereum, altcoin, NFT ve DeFi karşılaştırmaları	5
Şekil 3.1.1. Metin madenciliği adımları	16
Şekil 3.3.1. Makine öğrenmesi algoritmaları	19
Şekil 3.3.2. Sınıflandırma (soldaki) ve regresyon (sağdaki) problemleri	19
Şekil 3.3.3. Kümeleme analizi sonrası farklı renklerdeki dairelerin 3 kümeye ayrılması	20
Şekil 3.4. Yapay zeka, makine öğrenmesi ve derin öğrenme arasındaki ilişki...	21
Şekil 4.1.2.1. CBOW mimarisi	23
Şekil 4.1.2.2. Skip-gram mimarisi	24
Şekil 4.2.1. CNN mimarisi	26
Şekil 4.2.2.1. RNN mimarisi	28
Şekil 4.2.2.2. RNN mimarisi (açık gösterim)	28
Şekil 4.2.3.1. LSTM mimarisi (teorik gösterim)	30
Şekil 4.2.3.2. LSTM mimarisi (matematiksel gösterim)	30
Şekil 4.2.3.3. Cell state hattı	31
Şekil 4.2.3.4. Unutma kapısı	31
Şekil 4.2.3.5. Giriş kapısı	32
Şekil 4.2.3.6. Çıkış kapısı	32
Şekil 4.2.4. BLSTM mimarisi	33
Şekil 4.2.5.1. GRU mimarisi (teorik gösterim)	34
Şekil 4.2.5.2. GRU mimarisi (matematiksel gösterim)	34
Şekil 4.3.1. ABCDM mimarisi	35
Şekil 4.3.2. AC-BiLSTM mimarisi	36
Şekil 4.3.3. AGCNN mimarisi	36
Şekil 4.3.4. ARC mimarisi	37
Şekil 4.3.5. ATTPooling mimarisi	38
Şekil 4.3.6. CAT-BiGRU mimarisi	39
Şekil 4.3.7. CRNN mimarisi	40
Şekil 4.3.8. HAN mimarisi	41
Şekil 4.3.9. IWV mimarisi	41
Şekil 4.3.10. SS-BED mimarisi	42
Şekil 4.3.11. WCNLSTM mimarisi	43
Şekil 4.4.1. Son 12 ay içerisinde Facebook, Instagram, Twitter web arama karşılaştırmaları	44
Şekil 5.2.1.1. Doğru sınıflandırma oranı için temel etki grafiği	52
Şekil 5.2.1.2. Duyarlılık için temel etki grafiği	53
Şekil 5.2.1.3. Geri çağırma için temel etki grafiği	53
Şekil 5.2.1.4. F-ölçütü için temel etki grafiği	54
Şekil 5.2.2.1 Hibrit mimarilerde doğru sınıflandırma oranı için temel etki grafiği	56
Şekil 5.2.2.2. Hibrit mimarilerde duyarlılık için temel etki grafiği	56
Şekil 5.2.2.3. Hibrit mimarilerde geri çağırma için temel etki grafiği	57
Şekil 5.2.2.4. Hibrit mimarilerde F-ölçütü için temel etki grafiği	57

TABLO DİZİNİ

	Sayfa
Tablo 4.1. Manuel olarak etiketlenen tweet örnekleri.....	45
Tablo 4.2. Sınıflardaki tweet sayıları	46
Tablo 5.2.1.1. Derin öğrenme algoritmaları ve önerilen şema ile elde edilen sınıflandırma doğruluk değerleri.....	48
Tablo 5.2.1.2. Derin öğrenme algoritmaları ve önerilen şema ile elde edilen duyarlılık değerleri	49
Tablo 5.2.1.3. Derin öğrenme algoritmaları ve önerilen şema ile elde edilen geri çağırma değerleri	50
Tablo 5.2.1.4. Derin öğrenme algoritmaları ve önerilen şema ile elde edilen F-ölçütü değerleri	51
Tablo 5.2.2. Hibrit derin sinir ağı mimarileri ile tahmine dayalı performanslar.	55



TEŞEKKÜR

Çalışmamın her aşamasında bana destek olan, bilgi ve deneyimleri ile yol gösteren, tecrübeleri ile beni aydınlatan ve desteklerini hiç eksik etmeyen, kendilerini tanımaktan büyük onur duyduğum danışman hocam Sayın Dr. Öğr. Üyesi Mansur Alp TOÇOĞLU ve ikinci danışman hocam Sayın Doç. Dr. Aytuğ ONAN'a; öğrenim hayatım boyunca her zaman yanımda olduğunu hissettiren ve her konuda beni destekleyen değerli eşim Abdulkadir HABEK'e ve sevgi ve desteklerini hiç esirgemeyen hep yanımda olan aileme yürekten teşekkür ederim.

Gül Cihan HABEK
Manisa, 2022



ÖZET

Yüksek Lisans Tezi

Gül Cihan HABEK

**Manisa Celal Bayar Üniversitesi
Fen Bilimleri Enstitüsü
Yazılım Mühendisliği Anabilim Dalı**

Danışman: Dr. Öğr. Üyesi Mansur Alp TOÇOĞLU

İkinci Danışman: Doç. Dr. Aytuğ ONAN

İlk kripto para birimi olan Bitcoin'in çıkması ile birlikte hayatımıza giren kripto para kavramı finansal yönden pek çok yeniliği beraberinde getirmektedir. Daha önce bilinmeyen pek çok kavram kripto paraların çıkması ve yaygınlaşması ile birlikte konuyla ilgilenen insanlar tarafından merak edilip araştırılmaktadır. Kripto paralara olan yönelimlerin ve araştırmaların artması ile birlikte daha çok insan tarafından konu ile ilgili yazılar yazılmakta ve duygu ve düşünceler uzun metinlerden ziyade birkaç kelimeden oluşan kısa cümleler ile platformlarda anlık olarak paylaşılmaktadır. Bunun sonucunda kripto paralar ile ilgili metin verilerinin sayıları artmakta ve gün geçtikçe metin verileri büyük veri kaynağı haline gelmektedir. Büyük veri kaynağı haline gelen kripto para paylaşımları, verilerin duygularını analiz edip sınıflandırarak kripto para piyasası ile ilgilenen insanlar için piyasanın yönü hakkında bilgi verme isteğini doğurmuştur. Duygu analizi çalışmaları pek çok konu başlığı için tercih edilebilmektedir. Literatürdeki çalışmalara bakıldığında ise kripto paralar ile ilgili duygu analizi yapılan çalışmalar mevcuttur. Fakat yabancı metinler kullanılarak yapılan çalışmalar Türkçe dilinde yapılan çalışmalara kıyasla oldukça fazladır ve yapılan çalışmalarda ele alınan kripto para çeşitliliği genellikle çok azdır.

Bu tez çalışmasında Bitcoin, Ethereum, Solana, Ripple, Avalanche, Chiliz, Bitcointoken kripto paraları ile NFT-DeFi teknolojileri hakkında Twitter platformu üzerinden paylaşılan metin parçaları kullanılarak bir duygu analizi çalışması yapılmaya çalışılmıştır. Bunun için derin öğrenme algoritmalarından CNN, RNN, LSTM ve GRU algoritmaları kullanılarak sınıflandırma doğruluğu, duyarlılık, geri çağırma ve F-ölçütü metrikleri için sonuçlar alınmıştır. Daha sonra ABCDM, AC-BiLSTM, AGCNN, ARC, ATTPooling, CAT-BiGRU, CNN-GRU, CRNN, HAN, IWV, SS-BED, Ağaç-BLSTM, Ağaç-LSTM ve WCNNLSTM hibrit mimarileri kullanılarak sonuçlar alınmış ve elde edilen sonuçlar karşılaştırılmıştır.

Anahtar Kelimeler: Kripto Para, Duygu Analizi, Türkçe Metinler, Twitter'da Duygu Analizi, Makine Öğrenmesi, Derin öğrenme, Hibrit Mimariler

2022, 64 sayfa

ABSTRACT

M.Sc. Thesis

Gül Cihan HABEK

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

Supervisor: Asst. Prof. Mansur Alp TOÇOĞLU

Co-Supervisor: Assoc. Dr. Aytuğ ONAN

The concept of cryptocurrencies, which entered our lives with the emergence of Bitcoin, the first cryptocurrency, brings many innovations in financial terms. Many previously unknown concepts are wondered and explored by people who are interested in the subject with the emergence and spread of cryptocurrencies. Thanks to the increase in trends and research on cryptocurrencies, more and more people are writing articles on the subject; feelings and thoughts are shared instantly on platforms with short sentences consisting of a few words rather than long texts. As a result, the number of text data about cryptocurrencies is increasing and text data is becoming a big data source day by day. Cryptocurrency shares, which have become a source of big data, have led to the desire to provide information about the direction of the market for people interested in the cryptocurrency market by analyzing and classifying the sentiment of the data. Sentiment analysis studies can be preferred for many topics. When we look at the studies in the literature, there are studies on sentiment analysis about cryptocurrencies. However, studies using foreign texts are quite high compared to studies conducted in Turkish, and the variety of cryptocurrencies discussed in studies is generally very low.

In this thesis, a sentiment analysis study has been tried to be made by using text pieces shared on the Twitter platform about Bitcoin, Ethereum, Solana, Ripple, Avalanche, Chiliz, Bitcicoin cryptocurrencies and NFT-DeFi technologies. For this, results were obtained for classification accuracy, precision, recall and F-measure by using CNN, RNN, LSTM and GRU algorithms from deep learning algorithms. Then, the results were obtained using hybrid architectures ABCDM, AC-BiLSTM, AGCNN, ARC, ATTPooling, CAT-BiGRU, CNN-GRU, CRNN, HAN, IWV, SS-BED, Tree-BLSTM, Tree-LSTM and WCNNLSTM; and the results were compared.

Keywords: Cryptocurrency, Sentiment Analysis, Turkish Texts, Sentiment Analysis on Twitter, Machine Learning, Deep Learning, Hybrid Architectures

2022, 64 pages

1. GİRİŞ

1.1. Kripto Para Nedir?

2017 yılının sonlarında popülerliği artmaya başlayan ve günümüzde trend bir konu olmayı başaran kripto paralar, herhangi bir merkeze ya da üçüncü bir kuruluşa bağlı olmayan böylelikle denetlenmeyen ve sanal para birimi olarak internet ortamında kullanılabilen dijital varlıklardır [1]. Sistemlerinde aracılık yapan üçüncü bir kurum veya kuruluşun olmaması saklanması ve transfer edilmesi aşamalarında kripto paraları daha güvenilir yapar.

Günümüzde adından sıklıkla söz ettiren kripto paralar eşten eşe ağ kullanarak blok zinciri üzerinden çalışmakta ve kriptografi denilen güvenlik tekniği ile şifrelenmektedir [2].

Eşten Eşe Ağ: Eşten eşe veya diğer adı ile eşler arası ağ, verilerin iki veya ikiden fazla istemci arasında paylaşılması ve dağıtılması sırasında kullanılan iletişim protokollerinden biridir [3]. Eşler arası ağ protokolünde ağdaki bir bilgisayarın hata vermesi durumunda hata veren kullanıcının bağlantısı kaybolur, verinin paylaşım ve dağıtımını ise sadece bir noktaya bağlı olmaması sayesinde herhangi bir kesinti yaşanmadan devam eder [2].

Kriptografi: Düz metni çeşitli şifreleme algoritmaları ile kodlayıp şifreli metne dönüştürerek, iletilecek olan mesajın sadece alıcısına ulaşmasını sağlayan bir güvenlik tekniğidir. Şifreleme işlemlerinde temel amaç iletilmek istenilen mesajı sadece alıcısına ulaştırmak, farklı bir kişinin eline geçtiği takdirde ise mesajın anlaşılabilir olmasını sağlayacak şekilde güvenilirliği arttırmaktır.

Blok Zinciri: Merkezi bir yapıları olmayan kripto paraların kontrolleri blok zincir adı verilen bir teknoloji ile gerçekleştirilmektedir. Blok zinciri teknolojisi birbirine bağlanmış ve her biri belirli sayılarda işlem barındıran bloklardan oluşur. Birbirine kronolojik bir sıra ile bağlı olan ve içerisinde tüm verilerin tutulduğu bloklar matematiksel birtakım işlemler ve kriptografi tekniği kullanılarak onaylanıp zincire eklenir. Kripto paraların blok zincir teknolojisini kullanması ek maliyetleri ortadan kaldırır ve yüksek güvenilirlik sağlar.

1.1.1. Kripto Para Tarihi

Blok zinciri teknolojisinin temelleri iki kriptografi uzmanı Scott Stornetta ve Stuart Haber'ın 20. yy.'ın sonlarında yapmış oldukları çalışmalara dayanmaktadır [4]. İki uzman, Ralph Merkle tarafından patentlenmiş olan “hash ağacı” fonksiyonunu baz alarak çalışmalarını yürütmüş ve günümüzdeki anlamıyla blok zincir teknolojisine yakın bir tanım yaparak bu teknolojinin temellerini atmışlardır.

Modern anlamıyla ilk blok zinciri ise, ilk kripto para birimi olan Bitcoin ile 31 Ekim 2008 tarihinde ortaya çıkmıştır [5]. Satoshi Nakamoto takma isimli, kim olduğu henüz bilinmeyen, bir kişi ya da grup tarafından yayımlanan “Bitcoin: Eşten Eşe Elektronik Nakit Sistemi” isimli makalede Bitcoin kripto parasından bahsedilmiştir [3]. 2009 yılının başlarında ise Bitcoin yazılımı halka ilk kez açılmış ve böylelikle Bitcoin madenciliği başlamıştır.

Merkezi bir yapısı olmayan şifreli para birimleri benimsenip bu fikre gösterilen ilgi ve talep arttıkça farklı kripto para birimleri de altcoin adı altında ortaya çıkmaya başlamıştır. Altcoin terimi Bitcoin'den sonra ortaya çıkan alternatif coinlerin genel bir adıdır. Bunlar arasında ilk altcoinler olan Namecoin ve Litecoin 2011 yılında piyasada dolaşıma sunulmuştur.

En çok bilinen kripto para birimleri sırasıyla Tether, BNB, USD Coin, Cardano, Solana, XRP, Terra, Dogecoin, Polkadot, Avalanche, Binance USD, Polygon, Shiba Inu, TerraUSD, Crypto.com Coin, Wrapped Bitcoin, Cosmos, Dai, Litecoin ve benzeri şeklinde sıralanır [6]. Bu sıralama listeye yeni altcoinler eklendikçe veya mevcut altcoinlerin değerleri yükselip düştükçe her geçen gün değişebilmektedir.

1.1.2. İlk Kripto Para: Bitcoin

2022 yılı başları itibariyle 10.000 civarında kripto para çeşidi bulunmaktadır ve bunlar arasında en çok bilinen ve en önemli sayılan para ilk kripto para olması sebebiyle Bitcoin'dir. Satoshi Nakamoto Bitcoin için “birebir elektronik nakit sistem” diye bir tanımlama yapmış ve sistem maksimum sayı olarak 21 milyon Bitcoin'e izin verecek şekilde oluşturulmuştur. Ayrıca Bitcoin El Salvador Devleti tarafından yasal para birimi olarak tanınarak dünyanın ilk resmi sanal para birimi olmuştur.

1.1.3. Ethereum

En çok bilinen ikinci para birimi Ethereum'dur. Ethereum, kurumsal para birimi Ether olan merkeziyetsiz ve açık kaynak kodlu bir blok zinciridir. Genel kripto para piyasasına bakıldığında ikinci en büyük piyasa değerine sahip olan ve bu konumunu yıllardır koruyan Ethereum, kurucuları tarafından dijital bir para olarak değil; blok zincirinin güvenlik ve açıklık özelliklerini büyük çaplı uygulamalarda kullanabilen yeni bir platform olarak geliştirilmiştir.

1.1.4. Veri Seti İçerisinde Bulunan Diğer Altcoinler

Duygu analizi çalışması yapmak için Twitter platformu üzerinden belirli anahtar kelimeler aratılarak oluşturulan veri seti içerisinde ilk kripto para birimi Bitcoin ve ikinci para birimi olarak bilinen Ethereum dışında Solana, Ripple, Avalanche, Chiliz, Bitcointoken olmak üzere beş altcoin ve NFT ile DeFi teknolojileri ile ilgili kullanıcılar tarafından gönderilen tweetler de bulunmaktadır. Bir sonraki paragraftan itibaren kullanılan altcoinlerin açıklamalarına yer verilmiştir.

Solana, açık kaynak kodlu, merkeziyetsiz ve DeFi tabanlı bir projedir. Projenin resmi kripto para birimi SOL'dur [7]. Projenin çalışmalarına Solana Foundation tarafından 2017 yılında başlanılmış ve resmi açılış 2020 yılının başlarında yapılmıştır. Henüz yeni sayılabilecek bir altcoin olan Solana'nın saniyede ölçeklendirdiği işlem sayısının yaklaşık olarak 50.000 olması Solana projesinin oldukça hızlı olduğunu gösterir.

Ripple, Ripple Labs tarafından 2012 yılında geliştirilmiş ve para işlemleri yapmak için kullanılan ödeme ağı sistemidir. Kripto para değeri XRP'dir. Var olan birçok merkeziyetsiz kripto para birimlerinin aksine merkezi bir yapısı olan Ripple üzerindeki bütün işlemler kurumsal olan şirket tarafından yönetilmektedir. Ripple ödeme ağında farklı büyüklüklerdeki varlık ve para işlemlerinin oldukça hızlı, güvenilir ve düşük maliyetlerle blok zinciri üzerinden yapılmasına olanak sağlanmaktadır.

Emin Gün Sirer liderliğindeki Ava Labs tarafından geliştirilen Avalanche, 2020 yılının son aylarında piyasaya sürülmüş olan ölçeklenebilir ve güvenli bir ağdır. Resmi kripto para birimi Avalanche'tır ve AVAX sembolü ile gösterilmektedir. İzin gerektiren özel blok zincirler ve izin gerektirmeyen genel blok zincirlerin tek bir

ağ üzerinde çalışabilmelerine olanak tanır; çözümler alt ağlarda geliştirilebilir. Avax'ın diğer birçok blok zincirlerden en büyük farkı bir zincir üzerinden ilerlemeyip tüm unsurlar için bir alt ağın tasarlanmış olmasıdır.

Chiliz, kısaltılmış hali ile CHZ, 14 Şubat 2018 yılında Alexandre Dreyfus ve arkadaşları tarafından çıkarılmış olan bir taraftar token platformu coinidir. Ethereum ile aynı ağı kullandığı bilinen ve ERC-20 token tabanlı olan CHZ, dünya genelinde birçok spor kulübü ile projeler yapmakta ve anlaşmalar imzalamaktadır.

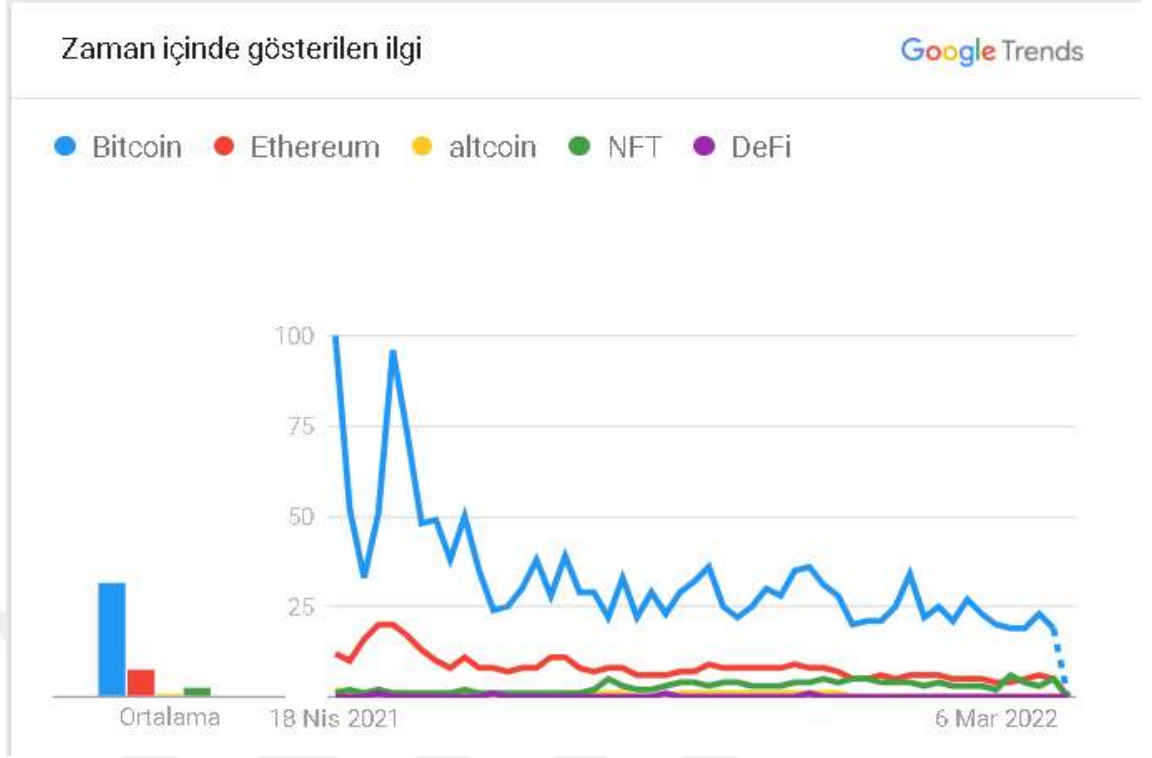
Son olarak Bitci kendisine ait bir blok zincir ağı olan Türkiye'nin ilk kripto borsasıdır ve bu yüzden de çalışmada veri seti oluştururken bu platformun coin olan Bitcicoın verilerinden de yararlanmak istenilmiştir. Bitci, Çağdaş Holding'in Ceo'su Burak Çağdaş Çağlar tarafından 2018 yılında kurulmuştur. Dünyada ilk defa kişiye özel kar-zarar grafiği sunması yönüyle de önemli olan Bitci bünyesinde çok sayıda kripto para barındırmaktadır.

1.1.5. NFT

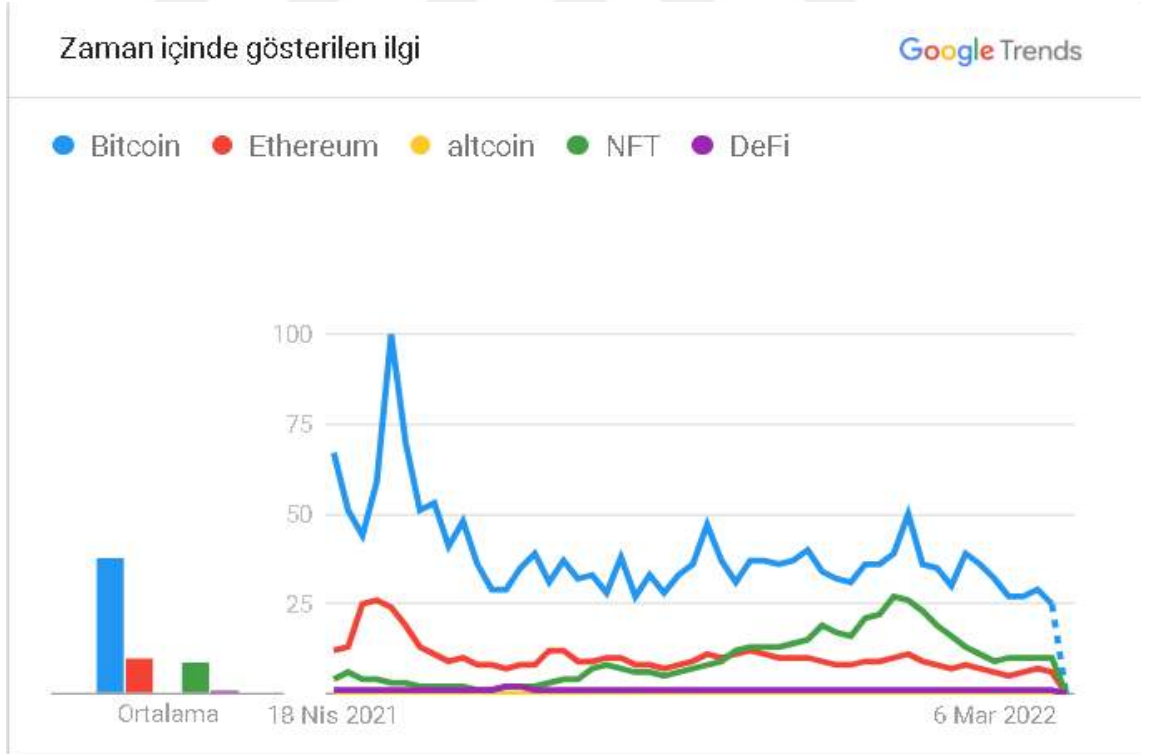
NFT, Non-Fungible Token ifadesinin kısaltılmış hali olan ve Türkçeye “Değiştirilemez Para” olarak çevrilen dijital varlıklardır. Yani NFT'ler yalnızca bir kişiye ait olan ve sadece bir tane bulunan tokenlara verilen addır. Örnek vermek gerekirse bir kişide bulunan Bitcoin, Ethereum veya herhangi bir altcoin bir başkasının cüzdanında da aynı şeyi ifade eden ve transfer aracı olarak kullanılabilen varlıklar iken NFT'ler bir varlıktan birden fazla bulunmasına rağmen her birinin farklı değerlerde olduğu tokenlardır. Herhangi bir resim veya görseli, Twitter'dan atılan bir tweeti, bir futbol veya futbolcu kartını, sosyal medya üzerinden oluşturulan bir hikâyeyi ve bunun gibi birçok şeyi temsil edebilir.

1.1.6. DeFi

DeFi, Decentralized ve Finance kelimelerinin kısaltmasından oluşan belirli bir merkezi ve bağlı olduğu bir otoritesi olmayan merkeziyetsiz finans işlemlerinin genel adıdır. Merkeziyetsiz Finance uygulamaları Ethereum ağı üzerine inşa edilmiştir ve herkese açık blok zincirlerindeki eşler arası finansal hizmetlerde kullanılır. Bankalarda verilen hizmetlerin çoğu DeFi ile daha hızlı bir şekilde ve üçüncü bir kurum olmadan doğrudan iki kişi arasında gerçekleştirilebilir.



(a) Türkiye genelinde web arama sonuçlarına göre



(b) Dünya genelinde web arama sonuçlarına göre

Şekil 1.1. Son 12 ay içerisinde Bitcoin, Ethereum, altcoin, NFT ve DeFi karşılaştırmaları [8]

1.2. Twitter ile Duygu Analizi

Duygu analizi bir yazı ya da yazarın belirli bir konudaki duygu ve düşüncelerini doğal dil işleme, sınıflandırma algoritmaları, istatistik ve programlama gibi çeşitli yöntemler aracılığıyla hesaplayarak görüş ve ifadelerinin olumlu (pozitif) veya olumsuz (negatif) olduğunu belirlemek için kullanılan yöntemin genel adıdır [9].

Çok güncel bir konu olan kripto paralar ile ilgili fikir ve görüşlerin sınıflandırılması oldukça önemlidir, piyasanın yönü hakkında bilgi verir. Çok sayıda aktif kullanıcıya ve büyük verilere sahip platformlardan biri olan Twitter, kripto paralar ile ilgili düşüncelerin belirlenebilmesi için elverişli bir veri kaynağı olarak görülebilir.

İstatistik verilere bakıldığı zaman dünya genelinde Twitter aylık aktif kullanıcı sayısı 2020 yılının sonunda 321 milyon olarak belirtilmiştir ve Similar Web’e göre en çok ziyaret edilen 7. site olmuştur. 2021 yılı itibarıyla Türkiye genelinde ise günlük atılan ortalama tweet sayısı 7 milyon olarak belirlenmiştir. Gönderilen tweetlerin içeriklerine bakılarak genel bir olumlu-olumsuz sınıflandırması yapıldığında %41 olumlu, %59 olumsuz tweet atıldığı görülmüştür [10]. Tüm bu istatistik verilere bakıldığında Twitter medyasının çok büyük bir veri kaynağı olduğu açıkça görülmektedir ve bu platform üzerinden gönderilen kripto paralar ile ilgili tweetlerin sınıflandırılması, kullanıcıların bu konu üzerindeki düşüncelerini ve piyasanın pozitif ya da negatif yönde olduğunu saptamak açısından oldukça önemlidir.

2. LİTERATÜR ARAŞTIRMASI

Çalışmanın literatür araştırması kısmında öncelikle farklı konular için gerçekleştirilen duygu analizi ile ilgili literatür çalışmaları ele alınmış, daha sonra, kripto paralar üzerinde duygu analizi ile ilgili literatür çalışmalarına yer verilmiştir.

2.1. Farklı Konularda Yapılan Duygu Analizi Çalışmaları

Farklı konularda yapılan duygu analizi çalışmalarından birinde Emel ve ark. (2022) çalışmalarında sosyal medya kullanıcılarının Covid-19 için geliştirilmiş olan Sinovac ve BioNTech marka aşılara olan bakış açıları ve yaklaşımlarını duygu analizi tekniğini kullanarak ortaya koymaya çalışmışlardır. Ayrıca yaptıkları çalışmada yeni bir yöntem olan metin analizi tekniğini kullanarak metodolojiye bir katkı sağlamayı hedeflemişlerdir. Kullanılan veri seti Twitter API kullanılarak Twitter platformu üzerinden gönderilen Sinovac ve BioNTech ile ilgili 15.03.2021 – 15.04.2021 tarihleri arasındaki Türkçe tweetler çekilerek oluşturulmuştur. Çekilen veriler RStudio programı aracılığıyla ön işleme aşamalarından geçirilmiş ve temizlenen veriler arasında en sık kullanılan kelimeleri bulmak için tweetler frekanslara ayrılmıştır. Çalışmanın bir sonraki aşamasında RStudio programında NRC duygu sözlüğü geliştirilmiş ve frekansları belirlenen tweetler ile eşleştirilerek kelime düzeyinde duygu analizi yapılmaya çalışılmıştır. Aşılarla ait tweetler üzerinde yapılan duygu analizine göre Twitter kullanıcılarının Sinovac markalı aşılara daha güvenli buldukları ve BioNTech markalı aşılara karşı tarafsız oldukları sonucuna ulaşılmıştır [11].

Diğer bir çalışmada, Özcan ve ark. (2021) yaptıkları çalışmada Covid-19 sürecinde hayata geçirilen uzaktan eğitim ile ilgili atılan tweetlerin duygu analizini sözlük tabanlı bir yaklaşım kullanılarak yapmayı hedeflemişlerdir. Çalışmada kullanılan veri seti Kaggle platformundan alınan 23.07.2020 23:51:34 – 14.08.2020 05:43:52 52 arasındaki 202.645 tweetten oluşmaktadır. Veriler Twitter platformundan Twitter API aracılığıyla çekilmiştir. Çekilen verilerden rastgele seçilen 999 tweet el ile pozitif ve negatif olmak üzere 2 duygu sınıfına ayrılacak şekilde etiketlenmiştir. Etiketli veri setinde pozitif tweet sayısı 949 ve negatif tweet sayısı 42 olarak elde edilmiştir. Duygu analizi için KNIME üzerinde bir model kurulmuş ve uygun düğümler kullanılarak ön işleme yapılmıştır. Daha sonra duygu analizi aşamalarından

geçirilmiş ve başarı oranı hesaplaması yapılmıştır. Çalışmada sözlük tabanlı yaklaşım kullanılarak bulunan doğruluk oranı %88.4 olarak bulunmuştur [12].

Mustafa ve ark. (2021) insanların koronavirüse karşı tepkilerini ölçmek için bir duygu analizi çalışması yapmışlardır. Çalışmada koronavirüs ile ilgili en sık kullanılan 10 anahtar kelime #Covid, #Corona, #CoronaVirus, Covid-2019, #Covid2019, #Corona, #CoronaVirusUpdate, Covid-19, #Covid19, #Wuhan olarak tespit edilmiş ve bu anahtar kelimeleri içeren 03.2020 ile 04.2020 tarihleri arasındaki 387.953 tweet Twitter Stream API kullanılarak anlık olarak çekilmiştir. Korona ile ilgili atılan tweetlerin genellikle olumsuz olması beklenilse de duygu analizi ve makine öğrenme teknikleri kullanılarak metinleri analiz edebilmek için 155.305 pozitif, 60.610 negatif ve 172.038 nötr tweet ile sınıflar oluşturulmuştur. Verilerin ön işleme aşamasında NLTK kütüphanesi kullanılmış ve tweetlerin pozitif, negatif ve nötr duygu bilgileri TextBlob kütüphanesi kullanılarak elde edilmiştir. Sınıflandırma algoritmalarından Lojistik Regresyon (LR), Çok Terimli Naive Bayes, Destek Vektör Makinesi (DVM) ve Karar Ağacı algoritmaları uygulanarak alınan sonuçlarda; LR, DVM ve karar ağacı algoritmalarının F1-Skor değerleri birbirlerine yakın değerler olup sırasıyla 0.96, 0.97 ve 0.95 olarak bulunmuştur. Çok terimli NB ile ise 0.65 F1-Skor değerine ulaşılmıştır [13].

Rıza ve ark. (2020) çalışmalarında Twitter üzerinden gönderilen futbol takımları ve futbol müsabakaları ile ilgili tweetlerin duygu analizini yapmayı planlamışlardır. Bu çalışma doğrultusunda konu ile ilgili Twitter’da paylaşılan 30.000 Türkçe tweet C# .NET programlama dili kullanılarak geliştirilen bir uygulama üzerinden Twitter Streaming API kullanılarak anlık olarak çekilmiştir. Çekilen tweetler “Beşiktaş, Galatasaray, Fenerbahçe, Trabzonspor” futbol takımları ile “futbol, maç, hakem, penaltı, gol” vb. spor ile ilgili anahtar kelimeler içermektedir. Çekilen tweetler belirli ön işlemlerden geçirildikten sonra geliştirilen bir uygulama yardımıyla el ile etiketlenmiş ve olumlu, olumsuz, tarafsız ve alakasız olmak üzere 4 ayrı duygu sınıfı oluşturulmuştur. Etiketli tweetlerin aynı kişi tarafından beş kez kontrolü sağlanarak eğitim setlerindeki olası sınıflandırma hataları minimize edilmeye çalışılmıştır. Oluşturulan dört duygu sınıfının da eşit sayıda tweet içermesi sağlanarak 7500’er tweet ile dengeli eğitim veri setleri oluşturulmuştur. Etiketli veriler kullanılarak farklı özniteliklere sahip 12 farklı eğitim seti oluşturulmuş ve oluşturulan

her eğitim seti için WEKA aracı kullanılarak makine öğrenme algoritmalarından NB, DVM, C4.5 ve K-En Yakın Komşu (KNN) algoritmaları ile test edilmiş ve modeller oluşturulmuştur. Algoritmalarının sınıflandırma doğrulukları NB için %84.30, DVM için %92.30, C4.5 algoritması için %89.60 ve KNN algoritması için %87.73 olarak bulunmuştur. Ayrıca futbol müsabakaları ile ilgili Twitter üzerinden gönderilen tweetleri anlık olarak toplayan, çalışmada oluşturulmuş olan modelleri kullanılarak sınıflarını tespit eden ve sınıflara ait sonuçları gerçek zamanlı olarak görselleştiren bir uygulama da geliştirilmiştir [14].

Son olarak, Buğra ve ark. (2019) Twitter üzerinden gönderilen tweetlerin İslamofobik olup olmadığını öğrenmek için bir duygu analizi çalışması gerçekleştirmişlerdir. Çalışmada geliştirilen PHP tabanlı yazılım yardımıyla Twitter API kullanılarak belirli anahtar kelimeler içeren 08-2018-05.09-2018 tarihleri arasındaki 162.000 İngilizce Tweet toplanmış ve MySQL veri tabanında saklanmıştır. Toplanan tweetler gönüllü beş kişi tarafından İslamofobi içeren tweetler pozitif ve İslamofobi içermeyen tweetler negatif olacak şekilde işaretlenmiştir. İşaretlenen pozitif tweet yüzdesi %51, negatif tweet yüzdesi %49 olmuştur. Eğitim ve test verileri %80-%20 oranında ikiye ayrılmış ve gözetimli öğrenme metoduyla eğitilmiştir. Kesinlik (Precision), Duyarlılık (Recall) ve F1 ölçütlerinde sonuçlar almak için Lineer Ridge Regresyonu ve NB sınıflandırma algoritması tercih edilmiş olup NB sınıflandırıcıda %95.3, Ridge Regresyonunda %96,3 oranında doğru sonuca ulaşılmıştır [15].

2.1. Kripto Paralar ile İlgili Duygu Analizi Çalışmaları

Literatürün ikinci kısmı olan kripto paralar ile ilgili duygu analizi çalışmalarına bakıldığında ise Burak ve ark. (2021) çalışmalarının birinci bölümünde ilk kripto para birimi olan Bitcoin ile ilgili Twitter kullanıcıları tarafından gönderilen tweetlerin duygu analizini yapmışlardır. Duygu analizi aşamasında Twitter API ve Python Dili kullanılarak Twitter platformundan gönderilen Bitcoin anahtar kelimesi içeren 4062 adet Türkçe tweet çekilmiştir. Çekilen tweetler ön işleme aşamalarından geçirilerek veriler temizlenmiş ve bu aşamalardan sonra 3737 tweet kalmıştır. Tweetler kelime bazlı duygu analizi yapılarak pozitif, negatif ve nötr olmak üzere üç sınıfa ayrılmıştır. Veriler %80 eğitim verisi ve %20 test verisi olacak şekilde ikiye ayrılmıştır. Modeller NB ve LR algoritmaları ile oluşturulmuştur. NB modelinin başarı oranı %72,19 ve LR

modelinin başarı oranı %75,53 olarak bulunmuştur. İkinci bölümde ise Bitcoin ile ilgili atılan günlük pozitif tweet oranları ve günlük Bitcoin açılış değerlerinden yola çıkarak Bitcoin için kapanış fiyatı tahmini yapmaya çalışılmıştır. Tahminleme için finans verileri Yahoo Finance web sayfasından çekilmiş ve Doğrusal Regresyon ve Rastgele Orman (Random Forest, RF) Regresyon algoritmaları ile modeller oluşturulmuştur. Doğrusal Regresyon kullanılan yöntemde %88,97, RF Regresyon kullanılan yöntemde %94,16 r^2 değerlerine ulaşılmıştır [9].

Cihan ve ark. (2020) yaptıkları çalışmada Twitter platformu üzerinden gönderilen tweetlerin Bitcoin fiyatları ile ilişkisi ortaya konulmaya çalışılmıştır. Çalışmada Bitcoin anahtar kelimesi içeren 06.10.2018-19.05.2019 tarihleri arasındaki İngilizce 2.819.784 tweet Python Dilinin Scraby Kütüphanesi yardımı ile çekilerek veri seti oluşturulmuştur. Tweetlerde yer alan birçok bilgi silinerek her tweetin sadece metin ve tarih bilgileri bir tablo halinde tutulmuş ve sınıflandırma yapmaya uygun hale getirilmiştir. Tweetler metin ön işleme aşamasından geçirilmiş ve 200.000 tweet arasından seçilen 1500 tweet 500 tweet pozitif, 500 tweet negatif ve 500 tweet nötr olacak şekilde sınıflandırılarak sınıflar arasında dengeli bir dağılım yakalanmıştır. Daha sonra bu etiketli 1500 veri eğitim verisi olarak kullanılmış ve yapay öğrenme işlemi gerçekleştirilmiştir. Metinlerin sınıflandırılması aşamasında ise veri setinde bulunan toplam 2.819.784 tweet 825.495 pozitif, 730.140 negatif ve 1.264.149 nötr olacak şekilde sınıflandırılmıştır. Çalışmada NB, DVM, LR ve Yapay Sinir Ağları (YSA) ile modeller oluşturularak sonuçlar karşılaştırıldığında en iyi doğruluk oranı YSA ile alındığı için bu yöntem kullanılmıştır. Sınıflandırılmış olan tweetler tarih bilgileri ile ele alınarak her gün için gönderilen pozitif tweet oranları tespit edilmiştir. Daha sonra belirli günlerdeki pozitif tweet oranları ve Bitcoin günlük kapanış değerleri ile veri seti oluşturulmuştur. İkili korelasyon yapmak için Bitcoin fiyatları ve pozitif/negatif sınıflardaki tweetlerin oranları ele alınmış ve korelasyon katsayısı 0.681 olarak bulunmuştur. Çalışmada bu değer ile pozitif yönde orta üstü kuvvetli ilişki bulunduğu sonucuna varılmıştır [16].

Valencia ve ark. (2019) yaptıkları çalışmada, Bitcoin, Litecoin, Ethereum ve Ripple kripto paraları için fiyat tahmini yapmayı hedeflemişlerdir. Bu çalışma doğrultusunda dört kripto para birimi için sosyal veriler Twitter platformu üzerinden toplanılan tweetler ile oluşturulmuştur. Duygu analizi için Twitter içeriğine özel olarak

uyarlanmış olan Vader aracından yararlanılarak tweetler pozitif, negatif, nötr ve birleşik duygular olacak şekilde sınıflara ayrılmıştır. Geçmiş piyasa verileri, 80 güne kadar geçmiş verilerin ücretsiz olarak talep edilmesini sağlayan cryptocompare.com genel API'sinden alınmıştır. YSA, DVM ve RF algoritmaları kullanılarak gerçekleştirilen tahminlemede DVM ile oluşturulan model Bitcoin, Litecoin ve Ripple kripto paralarının fiyatlarını tahmin etmede başarılı sonuçlar yakalamıştır. Fakat bu model Ethereum fiyatını tahmin etmede başarısız olmuştur. RF algoritması ile oluşturulan modelde ise Bitcoin ve Litecoin kripto paralarının fiyat tahmini sırasında başarılı sonuçlar alınmıştır [17].

Alghobiri (2018) kullanıcıların Bitcoin konulu tweetlerinin duygu analizini yapmıştır. Çalışmada kullanılan veri seti Kaggle'den alınmıştır. Veri seti Bitcoin ile ilgili farklı geçmişlere sahip kullanıcıların 50.859 tweetinden oluşan gerçek dünya verilerini içermektedir. Daha iyi bir karşılaştırmalı analizler yapmak adına veri seti beş ayrı bölüme ayrılmıştır. Bölümler tweet, retweet, bahsetme, harici link içeren tweet ve Bitcoin ile ilgili tartışan Twitter kullanıcılarının bilgilerini içeren kapsamlı veri analizi şeklindedir. Tweetleri pozitif ve negatif olarak sınıflandırmak için tweet veri setine NB Duygu Sınıflandırma modeli uygulanmıştır. Beş bölüme ayrılmış olan veri setinde her bölüm için ayrı ayrı analizler yapılarak detaylı metin analizi yapılmıştır. Sonuçlar, ortak kullanıcıların Bitcoin'e yönelik duygularının çoğunlukla olumlu olduğunu ve Twitter kullanıcılarının Bitcoin'i tartışırken çoğunlukla olumlu tweetler attıklarını göstermiştir [18].

Ceyhan ve ark. (2018) çalışmalarında Bitcoin kripto parasının hareket yönünü tahmin etmeyi amaçlamışlardır. Bu amaç doğrultusunda Günlük Haber Başlıkları ve Bitcoin Tarihi Verileri olmak üzere Temmuz 2010 ile Haziran 2016 tarihleri arasındaki veriler kullanılarak 1466 örnekten meydana gelen iki veri seti oluşturulmuştur. Günlük Haber Başlıkları veri seti günlük olarak çıkan haber başlıklarının birleşimlerinden oluşmaktadır. Bu veri setinde her günün haber başlığı farklı sayıda başlıktan oluşmaktadır ve birleştirilerek her gün için bir metin parçası halinde saklanmaktadır. İkinci veri seti olan Bitcoin Tarihi Verileri veri seti içerisinde açılış, kapanış, düşük, yüksek olmak üzere toplam dört öznitelik bulunmaktadır. Art arda gelen iki günün kapanış değerlerinin birbirlerinden çıkarılması ile hedef değer bulunmuştur. Örneğin ele alınan bir gün için kapanış değeri kendinden önceki değerden düşük ise hedef değer

0'dır ve Bitcoin düşmekte demektir; büyük ise hedef değer 1'dir ve Bitcoin yükselmekte demektir. Daha sonra oluşturulan veri setleri metin ön işleme adımlarından geçirilmiş ve YSA, DVM, LR, RF Algoritmaları ve Adaboost algoritmaları ile modeller oluşturularak Bitcoin hareketi tahmin edilmeye çalışılmıştır. Elde edilen sonuçlar yeterince başarılı bulunmadığı için SVM, LR ve ANN algoritmaları ile oluşturulan modellerin parametreleri üzerinde iyileştirmeler yapılarak Grid Search yöntemi uygulanmış ve sonuçlar alınmıştır. SVM, LR ve ANN modellerinin doğruluk oranları sırası ile 0.66, 0.65 ve 0.67 olarak bulunmuştur. Çalışmada oluşturulan iki veri seti birleştirilerek sonuç alındığında ise Adaboost modeli ile 0.91 değerine ulaşılmıştır [19].

Kinderis ve ark. (2018) projelerinde veri madenciliği tekniklerini kullanarak haber makaleleri ve tweetlerden çektikleri bilgiler ile kripto para birimi arasındaki ilişkiyi anlamaya çalışmışlardır. Bu çalışma doğrultusunda duygu analiz teknikleri ve tahmine dayalı makine öğrenmesinden oluşan karma bir model önerilmiş ve bunun için LSTM RNN (Uzun-Kısa Süreli Bellek Tekrarlayan Sinir Ağları) kullanılmıştır. Büyük miktarda finansal veri Quandl API kullanılarak toplanmıştır. 2013 yılı itibarıyla Bitcoin ile ilgili 8620 makale ve 7.000.000'dan fazla tweet çekilmiş olup tweet ve makaleleri harmanlamak için Selenium aracından yararlanılmıştır. Her gün için elde edilen veriler pozitif, negatif ve nötr olacak şekilde üç gruba ayrılmıştır. Bunun için TextBlob Kütüphanesi kullanılarak metinlerden bilgi veren veri kümeleri oluşturulmuş ve Doğal dil İşleme adımları uygulanmıştır. Günlerdeki duygular ve ertesi gün için finansal olarak tahminleri tanımlayacak değerler tespit edilmiş ve KNN, DVM, LR, Gaussian NB, Karar Ağaçları, Doğrusal Ayrımcılık Analizi algoritmaları uygulanarak Bitcoin kripto parasının fiyatının yönü tahmin edilmeye çalışılmıştır. Çalışmada kullanılan algoritmalar içerisinde en iyi sonuçları Gaussian NB (%67.4) ve Doğrusal Ayrımcılık Analizi (%67.6) oluşturulan modelleri vermiştir [20].

Pant ve ark. (2018) Twitter platformu üzerinden paylaşılan Bitcoin anahtar kelimeli tweetler ile ilgili duygu analizi ve Bitcoin için fiyat tahmini yapmayı hedeflemişlerdir. Duygu analizi yapmak için Twitter kullanıcıları arasından BitcoinMagazine (@BitcoinMagazine), BitcoinNews(@BTCTN), Bitcoin Forum (@BitcoinForums), CoinDesk (@coindesk), Cryptocurrency(@cryptocurrency), Roger Ver (@rogerkver) ve CryptoYoda (@CryptoYoda1338) olmak üzere belirli

kullanıcıların göndermiş olduğu 01.01.2015 ve 31.12.2017 tarihleri arasındaki tweetler çekilerek veri seti oluşturulmuş ve veriler pozitif-negatif olacak şekilde sınıflandırılmıştır. Duygu analizi çalışmasında %81,39 doğruluk değerine ulaşılmıştır. Duygu analizinden sonra Coinmarketcap sitesi üzerinden aynı tarih aralığındaki Bitcoin fiyat verileri çekilerek değişkenlik gösteren Bitcoin fiyatı Tekrarlayan Sinir Ağı (RNN) kullanarak tahmin etmek istenilmiştir. Çalışma kapsamında duygu analizi sırasında elde edilen pozitif ve negatif tweetlerin yüzdeleri alınmış ve Bitcoin geçmiş fiyat bilgileri RNN modeli ile beslenilerek bir sonraki zaman diliminde Bitcoin'in yeni değeri tahmin edilmeye çalışılmış ve %77,62 doğruluk oranı ile fiyat tahmini yapılmıştır [21].

Lamon ve ark. (2017) Bitcoin, Ethereum ve Litecoin kripto paralarının fiyat dalgalanmalarını tahmin eden bir çalışma ortaya koymayı hedeflemişlerdir. Bu çalışma için cryptocoinsnews.com sitesinden yaklaşık 3.600 haber makalesi başlığı ve her para birimi ile ilgili yaklaşık olarak 10.000'er olmak üzere toplam 30.000 civarında tweet Twitter platformu üzerinden çekilmiştir. Günlük haberler ve Twitter verileri üç para birimi için gerçek fiyat değişiklikleri göz önüne alınarak etiketlenmiştir. Seçilen üç kripto para için fiyat verileri ise Kaggle'dan alınarak modele veri girişleri yapılmıştır. Modeller Doğrusal DVM, LR ve NB algoritmaları ile oluşturulmuş ve sonuçlar karşılaştırılmıştır. LR algoritması Bitcoin ve Litecoin kripto paraları için en iyi sonucu vermiştir. NB algoritması ile de Ethereum için en iyi sonuç elde edilmiştir. Çalışma için oluşturulan modelin son sürümü ile ortalama olarak 67 gün boyunca yüzdelik en büyük artış ve yüzdelik en büyük düşüşlerin olduğu günler doğru bir şekilde tahmin edilmiştir [22].

Colianni ve ark. (2015) Twitter'dan çekilen kripto paralar ile ilgili tweetlerin kripto para ticaretinde strateji geliştirmek için kullanılıp kullanılmayacağını ortaya çıkarmak için bir çalışma yapmışlardır. Çalışmada kullanılan veri seti Twitter medyası üzerinden Bitcoin anahtar kelimesi içeren gerçek zamanlı tweetler toplanılarak oluşturulmuştur. Bitcoin değerinin belirli bir zaman dilimi içerisinde yükseleceğini veya düşeceğini tahmin edebilmek için duyarlılık analizi algoritmaları ve metin sınıflandırma algoritmalarından NB, DVM, LR algoritmaları ile model oluşturulmuştur. Veri seti %70 eğitim ve %30 test veri seti olmak üzere ikiye ayrıldıktan sonra model eğitim seti ile eğitilip test verisi ile doğru olarak

sınıflandırılma oranları hesaplanmıştır. Bu işlemler on defa tekrarlanarak her tekrar için doğru olarak sınıflandırma oranlarının toplamı alınmıştır. Bu çalışma kapsamında kullanılan algoritmalar arasında en iyi performans Bernoulli NB ile günlük doğruluk oranı %95 ve saatlik doğruluk oranı %76,23 olarak elde edilmiştir [23].

Literatür çalışmaları göz önünde bulundurulduğunda yapılan duygu analizi çalışmalarının genel olarak İngilizce metinler üzerinde yoğunlaştığı, Türkçe metinler üzerinde yapılan duygu analizi çalışmalarının ise diğer diller üzerindeki çalışmalara nazaran çok daha az olduğu görülmektedir. Buna ek olarak günümüzde oldukça popüler bir konu olan ve gündemde kalmayı başaran kripto paralar ile ilgili duygu analizi çalışmalarının henüz yeterli sayıda olmadığı düşünülmektedir.

Tez çalışmasının amacı literatürdeki Türkçe duygu analizi çalışmalarına katkı sağlamak ve güncel bir konu olan kripto paralar ile ilgili duygu analizi yaparak piyasa ile ilgilenen kullanıcılara piyasanın genel durumu ile ilgili bilgi verebilmektir. Bu amaç kapsamında Derin öğrenme algoritmalarından CNN, RNN, LSTM, GRU ve on dört hibrit mimari kullanılarak Kripto paralar, NFT ve DeFi teknolojileriyle ilgili duygu analizi yapılmış ve sonuçlar karşılaştırılmıştır.

Çalışmada kullanılan veri seti ilk kripto para birimi olan Bitcoin, Bitcoin'den sonra en çok bilinen ikinci para birimi Ethereum ve diğer en çok bilinen kripto para birimlerinden Solana, Ripple, Avalanche, Chiliz, Bitcicoine altcoinleri ve NFT-DeFi ile ilgili Twitter platformundan atılan belirli tarihlerdeki Türkçe tweetlerden oluşmaktadır. Farklı kripto para birimlerinden tweetlerin çekilmesinin amacı veri setinin çeşitlendirilmek istenmesidir. Oluşturulan veri seti tweetlerin anlamlarına göre pozitif ve negatif olmak üzere 2 ayrı sınıfa ayrılmıştır. Veri setinin oluşturulması süreci ayrıntılı olarak Materyal ve Yöntemler bölümünde anlatılmaktadır.

3. GENEL BİLGİLER

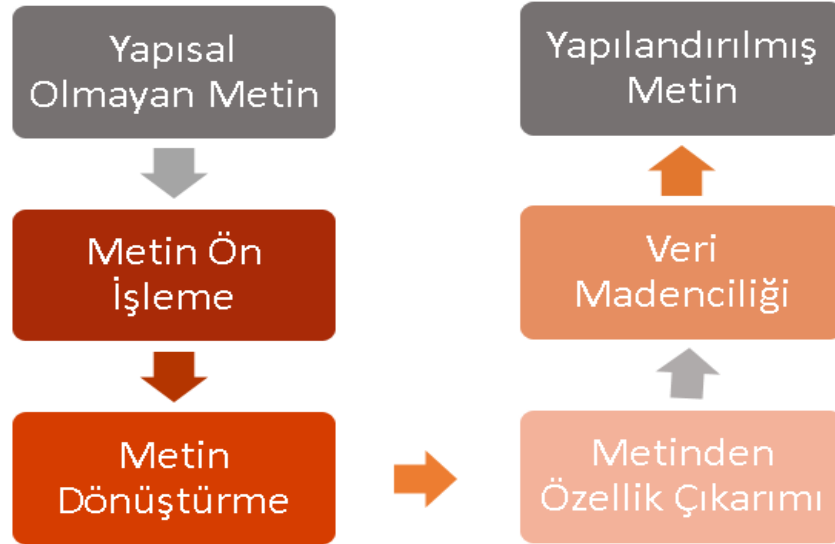
3.1. Metin Madenciliği

Metin madenciliği, yapısal olmayan düzensiz yazılı kaynaklara belirli işlemler uygulanarak metinlerden yapısal ve düzenli veri elde etme sürecidir [24]. Yapısal olmayan düzensiz yazılı kaynaklara örnek olarak sosyal medya verileri, anketler, blog yazıları, ürün yorumları, kitap özetleri gibi metin verileri verilebilir. Bu kadar çok yapısal olmayan metinlerin içerisinde birçok gereksiz bilgi ile birlikte kişi veya kurumların ihtiyacı olan birçok bilgi de barınabilmektedir. Metin verileri arasından sadece gerekli olan bilgiyi alarak metin analizi yapmak için metin madenciliği teknikleri kullanılmaktadır [25].

Veri madenciliğinin bir alt başlığı olan metin madenciliği ile veri madenciliği arasındaki temel fark işlenerek anlamlı bir hale getirilmek istenen verilerin metinlerden oluşmasıdır [26]. Veri madenciliği çalışmasına örnek olarak teknolojik ürünler satılan e-ticaret sitesinde bilgisayar alan bir müşteri için birlikte alınması muhtemel ürünler olan klavye ve fare model önerileri verilebilir. Aynı e-ticaret sitesinde müşterinin alışveriş sonrası yazdığı yorumlar için ise metin madenciliği teknikleri kullanılır. Metin madenciliği sürecinde bilgisayarlara doğal dil işleme ile metinleri nasıl analiz edeceklerini öğretmek metinlerden bilgi çıkarımı, metinlerin özetlenmesi, bilgilerin görselleştirilmesi, duygu analizi, metinlerin sınıflandırılması ve kümeleme işlemlerinde bilgisayarlardan yararlanılabilmektedir [27].

3.1.1. Metin Madenciliği Süreçleri

Metin analizi yapmanın en çok kullanılan yöntemi olan metin madenciliği yönteminde Şekil 3.1.1’de gösterildiği gibi temel altı işlem basamağından bahsedilebilir.



Şekil 3.1.1. Metin madenciliği adımları

Bu adımlar şunlardır:

- Sosyal medya, haber başlıkları, web sayfaları, kişisel bloglar vb. metin kaynaklarından ham (yapılandırılmamış) verilerin toplanılması.
- Metin ön işleme adımında metindeki bütün harflerin büyük veya küçük harfe çevrilmesi, noktalama işaretleri ve duruma göre sayıların silinerek metinde sadece kelimelerin bırakılması, gereksiz kelimelerin (stopwords) bulunup silinmesi, kelimelerin kök haline getirilmesi, en az geçen belirli sayıdaki (örneğin 1000) kelimelerin bulunup metinden silinmesi, yazım hatalarının bulunup düzeltilmesi.
- Yaygın olarak tercih edilen Kelime Çantası (Bag of Words) modeli kullanılarak öznitelik üretimi.
- Öznitelik seçimi için metinde geçen önemli kelimelerin belirlenip alakasız kelimelerin metinden çıkartılması. TF, TF-IDF, Word2Vec, GloVe, FastText gibi birçok yöntem arasından bir veya bir kaçının uygulanarak kelime gömme işleminin yapılması.
- Makine öğrenme veya derin öğrenme algoritmalarından bir veya birden fazla algoritma kullanılarak yapılandırılan metnin analiz edilmesi.
- Son olarak kullanılan algoritmalar sonucu elde edilen sonuçların performans değerlendirme ölçütleri kullanılarak değerlendirilmesi ve raporlanması.

3.2. Duygu Analizi

Teknolojinin ilerlemesi ve internetin yaygınlaşması ile birlikte insanlar herhangi bir konu üzerindeki duygu ve düşüncelerini uzun uzun yazmak yerine birkaç kelimeden oluşan kısa cümlelerle sosyal medya üzerinden ifade etmeye başlamışlardır [9]. Belirli bir düzen içerisinde olmadan dağınık bir şekilde sosyal medya üzerinden paylaşılan bu düşünce yazılarını farklı alanlardaki araştırmalarda kullanabilmek için pratik bir şekilde analiz etme ihtiyacı doğmuştur. Analiz etme yöntemleri içerisinde en çok tercih edilen yöntemlerden biri duygu analizi yöntemidir.

Duygu analizi, bir yazı ya da yazarın belirlenen bir konu ile ilgili duygu ve düşüncelerinin doğal dil işleme (NLP), metin analizi, sınıflandırma algoritmaları, istatistik ve programlama gibi çeşitli yöntemler kullanarak olumlu veya olumsuz olduğunu belirlemede kullanılan yöntemin genel bir adıdır [9]. Diğer bir ifadeyle duygu analizi metinler içerisinde gizlenmiş olarak bulunan fikir ve görüşlerin hangi duygu sınıfına ait olduğunun bulunması ve araştırmalarda kullanılmaya uygun bir biçime getirilmesi işlemidir.

Duygu analizi yapmak için tercih edilen konu başlıkları oldukça fazladır. Örneğin haber başlıkları, makaleler, anketler, kitap özetleri gibi birçok metin üzerinde sağlık, spor, magazin, ekonomi, bir kişi veya ürün hakkında duygu analizi çalışmaları yapılabilmektedir. Facebook, Twitter, Instagram gibi popüler sosyal medya platformları üzerinde yapılan duygu analizi çalışmaları ise en çok tercih edilenler arasındadır. Çünkü insanlar bir konu ile ilgili hissettiklerini; sevinç, öfke, kızgınlık, umutsuzluk, mutluluk gibi olumlu veya olumsuz duygularını genellikle anlık olarak sosyal medya üzerinden paylaşmaktadır. Bu sebepten dolayı sosyal medya üzerinden paylaşılan metinler duygu analizi çalışmalarında kullanılabilen elverişli veri kaynaklarıdır.

Literatür araştırması bölümünde de bahsedildiği gibi çalışmanın konusu ile ilgili Türkçe dilinde yapılan duygu analizi çalışmaları diğer dillere kıyasla oldukça azdır. Çalışmada yapılan duygu analizi çalışması için oluşturulan veri seti Twitter kullanıcıları tarafından gönderilen belirli anahtar kelimeli Türkçe dilindeki tweetlerden toplanılarak aynı zamanda literatüre katkı sağlamak hedeflenmiştir.

3.3. Makine Öğrenmesi

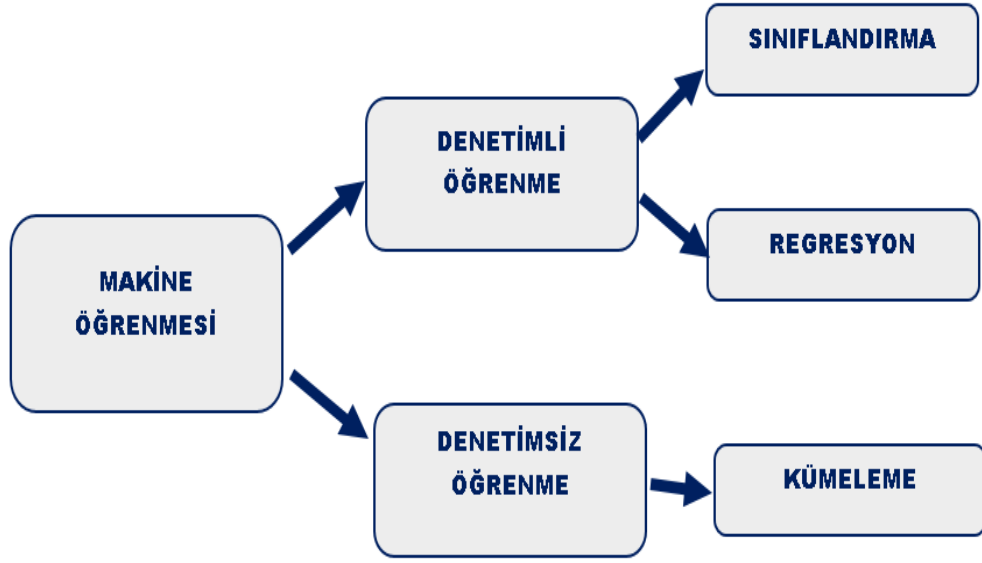
İnsanlar zekaları sayesinde anlama, öğrenme, analiz etme, yeni karşılaştığı bir sorunla başa çıkabilme gibi yeteneklere sahiptir. Bilgisayarların bu yeteneklere tümüyle sahip olması mümkün olmasa da algoritmalar ile birtakım öğrenimler sağlanabilir. Yazılımlar aracılığıyla bilgisayarların insan zekasını örnek alıp taklit ederek onlar gibi davranması durumu yapay zeka olarak adlandırılmaktadır [28].

Yapay zekanın bir parçası olan makine öğrenmesi, bilgisayarlara geçmiş verilerden yola çıkarak henüz karşılaşılmamış bir durum için otomatik öğrenme yeteneği kazandırılmaktadır. Diğer bir deyişle makine öğrenmesi bir önceki deneyimlere benzer bir durum ile karşılaşan bilgisayarın herhangi bir dış müdahale olmadan karar almasını sağlayan teknolojidir.

Makine öğrenmesinde deneyimler arttıkça doğru orantılı olarak performans da artar. Son yıllarda makine öğrenmesinde oldukça başarılı sonuçlar alınmasının en önemli sebeplerinden biri deneyimlerin yani toplanılmış geçmiş verilerin her geçen gün artması ve büyük veriler ile öğrenme gerçekleştirilmesidir [28].

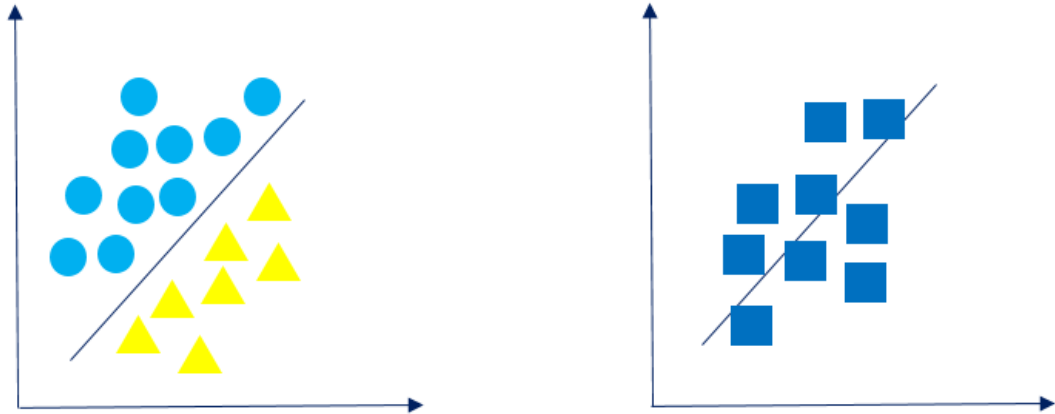
Makine öğrenmesi birçok alanda farklı görevler için kullanılmakta ve başarılı sonuçlar alınmaktadır. Örneğin mail adresine gelen gönderiler içerisinde spam mailleri tanıyarak onları ayrı bir klasörde tutmak için makine öğrenmesi teknikleri kullanılmaktadır. Veya ele alınan bir bölgedeki şehirlerin geçmiş yıllardaki yağış miktarı verilerinden yola çıkarak bu yılki yağış miktarı tahminini yapmak makine öğrenmesi teknikleri ile gerçekleştirilmektedir. Hastalık tanılarının konulması, ses tanıma işlemi, alışveriş yapan müşteriye yaptığı alışverişten yola çıkarak birlikte alınması muhtemel benzer ürünlerin önerilmesi, trafik yoğunluğunun tahmin edilmesi, hava durumu tahminleri ve duygu analizi gibi birçok örnekte makine öğrenmesi tekniklerinden yararlanılmaktadır.

Makine öğrenmesi Şekil 3.3.1’de gösterildiği gibi denetimli öğrenme ve denetimsiz öğrenme olmak üzere ikiye ayrılmaktadır.



Şekil 3.3.1. Makine öğrenmesi algoritmaları

Denetimli öğrenme eğitilmiş olan bağımsız veri seti kullanılarak eğitilmemiş bağımlı değişkenin tahmin edildiği yöntemdir. Denetimli öğrenme yönteminde eldeki veriler girdi olarak verilir ve bağımsız değişken ile bağımlı değişken arasında eşleme yapılarak bir fonksiyon üretilir [24]. Denetimli öğrenme yöntemi sınıflandırma ve regresyon problemlerinde kullanılmaktadır.



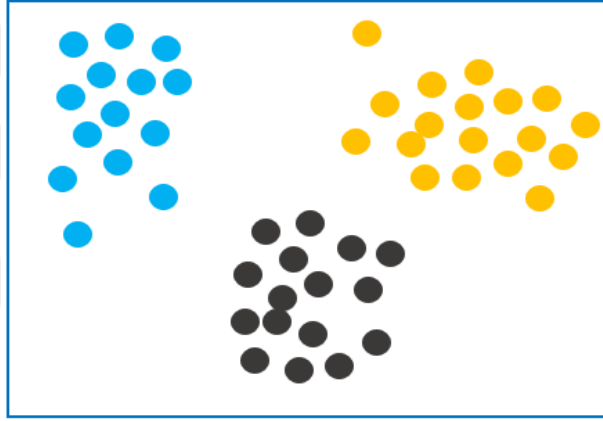
Şekil 3.3.2. Sınıflandırma (soldaki) ve regresyon (sağdaki) problemleri

Sınıflandırma yani kategorize etme etiketlenmiş veriler kullanılarak yeni verinin hangi sınıfa ait olduğunu bulma işlemidir. Bu tez çalışmasında Twitter'dan toplanan kripto paralar ve NFT-DeFi teknolojisi ile ilgili toplanan tweetler pozitif ve negatif olmak üzere iki kategoriye ayrılmış yani sınıflandırma yapılmıştır. Gelen

maillerin spam veya spam değil diye ayrılması da aynı şekilde bir sınıflandırma problemidir. Bir diğer sınıflandırma örneği satış sonrası müşteri yorumlarının pozitif, negatif ve nötr olarak sınıflandırılmasıdır.

Regresyon probleminde ise veriler sayısal değerdedir ve gelen veriler göz önünde bulundurularak tahmini sonuçlar üretilir. Örneğin bir arsanın büyüklüğü ve konumuna bağlı olarak fiyat tahmininin yapılması bir regresyon problemidir.

Denetimsiz öğrenme yönteminde denetimli öğrenmede olduğu gibi etiketlenmiş veriler ve eğitim verisi yoktur. Sadece etiketsiz veriler kullanılarak modelin bilgileri kendi başına keşfetmesi beklenir. Bu yöntemin en bilinen örneği Şekil 2.3.3’de gösterilen kümeleme problemleridir.



Şekil 3.3.3. Kümeleme analizi sonrası farklı renklerdeki dairelerin 3 kümeye ayrılması [29]

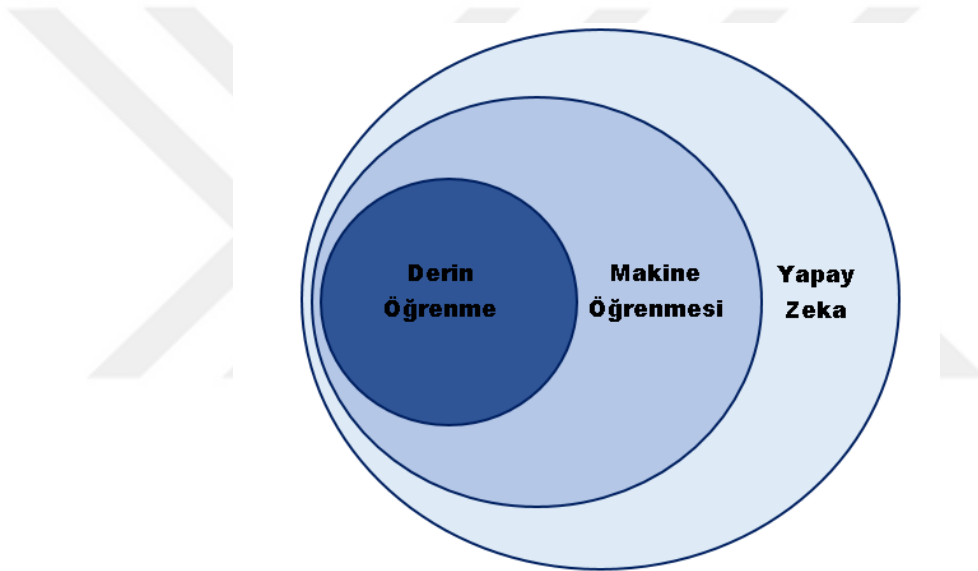
Kümeleme problemi verilen veri seti içerisinde birbirine benzer olan verileri aynı kümeye alarak verileri gruplandırma işlemidir. Google fotoğraf arşivinde bitki, hayvan ve insanların ayrı kümelere ayrılması kümeleme yöntemine güzel bir örnektir [30].

3.4. Derin Öğrenme

Veri kaynakları arasında önemli bir yeri olan metin belgeleri muazzam büyüklükteki verilerdir ve hacimleri her geçen gün artmaktadır. İnsan gözü ile büyük metin verileri arasından gerekli bilgiye ulaşılması ise ancak çok büyük emekler harcanarak uzun bir zaman dilimi içerisinde gerçekleştirilebilir. Gelişen teknolojiler ile birlikte bu işlemleri bilgisayar ortamında düşük hesap maliyetleri ile kısa bir zaman

dilimi içerisinde gerçekleştirmek mümkün hale gelmiştir. Henüz yeni bir teknoloji olan derin öğrenme yöntemi kullanarak büyük veriler üzerinde etkin öğrenme modeli oluşturma ile büyük başarılar elde edilmektedir [31].

Derin öğrenme Şekil 3.4’de gösterildiği gibi Tek Katmanlı Yapay Sinir Ağları ve Çok Katmanlı Yapay Sinir Ağları’ndan sonra geliştirilmiş olan makine öğrenme algoritmalarından biri ve yapay zekanın bir uygulamasıdır [32, 33, 34]. Yapay Sinir Ağları (YSA) insan beyninden esinlenilerek geliştirilmiştir ve bilgisayarın insan gibi kararlar alıp sonuçlar üretmesini mümkün kılar. YSA kullanılarak ses işleme, görüntü işleme, tanı ve teşhis koyma sınıflandırma vb. çok fazla problem bilgisayar aracılığıyla yapılabilmektedir.



Şekil 3.4. Yapay zeka, makine öğrenmesi ve derin öğrenme arasındaki ilişki [35]

Makine öğrenme yönteminin Derin Öğrenme yönteminden farkı algoritmaların insanlar tarafından oluşturulmuş olması ve sonradan öğrendiklerini kullanarak kararlar almalarıdır. Derin Öğrenme yönteminde ise insanların algoritmalara ne yapacağını söylemesine ihtiyaç yoktur, YSA kullanılarak algoritmaların öğrenmesi direkt olarak ses, görüntü veya metin verileri üzerinden gerçekleştirilmektedir [36]. Yani kısaca Derin Öğrenme eldeki veriler kullanılarak yeni bilgiler elde etmeye yarayan bir yaklaşımdır.

4. MATERYAL VE YÖNTEMLER

4.1. Kelime Gömme Teknikleri

Bu bölümde en çok tercih edilen kelime gömme teknikleri arasından TF-IDF, Wor2Vec, Google BERT, FastText ve GloVe algoritmalarının açıklamalarına yer verilmiştir.

4.1.1. Terim Frekansı ve Ters Doküman Frekansı (TF-IDF)

Kısaca TF-IDF olarak adlandırılan yöntemde Terim Frekansı bir doküman içerisinde geçen kelimelerin kaç defa geçtiğinin sayısını bulmaya yarar. Ters Doküman Frekansı ise kelimenin dokümanlar içerisinde kaç defa geçtiğini bulmaya yarar. Ters doküman frekansı yani IDF bulunurken doküman sayısının, bulunduğu doküman sayısına oranı bulunarak elde edilen değerin mutlak logaritmik değeri alınır. Çıkan sonuç TF değeri ile çarpılarak TF-IDF değeri elde edilir.

Örneğin 100 kelimeden oluşan bir dokümanda altcoin kelimesinin 5 kez geçtiğini varsayalım. Bu durumda $TF = \frac{5}{100} = 0.05$ olarak bulunur. Veri setindeki toplam doküman sayısı 10.000 ve altcoin kelimesinin bulunduğu dokümanların sayısı ise 100 olduğunu varsayarsak $IDF = \log\left(\frac{10000}{100}\right) = 2$ olarak bulunur. Sonuç olarak bu örnekte altcoin terimi için TF-IDF değeri $TF * IDF = 0.05 * 2 = 0.1$ 'dir. Aynı örnek için altcoin teriminin geçtiği doküman sayısı 1000 olsaydı TF değeri aynı kalacak fakat $IDF = \log\left(\frac{10000}{1000}\right) = 1$ olacak yani değeri düşecekti. Bu durumda $TF * IDF = 0.05 * 1 = 0.05$ olacak yani orantılı olarak TF-IDF değeri de düşecektir.

Örneklerde alınan sonuçlardan anlaşılacağı üzere kelimenin bir doküman içerisinde sık geçmesi ve tüm dokümanlar arasında az sayıda dokümanda bulunması o kelimeyi değerli yapar. TF-IDF değeri bir kelimenin dokümandaki önemi hakkında bilgi vermek için elverişli olabilir fakat eksik olan tarafı kelimenin semantik (anlamsal) bilgi vermemesidir [33].

4.1.2. Word2Vec

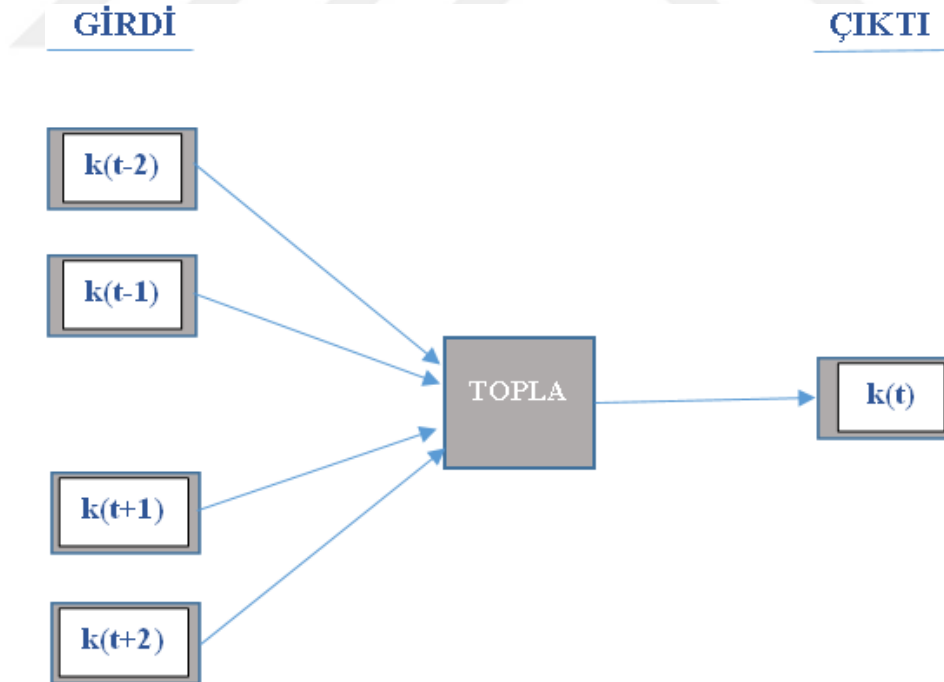
Word2Vec kelimeleri vektör uzayında göstermeye yarayan tahmin temelli ve denetimsiz bir modeldir [37]. 2013 yılında araştırmacı Tomas Mikolov ve ark. tarafından geliştirilmiştir [38]. Belirli bir dokümanda geçen her bir kelime için

kelimelerden vektörler oluşturarak semantik bilgi vermeye çalışır. Oluşturulan vektörler metin sınıflandırma, duygu analizi, metinden bilgi çıkarımı gibi uygulamalarda kullanılabilir.

Word2Vec modelinin Continuous Bag of Words yani kısaltılmış hali ile CBOW ve skip-gram olmak üzere birbirine benzer iki temel algoritması vardır. Her iki temel algoritmada da dokümanda bulunan diğer kelime veya kelimelerin bulunması için çalışılır. Word2Vec modelinin bu iki algoritmasının arasındaki fark ise girdi ve çıktıların farklı olmasıdır.

4.1.2.1. Sürekli Kelime Torbası (CBOW)

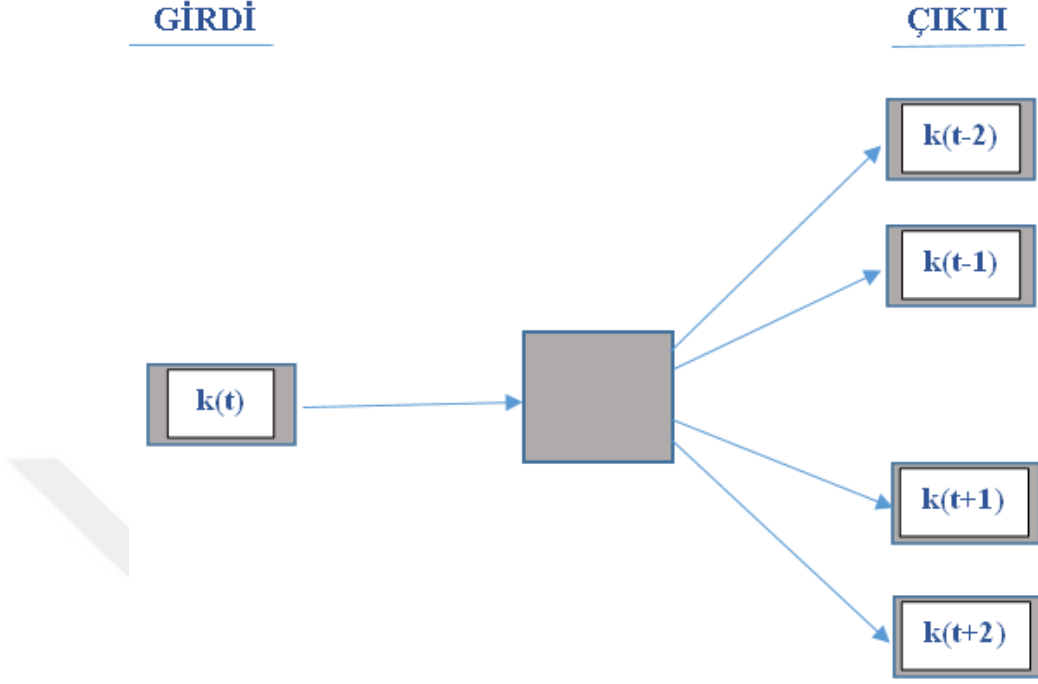
Bu öğrenme algoritmasında Şekil 4.1.2.1’de görüldüğü üzere bulunması istenilen kelimenin verilen pencere boyutu sayısı kadar sağ ve sol yanında bulunan kelimelere bakılarak istenilen kelime tahmin edilmeye çalışılır [37]. Pencere boyutu sayısı kelimenin sağında ve solunda yer alan kelimelerden kaç tanesine bakılması gerektiğinin bilgisinin tutulduğu değişkendir. CBOW algoritması az miktarda veri ile çalışıldığı takdirde güzel sonuçlar verebilir.



Şekil 4.1.2.1. CBOW mimarisi [39]

4.1.2.2. Skip-Gram

Bu algoritmanın çalışma mantığı ise Şekil 4.1.2.2’de gösterildiği gibidir:



Şekil 4.1.2.2. Skip-gram mimarisi [39]

Verilen şekilde görüldüğü gibi skip-gram algoritmasında girdi olarak verilen kelime kullanılarak sağında ve solundaki kelimeler tahmin edilmeye çalışılır. Skip-gram algoritmasıyla büyük boyutlardaki veri setlerinde oldukça başarılı sonuçlar alınabilmektedir.

$$\operatorname{argmax}_{\theta} \frac{1}{T} \sum_{t=1}^T \sum_{-C \leq j \leq C, j \neq 0} \log P_{\theta}(w_{t+j} | w_t) \quad (4.1)$$

Skip-gram modeli yukarıda verilen Denklem 4.1'e göre belirlenir [40]. Denklemde $w_1, w_2, \dots, w_T$ eğitim kelimelerinin bir dizisi T uzunluğunda gösterilmiştir. C, eğitim bağlamının boyutunu ve $P(w_{t+j} | w_t)$ ise θ ile gösterilen bir dizi parametreye sahip bir sinir ağını ifade eder.

4.1.3. Google BERT

Bidirectional Encoder Representations from Transformers kelime grubunun baş harflerinden oluşan BERT, Google tarafından 2018 yılında geliştirilen bir

algoritmadır. İngilizce BERT İngilizce Wikipedia'dan toplanan etiketlenmemiş 2500 milyon kelime ve BooksCorpus'tan alınan 800 milyon etiketlenmemiş kelime ile oluşturulan büyük boyutlara sahip veri kümesi kullanılarak eğitilmiştir [41].

BERT temel anlamda transfer makine öğrenme ve yapay zeka teknolojilerini bir arada kullanan doğal dil işleme tekniğidir. Transfer öğrenme yönteminde eğitilmiş olan ağ bilgileri benzer problemler içeren başka bir ağa aktarılarak çalışılır. Google'ın bu algoritmayı geliştirmesindeki temel amaç kullanıcılarının arama sorgularını analiz edip aramak istedikleri konu hakkında onlara daha doğru sonuçlar vermeyi sağlamaktır. BERT'in diğer modellerden farkı iki yönlü olarak çalışarak her bir kelimenin diğer kelimelerle olan bağlantısını göstermesidir.

4.1.4. FastText

FastText modeli 2016 yılında Facebook ekibi tarafından geliştirilmiş olan Word2Vec tabanlı bir modeldir ve Word2Vec modelinin temel algoritmalarından biri olan skip-gram algoritması baz alınarak geliştirilmiştir.

FastText modelinde kelimeler Word2Vec modelinde olduğu gibi en küçük birimler olarak doğrudan kullanılmaz. Bunun yerine kelimeler n-gramlarına yani alt kelimelerine ayrılarak öğrenme gerçekleştirilir. Örneğin Bitcoin terimin 3-gramları ['bit', 'itc', 'tco', 'coi', 'oin'] olacaktır.

N-gram kullanan bu yöntem ile kelimelerin yapıları dikkate alınmaktadır. Kelimelerin alt kelimelere ayrılması sayesinde bir terim kelime dağarcığında bulunmasa bile anlamsal benzerlik içeren kelimeler bulunup getirilecektir. Ek olarak yanlış yazılan bir kelime doğru yazılan kelimeye olan benzerliğinden dolayı kelime gösterimi çok benzer olabilecektir.

4.1.5. GloVe

Kelime gömme yöntemlerinden biri olan GloVe, "Global Vectors for Word Representation" kelime grubunun baş harflerinden oluşmaktadır [42]. Word2Vec algoritmasından sonra en çok tercih edilen ikinci yöntem olan GloVe, kelimelerin vektörel olarak temsil edilmesine yarayan denetimsiz bir derin öğrenme algoritmasıdır [42, 43].

4.2. Derin Öğrenme Algoritmaları

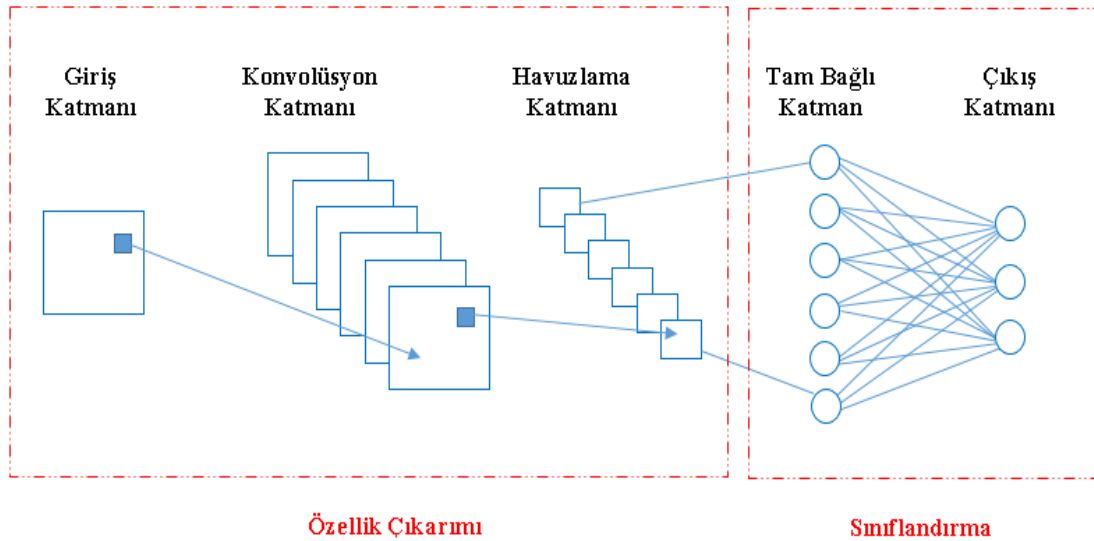
Bu başlık altında Evrişimli Sinir Ağları (CNN), Tekrarlayan Sinir Ağları (RNN) ile RNN yönteminden türetilmiş olan LSTM-BLSTM ve GRU yöntemlerinin açıklamalarına yer verilmiştir.

4.2.1. Evrişimli Sinir Ağları (CNN)

Convolutional Neural Networks kelimelerinin baş harflerinden oluşan ve Türkçeye Evrişimsel ya da Konvolüsyonel Sinir Ağı olarak çevrilen CNN algoritmaları, verileri ızgara benzeri bir topoloji kullanarak işleyen derin sinir ağları mimarileridir. Evrişim, doğrusal (linear) bir matematiksel işlemdir ve algoritma bu lineer matematiksel işlem süreçlerinden geçtiği için evrişimsel adını almıştır [44].

Genel olarak görselleri girdi olarak alarak cisimlerin tanınması, hastalıkların teşhisi, plaka tanıma, nesne takibi gibi görüntü işleme alanlarında tercih edilen ve oldukça yüksek performans gösteren CNN algoritması NLP alanında da tercih edilmeye ve bu alanda da CNN ile güzel sonuçlar alınmaya başlanmıştır.

CNN algoritmaları Şekil 4.2.1’de görüldüğü gibi beş katmandan oluşmaktadır:



Şekil 4.2.1. CNN mimarisi [37]

İşlenecek olan metin veya görüntü verileri yeniden boyutlandırılarak CNN ağına verilir. Yeniden boyutlandırma işlemi sırasında veri boyutu tutarlı ve dengeli seçilmelidir. Çünkü veri boyutu büyüdükçe CNN algoritmaları girdiden daha fazla

özellik çıkarır, sistemin öğrenme başarısı artar ve daha yüksek sınıflandırma başarıları yakalanır. Fakat bu kadar avantajlı durumun yanında büyük boyutlara sahip verinin girdi olarak verilmesi sistemin donanımsal ihtiyaçları ve sistem eğitim süresinin artması gibi dezavantajlı durumlara da sebep olur [37]. Ağa verilen girdinin özellik çıkarımı konvolüsyon katmanında gerçekleştirilerek bu katmandan alınan çıktı bir sonraki katman olan havuzlama katmanına çıktı olarak verilir. Havuzlama katmanında çıktının boyutu ortalama havuzlama, maksimum havuzlama veya L2-norm havuzlama tekniklerinden biri kullanılarak düşürülür. Böylece ağı hafıza ve hesaplama maliyetlerinin önüne geçilir. Son olarak tam bağlı katman kendisinden önce gelen katmanlardan çıkan veriyi girdi olarak alarak bir tahmin üretir.

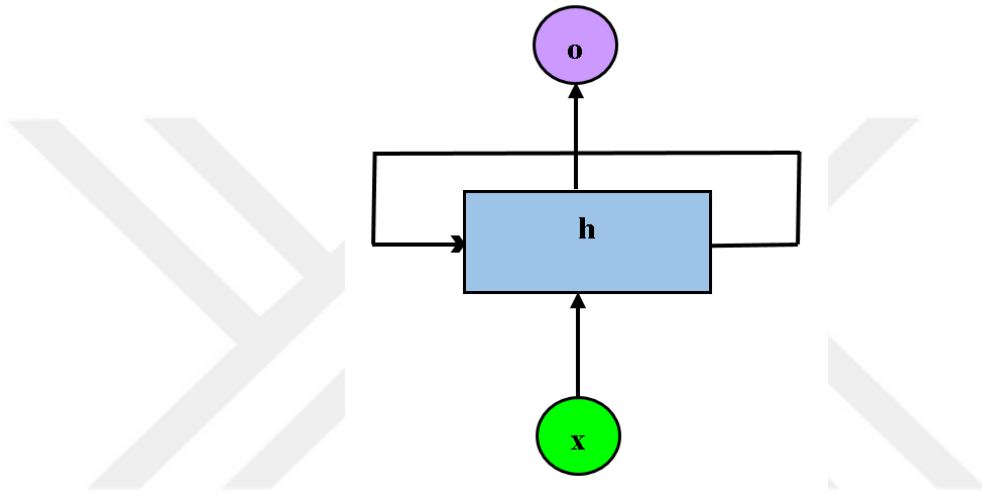
4.2.2. Tekrarlayan Sinir Ağları (RNN)

İnsan beyni geçmiş bilgileri gelecekte de hatırlayıp kullanabilmek için depolayan karmaşık işlemler zinciridir. Zeka sayesinde insanlar fark eder, kaydeder, hatırlar ve unutmaz. Örneğin bir şarkı duyduğumuzda bu şarkının adını, sözlerini, ne zaman ve nerede dinlediğimizi, dinlerken bizde bıraktığı hissi hatırlarız. Geçmişimize dair bu bilgiler gözümüzün önünden bir film şeridi gibi geçer. Aslında biz şarkıyı duyduğumuz anda olan şey beynimizin farklı bölgelerinde kaydedilen hatıraları yeniden yapılandırmasıdır.

Geleneksel YSA yöntemi beynin yaptığı gibi geçmiş bilgileri hatırlayamaz. Bu yüzden insan zekasından farklı olarak her girdi ve çıktı diğerlerinden bağımsızdır [33]. Bu dezavantaj sebebiyle bazı problemlerde doğru sonuca ulaşmak zorlaşmaktadır. Örnek olarak cümle içerisinde bulunan bir kelimenin tahmin edilebilmesi için o kelimeden önce yer alan kelimeleri bilmek gerekmektedir. Fakat girdilerin birbirinden bağımsız olduğu ve hatırlama özelliği olmayan bir zeka ile geçmiş kelimelere ulaşmak imkânsızdır. YSA yöntemine eklemeler yapıp yeni yaklaşımlar geliştirilerek bu sorun çözülmeye çalışılmıştır.

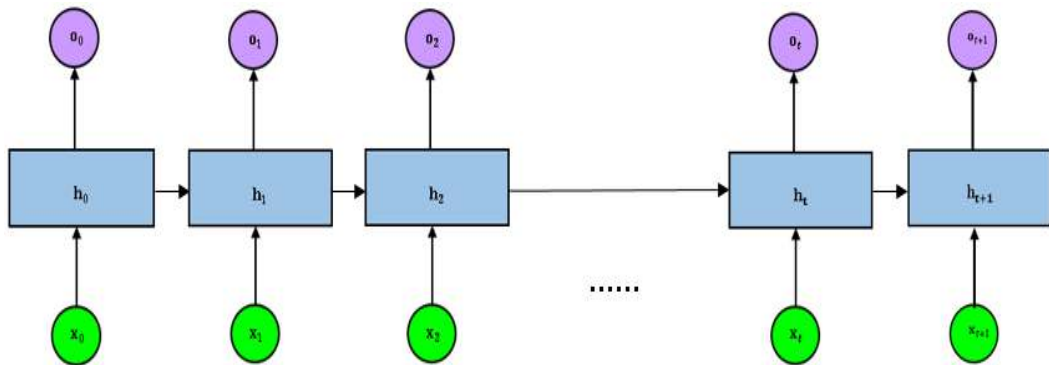
Tekrarlayan Sinir Ağları (Recurrent Neural Networks, RNN), Geleneksel Sinir Ağları yöntemine hatırlama özelliğinin eklenmiş halidir. Ayrıca Geleneksel Sinir Ağlarında her girdi birbirinden bağımsız iken RNN'lerde girdiler birbirleri ile bağımlıdır [35]. RNN yönteminde sıralı bir şekilde girilen bilgiler bir sonraki adımı tahmin etmek için kullanılır.

RNN algoritması sahip olduğu döngüler sayesinde öğrenilmiş geçmiş bilgileri gelecekte bir noktada tahmin yapmak için saklar [33]. Tezde kullanılan veri setinden örnek bir tweet ile bu durumu açıklamak gerekirse “Japonya Merkez Bankası eleştirileri sonrası Bitcoin geriledi.” cümlesinde ilk olarak Japonya kelimesi modele girdi olarak verilir ve bir çıktı üretilir. Bir sonraki adımda Merkez kelimesi ile birlikte bir önceki adımda üretilen çıktı girdi olarak modele verilir ve bu adımda da bir çıktı üretilerek bir sonraki adımın girdisi ile birlikte modele verilir. Son adıma kadar model bu şekilde bir önceki adımların çıktıları ile bağlantılı çalışarak girdiler üretir.



Şekil 4.2.2.1. RNN mimarisi [45]

RNN mimarisindeki döngü Şekil 4.2.2.1’de gösterilmiştir. Görselde x ile ifade edilen düğüm girdi ve o ile ifade edilen düğüm çıktıdır. h ağı x girdisinden o çıktısını üretmiştir. Döngü sayesinde alınan çıktı bir sonraki adımda kullanılmaktadır.



Şekil 4.2.2.2. RNN mimarisi (açık gösterim) [45]

RNN'in açık mimarisi ise Şekil 4.2.2.2'de gösterildiği gibidir. Şekilde h_0 ağı x_0 girdisini alıp çıktı olarak o_0 çıktısını üretmiştir. Bir sonraki adımda ise h_1 ağı hem x_1 'i hem de h_0 ağının üretmiş olduğu o_0 çıktısını girdi olarak alır. Diğer adımlarda da bu şekilde her bir çıktı için bir önceki adımda öğrendiği bilgiler ile yeni bilgiyi girdi olarak alıp yeni bir çıktı üreterek ilerler.

$$o_t = f(o_{t-1}, x_t) \quad (4.2.1)$$

Denklem 4.2.1'de o_t t . adımdaki o değerini, o_{t-1} bir önceki adımdaki o değerini, x_t ise t . adımdaki girdiyi ifade etmektedir. RNN mimarisinin açık denklemi ise Denklem 4.2.2'de verilmiştir.

$$o_t = \tanh(o_{t-1} * W + x_t * U) \quad (4.2.2)$$

Denklem 4.2.2'de f fonksiyonu yerine $\tanh$ kullanılmıştır. Bir önceki adımdaki o değeri ile W ağırlığının çarpımına t . adımdaki girdi ile şimdiki ağırlık olan U değerinin çarpımı eklenerek çıkan sonuç $\tanh$ fonksiyonuna konulmuş ve ağı t . adımdaki çıktısı bulunmuştur.

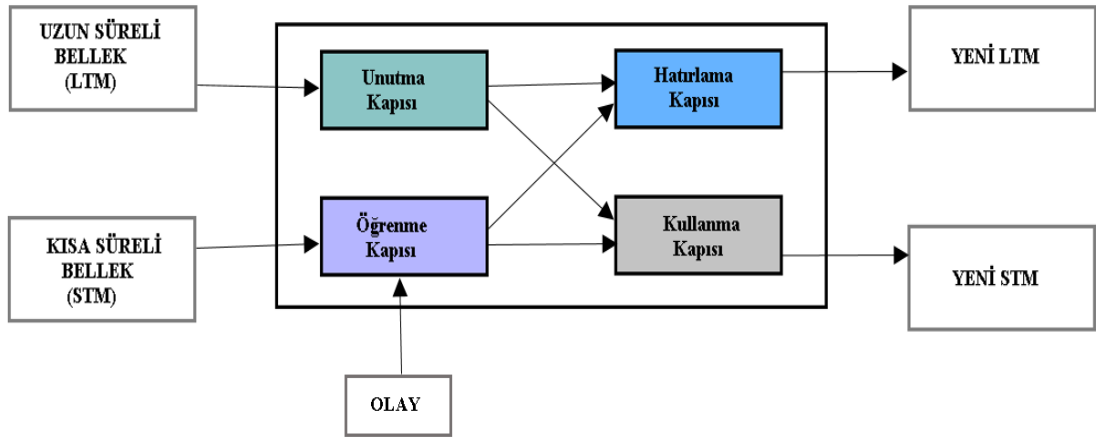
RNN algoritmasında bir önceki ağı çıktısı ile ilişki kurulup yeni girdi ile bağlantılı bir şekilde ilerlemesi önceki girdilerin hatırlanabilmesini sağlar. Hatırlama özelliği sayesinde RNN sıralı verilerin kullanıldığı problemlerde tercih edilmektedir. Fakat algoritmada çok fazla girdi değerinin olması hesaplama zamanını arttırmakta ve önemli olarak nitelendirilen geçmiş bilgilerin bir süre sonra kaybolmasına neden olmaktadır. RNN modelinin çalışma mantığına örnek olarak verilen “Japonya Merkez Bankası eleştirileri sonrası Bitcoin geriledi.” cümlesinde model son kelimeyi tahmin edebilir fakat kelime sayısı arttıkça modelin önceki çıktıları hatırlaması zorlaşır. Bu problemin çözümü için hangi verileri hatırlaması gerektiğini bilen yeni bir model önerilmiştir.

4.2.3. Uzun Kısa Süreli Bellek (LSTM)

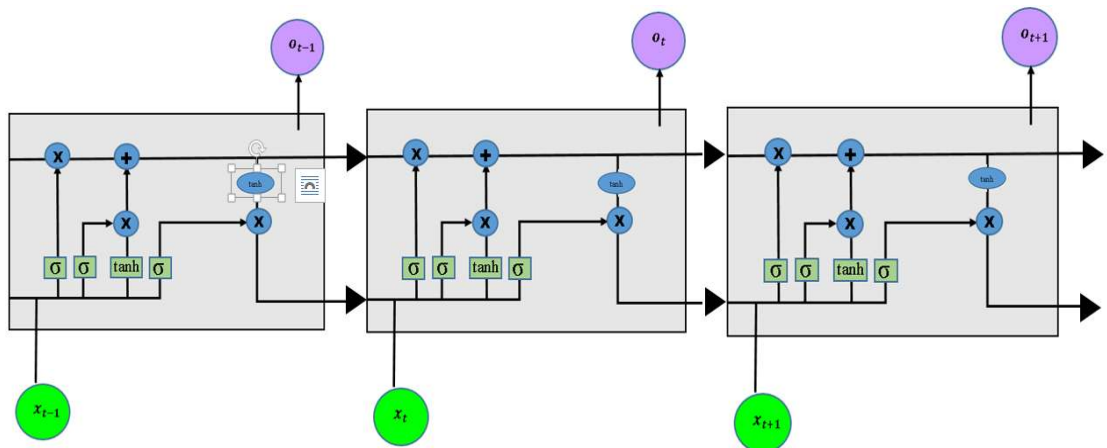
Uzun Kısa Süreli Bellek (Long-Short Term Memory, LSTM) RNN modelindeki dezavantajlı durumları ortadan kaldırmak için geliştirilen RNN modelinin bir çeşididir [37]. İnsanlardaki uzun süreli bellek ve kısa süreli bellek

modelinden esinlenilerek oluşturulan LSTM yönteminin teorik olarak mimarisi Şekil 4.2.3.1’de gösterilmiştir.

LSTM mimarisinin matematiksel gösterimi ise Şekil 4.2.3.2’deki gibidir. Şekil 4.2.3.2’de görüldüğü gibi ağların her biri 3 farklı girdi alır ve iki çıktı üretir. Girdilerin birincisi ağın kendi zamanındaki girdisi olan x girdisidir. İkinci girdi bir önceki ağın üretmiş olduğu çıktıdır. Üçüncü ve en önemli girdi ise bir önceki ağın belleğidir. Her bir ünitenin çıktıları ise ağın kendi çıktısı ve ağın üretmiş olduğu bellektir. Şekillerden de anlaşılacağı gibi LSTM yönteminde RNN’den farklı olarak karar verme sürecinde ağ bir önceki belleği de dikkate alır ve hafızasını sürekli günceller.



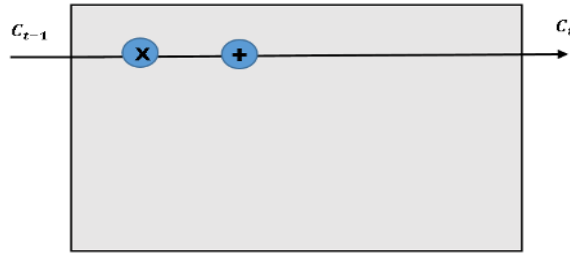
Şekil 4.2.3.1. LSTM mimarisi (teorik gösterim) [46]



Şekil 4.2.3.2. LSTM mimarisi (matematiksel gösterim) [47]

LSTM modeli bir girdinin kaydedilmesinin gerekli veya gereksiz olduğunu belirlemeye yarayan üç kapıya ve bir hücreye sahiptir [33] :

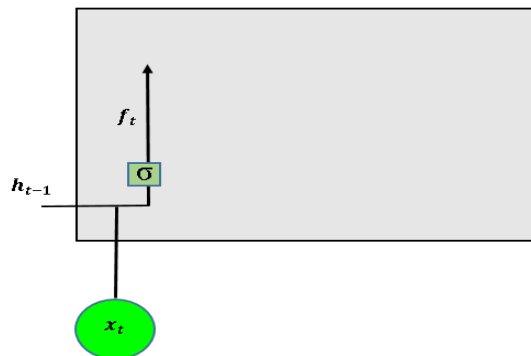
- **Cell State Hattı:** Şekil 4.2.3.3’de görülen yatay çizgi her ünitenin üst kısmında bulunan cell state hatlarıdır. Cell state’ler gerekli olan bilgileri hücreler boyunca taşımaya yarayan iletişim hatlarıdır [48].



Şekil 4.2.3.3. Cell state hattı [49]

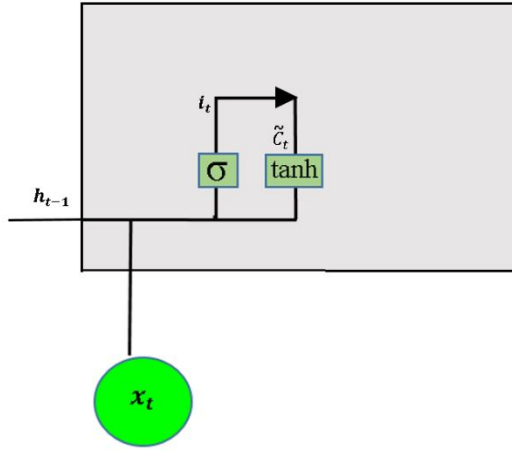
Her kapıda bulunan sigmoid fonksiyonu gelen her veri için 0 veya 1 sayısal değerlerinden birini üretir [34]. 0 değeri üretilen veri önemsiz veri olarak görülür ve silinip kaybolur. 1 değerini alan veri ise önemli veri olarak kabul edilip cell state hattı aracılığıyla hücrelerde ilerleyip en son kapıda tahmin üretmede kullanılır.

- **Unutma Kapısı:** Şekil 4.2.3.4’de gösterilen unutma kapısında bilgilerden hangilerinin tutulup hangilerinin silinmesi gerektiğine karar verilir. Kapıda bulunan sigmoid fonksiyonunun üretmiş olduğu çıktı 0 ise bilgi silinir, 1 ise cell state taşıma hattına gönderilir.



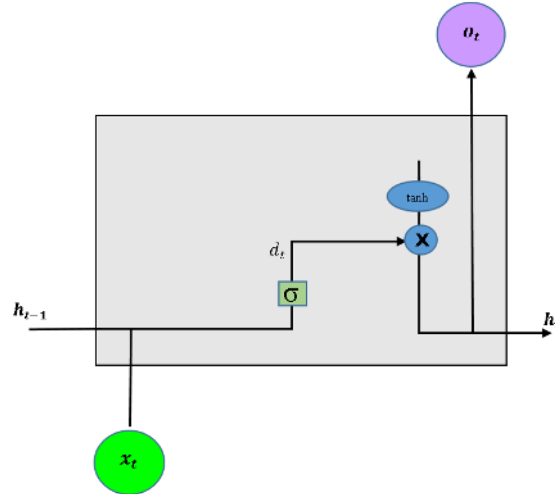
Şekil 4.2.3.4. Unutma kapısı [49]

- **Giriş Kapısı:** Şekil 4.2.3.5’de gösterildiği gibi önceki ve mevcut bilgiler sigmoid fonksiyonuna sokularak 0 ve 1 değerlerinden biri üretilir. Çıkan sonuca göre bilgi güncellemesi yapılır veya yapılmaz. -1 ve 1 arasında sonuçlar üreten tanh fonksiyonu ile sigmoid fonksiyonunun ürettiği değer çarpılarak çıkan sonuca göre hangi bilginin güncellenmesi gerektiğinin kararı verilir.



Şekil 4.2.3.5. Giriş kapısı [49]

- **Çıkış Kapısı:** Önceki ve mevcut bilgiler sigmoid fonksiyonuna ve cell state üzerinde taşınan bilgi ise tanh fonksiyonuna konulur. İki fonksiyondan elde edilen sonuçlar çarpılarak bir tahmin ve yeni bir kısa süreli bellek üretilir.

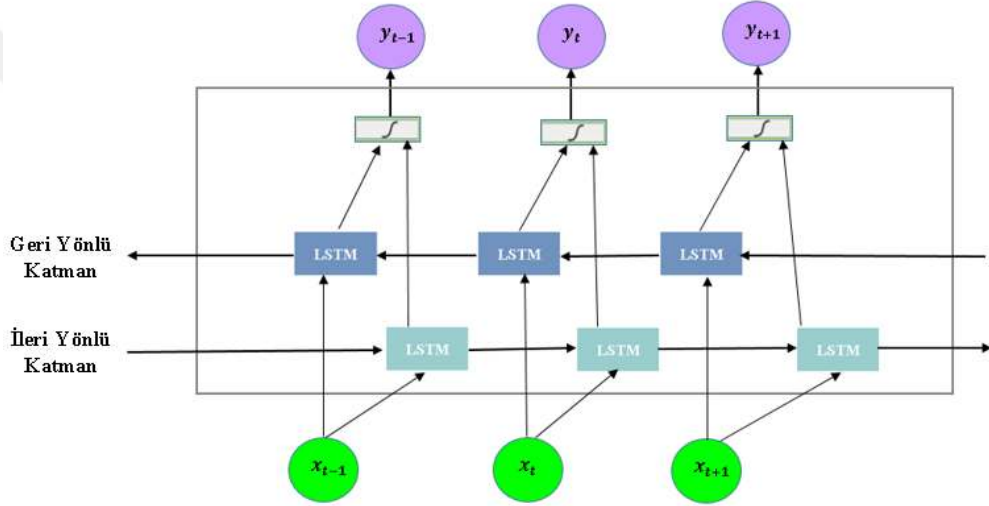


Şekil 4.2.3.6. Çıkış kapısı [49]

4.2.4. İki Yönlü Uzun Kısa Süreli Bellek (BLSTM)

İki yönlü LSTM (Bidirectional Long Short-Term Memory, BLSTM) LSTM performansını arttırmak için geliştirilen bir modeldir. Şekil 4.2.4’de çalışma mantığı gösterilen BLSTM modeli temel olarak LSTM’in sıralı bilgileri düz sıralı ve ters sıralı olarak aynı anda işleme alıp sonuçları birleştirmesi şeklinde çalışmaktadır [27].

NLP yönteminde cümle içerisindeki bir kelimenin tahmini için geçmiş kelime bilgileri kadar gelecek kelime bilgilerine de ihtiyaç vardır. BLSTM yöntemi hem geçmişten gelen kelimelerin bilgisine hem de gelecekte gelen kelimelerin bilgisine sahip olmayı mümkün kılar.



Şekil 4.2.4. BLSTM mimarisi [27]

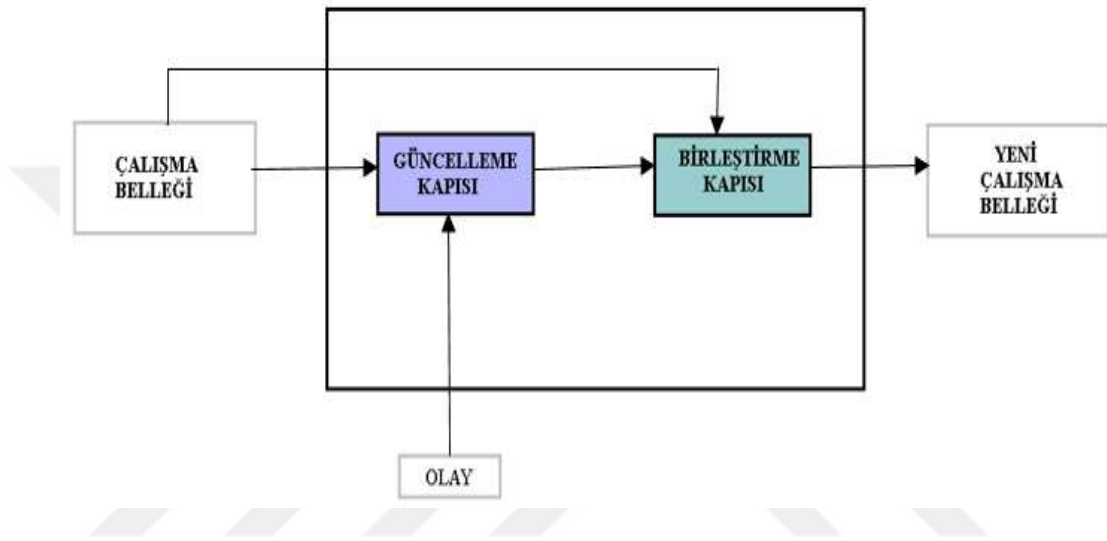
Şekil 4.2.4’de gösterildiği gibi t anındaki bilgi hem ileri yönlü LSTM katmanının hemde geri yönlü LSTM katmanının girdisidir. Bu iki katmandan üretilen çıktılar aktivasyon katmanına sokularak y_t çıktısı elde edilmiştir.

$$y_t = \vec{h}_t + \overleftarrow{h}_t \quad (4.2.4)$$

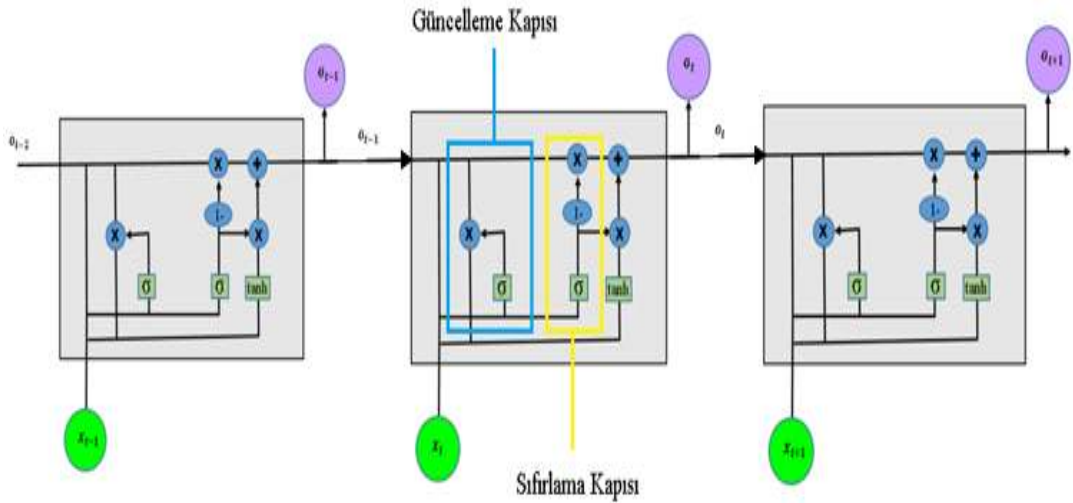
Denklem 4.2.4 y_t çıktısının nasıl bulunduğunu göstermektedir. Denklemdeki y_t sistemin çıktısını, $\vec{h}_t$ ileri yönlü LSTM katmanının çıktısını, $\overleftarrow{h}_t$ geri yönlü LSTM katmanının çıktısını ifade etmektedir.

4.2.5. Geçitli Tekrarlayan Birim (GRU)

Geçitli tekrarlayan birim (Gated Recurrent Unit, GRU), LSTM'deki işlem karmaşıklığını azaltmak için geliştirilen bir RNN alt modelidir. GRU mimarisi LSTM mimarisine oldukça benzer. Aralarındaki temel fark ise Şekil 4.2.5.1 ve Şekil 4.2.5.2'de görüldüğü gibi bu modelde LSTM'deki dört kapı yerine iki kapı bulunmasıdır. Algoritmanın daha az sayıda kapı ile çalışması işlem sayısını azaltarak GRU modelinin LSTM modelinden daha hızlı olmasını sağlar.



Şekil 4.2.5.1. GRU mimarisi (teorik gösterim) [46]



Şekil 4.2.5.2. GRU mimarisi (matematiksel gösterim) [46]

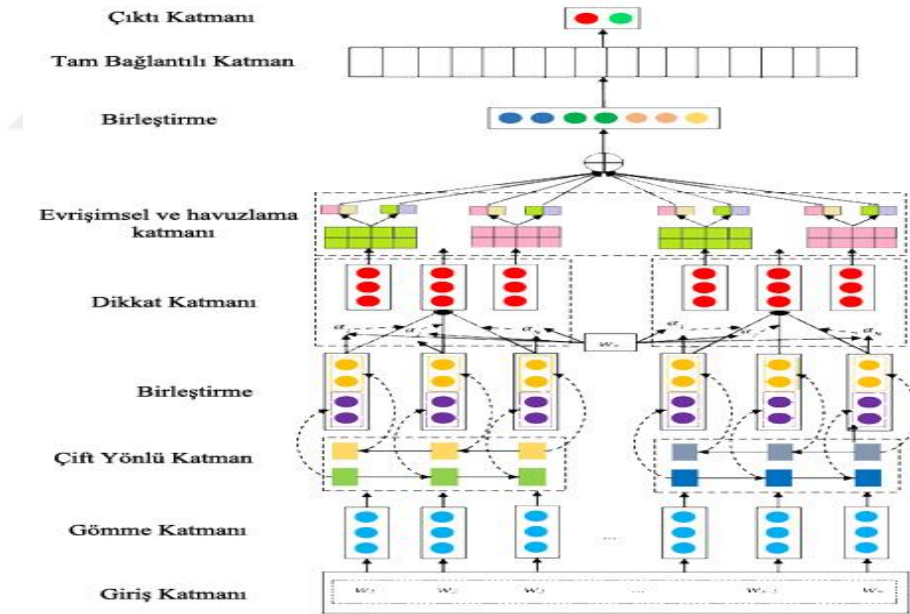
Kapılardan birincisi LSTM modelindeki unutma kapısı ve giriş kapısının birleşiminden oluşan güncelleme kapısıdır. Güncelleme kapısı gelen bilgilerden

hangilerinin silinip hangilerinin tutulacağına ve yeni gelen bilgilerden hangisinin eklenmesi gerektiğine karar verir. GRU mimarisinde ikinci kapı sıfırlama kapısıdır. Sıfırlama kapısı önceki bilgilerin ne kadarının kullanılması gerektiğine karar verir.

4.3. Hibrit Modeller

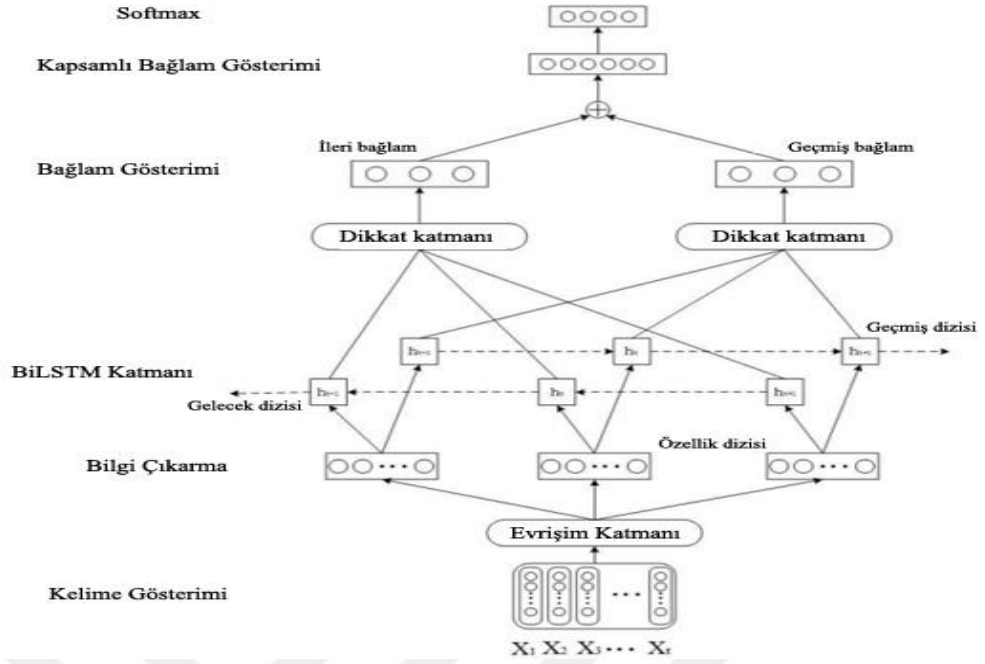
Bu başlık altında deneylerde kullanılan hibrit modeller ile ilgili özet bilgilere yer verilmiştir.

- **ABCDM mimarisi:** Şekil 4.3.1’de görüldüğü gibi model gömme katmanıya başlar ve ardından çift yönlü bir katmana geçiş yapar. Çift yönlü katmandaki metin belgelerinden genel özellikleri çıkarmak için çift yönlü LSTM’ler ve çift yönlü GRU’lar paralel olarak kullanılmıştır. Sonrasında dikkat mekanizması, ardından evrişim ve havuzlama katmanları kullanılır. Son olarak, gözden geçirme belgelerinin duyarlılığını belirlemek için dropout ve softmax ile tam bağlantılı katman kullanılmıştır [50].



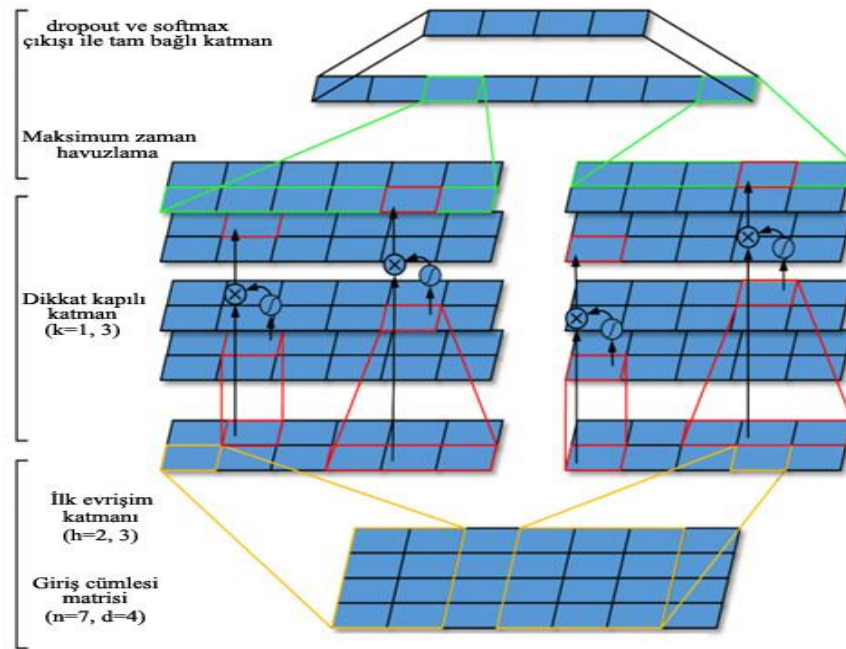
Şekil 4.3.1. ABCDM mimarisi [50]

- **AC-BiLSTM mimarisi:** Model, Şekil 4.3.2’de görüldüğü üzere çeşitli filtre boyutlarından oluşan tek boyutlu bir evrişimsel katmanla başlar ve bunu bir dikkat mekanizmasıyla donatılmış çift yönlü bir LSTM ağı izler. Son olarak, gözden geçirme belgelerinin duyarlılığını belirlemek için dropout ve softmax ile tam bağlantılı katman kullanılmıştır [51].



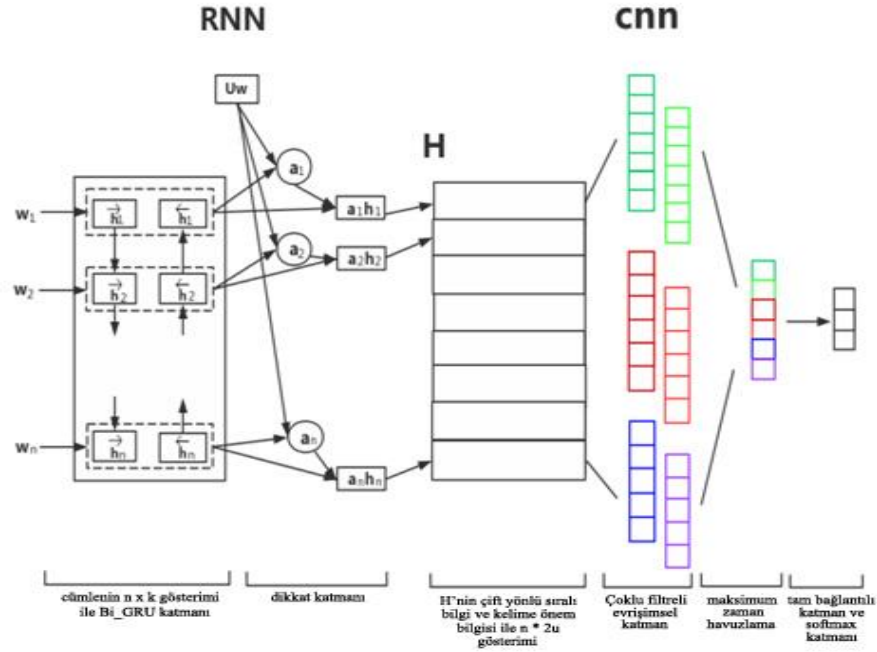
Şekil 4.3.2. AC-BiLSTM mimarisi [51]

- **AGCNN mimarisi:** AGCNN modeli Şekil 4.3.3’de görüldüğü gibi evrişimsel bir katmanla başlamış ve dikkat gerektiren bir katmana doğru ilerlemiştir. Maksimum havuzlama tercih edilmiş olup dropout ve softmax işlevine sahip tam bağlantılı bir katman kullanılmıştır [52].



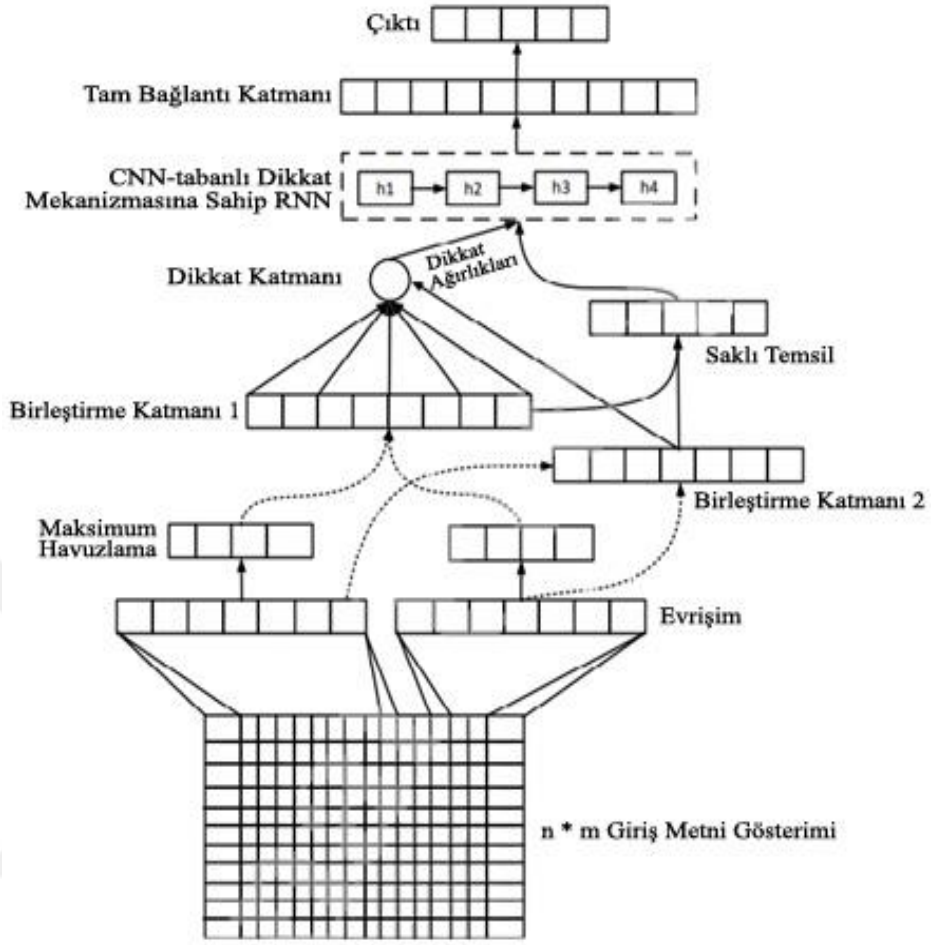
Şekil 4.3.3. AGCNN mimarisi [52]

- **ARC mimarisi:** Model Şekil 4.3.4’de verildiği gibi beş modülden oluşur: çift yönlü GRU, dikkat mekanizması, evrimsel katman, havuzlama katmanı ve tam bağlantılı katman [53].



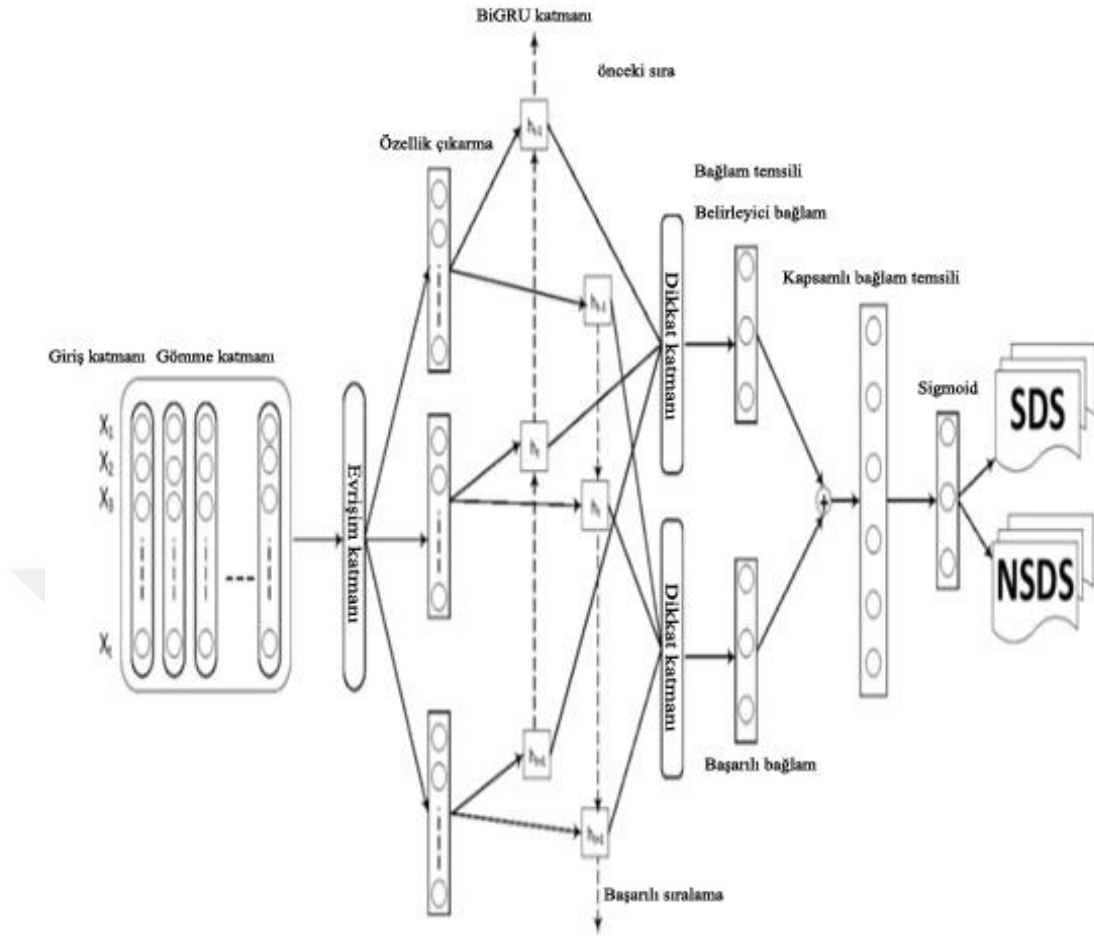
Şekil 4.3.4. ARC mimarisi [53]

- **ATTPooling mimarisi:** Şekil 4.3.5’te mimari şekli verilen model, CNN'e dayalı dikkat mekanizmasına sahip bir RNN'den oluşmaktadır. CNN, bölgenin ayırt edici özelliklerini yakalayarak başlar. Daha sonra dikkat mekanizması kullanılarak özellikler hesaplanmıştır. Son olarak, CNN tabanlı dikkat mekanizmasına sahip bir RNN ve tam bağlantılı bir katman kullanılmıştır [54].



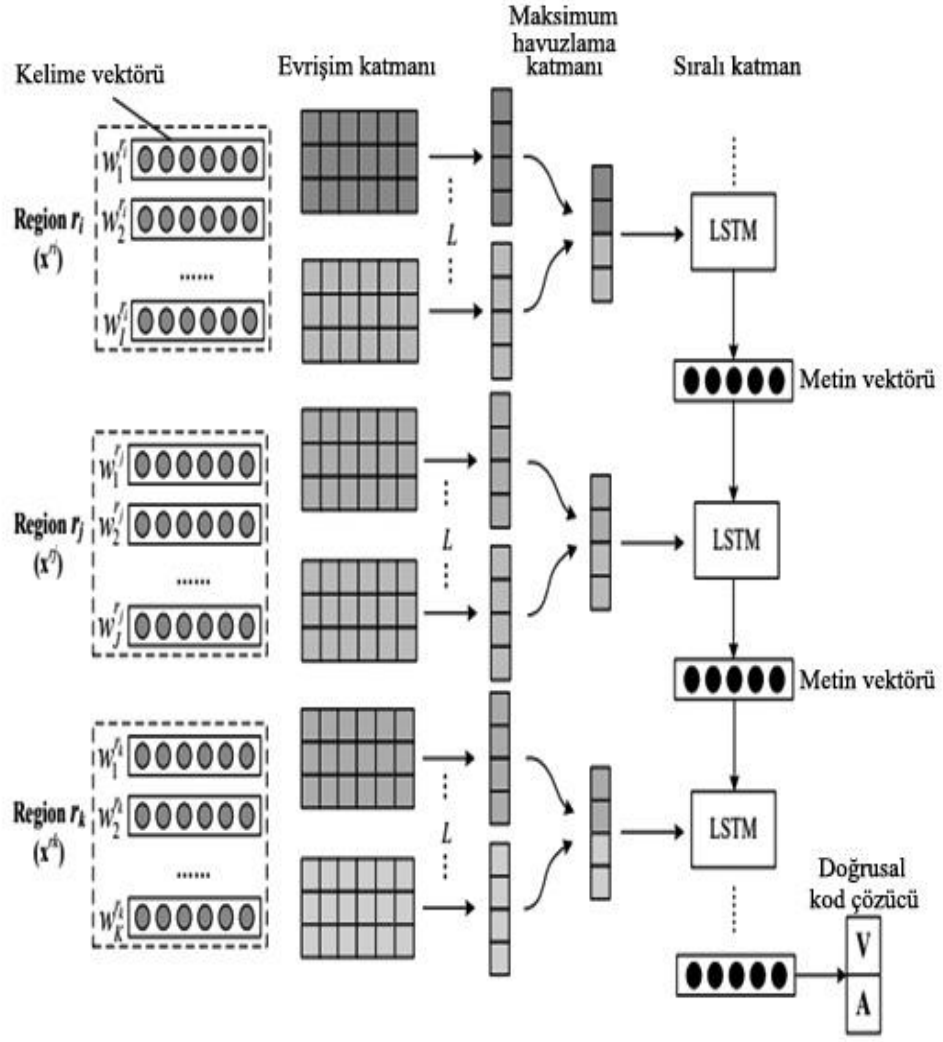
Şekil 4.3.5. ATTPooling mimarisi [54]

- **CAT-BiGRU mimarisi:** Model, çift yönlü GRU'ları kullanarak evrişim ve dikkati birleştirmiştir. Şekil 4.3.6'da görüldüğü gibi iki dikkat katmanı ile girdi, gömme, evrişim ve çift yönlü GRU'lardan oluşur [55].



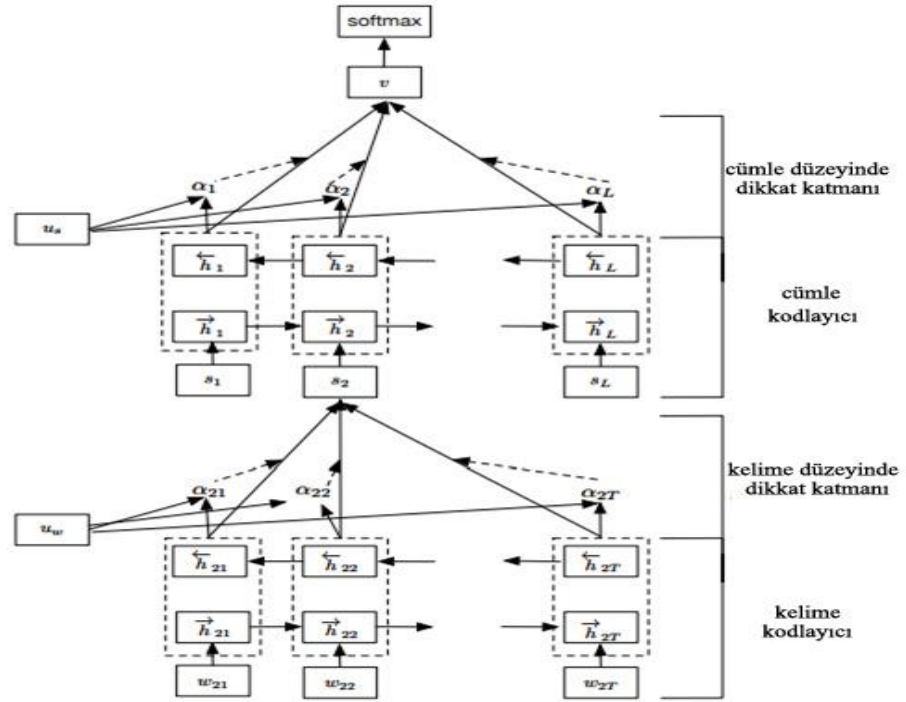
Şekil 4.3.6. CAT-BiGRU mimarisi [55]

- **CNN-GRU mimarisi:** Model; gömme katmanı (Word2Vec'te önceden eğitilmiş vektörleri kullanan), maksimum havuzlama katmanı ve GRU tekrarlayan birimi içerir [56].
- **CRNN mimarisi:** Şekil 4.3.7'de mimarisi verilen bu model, her cümleyi bir bölge olarak ele alır ve giriş kelime vektörlerini bölgesel CNN'e tabi tutar. Daha sonra, maksimum havuzlama kullanılarak yerel özellikler azaltılır. Uzun vadeli bağımlılıkları tanımlamak için ise bir LSTM katmanı kullanılır. Ayrıca gelecekteki değerlik ve uyarmayı belirlemek için ise doğrusal bir kod çözücü kullanılır [57].



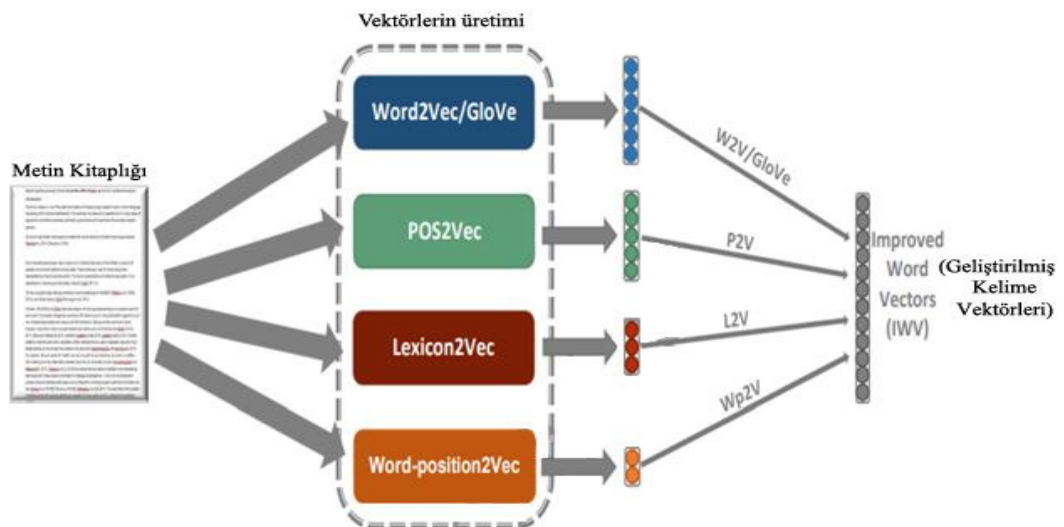
Şekil 4.3.7. CRNN mimarisi [57]

- **HAN mimarisi:** Model Şekil 4.3.8’de görüldüğü gibi dört modülden oluşmaktadır: sıra kodlayıcı, cümle kodlayıcı, kelime tabanlı dikkat mekanizması ve cümle tabanlı dikkat mekanizması. Bu şemadaki dizi kodlayıcı ve cümle kodlayıcı modülleri, çift yönlü GRU’ları kullanır [58].



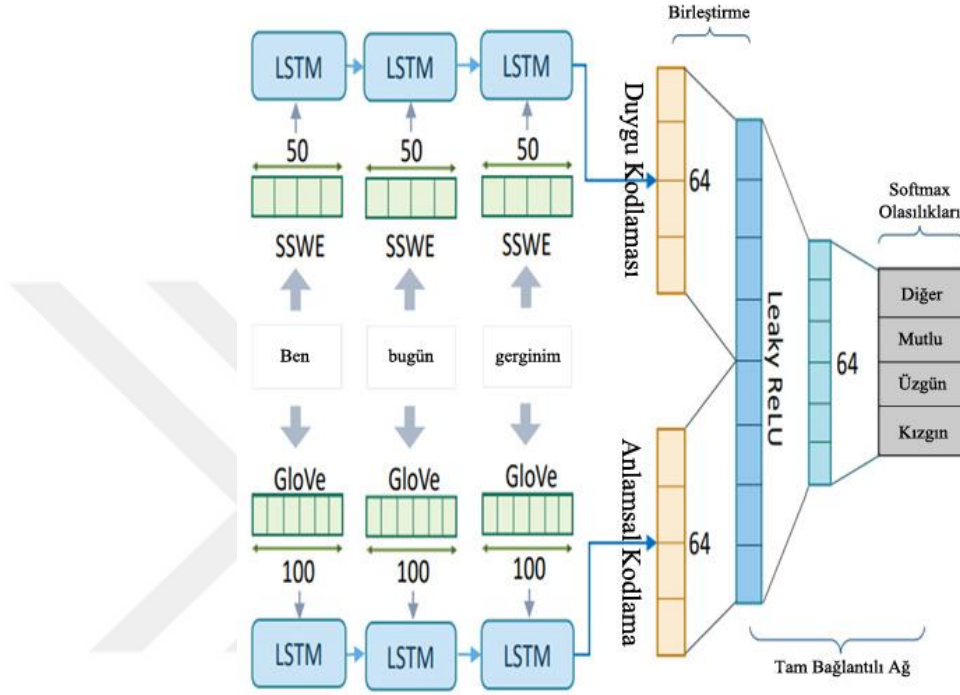
Şekil 4.3.8. HAN mimarisi [58]

- **Geliştirilmiş kelime vektörleri (IWV) mimarisi:** Bu model, önceden eğitilmiş duygu analizi kelime gömme şemalarının doğruluğunu geliştirecek şekilde tasarlanmıştır. Modele Şekil 4.3.9’da gösterildiği gibi maksimum havuzlamaya sahip üç evrimsel katman dahil edilmiştir ve tam bağlı katman ile devam etmektedir [59].



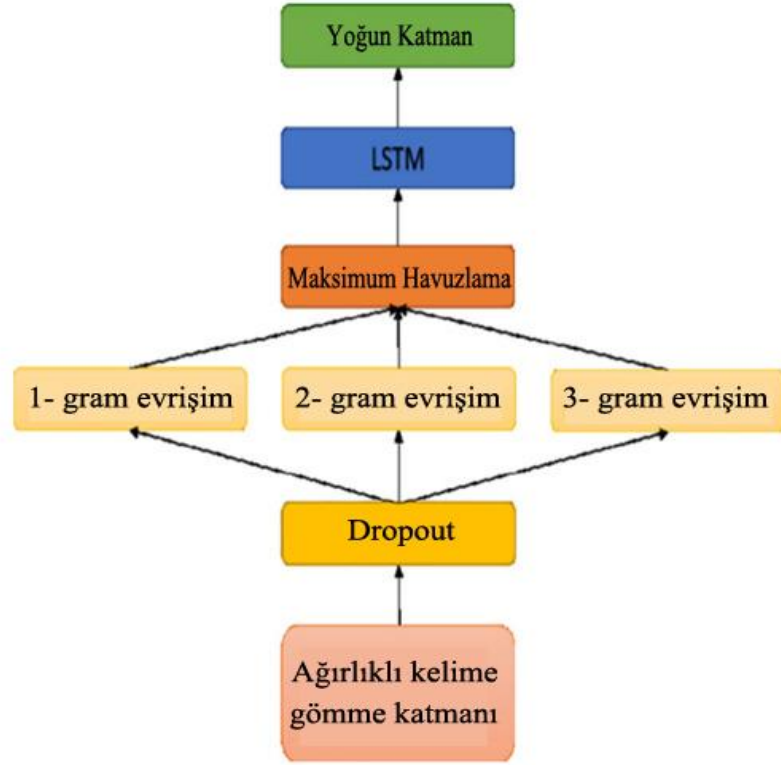
Şekil 4.3.9. IWV mimarisi [59]

- **SS-BED mimarisi:** Şekil 4.3.10’da verilen mimarisinde görüldüğü gibi iki paralel LSTM katmanı kullanan bu model, metin belgelerinden anlamsal özellikler çıkarmayı amaçlamaktadır. Modelde tamamen bağlı inceleme dokümanlarının duyarlılığını sınıflandırmak için tek gizli katmana sahip sinir ağı kullanılmıştır [60].



Şekil 4.3.10. SS-BED mimarisi [60]

- **Ağaç-BLSTM mimarisi:** Model, metin sınıflandırması için ağaç tabanlı çift yönlü bir LSTM mimarisi kullanmıştır [61].
- **Ağaç-LSTM mimarisi:** Model, geleneksel LSTM bloklarında bulunan kompozisyon katmanlarının yerini alan özyinelemeli ağaç tabanlı bir mimari aracılığıyla metin belgelerini anlamak amacıyla anlamsal bağlamdan yararlanmaktadır [62].
- **WCNNLSTM mimarisi:** Model Şekil 4.3.11’de görüldüğü gibi ağırlıklı gömme katmanı, dropout katmanı, n-gram evrişim katmanı (1 gram, 2 gram ve 3 gram evrişimlerle), maksimum havuzlama katmanı, LSTM katmanı ve tam bağlı katmandan oluşmaktadır [63].



Şekil 4.3.11. WCNNLSTM mimarisi [63]

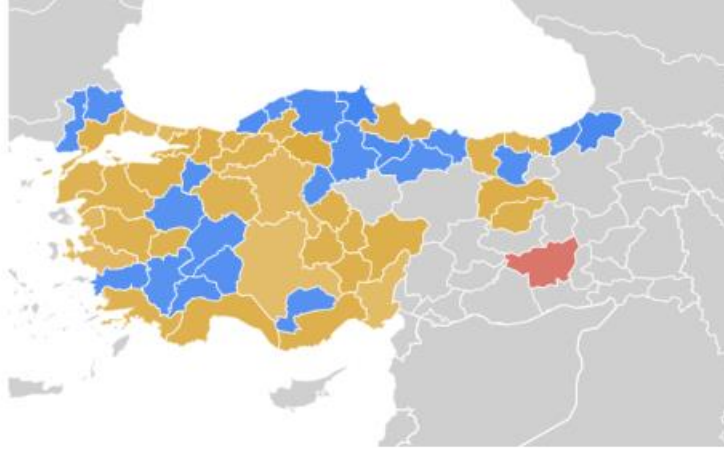
4.4. Veri Seti

4.4.1. Verilerin Toplanması

İnsanlar duygu ve düşüncelerini uzun metinlerle yazmak yerine Facebook, Instagram, Twitter gibi sosyal medya platformları üzerinden kısa metinlerle anlık olarak paylaşmayı daha fazla tercih etmekte ve bu tercih sonucunda sosyal medya kullanıcı sayısı her geçen gün hızla artmaktadır. Artan kullanıcı sayısı ile birlikte bu platformlar üzerinden paylaşılan herhangi bir konudaki duygu ve düşünceler de her geçen gün artmaktadır ve bu sayede sosyal medya platformları büyük veri kaynaklarına sahip hale gelmiş durumdadır.

Sosyal medya platformları arasından Facebook, Instagram ve Twitter platformlarının Google aramalarının yüzdelik karşılaştırmaları Şekil 4.4.1’de verildiği gibidir. Görüldüğü üzere Türkiye genelinde web üzerinde üç platform arasında en çok aranan sosyal medya haritada sarı renk ile temsil edilen Twitter olmuştur.

● Facebook ● İnstagram ● Twitter



Şekil 4.4.1. Son 12 ay içerisinde Facebook, Instagram, Twitter web arama karşılaştırmaları [8]

Tez çalışmasında kullanılan veri seti oluşturulurken Twitter platformundan yararlanılmıştır. Verilerin Twitter'dan toplanılmasının sebepleri Türkiye'de en fazla tercih edilen sosyal medya platformlarından biri olmasının yanı sıra aktif kullanıcı sayısının oldukça fazla olması ve bu nedenle büyük veri kaynaklarına sahip olmasıdır.

Tez çalışması için Twitter üzerinden gönderilen kripto paralar ile ilgili belirli kelimeler içeren veriler toplanılmıştır. Toplanan veriler 20.04.2021 ile 20.03.2022 tarihleri arasında Twitter kullanıcıları tarafından gönderilen Türkçe dilindeki tweetlerden oluşmaktadır. Veri çekilirken #avax, #avalanche, #bitci, #btc, #bitcoin, #chz, #chiliz, #eth, #ethereum, #solana, #xrp, #kripto, #nft, #defi etiketli 25000 tweet Spyder Anaconda ortamında Python dilinin Snsrape kütüphanesi kullanılarak çekilmiştir. Spyder ortamı Anaconda tarafından geliştirilen, Python ve R dilinde kod yazılmasını sağlayan basit arayüze sahip bir ortamdır. Snsrape ise açık kaynak kodlu bir kütüphane olup veri madenciliği için kullanılan bir sosyal ağ hizmetleri kazıyıcısıdır [64].

Twitter'dan toplanılan veriler Excel formatında bir dosyaya kaydedildikten sonra yinelenen veriler silinmiş, yanlış yazılan kelimeler düzeltilmiş, farklı dillerde yazılan kelimeler Türkçe'ye çevrilmiş ve metinleri etiketleme adımına geçilmiştir.

4.4.2. Verilerin Etiketlenmesi

Bitcoin, seçilen altcoinler ve NFT-DeFi ile ilgili 25000 veriden oluşan veri seti Excel formatında bir tabloda tutularak yaklaşık 10 ay gibi bir sürede tek tek el ile pozitif, negatif, nötr ve silinecek tweetler olmak üzere dört sınıftan oluşacak şekilde etiketlenmiştir.

Etiketleme yapılırken konu ile ilgili olumlu bir duygu içeren, piyasa ile ilgili olumlu bir beklentisi olan tweetler ve önemli kişi veya şirketlerin piyasaya girmesi durumu pozitif olarak; olumsuz duygu içeren, piyasa yönü ile ilgili olumsuz bir beklentisi olan tweet negatif olarak; herhangi bir duygu içermeyip kullanıcının tavsiye olarak yazdığı, anlık fiyat bildirimlerinin yazıldığı, güncel sepet paylaşımlarının yapıldığı veya piyasa ile ilgili tarafsız bilgi verildiği düşünülen tweetler nötr olarak; konumuz ile ilgili olmayan veya konu ile ilgili olup ağırlıklı olarak reklam içeren tweetler ise silinecekler olarak sınıflandırılmıştır. Silinecekler sınıfı daha sonra veri setinden silinmesi planlanarak oluşturulmuştur. Bu sınıfın oluşturulmasındaki amaç ise veri seti içerisinde olası bir karışıklığın önüne geçmektir. Sınıflandırma ile ilgili örnek bir durum Tablo 4.1’de verilmiştir.

Tablo 4.1. Manuel olarak etiketlenen tweet örnekleri

Metin ID	Metin (Tweet)	Etiket
1	Bitcoin dominance çok güzel yerde, Total2 direnç üstünde tutunmaya çalışıyor ayrıca çoğu altcoinin grafiği yukarıya gitmeye daha yatkın durumda.	1
2	Ethereum beklediğim yeri buldu ve 2000 USD üstündeki yerini aldı yeniden bu onunla ilgili tahminleri doğrular gibi Nisan ile birlikte demiştim oluyor	1
3	Bu haftanın yıldızı Chiliz olacak. Kırılmaları tamamladı. Alt yapıyı oluşturdu. Onumuzdeki hafta itibariyle sağlam olan projenin en az 2 usd seviyesine geleceğine inanıyorum. Sabreden kazanır.	1
4	Tüm kriptolarda bir düşüş bekliyorum. Ben önlemlerimi aldım, nakitte bekliyorum.	0
5	Elektrikli araç üreticisi Tesla'nın Bitcoin ile araç satın alımlarını askıya almasının ardından gelen yoğun satışlarla kripto para piyasalarının en yüksek hacimli birimi 43 bin dolara, tüm piyasaların toplam hacmi de 2 trilyon doların altına geriledi.	0
6	#DeFi protokolü @BadgerDAO'dan yaklaşık 10 milyon dolarlık kripto para çalındı. Haberin duyulmasının ardından \$BADGER fiyatı 28 dolar seviyesinden, 22 dolara kadar geriledi.	0

7	Adamin 70 bin takipçisi var. Ay başından beride 20x yaptım diyor. Hesabında 9000 TL var. Var olsun Traderler	2
8	NFT Library ekibi olarak, 3.000 takipçiye ulaştık! Tüm Topluluğa Teşekkürlerimizi İletiyoruz! Ve Bu Gelişmeyi NFT Çekilişiyile Kutlamak İstiyoruz! Takipte kalan kazanır!	2
9	Uluslararası Futbol Hukukçuları Kongresi bu yıl 4-5 Kasım arası İspanya/Sevilla'da düzenlenecek. Oturum Başkanlığını yapacağım “Yeni Teknolojilerin Futbol Devrimi” panelinde Espor, Blockchain, NFT, Cryptopara konularını ele alacağız.	3
10	Güncel sepetim: ALGO: 375 XRP: 185 adet, SXP: 135 adet, AVAX: 18 adet, DASH: 1 adet, BNB: 0.75 adet	3

Toplanan veri setindeki tweetler arasında çalışmada kullanılmaya uygun görülmeyen konu ile alakasız tweetler için silinecekler diye bir sınıf oluşturulmuş ve bahsedildiği üzere veri seti etiketleme işleminin tamamı bittikten sonra bu sınıf silinmiştir. Ayrıca veri setinde nötr olarak tespit edilen tweetler öğrenmeyi olumsuz etkilediği düşünüldüğü için daha sonra veri setinden çıkartılarak pozitif ve negatif olmak üzere iki sınıf ile çalışmaya devam edilmiştir. Son durumda veri setindeki toplam veri sayısı ve pozitif-negatif sınıflardaki tweet sayıları Tablo 4.2’de verilmiştir.

Tablo 4.2. Sınıflardaki tweet sayıları

Sınıf	Tweet Sayısı
Pozitif	6146
Negatif	3728
Toplam	9874

Twitter kullanıcıları kripto paralar ile ilgili genellikle olumlu tweetler paylaşmaktadır. Pozitif ve negatif sınıflardaki tweet sayılarını dengeleyebilmek adına piyasanın herhangi bir durumdan dolayı düşüşe geçtiği durumlarda tweet çekilerek olumsuz tweetler toplanılmaya çalışılmıştır. Fakat piyasanın düşüşte olduğu durumlarda bile olumsuz tweetler ile birlikte kullanıcıların bu durumu alım fırsatı olarak görerek olumlu beklentiler içerisinde olup pozitif içerikli tweetler paylaşımlarından dolayı dengeli sınıflar oluşturmak güçleşmiştir.

5. ARAŞTIRMA BULGULARI VE TARTIŞMA

5.1. Kullanılan Yöntem

Deneylerde kullanılan derin öğrenme mimarilerini uygulamak ve eğitmek için Tensorflow ve Keras kütüphanelerinden yararlanılmıştır. Tensorflow kütüphanesi derin öğrenme yönteminde kullanılan en yaygın kütüphanelerden birisidir. Açık kaynak kodlu bir kütüphane olan Tensorflow Google ekibi tarafından 2015 yılında geliştirilmiş ve ilk sürümü kullanıma sunulmuştur [65]. Projede kullanılan diğer bir kütüphane olan Keras Kütüphanesi de Tensorflow Kütüphanesi gibi derin öğrenme yönteminde sıklıkla tercih edilen; kolay ve hızlı bir şekilde model oluşturulmasını sağlayan bir kütüphanedir.

Modellerin tahmine dayalı verimliliğini arttırabilmek adına hiperparametre optimizasyonu kullanılmıştır. Hiperparametreler, kararı tasarlayıcı tarafından verilen; modelde kullanılan veri setine, ele alınan problemlere ve diğer etkenlere göre değişiklik gösteren parametrelerdir. Örneğin KNN algoritması kullanılarak yapılacak olan bir sınıflandırma için yeni bir değer geldiğinde bakılacak eleman sayısının tutulduğu k değeri, modeli tasarlayan kişi tarafından belirlenmektedir. Aynı şekilde derin öğrenme modeli tasarlanırken modelin nöron sayısı ve katman sayısı modelde kullanılan veri seti ve ele alınan problemler göz önünde bulundurularak geliştirici tarafından belirlenmektedir.

Hiperparametre optimizasyonu en uygun hiperparametre kombinasyonunu bularak işlem maliyetlerini azaltıp modelin performansını artırma işlemidir [66]. Optimizasyon işlemi için literatürde en çok tercih edilen yöntemler Bayes, Grid Arama, Rastgele Arama ve evrimsel algoritmalarlardır. Bu tez kapsamında hiperparametrelerin optimizasyonunu gerçekleştirebilmek için ise Bayes optimizasyonuna dayalı Gauss yönteminden yararlanılmıştır. Ayrıca son teknoloji mimarisi olan ABCDM ile doğru bir karşılaştırma sağlayabilmek adına deneysel sonuçlar için aynı padding boyutu ve parametreleri kullanılmıştır [50]. Tam bağlantılı katmanın ilk iki katmanı için RELU aktivasyon fonksiyonu ve tam bağlantılı katmanın son katmanı için softmax aktivasyon fonksiyonu kullanılmıştır. Evrişim filtrelerinin boyutları dört ve beş olarak ayarlanmıştır. CNN'nin gizli katmanının boyutu 100'dür. Bırakma (drop-out) katsayısı 0.30'dur ve mini-batch için örnek boyut 50'dir. Metin belgelerini temsil etmek için 300 boyutlu Word2Vec şeması tercih edilmiştir.

5.2. Deney Sonuçları

Bu kısımda dört derin öğrenme algoritması kullanılarak ve on dört hibrit mimari kullanılarak elde edilen sonuçlar ayrı başlıklar altında ele alınmaktadır.

5.2.1. Derin Öğrenme Algoritmalarıyla Elde Edilen Sonuçlar

Deneylerde Word2Vec, FastText ve GloVe olmak üzere üç farklı kelime gömme algoritması ile CNN, RNN, LSTM ve GRU olmak üzere dört derin öğrenme algoritması için farklı kombinasyonlar oluşturulmuş; doğruluk (accuracy), duyarlılık (precision), geri çağırma (recall) ve F-ölçütü metrikleri için sonuçlar alınmıştır. Sonuçlar alınırken vektör boyutu 200 ile 300 ve projeksiyon katmanının boyutu 100, 200, 300 olacak şekilde her kombinasyon için farklı değerler girilerek metrikler için alınan sonuçlar karşılaştırılmış ve en iyi sonuç elde edilmeye çalışılmıştır. Tablo 5.2.1.1, Tablo 5.2.1.2, Tablo 5.2.1.3 ve Tablo 5.2.1.4 sırasıyla doğruluk, kesinlik, duyarlılık ve F-ölçütü metrikleri için farklı parametreler ile elde edilen karşılaştırmalı sonuçları göstermektedir.

Tablo 5.2.1.1. Derin öğrenme algoritmaları ve önerilen şema ile elde edilen sınıflandırma doğruluk değerleri

Kelime Gömme	Vektör Boyutu	Projeksiyon Katmanı Boyutu	CNN	RNN	LSTM	GRU
Word2Vec (Skip-Gram)	200	100	84,9062	86,1300	86,5193	87,2704
Word2Vec (Skip-Gram)	200	200	85,3091	86,1505	86,9681	87,2791
Word2Vec (Skip-Gram)	300	100	85,5395	86,8367	87,0924	87,7042
Word2Vec (Skip-Gram)	300	300	85,8715	87,3834	88,5862	87,9131
Word2Vec (CBOW)	200	100	84,1355	84,6919	86,3301	86,5933
Word2Vec (CBOW)	200	200	84,3848	84,8301	86,3321	86,7954
Word2Vec (CBOW)	300	100	84,4691	85,2159	86,4670	87,2066
Word2Vec (CBOW)	300	300	84,6564	85,8096	86,4886	87,2103
FastText (Skip-Gram)	200	100	83,0755	83,9260	85,9995	86,0562
FastText (Skip-Gram)	200	200	83,4248	84,1585	86,1311	86,2161
FastText (Skip-Gram)	300	100	83,9981	84,1790	86,2756	86,2965

FastText (Skip-Gram)	300	300	84,1152	84,4471	86,2942	86,4113
FastText (CBOW)	200	100	82,7116	83,5779	84,1253	85,0902
FastText (CBOW)	200	200	82,7349	83,6589	85,1541	85,3837
FastText (CBOW)	300	100	82,8614	83,8124	85,2146	85,4376
FastText (CBOW)	300	300	82,9599	83,8509	85,6648	85,5584
GloVe	200	100	81,4875	82,6678	82,8163	83,2925
GloVe	200	200	81,5490	82,7000	83,0422	84,1660
GloVe	300	100	82,2485	82,9684	83,5552	84,5301
GloVe	300	300	82,4531	83,1941	83,9329	84,7005

Tablo 5.2.1.2. Derin öğrenme algoritmaları ve önerilen şema ile elde edilen duyarlılık değerleri

Kelime Gömme	Vektör Boyutu	Projeksiyon Katmanı Boyutu	CNN	RNN	LSTM	GRU
Word2Vec (Skip-Gram)	200	100	0,8576	0,8700	0,8739	0,8815
Word2Vec (Skip-Gram)	200	200	0,8617	0,8702	0,8785	0,8816
Word2Vec (Skip-Gram)	300	100	0,8640	0,8771	0,8797	0,8859
Word2Vec (Skip-Gram)	300	300	0,8674	0,8827	0,8948	0,8880
Word2Vec (CBOW)	200	100	0,8499	0,8555	0,8720	0,8747
Word2Vec (CBOW)	200	200	0,8524	0,8569	0,8720	0,8767
Word2Vec (CBOW)	300	100	0,8532	0,8608	0,8734	0,8809
Word2Vec (CBOW)	300	300	0,8551	0,8668	0,8736	0,8809
FastText (Skip-Gram)	200	100	0,8391	0,8477	0,8687	0,8693
FastText (Skip-Gram)	200	200	0,8427	0,8501	0,8700	0,8709
FastText (Skip-Gram)	300	100	0,8485	0,8503	0,8715	0,8717
FastText (Skip-Gram)	300	300	0,8496	0,8530	0,8717	0,8728
FastText (CBOW)	200	100	0,8355	0,8442	0,8498	0,8595
FastText (CBOW)	200	200	0,8357	0,8450	0,8601	0,8625

FastText (CBOW)	300	100	0,8370	0,8466	0,8608	0,8630
FastText (CBOW)	300	300	0,8380	0,8470	0,8653	0,8642
GloVe	200	100	0,8231	0,8350	0,8365	0,8413
GloVe	200	200	0,8237	0,8354	0,8388	0,8502
GloVe	300	100	0,8308	0,8381	0,8440	0,8538
GloVe	300	300	0,8329	0,8403	0,8478	0,8556

Tablo 5.2.1.3. Derin öğrenme algoritmaları ve önerilen şema ile elde edilen geri çağırma değerleri

Kelime Gömme	Vektör Boyutu	Projeksiyon Katmanı Boyutu	CNN	RNN	LSTM	GRU
Word2Vec (Skip-Gram)	200	100	0,8664	0,8789	0,8828	0,8905
Word2Vec (Skip-Gram)	200	200	0,8705	0,8791	0,8874	0,8906
Word2Vec (Skip-Gram)	300	100	0,8729	0,8861	0,8887	0,8949
Word2Vec (Skip-Gram)	300	300	0,8762	0,8917	0,9039	0,8971
Word2Vec (CBOW)	200	100	0,8585	0,8642	0,8809	0,8836
Word2Vec (CBOW)	200	200	0,8611	0,8656	0,8809	0,8857
Word2Vec (CBOW)	300	100	0,8619	0,8695	0,8823	0,8899
Word2Vec (CBOW)	300	300	0,8638	0,8756	0,8825	0,8899
FastText (Skip-Gram)	200	100	0,8477	0,8564	0,8775	0,8781
FastText (Skip-Gram)	200	200	0,8513	0,8588	0,8789	0,8798
FastText (Skip-Gram)	300	100	0,8571	0,8590	0,8804	0,8806
FastText (Skip-Gram)	300	300	0,8583	0,8617	0,8806	0,8817
FastText (CBOW)	200	100	0,8440	0,8528	0,8584	0,8683
FastText (CBOW)	200	200	0,8442	0,8537	0,8689	0,8713
FastText (CBOW)	300	100	0,8455	0,8552	0,8695	0,8718
FastText (CBOW)	300	300	0,8465	0,8556	0,8741	0,8730
GloVe	200	100	0,8315	0,8435	0,8451	0,8499
GloVe	200	200	0,8321	0,8439	0,8474	0,8588

GloVe	300	100	0,8393	0,8466	0,8526	0,8626
GloVe	300	300	0,8414	0,8489	0,8565	0,8643

Tablo 5.2.1.4. Derin öğrenme algoritmaları ve önerilen şema ile elde edilen F-ölçütü değerleri

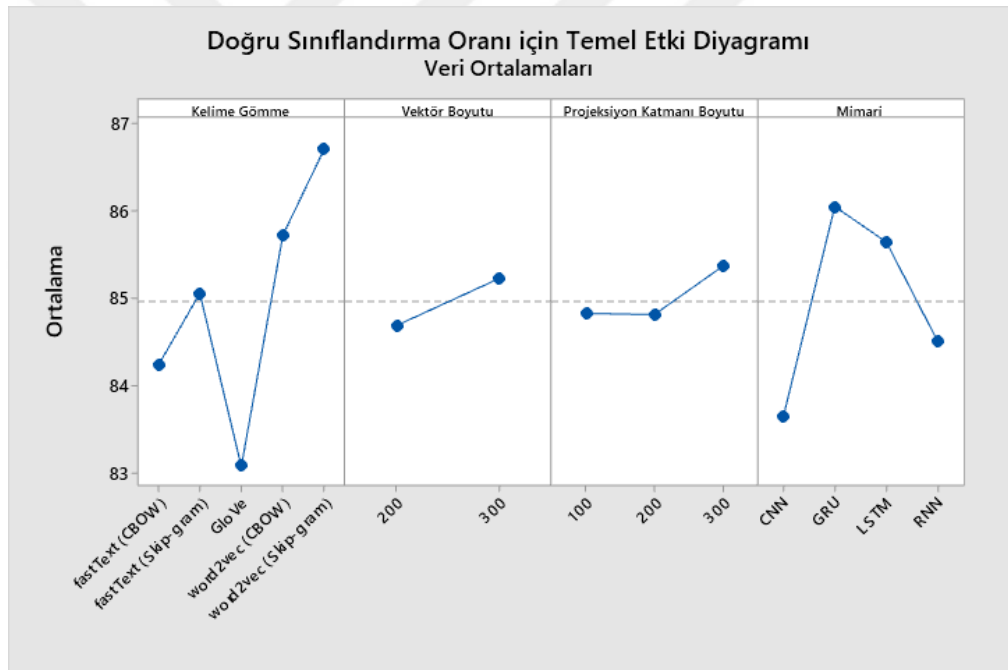
Kelime Gömme	Vektör Boyutu	Projeksiyon Katmanı Boyutu	CNN	RNN	LSTM	GRU
Word2Vec (Skip-Gram)	200	100	0,8620	0,8744	0,8784	0,8860
Word2Vec (Skip-Gram)	200	200	0,8661	0,8746	0,8829	0,8861
Word2Vec (Skip-Gram)	300	100	0,8684	0,8816	0,8842	0,8904
Word2Vec (Skip-Gram)	300	300	0,8718	0,8871	0,8994	0,8925
Word2Vec (CBOW)	200	100	0,8542	0,8598	0,8764	0,8791
Word2Vec (CBOW)	200	200	0,8567	0,8612	0,8765	0,8812
Word2Vec (CBOW)	300	100	0,8576	0,8651	0,8778	0,8853
Word2Vec (CBOW)	300	300	0,8595	0,8712	0,8781	0,8854
FastText (Skip-Gram)	200	100	0,8434	0,8520	0,8731	0,8737
FastText (Skip-Gram)	200	200	0,8470	0,8544	0,8744	0,8753
FastText (Skip-Gram)	300	100	0,8528	0,8546	0,8759	0,8761
FastText (Skip-Gram)	300	300	0,8540	0,8573	0,8761	0,8773
FastText (CBOW)	200	100	0,8397	0,8485	0,8541	0,8639
FastText (CBOW)	200	200	0,8399	0,8493	0,8645	0,8668
FastText (CBOW)	300	100	0,8412	0,8509	0,8651	0,8674
FastText (CBOW)	300	300	0,8422	0,8513	0,8697	0,8686
GloVe	200	100	0,8273	0,8393	0,8408	0,8456
GloVe	200	200	0,8279	0,8396	0,8431	0,8545
GloVe	300	100	0,8350	0,8423	0,8483	0,8582
GloVe	300	300	0,8371	0,8446	0,8521	0,8599

Karşılaştıran dört derin öğrenme algoritmasının sonuçlarına bakıldığında en düşük tahmin performansını CNN algoritması vermiştir. Algoritmalar arasında en

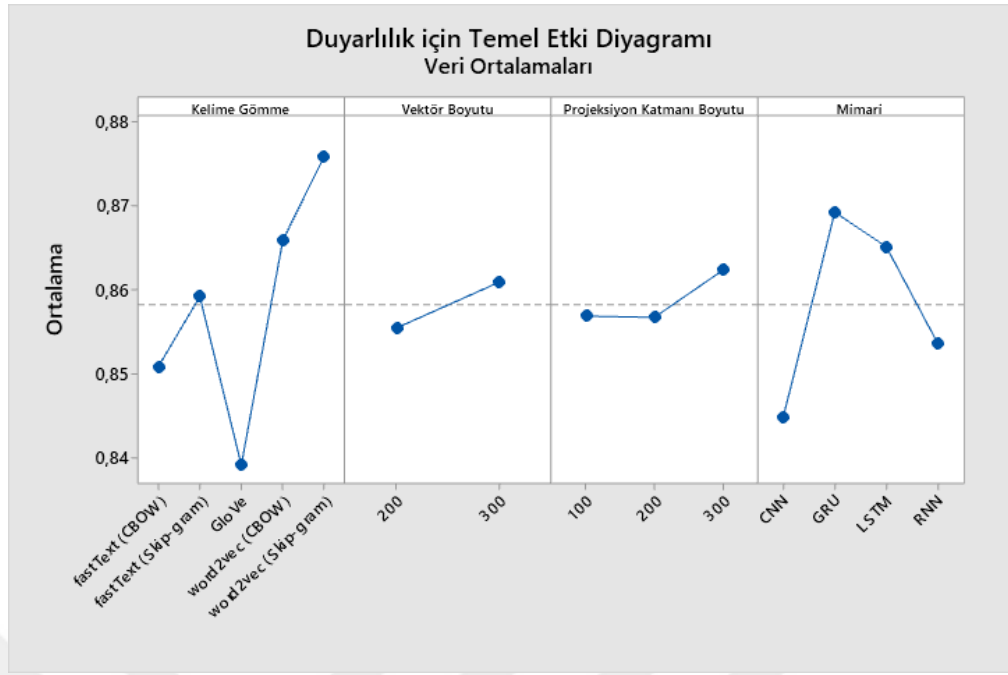
yüksek tahmin performansı ise GRU algoritması ile elde edilmiştir. Kelime gömme algoritmalarının performansları karşılaştırıldığında ise en düşük tahmin performansını GloVe algoritması ve en yüksek tahmin performansını Word2Vec(Skip-Gram) algoritması vermiştir.

Deneyisel analizde farklı parametreler girilerek elde edilen değerler karşılaştırıldığında iki vektör boyutu arasında 300 olarak girilen vektör boyutu daha yüksek tahmin performansı sağlamıştır. Projeksiyon katmanı boyutu için girilen üç değer arasında ise en iyi tahmin performansı 300 değeri ile elde edilmiştir.

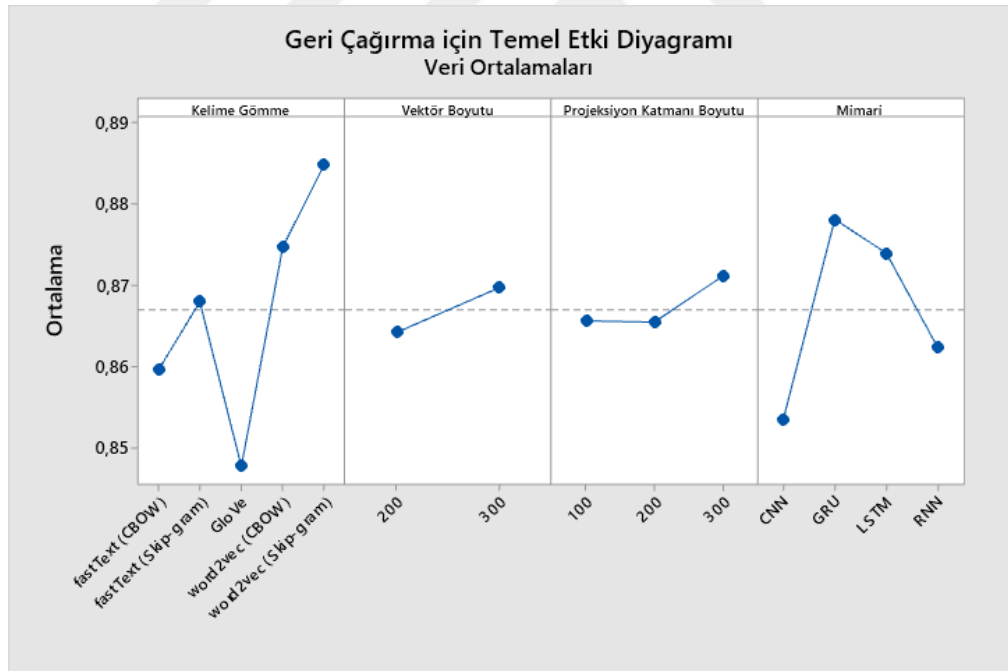
Analizde farklı kombinasyonlar kullanılarak dört metrik için elde edilen sonuçların ortalama değerleri Şekil 5.2.1.1, Şekil 5.2.1.2, Şekil 5.2.1.3 ve Şekil 5.2.1.4’de istatistiksel olarak gösterilmiştir.



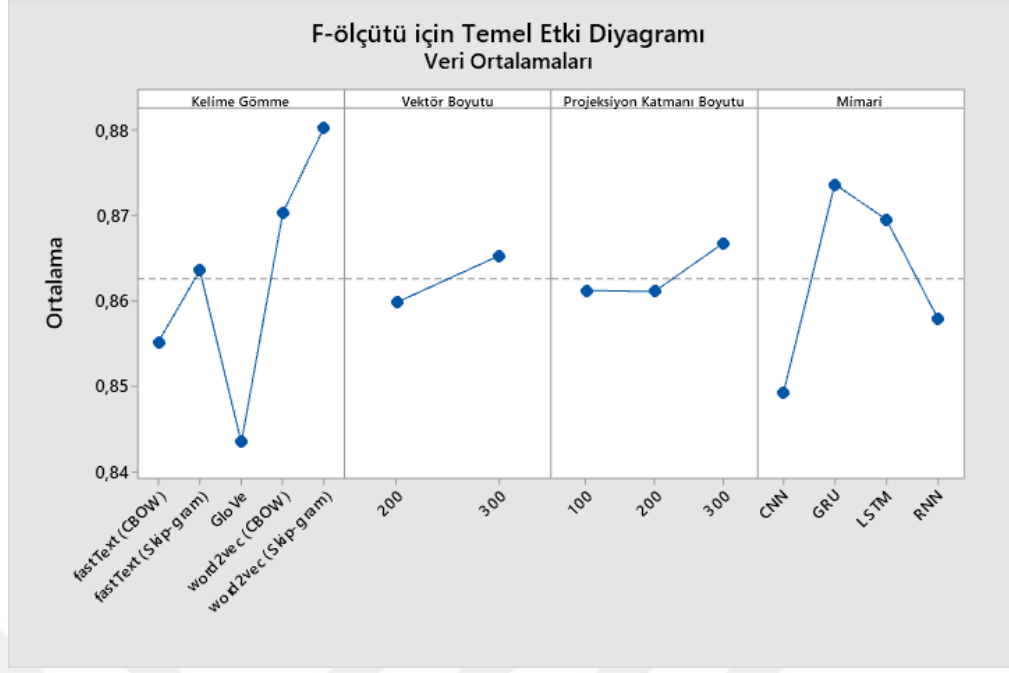
Şekil 5.2.1.1. Doğru sınıflandırma oranı için temel etki grafiği



Şekil 5.2.1.2. Duyarlılık için temel etki grafiği



Şekil 5.2.1.3. Geri çağırma için temel etki grafiği



Şekil 5.2.1.4. F-ölçütü için temel etki grafiği

5.2.2. Hibrit Modeller Kullanılarak Elde Edilen Sonuçlar

Veri kümesinin tahmine dayalı performansı; sınıflandırma doğruluğu, kesinlik, geri çağırma ve F-ölçütü metrikleri açısından on dört hibrit derin sinir ağı mimarisinde değerlendirilmiştir. Derin sinir ağı mimarileri kullanılarak elde edilen sınıflandırma doğruluk değerleri Tablo 5.2.2’de düşük performanstan yüksek performansa doğru olacak şekilde sıralı bir şekilde verilmiştir.

ABCDM mimarisi, %90.1395 sınıflandırma doğruluğu ile veri seti üzerinde en iyi tahmin performansını elde etmiştir. Alınan deneysel sonuçlar ABCDM mimarisinin, en gelişmiş mimariler kullanılarak alınan sonuçlar ile karşılaştırıldığında daha iyi performans gösterdiğini ortaya koymaktadır. Deneysel sonuçlar, ikinci en yüksek tahmin performansının %89.3312 sınıflandırma doğruluğu ile CAT-BiGRU mimarisi tarafından elde edildiğini ve bunu %88.5258 sınıflandırma doğruluğu ile CRNN mimarisi tarafından takip edildiğini göstermektedir.

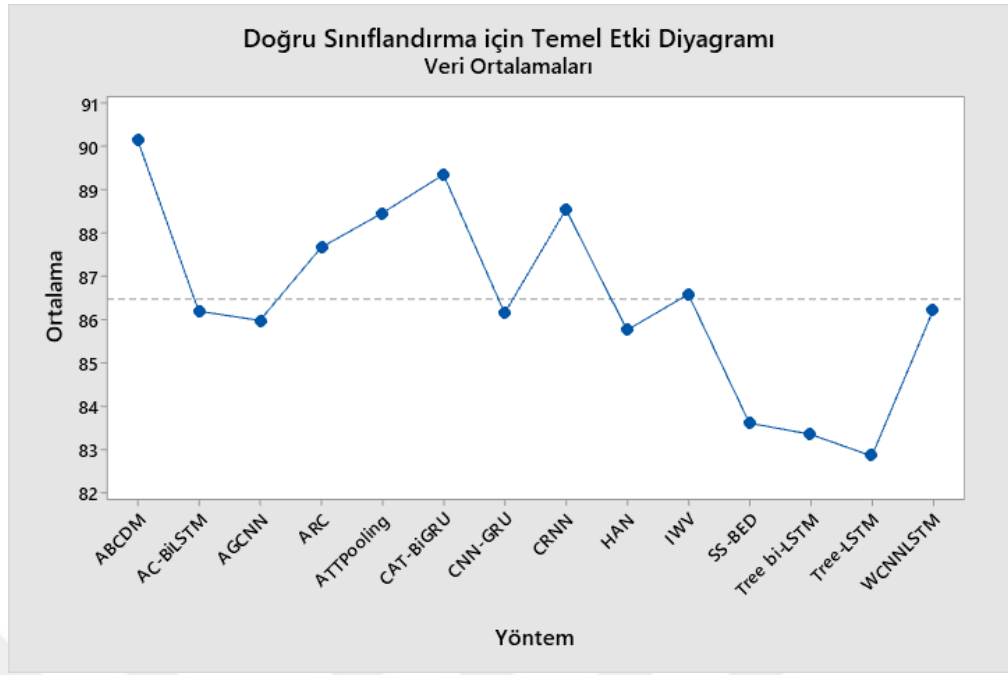
Tablo 5.2.2’de listelenen sonuçlardan görülebileceği gibi, ağaç-LSTM, ağaç-BLSTM ve SS-BED mimarileri veri seti üzerinde daha düşük tahmin performansları vermektedir. Ayrıca Tablo 5.2.2’de listelenen sınıflandırma doğruluk değerlerine göre, CNN’nin RNN ile birlikte kullanımı geleneksel derin sinir ağı mimarileri ile karşılaştırıldığında genellikle daha iyi performans göstermiştir.

Tablo 5.2.2. Hibrit derin sinir ağı mimarileri ile tahmine dayalı performanslar

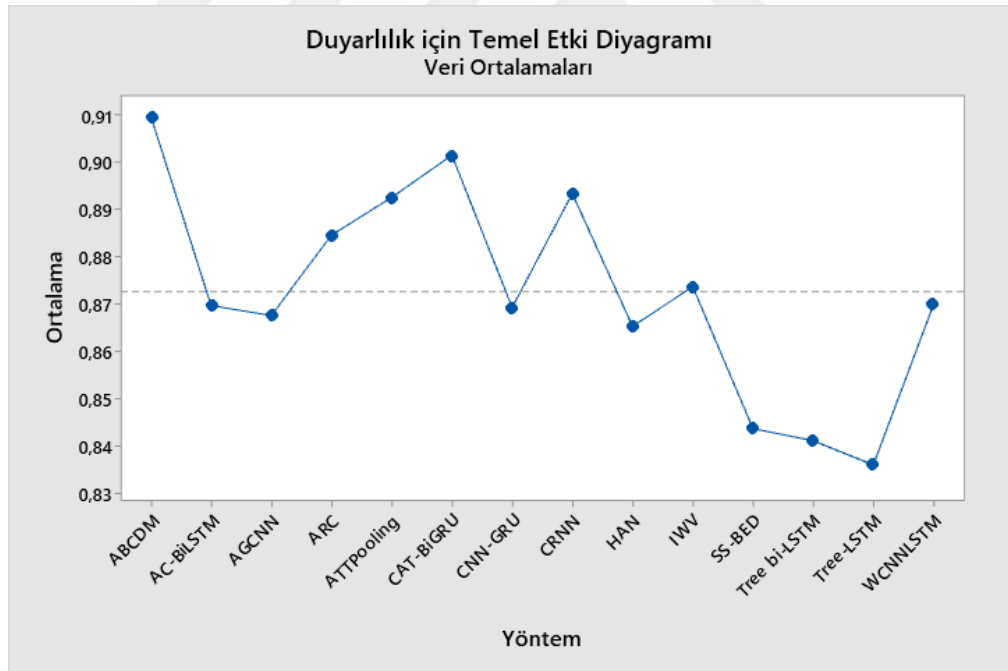
Mimari	Doğruluk	Kesinlik	Geri Çağırma	F-ölçütü
Ağaç-LSTM	82,8389	0,8359	0,8376	0,8368
Ağaç-BLSTM	83,3402	0,8410	0,8427	0,8418
SS-BED	83,5993	0,8436	0,8453	0,8444
HAN	85,7452	0,8652	0,8670	0,8661
AGCNN	85,9673	0,8675	0,8692	0,8684
CNN-GRU	86,1370	0,8692	0,8709	0,8701
AC-BiLSTM	86,1814	0,8696	0,8714	0,8705
WCNNLSTM	86,1943	0,8698	0,8715	0,8706
IWV	86,5722	0,8736	0,8754	0,8745
ARC	87,6503	0,8845	0,8863	0,8854
ATTPooling	88,4448	0,8925	0,8943	0,8934
CRNN	88,5258	0,8933	0,8951	0,8942
CAT-BiGRU	89,3312	0,9014	0,9032	0,9023
ABCDM	90,1395	0,9096	0,9114	0,9105

Tablo 5.2.2 ayrıca derin sinir ağı mimarilerinden elde edilen kesinlik, geri çağırma ve F-ölçütü metriklerinin değerlerini de sunmaktadır. Tablo 5.2.2’de gösterildiği gibi, sınıflandırma doğruluk değerlerine ilişkin deneysel bulgular kesinlik, geri çağırma ve F-ölçüleri için de geçerlidir. ABCDM, deneylerde kullanılan veri kümesi için sınıflandırma doğruluk değeri metriğinde olduğu gibi kesinlik, geri çağırma ve F-ölçütü metrikleri açısından da en gelişmiş mimarilerden daha iyi performans göstermiştir.

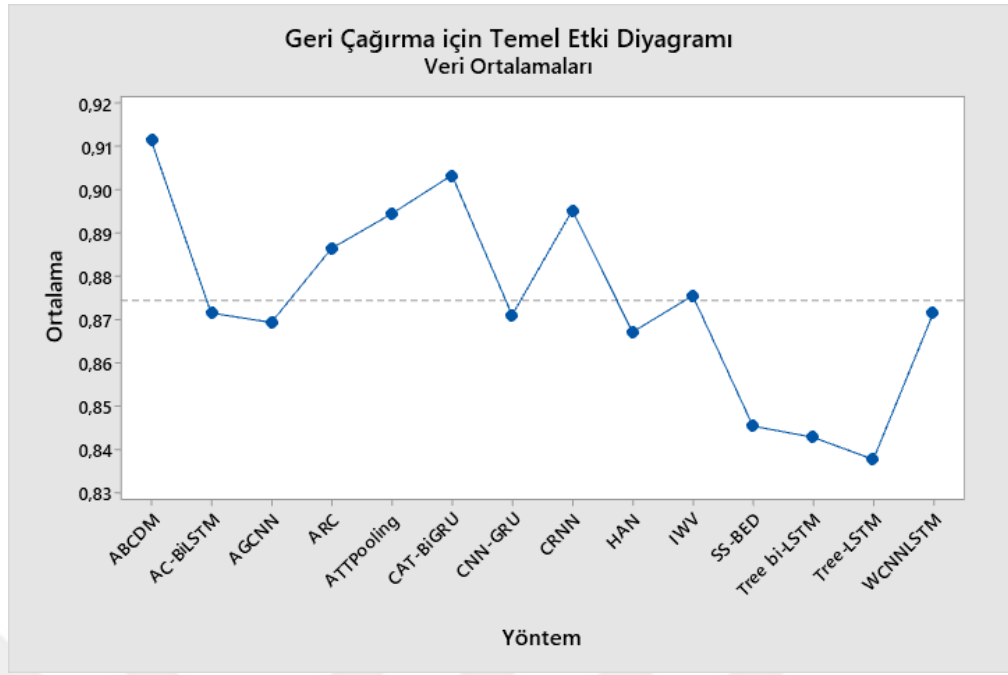
Hibrit mimariler kullanılarak dört metrik için elde edilen sonuçların ortalama değerleri Şekil 5.2.2.1, Şekil 5.2.2.2, Şekil 5.2.2.3 ve Şekil 5.2.2.4’de istatistiksel olarak gösterilmiştir.



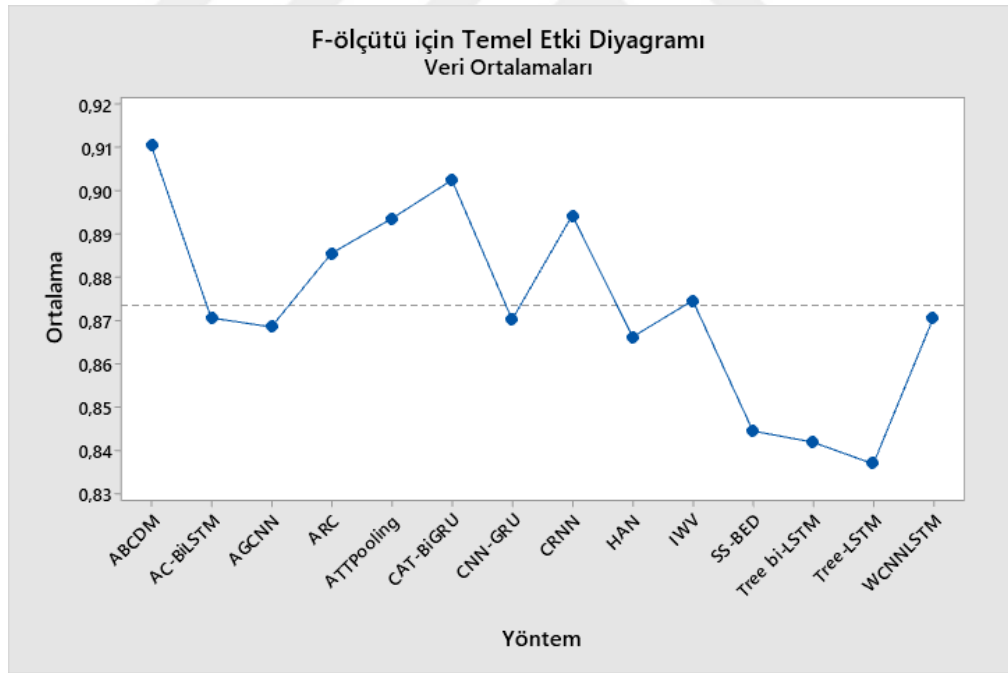
Şekil 5.2.2.1. Hibrit mimarilerde doğru sınıflandırma oranı için temel etki grafiği



Şekil 5.2.2.2. Hibrit mimarilerde duyarlılık için temel etki grafiği



Şekil 5.2.2.3. Hibrit mimarilerde geri çağırma için temel etki grafiği



Şekil 5.2.2.4. Hibrit mimarilerde F-ölçütü için temel etki grafiği

6. SONUÇ VE ÖNERİLER

Satoshi Nakamoto takma adlı bir kişi veya grup tarafından 2008 yılında yayınlanan makale ile isminden söz ettiren ve 2009 yılında halka açılan ilk ve en bilinen kripto para birimi Bitcoin ile birlikte kripto para kavramı hayatımıza girmiş olup ekonomik ve finansal anlamda pek çok değişiklik ve yenilikleri beraberinde getirmiştir. İnsanlar tarafından oldukça merak edilip araştırılan ve gün geçtikçe daha çok benimsenen kripto paralar ile ilgili yazılan makaleler, haberler ve anlık olarak paylaşılan kısa metin parçalarının sayısı da aynı şekilde artmakta ve konu ile ilgili metin verilerinin boyutu her geçen gün büyümektedir.

Kripto paralar ile ilgili paylaşılan metin verilerinin miktarı arttıkça metinlerde paylaşılan duyguların analiz edilmesi ve sınıflandırılması çalışmaları yapılmaya başlanmıştır. Literatürde bununla ilgili çalışmalar incelendiğinde ise Türkçe metinlerde konu ile ilgili yapılan çalışmaların yeterli olmadığı ve ele alınan kripto para çeşitlerinin sayı bakımından az olduğu dikkat çekmektedir. Bu tez çalışmasında literatürdeki eksiklikler göz önünde bulundurularak yedi farklı kripto para birimi ve NFT-DeFi ile ilgili 25.000 Türkçe tweet Twitter platformundan çekilmiş; tek tek el ile etiketleme işlemi yapılarak kalan 9874 tweet ile tez çalışması için veri seti oluşturulmuştur. Oluşturulan Türkçe veri seti ve tezde yapılan karşılaştırmalı deneysel çalışmalar ile literatüre katkı sağlamak ve kripto para piyasası ile ilgilenen insanlar için piyasanın yönü hakkında bilgi vermek amaçlanmıştır.

Çalışmada oluşturulan veri seti kullanılarak CNN, RNN, LSTM ve GRU olmak üzere dört derin öğrenme algoritması ile farklı parametreler ve kelime gömme algoritmaları için farklı kombinasyonlar oluşturulmuş ve sonuçlar alınmıştır. Alınan sonuçlar karşılaştırıldığında elde edilen deneysel bilgiler aşağıda sıralanmaktadır:

- Derin öğrenme algoritmaları arasında en yüksek tahmin performansını GRU algoritması, en düşük tahmin performansını ise CNN algoritması vermiştir.
- Word2Vec(Skip-Gram), Word2Vec(CBOW), FastText(Skip-Gram), FastText(CBOW) ve GloVe kelime gömme algoritmaları ile alınan sonuçlar karşılaştırıldığında en yüksek tahmin performansını Word2Vec(Skip-Gram), en düşük tahmin performansını GloVe algoritması vermiştir.

- Deneysel analizde farklı parametreler girilerek elde edilen değerler karşılaştırıldığında 200 ve 300 olan iki vektör boyutu arasında 300 olarak girilen vektör boyutu daha yüksek tahmin performansı sağlamıştır. Projeksiyon katmanı boyutu için girilen 100, 200, 300 değerleri arasında ise en iyi tahmin performansı 300 değeri ile elde edilmiştir.

Daha sonra aynı veri seti kullanılarak farklı çalışmalarda kullanılan ABCDM, AC-BiLSTM, AGCNN, ARC, ATTPooling, CAT-BiGRU, CNN-GRU, CRNN, HAN, IWV, SS-BED, Ağaç BLSTM, Ağaç LSTM ve WCNNLSTM hibrit mimarileri ile de sonuçlar alınmış ve elde edilen sonuçlar karşılaştırılmıştır. Seçilen on dört mimari ile alınan sonuçlar için elde edilen deneysel bilgiler şunlardır:

- ABCDM mimarisi kullanılarak elde edilen sonuçlar tüm hibrit mimariler arasında en iyi performansı göstermiştir. İkinci ve üçüncü en iyi performanslar ise sırasıyla CAT-BiGRU ve CRNN mimarileri kullanıldığında elde edilmiştir.
- Hibrit mimarilerin performansları iyiden kötüye doğru sıralandığında ağaç-LSTM, ağaç-BLSTM ve SS-BED mimarileri veri seti üzerinde daha düşük tahmin performansları vermektedir.

Deneysel çalışmada en gelişmiş mimariler ve hibrit mimariler ile alınan tüm sonuçlar karşılaştırıldığında ise alınan genel sonuçlar şunlardır:

- ABCDM mimarisi en gelişmiş mimarilerden daha iyi performans göstermiştir.
- İkinci ve üçüncü en iyi sonuçlar CAT-BiGRU ve CRNN mimarileri ile alınmıştır.
- En düşük performans sonucu ağaç-LSTM mimarisi kullanıldığında alınmıştır.

KAYNAKLAR

- [1] Çevrimiçi: DonanimHaber: <https://www.donanimhaber.com/AK-Partinin-Bitcoin-raporu-belli-oldu--96224> (Erişim Tarihi: 17.05.2022)
- [2] Özdoğan, B., Kargın, S. Blok Zinciri Teknolojisinin Muhasebe ve Finans Alanlarına Yönelik Yansımaları ve Beklentiler. Muhasebe ve Finansman Dergisi. 2018, 1(80), 161-166.
- [3] Durbilmez, S., Türkmen, S. Blockchain Teknolojisi ve Türkiye Finans Sektöründeki Durumu. Finans Ekonomi ve Sosyal Araştırmalar Dergisi. 2019, 4(1), 30-45.
- [4] Çevrimiçi: 360Koin: <https://360koin.com/blockchain-teknolojisi-nedir-blok-zinciri-nasil-calisir/> (Erişim Tarihi: 17.05.2022)
- [5] Çevrimiçi: Media3m: <https://media3m.com/tether-usdt-nedir/> (Erişim Tarihi: 17.05.2022)
- [6] Çevrimiçi: CoinMarketCap: <https://coinmarketcap.com> (Erişim Tarihi: 20.02.2022)
- [7] Çevrimiçi: WikiPedia: <https://tr.wikipedia.org/wiki?curid=3152607> (Erişim Tarihi: 21.02.2022)
- [8] Çevrimiçi: Google Trends: <https://trends.google.com/trends/?geo=TR>
- [9] Köksal, B., Erdem, G., Türkeli, C., Öztürk, Z. Twitter’da Duygu Analizi Yöntemi Kullanılarak Bitcoin Değer Tahminlemesi. Düzce Üniversitesi Bilim ve Teknoloji Dergisi, 2021, 9(3), 280-297.
- [10] Çevrimiçi: Branding Türkiye: <https://www.brandingturkiye.com/twitter-istatistikleri-guncel/>
- [11] Özdemir, Akcan, E., Sütütemiz, N. Analysis of Brand Perceptions of Covid-19 Vaccines by Sentiment Analysis on Social Media. Pamukkale Üniversitesi Sosyal Bilimler Enstitüsü Dergisi. 2022, 1(4), 145-162.
- [12] Özyurt, Ö., Nilgün, K. COVID-19 Salgını Sürecinde Uzaktan Eğitime İlişkin Tweetlerin Duygusal Analizi. Journal of Computer and Education Research. 2021, 9(18), 853-868.
- [13] Cingiz, M. Ö., Çelikleş, E. Keyword-based Sentiment Analysis of Covid-19 Related Tweets. Journal of Innovative Science and Engineering. 2021, 5(2), 173-182.
- [14] Korkusuz, R., Carus, A. Futbol Müsabakaları ile İlgili Tweetlerin Anlık Duygu Analizi. Avrupa Bilim ve Teknoloji Dergisi. 2020, 386-396.
- [15] Ayan, B., Kuyumcu, B., Ceylan, B. Twitter Üzerindeki İslamofobik Twitlerin Duygu Analizi ile Tespiti. Gazi Üniversitesi Fen Bilimleri Dergisi Part C: Tasarım ve Teknoloji. 2019, 7(2), 495-502.

- [16] Çılğın C., Ünal C., Alıcı S., Akkol E., Gökşen Y. Metin Sınıflandırmada YSA ile Bitcoin Fiyatları ve Sosyal Medyadaki Beklentilerin Analizi. Mehmet Akif Ersoy Üniversitesi Uygulamalı Bilimler Dergisi. 2020, 4(1), 106-126.
- [17] Valencia, F., Gómez-Espinosa, A., Valdés-Aguirre, B. Price Movement Prediction of Cryptocurrencies Using Sentiment Analysis and Machine Learning. Entropy. 2019, 21(6), 589.
- [18] Alghobiri, M. Using Data Mining Algorithm for Sentiment Analysis of Users' Opinions about Bitcoin Cryptocurrency. Journal of Theoretical and Applied Information Technology. 2018, 97(8), 2195-2205.
- [19] Ceyhan, K., Kurtulmaz, E., Sert, O. C., Özyer, T. Bitcoin Movement Prediction with Text Mining. 26th Signal Processing and Communications Applications Conference, 2-5 Mayıs, 2018, İzmir (IEEE, 1-4.)
- [20] Kinderis, M., Bezbradica, M., Crane, M. Bitcoin Currency Fluctuation. 3rd International Conference on Complexity, Future Information Systems and Risk, Mart, 2018, (31-41.)
- [21] Pant, D. R., Neupane, P., Poudel, A., Pokhrel, A. K., Lama, B. K. Neural Network Based Bitcoin Price Prediction by Twitter Sentiment Analysis. 3rd International Conference on Computing, Communication and Security, 25-27 Ekim, 2018, Kathmandu, Nepal (IEEE, 128-132.)
- [22] Lamon, C., Nielsen, E., Redondo, E. Cryptocurrency Price Prediction Using News and Social Media Sentiment. SMU Data Sci. Rev. 2017, 1(3), 1-22.
- [23] Colianni, S., Rosales, S., Signorotti, M. Algorithmic Trading of Cryptocurrency Based on Twitter Sentiment Analysis. CS229 Project. 2015, 1-5.
- [24] Karamanlı, E. Makine Öğrenmesi Algoritmaları Kullanarak, Metin Madenciliği Ve Duygu Analizi İle Müşteri Deneyiminin Geliştirilmesi. İstanbul Üniversitesi, Sosyal Bilimler Enstitüsü, İşletme Ana Bilim Dalı, İstanbul, 2019, 55 s. (Yüksek Lisans Tezi).
- [25] Yılmaz, G. E-Ticaret Sistemlerinde Yapılan Ürün Yorumlarının Metin Madenciliği Uygulaması İle İncelenmesi. İstanbul Aydın Üniversitesi, Lisansüstü Eğitim Enstitüsü, Bilgisayar Mühendisliği Ana Bilim Dalı, İstanbul, 2021, 77 s. (Yüksek Lisans Tezi).
- [26] Beşkirli, A., Gülbandır, E., Dağ, İ. Metin Madenciliği Yöntemleri ile Twitter Verilerinden Bilgi Keşfi. Estudam Bilişim Dergisi. 2021, 2(1), 21-25.
- [27] Karasu, E. Metin Madenciliği Tekniklerini Kullanarak Faaliyet Bazlı Proje Yönetiminde Bilgi Yönetimi Süreçleri için Derin Öğrenme Modeli. İstanbul Üniversitesi, Lisansüstü Eğitim Enstitüsü, Endüstri Mühendisliği Ana Bilim Dalı, İstanbul, 2019, 104 s. (Yüksek Lisans Tezi).
- [28] Emir, Ş. YSA ve DVM Yöntemlerinin Sınıflandırma Performanslarının Karşılaştırılması: Borsa Endeks Yönünün Tahmini Üzerine Bir Uygulama. İstanbul Üniversitesi, Sosyal Bilimler Enstitüsü, İşletme Anabilim Dalı, İstanbul, 2013, 257 s. (Doktora Tezi)

- [29] Çevrimiçi: Wikipedia: https://tr.wikipedia.org/wiki/K%C3%BCmeleme_analizi (Erişim Tarihi: 15.04.2021)
- [30] Çevrimiçi: VeriBilimiOkulu: <https://www.veribilimiokulu.com/makineler-nasil-ogrenir> (Erişim Tarihi: 15.04.2021)
- [31] Onan, A. Türkçe Metin Madenciliği için Dikkat Mekanizması Tabanlı Derin Öğrenme Mimarilerinin Değerlendirilmesi. Avrupa Bilim ve Teknoloji Dergisi. 2022, (34), 403-407.
- [32] Doğan, F., Türkoğlu, İ. Derin Öğrenme Algoritmalarının Yaprak Sınıflandırma Başarımlarının Karşılaştırılması. Sakarya University Journal of Computer and Information Sciences. 2018, 1(1), 10-21.
- [33] Karakuş, B. Derin Öğrenme ve Büyük Veri Yaklaşımları ile Metin Analizi. Fırat Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Ana Bilim Dalı, Elazığ, 2018, 246 s. (Doktora Tezi)
- [34] Çevrimiçi: Proente: <https://proente.com/makine-ogrenimi-derin-ogrenme-ve-yapay-zeka-arasindaki-fark> (Erişim Tarihi: 14.04.2022)
- [35] Konuş, S. Derin Öğrenme Yöntemleri Kullanılarak Çocukların Çizdiği Resimlerdeki Nesnelerin Tanınması. Trakya Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Ana Bilim Dalı, Edirne, 2022, 101 s. (Yüksek Lisans Tezi)
- [36] Feryad, V. Akıllı Elektrik Şebekeleri için Derin Öğrenme Kullanarak Elektrikli Ev Aletlerinin Sınıflandırması. Karadeniz Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Elektrik-Elektronik Mühendisliği Ana Bilim Dalı, Trabzon, 2021, 118 s. (Doktora Tezi)
- [37] Yılmaz, Ş. Word2Vec Temsillerini Kullanarak Türkçede Soru Sınıflandırmasında Derin Öğrenme Analizi. Düzce Üniversitesi, Fen Bilimleri Enstitüsü, Elektrik-Elektronik ve Bilgisayar Mühendisliği Ana Bilim Dalı, Düzce, 2020, 92 s. (Doktora Tezi)
- [38] Çelik, Ö., Koç, B. TF-IDF, Word2vec ve Fasttext Vektör Model Yöntemleri ile Türkçe Haber Metinlerinin Sınıflandırılması. Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi, 2020, 23(67), 121-127.
- [39] Landthaler, J., Walzl, B., Huth, D., Braun, D., Matthes, F., Stocker, C., Geiger, T. Extending Thesauri Using Word Embeddings and the Intersection Method. Second Workshop on Automated Semantic Analysis of Information in Legal Text, Haziran, 2017, Londra.
- [40] Onan A. Mining opinions from instructor evaluation reviews: A deep learning approach. Computer Applications in Engineering Education. 2020, 28(1), 117-138.
- [41] Çevrimiçi: Wikipedia: [https://en.wikipedia.org/wiki/BERT_\(language_model\)](https://en.wikipedia.org/wiki/BERT_(language_model)) (Erişim Tarihi: 13.04.2022)
- [42] Aydoğan, M., Karcı, A. Kelime Temsil Yöntemleri ile Kelime Benzerliklerinin İncelenmesi. Çukurova Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi. 2019, 34(2), 181-196.

- [43] Çevrimiçi: Wikipedia: <https://en.wikipedia.org/wiki/GloVe> (Erişim Tarihi: 15.05.2022)
- [44] Çevrimiçi: Wikipedia: https://tr.wikipedia.org/wiki/Evri%C5%9Fimli_sinir_a%C4%9Flar%C4%B1 (Erişim Tarihi: 15.05.2022)
- [45] Çevrimiçi: WikiMedia : https://commons.wikimedia.org/wiki/File:Recurrent_neural_network_unfold.svg (Erişim Tarihi: 15.04.2021)
- [46] Çevrimiçi: Medium: <https://medium.com/@mcvarer/rnn-lstm-ve-gru-modellerinin-incelemesi-f59a73499edb> (Erişim Tarihi: 18.04.2022)
- [47] Çevrimiçi: Mlreview: <https://blog.mlreview.com/understanding-lstm-and-its-diagrams-37e2f46f1714> (Erişim Tarihi: 18.04.2022)
- [48] Çevrimiçi: Medium : <https://medium.com/deep-learning-turkiye/uzun-k%C4%B1sa-vadeli-bellek-lstm-b018c07174a3> (Erişim Tarihi: 18.04.2022)
- [49] Çevrimiçi: Github: <https://colah.github.io/posts/2015-08-Understanding-LSTMs/> (Erişim Tarihi: 18.04.2022)
- [50] Basiri, ME., Nemati, S., Abdar, M., Cambria, E., Acharya, UR. ABCDM: An attention-based bidirectional CNN-RNN deep model for sentiment analysis. Future Generation Computer Systems. 2021, 115(1), 279-294.
- [51] Liu, G., Guo, J. Bidirectional LSTM with attention mechanism and convolutional layer for text classification. Neurocomputing. 2021, 337(14), 325-338.
- [52] Liu, Y., Ji, L., Huang, R., Ming, T., Gao, C., Zhang, J. An attention-gated convolutional neural network for sentence classification. Intelligent Data Analysis. 2019, 5(23), 1091-1107.
- [53] Wen, S., Li, J. Recurrent Convolutional Neural Network with Attention for Twitter and Yelp Sentiment Classification: ARC Model for Sentiment Classification. In Proceedings of the 2018 International Conference on Algorithms, Computing and Artificial Intelligence, 21 Aralık, 2018, (Bildiriler Kitabı, 1-7).
- [54] Usama, M., Ahmad, B., Song, E., Hossain, M.S., Alrashoud, M., Muhammad, G. Attention-based sentiment analysis using convolutional and recurrent neural network. Future Generation Computer Systems. 2020, 113(1), 571-578.
- [55] Kamal, A., Abulaish, M. CAT-BiGRU: Convolution and Attention with Bi-Directional Gated Recurrent Unit for Self-Deprecating Sarcasm Detection. Cognitive Computation. 2021, 1-9.
- [56] Zhang, Z., Robinson, D., Tepper, J. Detecting Hate Speech on Twitter Using a Convolution-GRU Based Deep Neural Network. Lecture Notes in Computer Science. 2018, DOI:10.1007/978-3-319-93417-4_48.

- [57] Wang, J., Yu, L.C., Lai, K.R., Zhang, X. Dimensional sentiment analysis using a regional CNN-LSTM model. Hesaplamalı Dilbilim Derneği'nin 54. Yıllık Toplantısının Tutanakları (Cilt 2: Kısa Yazılar), Ağustos, 2016, Berlin, Almanya (225-230).
- [58] Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., Hovy, E. Hierarchical attention networks for document classification. Hesaplamalı Dilbilim Derneği Kuzey Amerika Bölümünün 2016 Konferansı Tutanakları: İnsan Dili Teknolojileri, Haziran, 2016 San Diego, California (1480-1489).
- [59] Rezaeinia, S.M., Rahmani, R., Ghodsi, A., Veisi, H. Sentiment analysis based on improved pre-trained word embeddings. Expert Systems with Applications. 2019, 117(1), 139-147.
- [60] Chatterjee, A., Gupta, U., Chinnakotla, M.K., Srikanth, R., Galley, M., Agrawal, P. Understanding emotions in text using deep learning and big data. Computers in Human Behavior. 2019, 93(1), 309-317.
- [61] Li, J., Luong, M.T., Jurafsky, D., Hovy, E. When are tree structures necessary for deep learning of representations?. Doğal Dil İşlemede Ampirik Yöntemler Konferansı Tutanakları, 28 Şubat, 2015, Lizbon, Portekiz, (2304-2314).
- [62] Zhu, X., Sobhani, P., Guo, H. Long short-term memory over tree structures. ICML'15: Makine Öğrenimi Üzerine Uluslararası Konferans Üzerine 32. Uluslararası Konferans Bildirileri, Temmuz, 2015, (1604-1612).
- [63] Onan A. Sentiment analysis on product reviews based on weighted word embeddings and deep neural networks. Concurrency and Computation: Practice and Experience. 29 Haziran, 2020.
- [64] Küçüktaş, Ü. T., Uysal, F., Hardalaç, F., Biri, İ. Detection of Increased Time Intervals of Anti-Vaccine Tweets for COVID-19 Vaccine with BERT Model. 1. Uluslararası Yapay Zeka ve Veri Bilimi Kongresi, 26-28 Kasım, 2021, İzmir (Bildiriler Kitabı, 19-24.)
- [65] Çevrimiçi: Wikipedia: <https://tr.wikipedia.org/wiki/TensorFlow> (Erişim Tarihi: 16.05.2022)
- [66] Çevrimiçi: Medium: <https://medium.com/bili%C5%9Fim-hareketi/hiperparametre-optimizasyonu-9ba0e7f32e6f#:~:text=Hiperparametre%20optimizasyonu%2C%20bir%20makine%20%C3%B6%C4%9Frenmesi,overfitting%20ve%20underfitting%20dengesi%20sa%C4%9Flanabilir>. (Erişim Tarihi: 01.06.2022)