

T.C.
İSTANBUL ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
SAYISAL YÖNTEMLER
ANABİLİM DALI

YÜKSEK LİSANS TEZİ

KRİPTO PARA PİYASASINA YÖNELİK DUYGU
ANALİZİ UYGULAMASI

Sezin ÖZTÜRK

2501180755

TEZ DANIŞMANI

Prof. Dr. Bilge ACAR BOLAT

İSTANBUL – 2022

ÖZ

KRİPTO PARA PİYASASINA YÖNELİK DUYGU ANALİZİ

UYGULAMASI

Sezin ÖZTÜRK

Teknoloji alanında yaşanan yenilikler hayatın her alanında yenilikleri beraberinde getirmektedir. Ödeme sistemleri ve yatırım araçları da teknolojiye yaşanan yenilikler ile birlikte yeni bir boyut kazanmıştır. Bununla birlikte sanal para kavramı da toplumlar tarafından benimsenmiş ve önemli bir yatırım aracı haline gelmiştir. Bu sanal paralar temelinde yatan kriptografik teknoloji sayesinde şifreler ile korunmaktadır. Bu paralara en iyi örnek 2009 yılından itibaren piyasalarda işlem görmeye başlayan Bitcoin'dir.

Kripto paraların günümüzde oldukça popüler olmasının sebebi blok zinciri teknolojisidir. Kripto paraların temelinde böyle bir teknoloji olmasaydı yatırımcılar bu sisteme güven sorunu yaşayabilirdi. Bitcoin gibi kripto paralar bugün ulaştığı yüksek değerlere ulaşamayabilir ve bu kadar popüler olmayabilirdi. Son yıllarda kripto para piyasalarındaki çok yüksek artış ya da azalış yaşanması da yatırımcıların oldukça dikkatini çekmektedir.

Bu çalışmada kripto para piyasasında yaşanan aşırı yükseliş ve düşüşlerin yaşandığı iki farklı dönem karşılaştırılmıştır. Bu karşılaştırılmanın yapılabilmesi için en büyük pazar payına sahip olan ve sistemi domine eden kripto para olan Bitcoin tercih edilmiştir.

Günümüzde farklı türdeki verilerin derlenip saklanabilmesi ve bu verilerin işlenebilmesi ile sosyal medya verileri de analizlerde kullanılmaya başlanmıştır. Sosyal medyadan elde edilen veri, yapılandırılmamış veri olarak karşımıza çıkmaktadır. Geleneksel analiz yöntemlerinde bir verinin işlenebilmesi için verinin yapılandırılmış olması koşulu varken, günümüzde yapılandırılmamış veriler kullanılarak da analizler gerçekleştirilmektedir. Bu analizlerin gerçekleştirildiği en etkili alan metin madenciliğidir. Metin madenciliği yöntemleri sayesinde yapılandırılmamış olan veriler de işlenebilir hale gelmiştir. Metin madenciliğinde duygu analizi uygulamaları son yıllarda popüleritesi hızla artan bir araştırma alanı

haline gelmiştir. Duygu analizinin temel amacı metinlerde yer alan duygunun ortaya çıkarılmasıdır.

Sosyal medyanın hayatımızda önemli bir yer haline gelmesiyle, kullanıcılar kendilerine sunulan içerikleri tüketen pasif kullanıcılar değil, içerik üretebilen aktif kullanıcılar haline gelmiştir.

Sosyal medyada oluşan bilgi yığınlarını da analiz etmek için en yaygın kullanılan yöntem duygu analizidir. Bu çalışmada kripto para piyasasında meydana gelen aşırı yükseliş ve düşüşlerin yaşandığı iki farklı dönem seçilmiştir. Seçilen bu dönemlerde sosyal medya kullanıcılarının #bitcoin ve #btc hashtag kullanarak paylaştıkları yorumlar duygu analizi aracılığıyla analiz edilmiştir. Yatırımcıların kripto para piyasalarına olan ilgisi farklı koşullarda değerlendirilmiştir.

Anahtar Kelimeler: kripto para, bitcoin, duygu analizi, metin madenciliği, sosyal medya



ABSTRACT

**SENTIMENT ANALYSIS APPLICATION FOR THE
CRYPTOCURRENCY MARKET**

Sezin ÖZTÜRK

Innovations in the field of technology bring innovations in all areas of life. Payment systems and investment instruments have also gained a new dimension with the innovations in technology. However, the concept of virtual money has also been adopted by societies and has become an important investment tool. These virtual currencies are protected by passwords thanks to the underlying cryptographic technology. The best example of these currencies is Bitcoin, which started to be traded in the markets since 2009.

The reason why cryptocurrencies are very popular today is the blockchain technology. If such a technology were not at the base of cryptocurrencies, investors would have a problem of trust in this system. Cryptocurrencies such as Bitcoin might not have reached the high values they have reached today and might not have been as popular. The very high increase or decrease in the crypto money markets in recent years also attracts the attention of investors.

In this study, two different periods of extreme rise and fall in the crypto money market were compared. In order to make this comparison, Bitcoin, which has the largest market share and dominates the system, was preferred.

Today, with the ability to compile and store different types of data and to process these data, social media data has started to be used in analysis. Data obtained from social media appears as unstructured data. While in traditional analysis methods, there is a condition for the data to be structured in order for a data to be processed, nowadays analyzes are carried out using unstructured data. The most effective area in which these analyzes are carried out is text mining. Thanks to text mining methods, unstructured data has also become workable. Sentiment analysis applications in text mining have become a rapidly increasing research area in recent years. The main purpose of sentiment analysis is to reveal the emotion in the texts.

With the social media becoming an important place in our lives, users have become active users who can produce content, not passive users who consume the content presented to them.

Sentiment analysis is the most widely used method to analyze the information stacks in social media. In this study, two different periods of extreme rise and fall in the crypto money market were selected. In these selected periods, the comments shared by social media users using #bitcoin and #btc hashtags were analyzed through sentiment analysis. Investors' interest in cryptocurrency markets has been evaluated under different conditions.

Keywords: cryptocurrency, bitcoin, sentiment analysis, text mining, social media

ÖNSÖZ

Bu çalışma İstanbul Üniversitesi Sosyal Bilimler Enstitüsü İşletme Anabilim Dalı Sayısal Yöntemler Bilim Dalı Yüksek Lisans programında yüksek lisans tezi olarak hazırlanmıştır.

Bu tezin ortaya çıkma sürecinde her durumda ilgisiyle destek olan değerli danışman hocam Prof. Dr. Bilge ACAR BOLAT’a teşekkürlerimi sunarım.

Eğitim hayatımın boyunca maddi ve manevi desteğini asla esirgemeyen ablam İpek Öztürk, babam Ramazan Öztürk, ağabeyim Ahmet Öztürk, eşim Çağlar Uşun ve benim bugünlere gelmemi sağlayan, her zaman uzaklarda bir yerlerde beni izlediğini bildiğim canım annem Saadet Öztürk’e sonsuz teşekkürlerimi sunarım.

Sezin ÖZTÜRK

İSTANBUL, 2022



İÇİNDEKİLER

ÖZ	ii
ABSTRACT	iv
ÖNSÖZ	vi
TABLolar LİSTESİ	x
ŞEKİLLER LİSTESİ	xi
RESİM LİSTESİ	xii
KISALTMALAR LİSTESİ	xiii
GİRİŞ	1

BİRİNCİ BÖLÜM

KRİPTO PARA

1.1. Literatür Taraması	3
1.2. Kripto Para	6
1.2.1. Kripto Paraların Çalışma Prensipleri (Özellikleri).....	8
1.2.2. Kripto Para Yapıları.....	10
1.2.3. Kripto Para Teknolojisi.....	11
1.2.3.1. Kriptolojik Özet Fonksiyonlar	12
1.2.3.2. Dijital İmza.....	13
1.2.3.3. İş İspatı	14
1.2.3.4. Eşler Arası Bağlantı	15
1.2.3.5. Akıllı Sözleşmeler.....	16
1.3. Blok Zinciri	17
1.3.1. Blok Zinciri Teknolojisi.....	20
1.3.1.1. Kriptoloji.....	22
1.3.1.1.1. Kriptografi.....	22

1.3.1.1.2. Kriptoanaliz.....	23
1.4. Kripto Para Avantaj ve Dezavantajları.....	24
1.5. Kripto Para Örnekleri	26
1.5.1. Bitcoin.....	27
1.5.2. Ethereum.....	28
1.5.3. Ripple.....	30

İKİNCİ BÖLÜM

VERİ MADENCİLİĞİ

2.1. Metin Madenciliği	33
2.1.1. Doğal Dil İşleme	34
2.1.2. Metin Madenciliği Kullanım Alanları	35
2.1.3. Metin Madenciliği Aşamaları	36
2.1.3.1. Veri Setinin Oluşturulması.....	36
2.1.3.2. Ön İşleme	36
2.1.3.2.1. Parçalara Ayırma.....	37
2.1.3.2.2. Durak Kelimeleri Çıkarma	37
2.1.3.2.3. Filtreleme	38
2.1.3.2.4. Kök Bulma	38
2.1.3.2.5. N-Gram	39
2.1.3.3. Öznitelik Seçme	39
2.1.3.4. Duygu Sınıflandırma Yöntemleri.....	40
2.1.3.4.1. Karar Ağaçları.....	41
2.1.3.4.2. K En Yakın Komşu	45
2.1.3.4.3. Yapay Sinir Ağları	46
2.1.3.4.4. Lojistik Regresyon	48
2.1.3.4.5. Naive Bayes	49

2.1.3.4.6. Destek Vektör Makineleri	51
2.1.3.5. Sınıflandırma Algoritmalarının Karşılaştırılması	53
2.3. Duygu Analizi	55
2.3.1. Duygu Analizi İçin Makine Öğrenmesi Yaklaşımı	58
2.3.2. Duygu Analizi için Sözlük Temelli Yaklaşım	59
2.3.3. Sosyal Medya Verileri ile Duygu Analizi.....	60

ÜÇÜNCÜ BÖLÜM

UYGULAMA

3.1. Araştırmanın Amacı	61
3.2. Araştırmanın Örneklemi.....	62
3.3. Araştırmanın Kısıtları.....	63
3.4. Veri Toplama.....	63
3.4.1. Aplikasyon Oluşturma	64
3.5. Araştırmanın Yöntemi ve Araçlar	65
3.6. Veri Ön İşleme Süreci	66
3.6.1. Verilerde Karakter Düzenlemesi	69
3.6.2. Sözlük Oluşturma	73
3.7. Duygu Skoru.....	73
3.7.1. Betimsel İstatistikler	76
3.7.2. Kelime Frekanslarının Değerlendirilmesi.....	78
3.7.3. Bitcoin Fiyat ve Hacim Verilerinin Değerlendirilmesi.....	82
3.7.3.1. Türk Kripto Para Borsalarında Bitcoin	83
3.7.3.2. Bitcoin Hacmi ile Google Trends Verileri Arasındaki İlişki	86
3.7.4. Makine Öğrenmesi İle Duygu Analizi.....	89
SONUÇ	92
KAYNAKÇA	97

TABLÖLAR LİSTESİ

Tablo 1-1: Kriptolojik Özet Fonksiyonu	12
Tablo 2-1: Araba Alımı Örnek Dizisi	42
Tablo 2-2: Naive Bayes Algoritması Örnek Veri Seti	50
Tablo 2-3: Kontenjans Tablosu	54
Tablo 3-1: Duygu Skorlarının Aritmetik Ortalaması	74
Tablo 3-2: Bitcoin Yükseliş Trendi Gösterdiği Zamana İlişkin Tweet Puanları ve Frekansları	74
Tablo 3-3: Bitcoin Düşüş Trendi Gösterdiği Zamana İlişkin Tweet Puanları ve Frekansları	75
Tablo 3-4: Makine Öğrenmesi Sonuçları	101

ŞEKİLLER LİSTESİ

Şekil 1-1: 2013-2022 Yılları Arası Kripto Para Piyasası.....	7
Şekil 1-2: Toplam Piyasa Değeri Yüzdesi.....	8
Şekil 1-3: Kripto Paraların Çalışma Düzeni	9
Şekil 1-4: Merkezi, Çok Merkezli ve Dağınık Ağ Yapıları	10
Şekil 1-5: Dijital İmzalama ve İmza Doğrulama.....	13
Şekil 1-6: İş İspatı Algoritması.....	15
Şekil 1-7: Akıllı Sözleşmeler Sistemi.....	17
Şekil 1-8: Blok Zinciri.....	18
Şekil 1-9: Blok Zincirinin Çalışma Prensibi.....	19
Şekil 1-10: Blok Zinciri Çalışma Sistemi.....	21
Şekil 1-11: Geleneksel ve Yeni Gizlilik Modeli.....	23
Şekil 1-12: Kripto Para Örnekleri.....	27
Şekil 1-13: Yıllar İçerisinde Bitcoin'de Gerçekleşen Fiyat Değişimi.....	28
Şekil 1-14: Yıllar İçerisinde Ethereum'da Gerçekleşen Fiyat Değişimi.....	29
Şekil 1-15: Yıllar İçerisinde Ripple'da Gerçekleşen Fiyat Değişimi.....	30
Şekil 2-1: Metin Ön İşleme Süreci	37
Şekil 2-2: Karar Ağacı Yapısı.....	42
Şekil 2-3: K- En Yakın Komşu Algoritması	46
Şekil 2-4: Yapay Sinir Ağı Modeli.....	47
Şekil 2-5: Lojistik Regresyon Modeli	49
Şekil 2-6: Duygu Analizi Yöntemleri.....	58
Şekil 3-1: Twitter Verilerinin Elde Edilmesi ve İşlenmesi Süreci	67
Şekil 3-2: 'veri1' Tweetlerin Retweet Edilme Oranları	76
Şekil 3-3: 'veri1' Tweetlerin Favori Edilme Oranları	77
Şekil 3-4: 'veri2' Tweetlerin Retweet Edilme Oranları	77
Şekil 3-5: 'veri2' Tweetlerin Favori Edilme Oranları	78
Şekil 3-6: 06-12 Kasım 2021 Tarih Aralığında Elde Edilen Tweetlerde En Sık Kullanılan Kelimelerin Kelime Bulutu	79
Şekil 3-7: 06-12 Kasım 2021 Tarih Aralığında Elde Edilen Tweetlerde En Sık Kullanılan Kelimelerin Frekans Dağılımı.....	80
Şekil 3-8: 05-11 Mayıs 2022 Tarih Aralığında Elde Edilen Tweetlerde En Sık Kullanılan Kelimelerin Kelime Bulutu	80

Şekil 3-9: 05-11 Mayıs 2022 Tarih Aralığında Elde Edilen Tweetlerde En Sık Kullanılan Kelimelerin Frekans Dağılımı.....	81
Şekil 3-10: İlgili Tarihlerde Bitcoin Fiyat Grafiği.....	82
Şekil 3-11: İlgili Tarihlerde Bitcoin Hacim Grafiği	83
Şekil 3-12: İlgili Tarihlerde Paribu Platformunda Gerçekleşen Bitcoin Hacim Grafiği	84

RESİM LİSTESİ

Resim 3-1: Veri Çekebilmek İçin Gerekli R Kodları.....	65
Resim 3-2: R Kütüphanesinden Gerekli Olan R Paketlerinin Çağırılması	66
Resim 3-3: RapidMiner Kullanılarak Ön İşleme Sürecinin Gerçekleştirilmesi.....	69
Resim 3-4: Verilerde Karakter Düzenlemesi	70
Resim 3-5: İşlenmiş Veri Setinin Kaydedilmesi	71

KISALTMALAR LİSTESİ

BTC	: Bitcoin
XRP	: Ripple
ETH	: Ethereum
DDİ	: Doğal Dil İşleme
YSA	: Yapay Sinir Ağları
K-NN	: K En Yakın Komşu
DVM	: Destek Vektör Makinaları
TP	: Doğru Pozitif
FP	: Yanlış Pozitif
FN	: Yanlış Negatif
TN	: Doğru Negatif
RAM	: Random Access Memory
API	: Application Programming Interface

GİRİŞ

Tarih boyunca insanlar tarafından para, ödeme aracı olarak kullanılmıştır. Para ile gerçekleştirilen işlemler teknoloji geliştikçe değişime uğramıştır. Günümüzde her geçen gün yeni teknolojik gelişmeler gerçekleşmektedir. Bu gelişmeler bilim ve teknolojinin yanı sıra, insanların hayatını da kolaylaştırarak ve hızlandırarak olumlu katkılar sağlamaktadır. Dünyadaki her bireyin, şirketin, devletin kullandığı ödeme araçları da teknolojik gelişmeler ile birlikte farklılaşmaktadır. Teknolojinin gelişmesi ile birlikte paranın evrimleşmesi hızlanmıştır. Bu evrimleşme sonucunda elektronik para, dijital para ve sanal paralar kullanılır hale gelmiştir.

Sanal paralar ile hayatımıza kripto para kavramı da girmiştir. Kripto paralar, temelinde kriptografi teknolojisi olan ödeme araçlarıdır. Kriptografi teknolojisi sayesinde yeni kripto para birimlerinin güvenli ve kontrollü bir şekilde oluşturulması sağlanmaktadır. Güvenli bir ortamın sağlanması için şifreleme kullanılan bir sisteme sahiptir. Sistemin kontrolü blok zinciri adı verilen işlem veri tabanları aracılığıyla sağlanmaktadır. Kripto para piyasası aktif olarak ilk 2009 yılında Bitcoin üretilmesi ile kullanılmaya başlanmıştır. Bitcoin'den sonra her geçen gün yeni bir kripto para ortaya çıkmıştır. Günümüzde kripto para piyasalarında binlerce kripto para işlem görmektedir. Sayıları hızla artan kripto paralar günlük hayatta, ticari faaliyetlerde ödeme aracı olarak kullanılmaktadır.

Özellikle 2017 yılından sonra kripto paralar oldukça popüler hale gelmiştir. Şirketlerin ve devletlerin kripto paralara bakış açısı önemli ölçüde değişmiştir. Bazı ülkeler kripto paraları yasal ödeme aracı olarak kabul ederken hala bu konuda çekimser yaklaşım sergileyen ülkeler vardır. Kripto paralar dünya çapında birçok büyük şirket tarafından ödeme aracı olarak kabul edilmeye başlanmıştır.

Günümüzde en yüksek pazar payına sahip kripto para olan Bitcoin, 2022 yılı ikinci yarısı itibariyle yaklaşık 30,200,000,000 Dolar işlem hacmine sahiptir. Ayrıca 2021 yılı içerisinde tarihinde en yüksek fiyat olan 65.000 Dolar seviyelerine ulaşmasıyla tüm dünyanın ilgisini çekmiştir.

Bu çalışma kapsamında sosyal medyada yer alan verilerin yatırımcılar için oldukça önemli olduğundan sosyal medyadan elde edilen verileri analiz eden bir model kurulmuştur. Bir sosyal medya platformu olan Twitter kullanılarak kripto para

piyasasına ilişkin Twitter kullanıcıları tarafından paylaşılan yorumlar toplanmıştır. Kripto para piyasasının yükseliş ve düşüş trendinde olduğu iki farklı dönem seçilmiş ve seçilen zaman dilimlerinde Twitter kullanıcılarının kripto para piyasasına ilişkin duygu analizi gerçekleştirilmiştir. İki farklı dönemin karşılaştırılması ile yatırımcıların kripto para piyasalarına farklı durumlarda sergiledikleri bakış açıları değerlendirilmiştir.

Tezin ilk bölümünde kripto para piyasası üzerine gerçekleştirilen duygu analizi çalışmaları içeren literatür taraması sunulmuştur. Ayrıca kripto para kavramı detaylı bir şekilde açıklanarak kripto para teknolojisinden bahsedilmiştir. Kripto para teknolojisi olan blok zincirinin sağladığı kolaylıklar ve blok zinciri çalışma sistemi açıklanmıştır. Kripto paraların avantaj ve dezavantajları da belirtilmiştir.

Çalışmanın ikinci bölümünde gerçekleştirilen uygulamaya yönelik kavramlar üzerinde durulmuştur. İkinci bölümde metin madenciliği kavramı ve metin madenciliği aşamaları detaylı bir şekilde aktarılmıştır. Metin madenciliği yöntemi olarak gerçekleştirilen duygu analizi ve duygu analizinde kullanılan sınıflandırma algoritmaları da bu bölümde açıklanmıştır.

Üçüncü bölümde ise çalışmanın uygulama bölümü sunulmuştur. Bu bölümde verilerin elde edilme aşaması detaylı bir şekilde aktarılmıştır. Verilerin elde edilmesinden sonra çalışmanın en fazla zaman alan kısmı olan veri ön işleme aşaması açıklanmıştır. Verilerin ön işleme aşaması gerçekleştirildikten sonra oluşturulan pozitif ve negatif duygu sözlükleri aracılığıyla tweetlerin duygu skorları hesaplanmıştır. Oluşturulan her iki veri seti için de duygu skorları ortalamaları ve frekansları paylaşılmıştır. Verilerin elde edildiği dönem Bitcoin fiyat ve hacim grafikleri ile duygu skorları karşılaştırılmıştır.

Son yıllarda kripto para piyasasının oldukça popüler hale gelmesiyle sosyal medyada da oldukça fazla gündem olan bir konu haline gelmiştir. Bu çalışma ile yatırımcıların, kripto para piyasasındaki dalgalanmalara nasıl tepki verdiğini gözlemlenmiştir. Duygu analizi ile yatırımcıların kripto para piyasasındaki dalgalanmalar karşısındaki tutumunun pozitif, negatif veya nötr olarak değerlendirilmek istenmiştir. Çeşitli algoritmalar aracılığıyla metinlerde paylaşılan görüş ve ifadeler üzerine sınıflandırma işlemi gerçekleştirilmiştir.

BİRİNCİ BÖLÜM

KRİPTO PARA

1.1. Literatür Taraması

Ray Li ve ark. (2019), ZClassic (ZCL) isimli alternatif kripto para biriminin fiyat dalgalanmalarını tahmin edebilmek için Twitter kullanıcılarının yorumlarını bir araç olarak kullandılar. ZCL ile ilgi yirmi dört gün boyunca paylaşılan tweetleri saatlik olarak veri seti haline getirip olumlu, olumsuz ve nötr olarak sınıflandırmışlardır. Bu sınıflandırmayı her tweet için -1 ile +1 arasında bir polarite değeri sağlayarak oluşturmuşlardır. Twitter gibi güçlü etkileşim sağlayan sosyal medya ağlarının fiyat hareketinin tahmininde önemli rol aldığı sonucuna ulaşılmıştır.

Kim ve ark. (2016), kripto para piyasasında işlem sayısının kullanıcı yorumlarının ve yanıtlarının kripto para birimi üzerinde etkisi olup olmadığını Granger nedensellik testi yardımıyla araştırmışlardır. Araştırma kapsamında Bitcoin (BTC), Ethereum (ETH) ve Ripple (XRP) olmak üzere üç kripto para birimine odaklanmışlardır. Olumlu kullanıcı yorumlarının Bitcoin fiyatı üzerinde olumlu bir etkisi varken olumsuz yorumların hiçbir etkisinin olmadığı sonucuna ulaşılmıştır. Ripple ve Ethereum fiyatında ise olumsuz kullanıcı yorumlarının önemli derecede fiyatı etkilediği sonucuna ulaşılmıştır.

Inamdar ve ark. (2019), Bitcoin (BTC) yatırımcıların minimum kayıpla yatırım yapmasına yardımcı olmak amacıyla piyasayı ve sosyal medya trendlerini analiz etmişlerdir. Bu analizi gerçekleştirmek için Twitter kullanıcı yorumları veri olarak kullanılmıştır. Elde edilen veriler üzerinde duygu analizi, haritalama ve tahmin işlemi yapılmıştır. Kullanıcıların paylaştığı tweetlerin duygu polaritesi (duygu puanı) ile birlikte rastgele orman algoritması kullanılarak bir sonraki 2 gün için tahmin yapılmıştır. Tweetlerin ve haber akışı verilerinin duygu polariteleri çoğunlukla nötr olduğundan yapılan birçok çalışmanın aksine sosyal medya gönderilerinin fiyatlar üzerinde etkisi olmadığı sonucuna ulaşılmıştır.

Lamon ve ark. (2017), Bitcoin, Ethereum ve Litecoin olmak üzere üç kripto para birimi için haber ve sosyal medya verilerini kullanarak fiyat dalgalanmalarını tahmin etmek istemişlerdir. Zaman serisi yapısını bozmamak için gün bazında

toplanan haber metinleri ve sosyal medya platformu Twitter'dan elde edilen tweetler çalışmanın veri setini oluşturmaktadır. Metin tabanlı duygu sınıflandırılması için denetimli öğrenme algoritmalarını kullanmışlardır. Bu algoritmaları kullanırken günlük haber ve sosyal medya verileri pozitif ve negatif duygulardan ziyade, her bir kripto para için bir sonraki günün gerçek fiyat değişikliklerine göre duygu polaritesi belirlenmiştir. Böylece ilk önce duyguyu tahmin etmek yerine fiyat dalgalanmaları doğrudan tahmin edilmek istenmiştir.

Kraaijeveld ve Smedt (2020), piyasa değeri en yüksek olan dokuz kripto para biriminin gelecekteki fiyatını tahmin etmek için Twitter duyarlılığının kullanılıp kullanılmayacağını incelemişlerdir. Çalışma kapsamında Bitcoin, Ethereum, XRP, BitcoinCash, EOS, Litecoin, Cardano, Stellar ve TRON kripto para birimleri üzerinden tahminler gerçekleştirilmiştir. Sözlük tabanlı duygu analizi yaklaşımı, finansal veriler ve Granger nedensellik testi kullanılarak gerçekleştirilen çalışmada Twitter duyarlılığının Bitcoin, BitcoinCash ve Litecoin'in getirileri için tahmin gücünün yüksek olduğu bulunmuştur.

Sattarov ve ark. (2020), Twitter kullanıcılarının Bitcoin getirilerinin ne ölçüde tahmin edilebileceğini incelemişlerdir. Tweetlerin duygu polaritesinin Bitcoin'in gelecek getirisini tahmin etmede etkili olup olmadığının araştırıldığı çalışmada bir korelasyonel ilişki gözlenmiştir. Bitcoin ile ilgili tweet duyarlılığına ve geçmiş Bitcoin fiyatına dayanarak yapılan tahminlerin rastgele orman algoritması ile %62,48 doğruluk payı olduğu sonucuna ulaşmışlardır.

Abraham ve ark (2018), Twitter ve Google Trends verilerini kullanarak piyasa değeri en yüksek iki kripto para olan Bitcoin ve Ethereum fiyatlarındaki değişiklikleri tahmin etmeye çalışmışlardır. Fiyatların düştüğü bir ortamda, tweetlerin nesnel olması ve fiyat değişikliklerinden bağımsız olarak olumlu duygu içerme eğiliminde olduğundan kripto para birimi fiyat değişikliklerinde duygu analizinin daha az etkili olduğu gözlenmiştir. Tweetlerin duygu polaritesinden çok tweet hacminin fiyat yönünün bir öngörücüsü olarak değerlendirilebileceği gözlenmiştir.

Hassan vd. (2021), makine öğrenimi yoluyla duygu teorisi ve sözlük duyarlılık analizi merceği altında Twitter kullanıcılarının kripto paralar ile ilgili görüşlerini keşfetmeyi amaçlamaktadır. Duygu analizi olumlu ve olumsuz iki duyguyu test etmek

için yapılmıştır. Kullanıcıların olumlu tweetleri piyasadaki yatırımın artışını tetiklediği gözlenmiştir.

Rognone vd. (2020), Bitcoin haberlerinin kripto para birimi Bitcoin ve geleneksel para birimlerinin getirileri, hacmi ve dalgalanması üzerindeki etkisini araştırmak üzere bir çalışma gerçekleştirmişlerdir. Haber metinlerinin duygu polaritesi yüksek olduğunda geleneksel para birimleri tipik olarak artış eğilimi gösterirken, düşük polarite olduğunda düşüş eğilimi göstermektedir. Bitcoin ise haberlerin duygu polaritesinin düşük ya da yüksek olduğu her durumda haberlere olumlu tepki verdiği gözlenmiştir. Haber içeriğinden çok Bitcoin ile ilgili yapılan haberlerin hacmi önemlidir. Sadece siber saldırı ve dolandırıcılık haberleri bu etkiyi azaltarak Bitcoin getirisini olumsuz etkilediğini gözlenmiştir.

Matta ve ark. (2015), Bitcoin fiyatındaki dalgalanmanın konu ile ilgili tweet hacimleri ve Google arama sonuçları ile ilgili olup olmadığını öğrenmek amacıyla bir çalışma gerçekleştirmişlerdir. Araştırma kapsamında 2 aylık süreci kapsayan Bitcoin ile ilgili tweetler elde edilmiş ve tweetler üzerinden duygu analizi gerçekleştirmişlerdir. Bitcoin fiyatındaki dalgalanmanın tweet hacmi, pozitif duygu içeren tweetler ve Google trends verileri ile arasındaki ilişkiyi çapraz korelasyon analizi ile incelemişlerdir. Hem tweetlerin hem de Google trends verilerinin Bitcoin için birkaç gün içindeki fiyatını tahmin etmek için kullanılabileceği sonucuna ulaşılmıştır.

Kaminski (2016), Bitcoin piyasa fiyatı ile Bitcoin ile ilgili duygusal sinyaller içeren Twitter gönderileri arasındaki korelasyonu ve nedenselliği analiz etmiştir. 104 günlük zaman dilimini içeren olumlu, olumsuz ve nötr duygu sinyallerini içeren 160.000 Twitter gönderisine erişim sağlamıştır. Pearson korelasyon sonuçlarına göre tweetlerin duygu içerikleri ile Bitcoin kapanış fiyatı, işlem hacmi ve gün içi fiyat dağılımı arasında önemli bir pozitif korelasyon olduğunu saptamıştır. Fakat Granger nedensellik analizi tweetlerin duygu içerikleri ile Bitcoin piyasa değerleri arasında bir nedensellik ilişkisi olmadığını göstermiştir. Bu sonuç duygu durumunun piyasayı tahmin etmekten çok piyasayı yansıttığını göstermiştir.

Colianni ve ark. (2015), Bitcoin ile ilgili Twitter verilerinin kripto para ticareti stratejileri geliştirmek için kullanılıp kullanılamayacağını araştırmışlardır. Denetimli öğrenme algoritmalarından lojistik regresyon, naive bayes ve destek vektör

makinelerinin kullanıldığı çalışmada saatten saate ve günden güne yapılan tahminlerin doğruluk oranının %90'dan fazla olduğu saptanmıştır. Verileri temizleme ve lojistik regresyon, Naive Bayes ve destek vektör makineleri gibi denetimli öğrenme algoritmalarını uygulamaya yönelik yaklaşım, son bir saatten saate ve günden güne tahmin doğruluğunun %90'ı geçmesini sağlamaktadır. Bu sonuca ulaşmak için, modelin her adımında doğru girdilerin kullanılmasını sağlamak için titiz hata analizi yapılır. Bu analiz ortalama olarak %25 doğruluk artışı sağlar.

1.2. Kripto Para

Kripto paralar esasen blok zinciri teknolojisine dayanan ve kriptografi ile güvence altına alınan, merkezi olmayan elektronik paralardır. Merkezi bir kuruma başvurmadan işlemleri kolaylaştırmak amacıyla tasarlanmıştır. Kripto paralar herhangi bir değerli metal tarafından desteklenmemektedir. Kripto para birimleri, dolar banknotları veya altın madeni gibi geleneksel para birimlerine benzer olarak belirli bir finansal aracıya ihtiyaç duymadan bireyler arasındaki işlemleri kolaylaştırmak için kullanılmaktadır. Merkezi olmayan bu yapının kontrolü Merkez Bankası gibi bir otorite tarafından değil, blok zinciri teknolojisi sayesinde gerçekleştirilmektedir. Kriptografi, hem bireysel işlemleri hem de hesap defterlerini güvence altına almak amacıyla kullanılmaktadır. Tüm işlemlerin kaydedildiği hesap defterleri yatırımcıların kolayca erişebileceği şekildedir. Fakat bu sistemde belirli bir adresi belirli bir bireye bağlayan merkezi bir otorite olmadığı için kripto para sahipleri anonim olarak işlem yapabilmektedir. Dijital para birimi fikri internetin doğuşu ile birlikte ortaya çıkmasına rağmen çok sonra başarıya ulaşmıştır. 1990'larda Cybercash ve Digicash gibi birçok elektronik para birimi dolaşıma girmiştir. Ancak internet üzerinden dolandırıcılık vakaları ile çok fazla karşılaşıldığından hiçbir zaman güvenilir bulunmamıştır (Cox, 2013).

2008 yılının sonlarında Satoshi Nakamoto takma adlı kişi veya kişilerin Metzdowd.com internet adresinde "Bitcoin: A Peer-to-Peer Electronic Cash System" adlı makale yayınlamaları ile günümüzdeki kripto paraların temelini atmışlardır (Nakamoto, 2008). Satoshi Nakamoto, piyasadaki ilk ve lider konumda olan kripto para birimi olan Bitcoin'i ortaya çıkarmıştır. 2009 yılında ise Bitcoin tüm dünyaya

sunulmuştur. Ancak Bitcoin 2012 yılında önde gelen iş insanları arasında resmi bir ödeme aracı olarak kabul görmeye başlamıştır (Surana vd., 2022).

Bitcoin kodunun açık kaynak olması nedeniyle kripto para birimlerinin sayısı günden güne hızla artmaktadır. Çünkü kripto para birimi oluşturmak için program kodları kolayca kopyalanıp değiştirilebilmektedir. Bu avantaj ile birlikte Şekil 1-1’de görüldüğü üzere kripto para üretimi artmıştır (Smith & Kumar, 2018: 1533). 2014 yılında yaklaşık 500 adet kripto para piyasada işlem görürken 2022 yılı itibari ile CoinMarketCap verilerine göre 15.000’in üzerinde kripto para yaklaşık 600 farklı kripto para piyasalarında işlem görmektedir. İşlem gören kripto paraların piyasa değeri ise 2022 yılı itibariyle 2.5 Trilyon Dolar’ın üzerindedir. İşlem hacimleri nedeniyle likiditelerinin de yüksek olması yatırımcılar için bir fırsat oluşturmaktadır. Kripto para piyasasındaki çeşitlilik ve avantajlar yatırımcılar için tercih sebebi haline gelmiştir (coinmarketcap.com, E.T: 04.2022).

Şekil 1-1: 2013-2022 Yılları Arası Kripto Para Piyasası

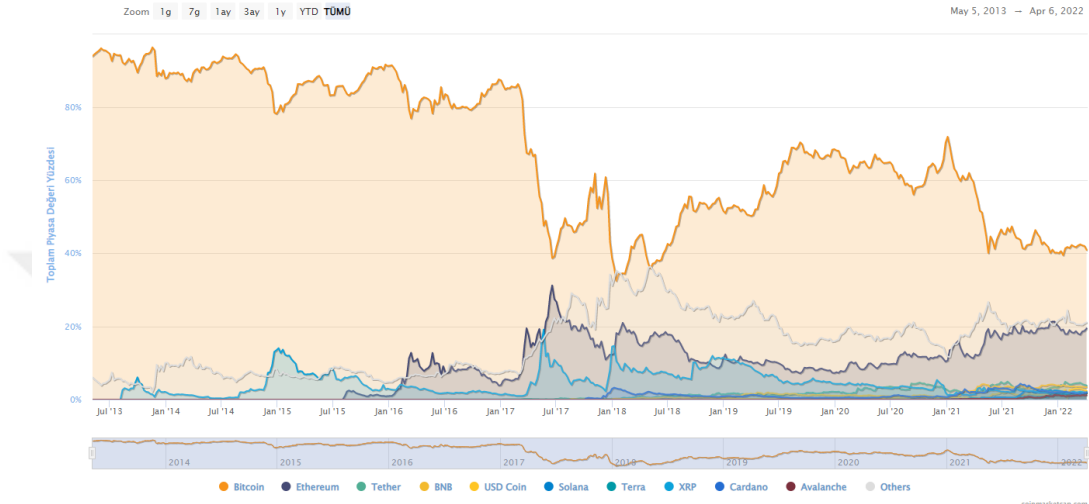


Kaynak: coinmarketcap.com, E.T.: 04.2022

Bitcoin, kripto para piyasasının lideri konumdadır. İlk ve piyasadaki en pahalı kripto para olması sebebiyle popülerliği yüksektir. Ayrıca yatırımcılara daha fazla güven vermektedir. Fakat zamanla farklı kişi ve kişiler tarafından birçok kripto para piyasalara dahil edilmiştir. Kripto paraların bir kısmı Bitcoin veya Ethereum gibi merkeziyetsiz yapıdayken bir kısmı da tersine merkeziyetsiz bir yapıda değildir. Belirli grup veya şirketlerin çıkardığı, farklı amaçlara hizmet eden birçok kripto para çeşitli

piyasalarda işlem görmektedir. 2013 yılından bu yana farklı kripto para birimlerinin borsalardaki sahip oldukları hacimler Şekil 1-2’de görüldüğü gibidir. Piyasalardaki kripto para sayısı ve hacmi artış gösterirken yıllar içerisinde Bitcoin’in piyasa hacminde azalış görünmesine rağmen piyasanın hala %40’ını domine etmektedir (Coinmarketcap.com).

Şekil 1-2: Toplam Piyasa Değeri Yüzdesi



Kaynak: coinmarketcap.com, E.T.: 06.04.2022

Günümüzde yatırımcılar merkezi borsalardan veya bireysel para birimi sahiplerinden kripto para birimlerini satın alabilir veya onlara satabilir. Binance gibi borsalar kripto para birimleri satın almanın veya satmanın yatırımcılar için en kolay yolu olarak görülmektedir. Bitcoin, Ethereum gibi kripto paralar sadece akıllı telefon kullanarak bir dijital cüzdandan bir diğerine kolayca aktarılabilir. Bu paralar mal veya hizmet satın alırken, ticaret yaparken kullanılabilir.

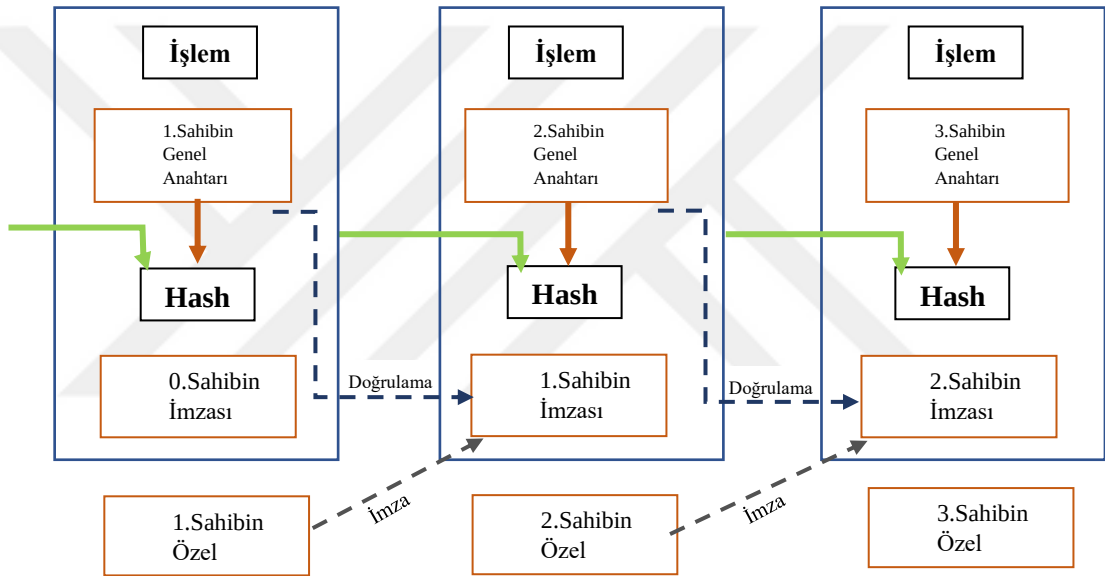
1.2.1. Kripto Paraların Çalışma Prensipleri (Özellikleri)

Kripto paralar hükümet ve merkezi otoriteler tarafından kontrol edilmezler. Farklı isimde ve türde madeni paralar kullanarak bankacılık sisteminin dışında çalışan dijital, şifreli ve merkezi olmayan varlıklardır. ABD Doları gibi değerini yöneten ve koruyan merkezi bir otorite yoktur. Güven yerine kriptografik kanıtlara dayalı bir elektronik ödeme sistemidir (Nakamoto, 2008: 2).

Kripto paralar ‘madencilik’ denilen tamamen dijital olarak yürütülen bir süreç sonunda üretilmektedirler. Kripto para madencilerinin, bu karmaşık süreç karşılığında kripto para çıktısı elde edebilmek için donanımlı bilgisayar sistemleri üzerinde belirli matematiksel algoritmaları çözmeleri gerekmektedir (Vaidhyanathan & Jain, 2022).

Kripto paralar dijital imzalar zinciri olarak tanımlanmaktadır. Her kullanıcı bir önceki işlemin bir özetini ve bir sonraki kullanıcının açık anahtarını dijital olarak imzalayarak ve bunları işlemlerin sonuna ekleyerek parayı bir sonrakine kullanıcıya aktarmaktadır. Her bir kullanıcı mülkiyet zincirini doğrulayabilmektedir. Şekil 1-3’te kripto paraların çalışma düzeni gösterilmektedir.

Şekil 1-3: Kripto Paraların Çalışma Düzeni



Kaynak: Nakamoto, 2008: 2

Sahiplerden birisinin parayı iki kez harcamadığını doğrulayabilmek ve bu durumun önüne geçebilmek adına her işlemi çift harcama için kontrol eden güvenilir bir merkezi otorite sunmak gerekmektedir. Bunun için bir alacaklının önceki sahiplerin işlemleri imzalayıp imzalamadığını bilmesi gerekmektedir. Bunu teyit etmenin tek yolu tüm işlemlere ulaşabilmek ve bu işlemlerden haberdar olmaktır. Bunu darphane veya merkezi bir otorite olmadan güvenilir bir şekilde başarabilmek için işlemlerin kamuya duyurulması ve şeffaf olması gerekmektedir (Nakamoto, 2008: 2).

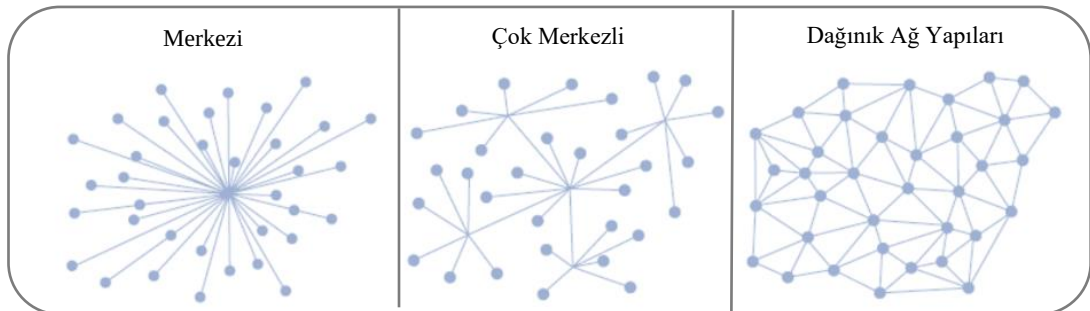
1.2.2. Kripto Para Yapıları

Her kripto paranın arz/talep dengesi ile oluşan maddi bir değeri vardır. Kripto paraların maddi değerlerinin kontrol edilebilir olması amacıyla üretilen kripto para belirli bir miktar ile sınırlandırılmaktadır. Örneğin Bitcoin (BTC) 21 milyon ile sınırlandırılmıştır. Nisan 2022 verilerine göre ise dolaşımdaki arzı 19 milyon civarındadır. Fakat bazı kripto paraların (SHIBA, Dogecoin gibi) üretim miktarları sınırlandırılmamıştır. Bu kripto paralarda öngörü sağlamak daha zor olmaktadır.

Kripto paraların hayatımıza girmesi ile birlikte blok zinciri teknolojisi de çağımızın önemli teknolojilerinden birisi olmuştur. Merkezi olmayan kripto para birimleri, merkezi bir otoriteye başvurmadan işlemleri daha kolay hale getirebilmek için tasarlanmıştır. Bu ademi merkeziyetçi yapı işlemlerin onaylanmasını sağlayan ve kullanıcıların kripto para cüzdanlarını takip etmeyi mümkün hale getiren dağıtılmış defter olan blok zinciri aracılığıyla elde edilmektedir. Dağıtılmış bir defter bir bilgisayarın ağının birden fazla düğümü arasındaki dağıtılmış finansal kayıtlardır. Bir kurum gibi tek bir otorite düğümlerin ve defterlerin işleyişinden tek başına sorumlu değildir. Defterler, dijital para birimleri aracılığıyla gerçekleştirilen tüm işlemlerin sıralı olarak kaydını içermektedir. Kullanıcıların kripto para birimleri ile gerçekleştirilen işlem bakiyeleri örtülü olarak kaydedilmektedir. Bu sayede kullanıcıların daha önce sahip olmadıkları veya kazanmadıkları para birimlerini harcamaları engellenmektedir.

Kullanıcıların kripto para birimleri ile gerçekleştirdikleri işlemlerin verileri farklı yöntemler ile depolanmaktadır. Bunlar Şekil 1-4'te görüldüğü üzere merkezi, çok merkezli ve dağınık ağ yapılarıdır.

Şekil 1-4: Merkezi, Çok Merkezli ve Dağınık Ağ Yapıları



Kaynak: İşler v.d., 2019:76

Merkezi ağ yapısında merkezde bulunan otoritenin verileri silme veya değiştirme yetkisi bulunmaktadır. Merkezde oluşacak hata, yetkiyi kötü kullanma gibi davranışlar tüm kullanıcıları etkilemektedir. Fakat tüm veriler tek merkezde toplandığından yönetilmesi en kolay olan ağ yapısıdır. Çok merkezli ağ yapısında ise birden çok merkez bulunmaktadır. Merkezi ağ yapısında olduğu gibi merkezi bir otorite bulunmamaktadır. Her kullanıcı birbiriyle iletişim kurarak düğümlerdeki bağlantıları yönetebilmektedir. Dağınık ağ yapısında ise her işlem ortak bir hedef doğrultusunda etkileşim sağlamaktadır. Merkezi bir otorite olmadan direkt kullanıcılar birbirleri ile etkileşime girebilmektedir. Kontrol ve güvenlik ağıdaki birimler üzerinde dağıtılmaktadır. Dağınık ağ yapısında sistemde gerçekleşen her işlemin ve kullanıcının verilerinin kopyası bulunmaktadır. Merkezi ve çok merkezli ağ yapılarına kıyasla daha güvenli ve daha az maliyetli bir yapıdır. Kullanıcı işlemlerini gerçekleştirebilmek için üçüncü bir aracıya gerek duymamaktadır. Veri tabanındaki her değişiklik kullanıcıların çoğunluğu tarafından doğrulanmaktadır. Böylece sistem kırılmaz bir zincire dönüşmektedir. Sisteme eklenen her bir zincir işlem sırasına göre oluşturulur ve oluşturulan zincir daha sonradan değiştirilemez (İşler v.d., 2019:76).

1.2.3. Kripto Para Teknolojisi

Kripto para teknolojisi ekonomi literatüründe oldukça yeni bir kavramdır. Bu alanda yapılan ilk çalışmalardan birisi 2012 yılında Avrupa Merkez Bankası'nın sanal para birimleri ile ilgili yayınlanan raporudur. Yatırımcılar için henüz yeni bir kavram olduğundan ilk zamanlar fazla tercih edilmeyen bir yatırım aracıydı. Fakat günümüzde oldukça popüler hale gelmiştir. Bunun en büyük sebebi kripto paraların sahip olduğu teknolojik arka planın güvenilirliğidir.

Kripto paralar, arka planında çalışan teknolojik bir altyapıya sahiptir. Kriptolojik özet fonksiyonu, dijital imza, iş ispatı, eşler arası bağlantı ve akıllı sözleşmeler örnek olarak verilebilir.

1.2.3.1.1. Kriptolojik Özet Fonksiyonlar

Herhangi bir ağ üzerinden veri aktarırken, aktarım süresince verinin bozulmadığından emin olmak gerekmektedir. Kriptolojik özet fonksiyonları bu sorunu çözmeye yardımcı olmaktadır. Her bir girdi için özet (hash) değeri hesaplanmaktadır. Girdideki herhangi bir biti değiştirmek tamamen farklı bir çıktı üretmektedir. Aynı çıktıya sahip iki farklı girdiye karşılık gelen bir girdi bulmak imkansızdır (Clark, 2013).

Bir veri parçasının özet (hash) değeri verildiğinde veriler üzerinde yeniden özetleme (hash) işlemi yapılarak ve özet değerinin aynı çıkmasını sağlayarak verilerin doğruluğunu test etmek mümkündür. Sağlanan çıktı, kullanıcının dijital imzasıdır. Bu işlem sayesinde mesaj matematiksel olarak imzalanmış bir sayı dizisine dönüştürülmüş olur (Kelly, 2015: 79).

Özet fonksiyon işlemleri giriş verilerine müdahale edilmesini de imkansız hale getirmektedir. Kullanıcıların birbirinden haberdar olmadığı için işlemler tarafsız bir şekilde yürütülecek ve kullanıcı giriş verisini değiştiremeyecektir (Clark, 2013).

Örneğin Bitcoin’de kullanılan özet fonksiyonu SHA-256 fonksiyonudur. SHA-256 çıktısı 256 bit veya 32 bayttır. Aşağıdaki ifade fonksiyon çıktısına bir örnektir.

SHA256(‘Bitcoin’)=0xb4056df6691f8dc72e56302ddad345d65fead3ead9299609a826e2344eb63aa4

Çıktının başındaki 0x, sayının onaltılık gösterimde ifade edildiği anlamına gelmektedir. Eğer girdide tek bir harf durumunda değişiklik olursa çıktı tamamen farklı olacaktır. (Clark, 2013)

Kriptografik özet fonksiyonu girdideki herhangi bir değişikliğin ‘digit’ olarak da adlandırılan tamamen farklı bir çıktı oluşturmaktadır. Oluşturulan çıktılara örnek Tablo 1-1’de gösterilmektedir.

Tablo 1-1: Kriptolojik Özet Fonksiyonu

GİRDİ	KRİPTOLOJİK ÖZET FONKSİYONU	ÇIKTI
10 €	SHA-256	46ab27f445d603f5c33f2153f1faabdc9064fc72e503ec4ae9234c96eec651a6

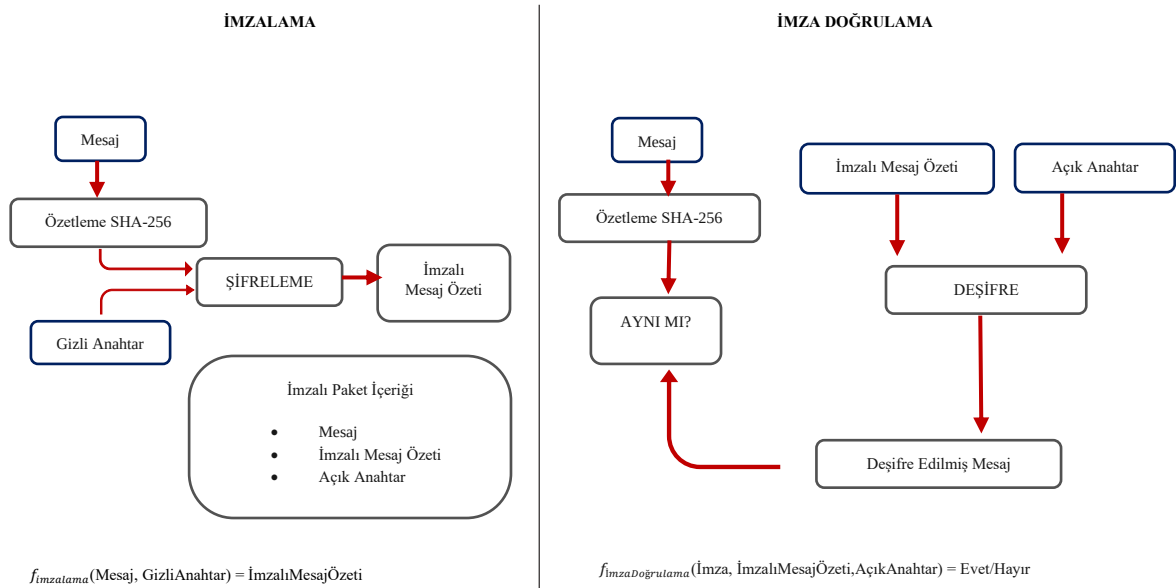
20 €	SHA-256	d84d6c04b78f6f3ba2ab62426dc741e57e917c3 42461a0e54b2c6046f431796a
300 €	SHA-256	bfa7dfc4590dfd62647fca561103e4fee5631e9d0 ff9329ff969e0e2a3146556

Kaynak: Kelly, 2015: 81

1.2.3.2. Dijital İmza

Block zinciri ve kripto para birimleri için ihtiyaç duyulan ikinci kriptografik ilke ise dijital imzadır. Dijital imza, elle kağıt üzerine atılan bir imzanın dijital analogudur. Dijital imza da manuel imzada olduğu gibi eşsizdir ve kişiye özeldir. Dijital imza bir belgeyi veya anlaşmayı onaylayabilmek için kişinin bilgisi dışında kullanılamaması için dijital imza belirli araçlarla belgeye bağlanmaktadır. Dijital imzanın bu özelliği manuel imzanın bir belgeden kopyalanıp başka bir belge için kullanılamamasına benzemektedir (Çarkacıoğlu, 2016: 22). Dijital imzalama süreci şekil 1-5'te gösterilmektedir.

Şekil 1-5: Dijital İmzalama ve İmza Doğrulama



Kaynak: Çarkacıoğlu, 2016: 22

Bir dijital imza planı üç alitmadan oluřmaktadır.

- (sk, pk) : createKeys (keysize) CreateKeys yöntemi bir anahtar hacmi alır ve bir anahtar çifti oluřturur. ‘sk’ özel olarak tutulmaktadır ve mesajları imzalamak için kullanılmaktadır. ‘pk’ ise herkesin ulaşabildiğı genel doğrulama anahtarıdır. Bu anahtara sahip olan herkes dijital imzayı doğrulayabilmektedir.
- Sig : sign(sk, message) Sign yöntemi bir mesaj ve gizli bir anahtar olan ‘sk’yı girdi olarak alır ve ‘sk’ altındaki mesaj için bir imza çıktısı verir.
- isValid : verify(pk, message, sig) Doğrulama yöntemi ise girdi olarak mesaj, imza ve bir ortak anahtar alır. Dijital imza, açık anahtar olan ‘pk’ altındaki ileti için geçerli bir imza ise ‘TRUE’, geçerli değil ise ‘FALSE’ olacak şekilde isValid çıktısı vermektedir.

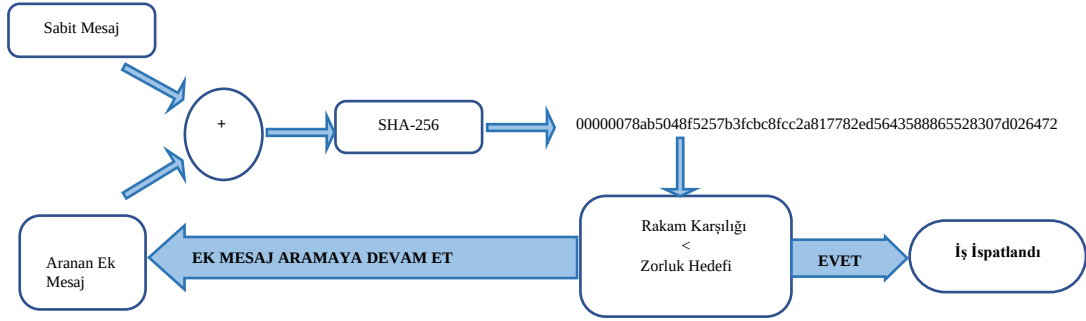
Geçerli bir imza “verify(pk, message, sign(sk, message)) == TRUE” şeklinde doğrulanmalıdır (Narayanan vd., 2006).

1.2.3.3. İş İspatı

Bir işlemin geçerliliğini doğrulamaya yardımcı olacak olan iş ispatı protokolü Adam Back isimli kriptograf tarafından 1997 yılında tanıtılmıştır. Kripto paralar için kullanılmaya başlanmadan önce bir spam azaltma tekniğı olarak kullanılmaktaydı.

İş ispatı, bir bilgisayarın bir iş için çalışıp çalışmamasını ispatlamak amacıyla kullanılan bir yöntemdir. İş ispatı, bir algoritmanın bilgisayarların çözmek için yarıştığı matematiksel bir problem sağladığı bir blok zincirindeki işlemleri doğrulama yöntemi olarak tanımlanmaktadır. Kripto para madenciliğı yapan her bilgisayar, blok olarak adlandırılan her bir işlemin doğrulanmasına yardımcı olan matematiksel bulmacayı çözmektedir. Çözülen bulmaca daha sonrasında blok zincirine eklenmektedir. Bunu başarılı bir şekilde gerçekleştiren ilk bilgisayar kripto para ile ödüllendirilmektedir. Bu bulmacaları çözmeye yarışı çok yüksek miktarda bilgisayar gücü ve elektrik enerjisi gerektirmektedir. Bu kaynakların maliyetleri göz önünde bulundurulduğunda madenciler kazanan kripto para ile bu zorlukları karşılayabilmektedir (Ashford & Schmidt, 2022).

Şekil 1-6: İş İspatı Algoritması



Kaynak: Çarkacıoğlu, 2016: 26

İş ispatı algoritması, Şekil 1-6’da görüldüğü üzere genellikle SHA-256 fonksiyonunu kullanarak çalışmaktadır. SHA-256 fonksiyonu 0 ve 1’den oluşan ve rastgele olan 256 bitlik çıktı üretmektedir. Bu çıktı tam sayı olarak ifade edilebilmektedir. İş ispatı yapılmak istenen bilgisayardan sabit olan mesaja ek mesaj eklenerek SHA-256 fonksiyonu özetlenmektedir. Elde edilen özetin rakam olarak bir karşılığı vardır. Bu rakamın zorluk hedefinden daha büyük olmaması beklenmektedir. Sistem defalarca deneme yaparak aranan ek mesajı bulmaya çalışmaktadır. Yani zorluk hedefinin küçük olması iş ispatı algoritmasını zorlayacaktır. Sonucu, mesaja ek mesajı da ilave ederek özetleme algoritmasından geçirir ve elde edilen 256 bitlik özetin rakam karşılığının zorluk hedefinden küçük olup olmadığına bakarak ispatın doğruluğunu kontrol etmektedir (Çarkacıoğlu, 2016: 26-27).

1.2.3.4. Eşler Arası Bağlantı

Kripto para teknolojisinde ihtiyaç duyulan şey kriptografik kanıtlara dayalı elektronik ödeme sistemidir. Bu elektronik ödeme sistemi, herhangi iki kullanıcının güvenilir bir üçüncü tarafa ihtiyaç duymadan doğrudan birbirleriyle işlem yapmasını sağlamaktadır. İşlemlerin kronolojik sırasının hesaplamalı olarak kanıtını oluşturmak amacıyla eşler arası bağlantı kullanılarak çift harcama soru ortadan kaldırılmaktadır. İki farklı kullanıcının yapmak istediği işlemi eşler arası ağda yayınlaması ile işlemler işlem havuzunda öğrenilebilmektedir. Doğrulanmış işlemler düğümler tarafından çağrılarak ağda kullanılan protokole göre işlemler toplu olarak bir bloğa yazılmaktadır.

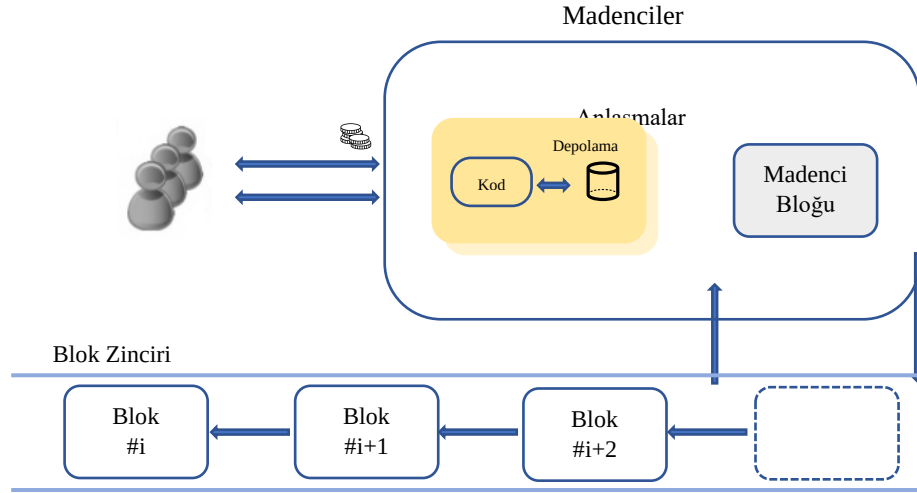
Böylece yeni blok oluşturulmaktadır. Doğrulama için eşler arası bağlantıdaki kullanıcılara yayın yapılır ve doğrulama bilgisinin tamamlandığı bilgisi ağ içerisinde iletilmektedir. Eşler arası bağlantıda konsensüs protokolü kullanılarak bir madenci düğümü seçilmektedir. Seçilen madenci düğümü yeni bloğu blok zincirine ekler ve talep edilen işlemin tamamlanması bilgisi işlemi gerçekleştiren kullanıcılara iletilmektedir (Karaarslan & Akbaş, 2017: 17).

1.2.3.5. Akıllı Sözleşmeler

Akıllı sözleşme, bir sözleşmenin şartlarını yerine getiren bilgisayarlarla işletilen işlem protokolü olarak tanımlanmaktadır. İlk kez 1993 yılında Nick Szabo tarafından tanımlanmıştır. Akıllı sözleşme, blok zincirinin ortaya çıkmasıyla birlikte gelişmiştir. Akıllı sözleşme sayesinde blok zinciri teknolojisi gelişmiş ve yaygınlaşmıştır. Akıllı sözleşmeler, blok zinciri sistemleri üzerinde olaylar zincirinin yürütülmesi amacıyla uzlaşma protokollerini kullanarak çalışan programın parçalarıdır. Akıllı sözleşmeler olay odaklı programlardır. Oluşturulduktan sonra takip etmeye gerek kalmadan otomatik olarak çalışmaktadır. Akıllı sözleşmeler merkezi otoriteden bağımsız ve dağıtılmış yapıdadır. Bu nedenle bir sözleşmenin şartlarının uygulanması ve yürütülmesi üçüncü bir taraf gerektiren geleneksel yöntemlere kıyasla daha düşük maliyetlidir (Tanrıverdi v.d., 2019: 210).

Stark, akıllı sözleşmeye dair literatürde yapılan tanımları ‘akıllı sözleşme kodu’ ve ‘akıllı yasal sözleşme’ olarak iki kategoride sınıflandırmıştır. Akıllı sözleşme kodu, bir blok zincirinde depolanan, doğrulanan ve yürütülen kod olarak tanımlanmaktadır. Akıllı yasal sözleşme ise yasal sözleşmeleri tamamlamak veya değiştirmek için gerekli olan kod anlamına gelmektedir. Akıllı yasal sözleşmenin kapasitesi teknolojiye değil yasalara bağlıdır (J. Stark, 2022).

Şekil 1-7: Akıllı Sözleşmeler Sistemi



Kaynak: Alharby & Moorsel, 2017: 127

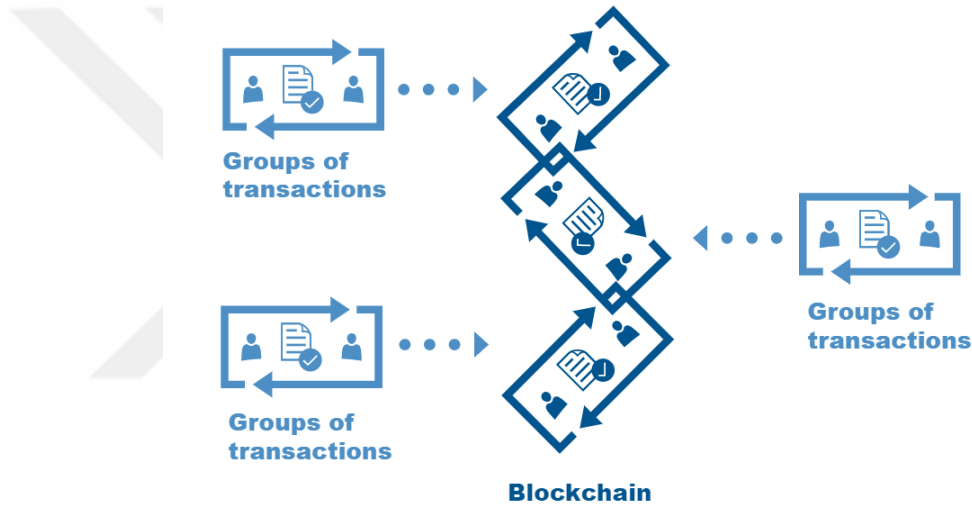
Akıllı sözleşmede hesap bakiyesi, özel depolama ve yürütülebilir kod bulunmaktadır. Sözleşmenin durumu, sözleşmenin depolanması ve dengesini içermektedir. Bu durum blok zincirinde saklanır ve sözleşme her çağırıldığında güncellenmektedir. Şekil 1-7’de akıllı sözleşme sisteminin işleyişi gösterilmektedir. Her sözleşme 20 baytlık eşi olmayan bir adrese atanmaktadır. Sözleşme blok zincirine yerleştirildikten sonra sözleşme kodu değiştirilememektedir. Bir sözleşme yürütmek için kullanıcıların sözleşme adresine bir işlem göndermeleri yeterli olmaktadır. Bu işlem daha sonra çıktı üzerinde fikir birliğine varabilmek adına zincirdeki her bir madenci tarafından yürütülmektedir. Sözleşmenin durumu daha sonra bu mutabakata göre güncellenmektedir. Sözleşme, aldığı işleme göre hesap bakiyesine para depolayabilir, kullanıcılardan mesaj veya para gönderebilir ve hatta yeni sözleşmeler oluşturabilmektedir (Alharby & Moorsel, 2017:128).

1.3. Blok Zinciri

Kripto para açısından değerlendirilecek olursa blok zinciri, erişimi yetkili kullanıcılar arasında dağıtılan ve halka açık bir veri tabanında depolanan dijital defter olarak tanımlanmaktadır. Bu defterin içeriğini üç anahtar bölüm oluşturur. İlk olarak her blok varlıklar ile ilgili tüm satın alma işlemlerinin tutar, saat ve tarih bilgilerini

içermektedir. Ayrıca işlemlere dahil olan kullanıcıların bilgilerini de kaydetmektedir. Blok zinciri, gerçek ad kullanmak yerine bir kullanıcı adı (dijital imza) oluşturularak herhangi bir tanımlayıcı bilgi ifşa edilmeden bilgilerin kaydedilmesine imkan sağlamaktadır. Son olarak her bir blok onu diğer bloklardan ayırt etmeye yardımcı bilgiler sağlamaktadır. Örneğin bir market satışları kaydederken müşterilerin adını her bir işlemi ayırt etmek için kullanır. Ancak blok zincirinde her bir blok “özet fonksiyon” (hash) olarak bilinen benzersiz bir kod kullanmaktadır. Bu sayede her bir işlemi diğerlerinden ayırt edilmesine imkan sağlamaktadır (Torres, J.A., 2021).

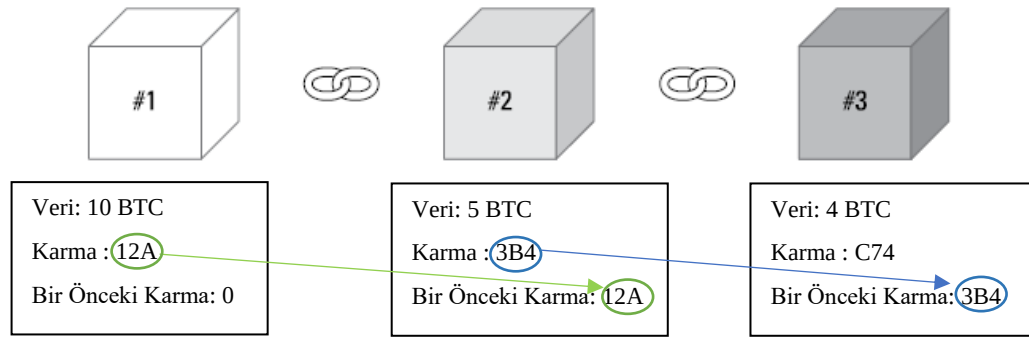
Şekil 1-8: Blok Zinciri



Kaynak: www.ecb.europa.eu

Block zincirinde kullanılan veri türü blok zincirinin ne için kullanıldığı ile ilişkilidir. Örneğin Bitcoin’de bir bloğun verileri gönderici, alıcı, dijital para (coin) sayısı vb. işlemler ile ilgili ayrıntıları içermektedir. Bir blok zincirindeki özet fonksiyonu (hash) ise parmak izi, imza gibi benzersiz bir kimliktir. Şekil 1-9’da görüldüğü üzere her blok kendisinden önceki bloğun bilgisini taşır. Bu sayede zincir çok güvenli hale gelir (Dupont, 2019: 50-51).

Şekil 1-9: Blok Zincirinin Çalışma Prensibi



Kaynak: Dupont, 2019: 52

Block zincirindeki herhangi bir bloğa müdahale etmek neredeyse mümkün değildir. Hashing işlemi yapılarak güvenlik düzeyi sağlanmalıdır. Block zinciri içindeki bir blok üzerinde değişiklik yapmak bloğun özet fonksiyonu (hash) değerinde değişikliğe neden olmaktadır. Bu değişiklik ise tüm blokları geçersiz kılmaktadır (Dupont, 2019: 52).

Merkezileştirilmiş para, Türkiye Cumhuriyeti Merkez Bankası gibi yetkililer tarafından yönetilen normal parayı ifade etmektedir. Fakat kripto para biriminin yükselişini veya düşüşünü denetlemekten sorumlu tutulabilecek bir kurum yoktur. Bu durum merkezi paraya göre bazı avantajlar sunmaktadır. Ağdaki herkes değiştirilemeyecek aynı bilgilere sahip olduğundan para birimi sahiplerinin bir kurum veya yönetime güvenmesine ihtiyaç görülmemektedir. Veriler sadece ağ kullanıcıları tarafından erişilebilir olduğundan sistem büyük ölçüde güvenlidir. Ayrıca kullanıcıların sahip olduğu para miktarı paylaşıldığından tüm kullanıcıların verilerin doğruluğunu onaylamasını daha mümkün kılmaktadır. Sonuç olarak yanlış veri yönetimi ve yanlış iletişim pek mümkün olmamaktadır (Rodeck & Curry, 2022).

Block zinciri teknolojisi günümüzde bankacılık, sigorta sektörlerinde kullanılmaktadır. 2021'den bu yana yıllık %12,8'lik bir bileşik büyüme oranıyla büyüyen kripto para piyasasının, günümüzün ödeme sistemlerinin verimliliğini iyileştirme ihtiyacı, küresel havalelerdeki artış ve artan veri güvenliği ihtiyacı sayesinde 2030 yılına kadar 4,94 milyar dolara ulaşacağı tahmin edilmektedir.

Block zincirine erişim kullanıcılar arasında gerçekleşmektedir. Paylaşılan herhangi bir bilgi şeffaf, anında ve değişmez olduğundan kullanıcıların güvenliğini

kazanmış bir teknolojidir. Bir blok zincirinin en temel parçası güvenlidir. Blok zincirinde verilerin kopyası tüm kullanıcılar tarafından kaydedilmektedir ve tüm kullanıcılar bu verilere erişebilmektedir. Verilerin bu şekilde saklanması sayesinde veri kaybı ve verilerin deforme olması önlenmektedir. Dijital imza ve imzaların doğrulanması sayesinde üçüncü taraflara gerek kalmadan kullanıcılar kendilerini güvende hissetmektedirler. Blok zincirinin dağınık yapısı sayesinde merkezi bir otorite veya üçüncü kişiler tarafından kontrol edilemez veya kapatılamaz. Ayrıca akıllı sözleşmeler sayesinde belirli işlemle otomatikleştirilebilmektedir. Blok zincirinin bu avantajlarının yanı sıra bazı dezavantajları da göze çarpmaktadır. İş ispatı kullanılan blok zincirlerinde çok yüksek oranda enerji tüketilmektedir. Ayrıca maliyeti yüksek bilgisayar sistemleri ile çalışılmaktadır. Bir blok zincirinde veriler her bir düğümde ayrı ayrı saklanmaktadır. Her işlem sonrasında ise düğümlerdeki verilerin tutarlılığı sağlanmaktadır. Zincirdeki her düğümün verilerin kopyasını saklayabilmesi ve içeriğin tamamen erişilebilir olması kullanıcıların mahremiyetine zarar verebilir. Bu nedenle kullanıcılar genellikle takma ad kullanarak gerçek kimliklerini saklamaktadırlar (Tanrıverdi v.d., 2019: 205).

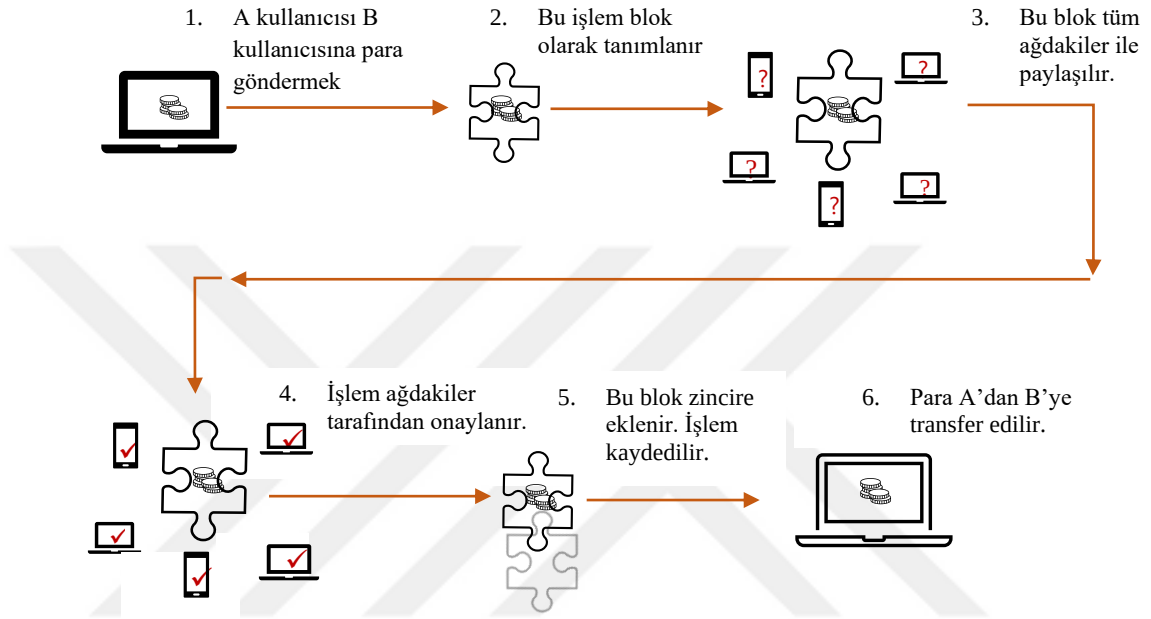
1.3.1. Blok Zinciri Teknolojisi

Blok zinciri, bir bilgisayar kodu zincirinde her bir işlem için işlemin gerçekleştiğinin onaylandığı bir bilgisayar ağıdır. İşlemlerin güvenliğinin sürdürülebilmesi için kriptografi kullanılmaktadır. Güvenliğin sağlanması için katılan maliyet ağıdaki kullanıcılar arasında paylaşılmaktadır. Transferlerin ayrıntıları ağıdaki tüm kullanıcıların görebileceği halka açık deftere kaydedilmektedir. Blok zincirinde bilgi, aracı bir kişi ya da kurum olmaksızın şeffaf olarak paylaşılan bir veri tabanında tutulmaktadır (Wild, 2015).

Kripto paralar blok zinciri teknolojisi içerisinde işlenen dijital veriler aracılığı ile kullanılmaktadır. Bu dijital veriler, kullanıcıların karşılıklı olarak birbirine transfer ettiği maddi değeri karşılamaktadır. Bu işleme üçüncü kişiler dahil olmadan eşler arasında (peer-to-peer) gerçekleşmektedir. İşlemlerin tamamen dijital olarak gerçekleşmesi süreyi kısaltarak hem maliyetleri azaltmakta hem de kullanıcılara kolaylık sağlamaktadır. Blok zinciri teknolojisi Şekil 1-10'da gösterildiği gibi birçok

katılımcının yeni işlemler girişi, yeni işlem bloklarını tanımlanması ve bunların işlem bloklarını zincire eklemesi şeklinde çalışmaktadır.

Şekil 1-10: Blok Zinciri Çalışma Sistemi



Kaynak: Wild vd., 2015)

Blok zinciri sisteminde madenciler zincirde gerçekleşen bütün işlemleri kaydetmektedirler. Blok zinciri yapısı kullanan uygulamalar aracılığı ile iki kullanıcının bilgisayarları arasında yapılacak olan işlem için yeni bir blok oluşturulur ve blok zincirine eklenmektedir. Yukarıdaki Şekil 1-10'da görüldüğü üzere A kullanıcısı yapılacak olan işlemi B kullanıcısını da dahil ederek eşler arası ağda yayınlamaktadır. Ağda kullanılan protokole göre n adet işlem toplu olarak bir bloğa yazılabilmektedir. Düğümler tarafından yeni blok oluşturularak doğrulama için eşler arası ağdaki kullanıcılara yayın yapılmaktadır. Doğrulama işleminin tamamlanmasıyla birlikte bu bilgi ağ içerisinde iletilmektedir. Eşler arası ağda bir madenci düğümü seçilerek yeni blok, blok zincirine eklenmektedir. A kullanıcısı tarafından talep edilen işlemin tamamlanmasının bilgisi işlemi gerçekleştiren makinelere iletilmektedir (Karaaslan & Akbaş, 2017: 18).

1.3.1.1. Kriptoloji

Kriptoloji kelime anlamı olarak şifre bilimi demektir. Çeşitli iletilerin belirli bir sisteme göre şifrelenmesi ve bu mesajların güvenli hale getirilmesi ve verilerin okunup anlaşılmasının imkansız hale getirilmesiyle alıcıya iletilmesi olarak tanımlanabilir. Kriptoloji, kriptografi ve kriptanaliz olarak iki farklı kola ayrılmaktadır.

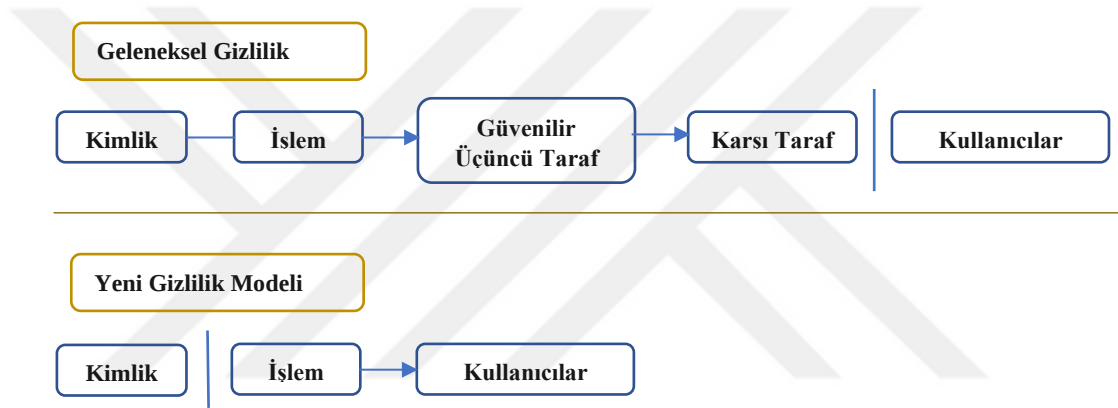
1.3.1.1.1. Kriptografi

Kriptoloji bir bilim dalı iken kriptografi bu bilim dalı içerisinde kullanılan şifreleme yöntemlerinden birisidir. Kriptografi, insanlar tarafından mesajları algılamadan iletmek için binlerce yıldır kullanılmaktadır. Kriptografinin ilk kullanımı M.Ö. 1900 yıllarına dayanmaktadır. İlk olarak Mısır'daki eski bir kralın mezarında örneğine rastlanmıştır. Modern toplumda da her zaman var olmuştur. Kriptografide şifreleme en kritik araçtır. Şifreleme ile bir mesaj istenmeyen bir okuyucu tarafından okunamaz hale getirilmektedir. Bu mesaj yalnızca gönderici ile alıcı arasında anlaşılabilir (Rosic, 2020) Block zinciri teknolojisinin gizlilik ve değişmezlik özellikleri kriptografi aracılığı ile etkinleştirilmektedir. Kriptografi, şifreleme tekniklerini kullanarak verileri yetkisiz erişimden korunmasını sağlayan bir yöntemdir. Modern bankacılıkta da sistemlerin güvenliğini sağlamak ve siber saldırılardan koruyabilmek için kriptografi esastır. Kara para aklamayı önleme hükümlerinin ihlal edilmediğinden emin olmak için güvenilir üçüncü tarafların da şifrelenen verilere belirli erişimlerinin olması gerekmektedir. Kripto paralar gibi dijital ödeme sistemlerinin dolandırıcılığa karşı güçlü bir güvenlik sistemi ile korunması gerekmektedir. Bitcoin bunlar için bazı yöntemleri kullanmaktadır.

Şifreleme, erişimi sınırlandırmak ve kısıtlamak için kullanılmaktadır. Bir dosya şifrelendiğinde yalnızca o dosyaya ait olan anahtar aracılığıyla erişim sağlanabilmektedir. Modern bankacılıkta, sistemlerin güvenliğinin sağlanması amacıyla şifreleme esastır. Kara para aklamayı önleme hükümlerinin ihlal edilmediğinden emin olmak için güvenilir üçüncü tarafların da şifrelenen verilere belirli erişimlerinin olması gerekmektedir. Bitcoin, bireylerin işlem günlüklerini ve alışverişlerini şifrelemek zorunda kalmadan kimlik verilerini tutması ve sürdürmesine

izin verdiği ölçüde farklılaşmaktadır. Bitcoin’de tüm işlemler açık metin olarak değiştirilir ve doğrulanır olarak gerçekleşmektedir. Dijital imza algoritmaları kullanılmasına rağmen hiçbir noktada şifreleme teknikleri kullanılmaz. Bütünlüğü doğrulamak için kullanılabilecek bir dosyaya izin üreterek güvenli bir yöntem sunmaktadır. Dosya bütünlüğü, kullanıcıların hiçbir şeyin değiştirilmediğinden emin olmalarını sağlamaktadır. Bu noktada Bitcoin ve benzer kripto paraların gizlilik ve güvenlik modelleri Şekil 1-11’de görüldüğü gibi geleneksel gizlilik anlayışından farklılaşmaktadır (Wright, 2020).

Şekil 1-11: Geleneksel ve Yeni Gizlilik Modeli



Kaynak: Wright, 2020

1.3.1.1.2. Kriptoanaliz

Şifrelenmiş verilerin okunabilir ve anlaşılabilir hale getirilmesine kriptoanaliz denilmektedir. Kriptolojinin alt dalı olan kriptoanaliz, kriptograflar tarafından şifreli hale getirilen metinlerin analizi ve şifrelerin çözümü ile ilgilenmektedir. Bu işi yapanlara kriptoanalist denilmektedir. Kriptoanaliz, kriptolojide kullanılan geleneksel yöntemlerin aksine modern, yeni ve daha güvenilir bir yöntem olarak kullanılmaktadır. Şifreli metinlerden orijinal metni elde etmek amacıyla kullanılmaktadır.

Düz Metin: MERHABA

Şifreli Metin: ĞKÖÖÖKG

Kriptoanaliz teknikleri şifreli olan bir metinden düz metni ortaya çıkarabilmek için uygulanmaktadır. Kullanılacak olan anahtar bilinmeden şifrelenmiş olan mesaj çözülmeye çalışılmaktadır (Coşkun & Ülker, 2013: 33).

1.4. Kripto Para Avantaj ve Dezavantajları

Bitcoin ile birlikte son yıllarda popüler hale gelen kripto paralar blok zinciri teknolojisi sektöründe yerini almıştır. Blok zinciri teknolojisi güvenli ve tutarlı bir veri tabanı misyonu ile finans sektörü başta olmak üzere birçok alanda kullanılmaya başlanmıştır. Blok zinciri teknolojisinin gelişimiyle sağladığı yenilik ve avantajlar veri sahipleri için önem arz etmektedir. Son yıllarda kullanıcı sayısının artmasıyla ve yaygınlaşmasıyla blok zinciri teknolojisinin ne kadar güvenilir olduğu da araştırmalara konu olmaya devam etmektedir. Blok zinciri sisteminde, verileri kontrol edecek merkezi bir otorite bulunmamaktadır. Geleneksel veri tabanları ise kontrollüdür. Bitcoin sistemi, şeffaflığı en fazla olan kripto para sistemidir. Kullanıcı kimliği hariç tüm işlem bilgisi isteyen herkese ulaşmaktadır. Şeffaflık aranan sistemlerde bu durum bir avantaj olarak görülmektedir. Fakat bankacılık gibi kişisel verilerin korunmasının oldukça önemli olduğu sistemler için dezavantaj olarak görülmektedir. Bazı blok zinciri sistemlerinde bu durumun üstesinden gelebilmek için işlemler şifreleme algoritmaları ile şifrelenmektedir. Fakat şifreleme işlemleri blok zincirinde yüksek işlem gücü ve zaman gerektiren bir yöntemdir. Bu noktada geleneksel veri tabanları bir adım öne çıkabilmektedir. Blok zincirinde gerçekleştirilen işlem düğümleri gizli bir kimliğe sahiptir. İşlemi gerçekleştiren kullanıcı bu sayede gizli kalır. Geleneksel yapılarda ise her kullanıcı belirli bir kimliğe sahiptir. Geleneksel ve yeni yapılardaki bu farklılık bazen avantaj sağlarken bazen de dezavantaj yaratmaktadır (Ünal & Uluyol, 2020: 172).

Kripto para piyasası son yıllarda muazzam bir şekilde gelişmektedir. Kripto para birimleri üzerine yapılan birçok araştırma kripto para ödeme sisteminin sınırlamalarını ve tuzaklarını ortaya çıkararak bunların üstesinden gelmenin yollarını önererek katkıda bulunmaktadır. Kripto para birimlerinin başlıca avantajları anonimlik, mahremiyet, gizlilik ve en önemlisi şeffaflık olduğu söylenebilir. Geleneksel banka ödeme sistemlerinden farklı olarak sadece kendi hesabı hakkındaki bilgiye sahip olmak kripto para teknolojisinin şeffaflık ilkesi sayesinde farklılık

göstermektedir. Kripto para ödeme sisteminde herkes diğer katılımcıların finansal işlemlerini görebilmektedir. Bağımsız bir otorite tarafından desteklenmese de kripto para birimlerini kullanıcıları için kabul edilebilir kılan özellik yüksek düzeydeki şeffaflık olarak görülmektedir (Rotman, 2014: 2).

Bunjaku v.d. gerçekleştirdikleri literatür çalışmasında kripto paraların avantaj ve dezavantajları üzerine yoğunlaşmışlardır.

- Kripto para madenciliği için çevrimiçi bankacılıkta kullanılan algoritmalar kullanılmaktadır. Fakat farklı olarak kripto para madenciliği kullanıcılara ait olan bilgileri açıklamaktadır. Bitcoin (BTC) ağındaki işlemlerin nasıl ve ne zaman gerçekleştiği paylaşılrken alıcısı ve göndericisi hakkında veriler paylaşılmaz.
- Kripto paraların maksimum üretilebilecek miktarı sınırlıdır. Bu düzen siyasi güçler veya kurumlar tarafından değiştirilemeyeceğinden sistemde enflasyonun gelişmesi mümkün olmamaktadır.
- Eşler arası kripto para ağında tüm işlemlerden sorumlu merkezi otorite yoktur. Kullanıcılar tarafından yüklenen tüm cüzdanlar bir kripto para ağının parçasıdır.
- Kripto para piyasası sınırsız işlem imkanı sağlamaktadır. Kullanıcılar başka bir kullanıcıya, herhangi bir miktarda ödeme yapabilmektedir. İşlemler engellenemez.
- Kripto paralar taklit edilemez, kopyalanamaz ve iki kere harcanamamaktadır. Sistemin bu özelliği sayesinde sistemin bütünlüğü garanti altındadır.
- Kripto para kullanılırken bankalara ve benzeri kuruluşlara komisyon ve ücret ödenmemektedir.
- Kripto para ağında merkezi bir otorite olmadığından kripto para madenciliği yapan her kullanıcı sistemin bir üyesidir. Bu durumda merkezi otoritenin kripto para sahipleri için kurallar belirleme ve yürütme yetkisi yoktur.
- Kripto para sisteminin kullanımı oldukça kolaydır. Bir şirketin veya kişinin kripto para cüzdanı oluşturmak için yaklaşık beş dakikaya ihtiyacı vardır.
- Kullanıcıların anonim olarak sistemde bulunması ve sistemin şeffaflığı kullanıcılar için özgür bir yatırım ortamı sunmaktadır (Bunjaku v.d., 2017: 37-38).

Kripto para piyasasının oldukça güçlü avantajları varken bazı dezavantajları da mevcuttur. Kripto para piyasasında çok güçlü dalgalanmalar meydana gelmektedir. Kripto para değerlerinde yüksek seviyede düşüş ve yükselişler yaşanabilmektedir. Bu yüksek dalgalanma nedeniyle orta ve uzun vadede yatırım yapmanın riskleri vardır. Yüksek dalgalanma riski bilgisayar korsanlığı risklerinin ve kurumsal yedekleme eksikliğinin kripto para teknolojisinin geleceği için risk oluşturduğu da düşünülmektedir. Sistem yasal bir otorite tarafından güvence altına alınmadığı için para aklama, terör ve diğer yasadışı faaliyetlere finansman sağlama riski mevcuttur. Ayrıca;

- Ödeme sisteminde işlemlerin geri alınamaması kullanıcılar için dezavantaj olabilmektedir.
- Kripto para birimlerinin hareketlerini izlemek döviz kurlarının takibine göre daha az fikir vermektedir.
- Kripto para birimleri ile alışveriş günümüzde yaygın olarak kullanılmamaktadır (Bondarenko v.d., 2019: 12).

1.5. Kripto Para Örnekleri

Bugünkü hali ile kripto para ilk kez 2008 yılında Satoshi Nakamoto tarafından yayınlanan “Bitcoin: A Peer-to-Peer Electronic Cash System” isimli makale ile birlikte hayatımıza girmiştir. Daha öncesinde de denenmesine rağmen girişimler başarısızlık ile sonuçlanmıştır. Bitcoin kripto para piyasasında lider ve yol gösterici bir görev üstlenmiştir. Zaman içerisinde de bazı değişiklikler ve düzenlemeler yaparak yeni birçok kripto para birimi ortaya çıkmıştır. Bitcoin’e benzetilerek Bitcoin’in alt para birimleri olarak oluşturulan kripto paralar ‘altcoin’ olarak adlandırılmaktadır. Günümüzde takip edilmesi zor olabilecek miktarlarda yeni kripto paralar oluşturulmaktadır. Bireysel kullanıcılar dışında kurum ve kuruluşlar da finans alanındaki çalışmalarında kripto para gibi yeni finansal ürünler kullanılmaya başlanmıştır. 2022 yılı itibariyle farklı piyasalarda işlem gören 9000 civarı kripto para mevcuttur (İşler v.d.,2019: 79).

Şekil 1-12: Kripto Para Örnekleri



Kaynak: İşler v.d., 2019: 79

Son yıllarda popüler olan ve çok kullanılan kripto para birimlerine Bitcoin (BTC), Ethereum (ETH), Ripple (XRP), Litecoin (LTC), BinanceCoin (BNB), Cardano (ADA), Avalanche (AVAX) örnek olarak sıralanabilir.

1.5.1. Bitcoin

Günümüzde kullanılan en yaygın kripto para birimi Bitcoin'dir (BTC). Kripto para piyasasının %40'ını Bitcoin varlıkları oluşturmaktadır. Bitcoin blok zinciri teknolojisinin ilk uygulaması olarak 2009 yılında işlem görmeye başlamıştır. İlk Bitcoin bloğu Nakamoto tarafından oluşturulmuştur. Bitcoin'in kendisinden önce üretilen kripto paralardan farklı olarak başarılı bir proje olmasının en büyük sebebi eşler arası bağlantı (peer-to-peer) ile transfer yapılıyor olmasıdır.

Bitcoin'in üretimi 21.000.000 BTC ile sınırlandırılmıştır. Nisan 2020 tarihi itibarıyla çeşitli piyasalarda yaklaşık 19.000.000 BTC işlem görmektedir. BTC piyasa değeri ise 753.432.531.061\$'dır. Bitcoin birim fiyatı 2016 yılının Mayıs ayında yaklaşık olarak 500\$ iken Nisan 2022 tarihine gelindiğinde 46.000\$'ın üzerine çıkmıştır. Sadece son altı yıl içerisinde Bitcoin %9000 büyümeye göstermiştir. Şekil 1-13'te 2014 yılından bu yana Bitcoin'in birim fiyatındaki meydana gelen değişim Dolar cinsinden gösterilmektedir (çevrimiçi: <https://coinmarketcap.com/tr/currencies/bitcoin/> E.T.: 06.04.2022).

Şekil 1-13: Yıllar İçerisinde Bitcoin'de Gerçekleşen Fiyat Değişimi



Kaynak : <https://coinmarketcap.com/tr/currencies/bitcoin/> E.T.: 06.04.2022

Bitcoin, kripto para piyasasında öncü olmasından dolayı yatırımcılar tarafından kripto para piyasasının rezervi olarak görülmektedir. Bu nedenle altcoinlerin fiyatı da Bitcoin değerine bağlı olarak hareket etmektedir. Bitcoin, iş ispatı mekanizması ile güvence altına alınmıştır. İş ispatı, milyonlarca madenci, merkezi otoriteye bağlı olmayan ağın güvenliğini sağlamak için birlikte çalışmasıdır (çevrimiçi: <https://www.binance.com/tr/price/bitcoin>, E.T.: 04.2022).

1.5.2. Ethereum

Ethereum (ETH), 2015 yılının ortalarında ortaya çıkan merkezi olmayan bir kripto para birimidir. Ethereum, Vitalik Buterin tarafından ‘Ethereum White Paper’ adlı çalışmanın yayınlanması ile hayatımıza girmiştir. Ethereum, 2016 yılından beri Bitcoin’den sonra piyasalarda işlem yapılmaya başlanan ikinci kripto para birimidir. Nisan 2022 verilerine göre Ethereum’un piyasa değeri 354.706.256.880\$’dır. Çeşitli kripto para piyasalarında 120.513.044 ETH işlem görmektedir. Ethereum maksimum arz miktarı ile sınırlandırılmamıştır bir kripto para birimidir (<https://www.binance.com/tr/currencies/ethereum>, E.T.: 10.04.2022). Şekil 1-14’te Ethereum’un yıllar içerisindeki fiyat değişimi gösterilmektedir.

Şekil 1-14: Yıllar İçerisinde Ethereum'da Gerçekleşen Fiyat Değişimi



Kaynak : <https://coinmarketcap.com/tr/currencies/ethereum/> E.T.: 10.04.2022

Bitcoin ve Ethereum her işlemde merkezi olmayan bir ağ ve blok zinciri teknolojisi kullanmaları ile benzerlik göstermektedir. Ancak tasarımları ve kullanımları tamamen farklılık göstermektedir. Bitcoin'in birincil işlevi ödeme para birimi olarak kullanılması iken Ethereum blok zinciri ve iş dünyasına yarar sağlayabilecek çok daha fazla işleve izin verecek şekilde tasarlanmıştır. Ethereum'un amacı bir komut dosyası oluşturarak altcoinler ve blok zinciri protokollerini bir araya getirmek, geliştirmek ve geliştiricilerin standardizasyonunu sağlamaktır. Ethereum, hesap denilen nesnelerin bir araya gelmesiyle oluşmaktadır. Her bir hesabın adresi 20 bayttan oluşmaktadır. Ethereum, her işlemin yalnızca bir kez işlenebilmesi için bir sayaç kullanmaktadır. Ethereum Akıllı sözleşmeler yapabilen bir blok altyapısına sahiptir. Bu sistemde madencilere kendi blok zinciri üzerinde merkeziyetsiz bir uygulama geliştirme imkanı sunmaktadır. Ethereum blok zinciri üzerinde bulunan her akıllı sözleşme, kendi kendini işleyerek otomatik işlem yapan bir bilgisayar programına benzer çalışmaktadır. Blok zinciri, akıllı sözleşme kodlarının bir aksamı, sahtecilik veya herhangi bir müdahale olmadan tam olarak programlandığı gibi çalışmasını sağlamaktadır. Ethereum platformunun kendi dijital para birimi 'Ether' olarak adlandırılmaktadır. Bu para birimi işlem tutarlarını ödemek amacıyla kullanılmaktadır. Özel anahtarlar tarafından kontrol edilebilen harici olarak sahip olunabilen hesaplar ve sözleşme kodları tarafından kontrol edilen sözleşme hesapları olmak üzere iki farklı hesap vardır. Harici olarak sahip olunan hesabın kodu yoktur ve harici olarak sahip olunan bir hesaptan bir işlem oluşturularak ve imzalanarak mesaj

gönderilebilmektedir. Böylece bu hesapların dahili depolama alanına okuyup yazmasına, başka mesajlar gönderebilmesine ve sırasıyla sözleşmeler oluşturmaya olanak sağlamaktadır (Buterin, 2014: 13).

1.5.3. Ripple

Ripple (XRP), 2012 yılında Chris Larsen tarafından ortaya çıkarılmıştır. Ethereum'dan sonra en büyük pazar payına sahip bir altcoindir. Kripto para piyasasındaki yer alan kripto paralara göre düşük bir fiyat seyri izlemiş olmasına rağmen 2017 yılında gösterdiği yükseliş sayesinde kurucusu dünyanın en zengin iş insanları arasına girmiştir. Şekil'de Ripple'ın 2014 yılından bu yana meydana gelen fiyat değişimi görülmektedir. Ripple'ın Nisan 2022 tarihi itibarıyla piyasa değeri 33.784.010.506\$'dır. XRP üretimi 100.000.000.000 XRP Coin ile sınırlandırılmıştır. Kripto para piyasalarında 48.105.234.849 adet XRP Coin işlem görmektedir (çevrimiçi: https://www.binance.com/tr/trade/XRP_USDT, E.T.: 10.04.2022).

Şekil 1-15: Yıllar İçerisinde Ripple'da Gerçekleşen Fiyat Değişimi



Kaynak : <https://coinmarketcap.com/tr/currencies/xrp/> E.T.: 04.2022

Ripple (XRP), RippleNet isimli dijital ödeme platformunda kullanılan bir kripto para birimidir. RippleNet, blok zinciri tabanlı çalışmamaktadır. Onun yerine XRP Ledger isimli bir dağıtık defter ağı veri tabanı üzerinde çalışmaktadır. RippleNet, Ripple isimli bir şirket tarafından yönetilmektedir. Ripple'ın oluşturulma amacı,

bankaların ve müşterilerin gerçekleştirmek istedikleri maddi varlık transferlerini daha kolay bir hale getirmektir. Böylece düşük masraflar ile para transferi gerçekleştirilmesine imkan sağlanmaktadır.



İKİNCİ BÖLÜM

VERİ MADENCİLİĞİ

Teknoloji alanındaki hızlı gelişmeler verinin saklanıp depolanmasını hem kolay hem de ucuz hale getirmiştir. Verilerin depolanma kapasitesi arttıkça çok karmaşık, anlaşılması güç veri setleri ortaya çıkmıştır. Geleneksel veri analiz yöntemleri bu karmaşık verileri analiz etmek için yetersiz kalmıştır. Büyük miktardaki ve karmaşık yapıda olan veri setlerini daha anlamlı hale getirebilmek için birtakım analizlere ihtiyaç duyulmaktadır. Bu veriler işlenmediği sürece anlamsız bir yığın olarak veri tabanlarında depolanır. Veri setlerinden geçerli, anlamlı, anlaşılır örüntüler ve kurallar ortaya çıkarabilmek için çalışılır. Bunun için bazı bilgisayar teknolojileri, istatistik gibi farklı disiplinlerden faydalanılır. Bu süreç veri madenciliği olarak tanımlanır. Veri madenciliğinin hedefi etkin kararlar verilmesini sağlamaktır (Ravindran ve Garg, 2015:4).

Veri madenciliği uygulamaları sağlık, ekonomi, pazarlama, sosyoloji, bankacılık gibi birçok farklı alanda kullanılabilir. Veri madenciliğinin özellikle son yıllarda ticari kullanımı baskın hale gelmektedir. Örneğin bankalar, dolandırıcılığı tespit etmek, kredi kartlarını takip etmek için sıklıkla kullanılır. Ayrıca tüketici ve yaşam tarzı verileri, işletmelerin pazarlama kampanyalarını özelleştirmelerine olanak sağlar. Bu şekilde müşteri profilleri ortaya çıkarılarak ürünlerden etkilenen müşteri alt kümeleri elde edilir. Böylece işletme için karlı satışlara yol açmanın yanı sıra, en fazla kar elde edilebilecek müşterilerin özelliklerini belirleyerek onları hedeflemeye olanak sağlar (Özbay, 2015:263-264). Ekonomik sistemlerin işleyişi, elektronik iletişimde kullanılan e postalar, bloglar, otomasyon sistemleri çok sayıda yapılandırılmış ve yapılandırılmamış veriyi ortaya çıkarır (Voican,2020: 159-160).

Veri madenciliği farklı yöntemler ile uygulanabilir. Kullanılan yöntemler bu bölümde detaylı olarak anlatılmıştır. Bu çalışma kapsamında veri madenciliğinin bir alt dalı olan metin madenciliği gerçekleştirilmiştir. Metin madenciliği kapsamında elde edilen veriler üzerinde bir duygu analizi uygulaması gerçekleştirilmiştir.

2.1. Metin Madenciliği

Metin tabanlı veri setleri üzerinden analizler yapılmak istendiğinde metin madenciliği teknikleri kullanılır. Metin veri madenciliği, akıllı metin analizi, metinde bilgi keşfi olarak da bilinen metin madenciliği, metinsel dökümanlar içerisinde yapılandırılmamış veriyi bilgisayar sistemleri kullanılarak yapılandırılmış, anlaşılır veri haline getirme işlemi olarak tanımlanmaktadır. Metin madenciliğini veri madenciliğinden ayıran en temel özellik yapılandırılmamış veri ile çalışılmasıdır. Metinlerin işlenip yapılandırılmış veri haline getirilebilmesi için veri ön işleme, özellik seçme gibi adımlar uygulanır. Bu işlemler ile yapısal olmayan veri bilgisayar tarafından işlenebilen yapılandırılmış veri haline getirilmektedir. Bu yapısal veriler metin madenciliği uygulamalarında kullanılmaktadır (Feldman & Sanger, 2007, 58).

Hotho, metin madenciliğini verilerdeki geçerli, yeni, potansiyel olarak faydalı ve anlaşılabilir kalıpları tanımlamanın oldukça önemli olan bir işlemi olarak belirtmektedir. Verilerin analizi verilerdeki gizli kalıpları ve bağlantıları bulmayı amaçlamaktadır (Hotho, 2005: 2).

Metin madenciliği, veri madenciliği araçlarının veri tabanlarından yapılandırılmış verileri işlemek üzere tasarlanması dışında veri madenciliğine benzemektedir. Fakat aralarındaki en büyük fark metin madenciliğinde veri tabanının yapılandırılmamış doğal dil metinlerinden çıkarılmış olmasıdır. Bilgi depolamanın en doğal şekli metin olduğundan, bilgi açısından metin madenciliği veri madenciliğinden daha yüksek potansiyele sahip olduğu görülmektedir. Ayrıca metin madenciliği yapılandırılmamış ve bulanık olan dil metinleri ile analizi içerdiğinden veri madenciliğine göre çok daha karmaşık olabilmektedir. Metin madenciliği doğal dil işleme (DDİ), makine öğrenmesi (MÖ), istatistik ve dilbilimi gibi disiplinler arası yöntemlerden yararlanmaktadır. Bu yöntemler sayesinde yapılandırılmamış veriler olan doğal dil metinleri yapılandırılmış veriler haline getirilmektedir. Bu işlemler bilgisayar sistemlerinin yardımıyla gerçekleştirilmektedir (İlhan v.d., 2008: 26).

Doğal dil insanların birbirleriyle iletişim kurmasına ve bilgi kaydetmesine yardımcı olmak için geliştirilmiştir. İnsanlar metinleri dilbilimsel olarak ayırt edebiliyor ve engelleri aşabiliyorken bilgisayar programları doğal dili anlamaktan çok uzaktır. Metin madenciliğinin anahtarı, bir insanın dilbilimsel yeteneklerini bir

bilgisayarın hızı ve doğruluğu ile birleştiren bir teknoloji haline getirebilmektir (Fan v.d., 2006:78).

Metin madenciliği metinlerin sınıflandırılması, kümelenmesi, duygu analizi, metin özetleme gibi amaçlar doğrultusunda uygulanmaktadır. Bu amaçlara ulaşabilmek için işaretleme, filtreleme, gövdeleme, N-gram yöntemleri gibi metin ön işleme tekniklerinden faydalanılmaktadır. Bu teknikler ön işleme başlığı altında anlatılmıştır.

2.1.1. Doğal Dil İşleme

İnsanlar kendi aralarında kurdukları iletişimi zamanla bilgisayarlar ile de sağlamak istemişlerdir. Bunun amacı, insanların birbirleri ile iletişim kurarken sergiledikleri davranışların makine tarafından da yapılmasının istenmesidir. Bu nedenle yapay zeka kuramından yararlanılmıştır. Yapay zekanın bir alanı olan doğal dil işleme bu amaç doğrultusunda geliştirilmiştir (Çabuk v.d., 2003: 679). Doğal dil işleme (DDİ), insan dilinin otomatik analizi ve temsili için teoriye dayalı bir dizi hesaplama tekniğidir. DDİ süreci, bir cümlemin analizinin Google ve buna benzer milyonlarca web sayfasının bir saniyeden daha kısa sürede işlenebildiği bir süreçtir. Doğal dil işleme, bilgisayarların bir konuşma parçası etiketlemeden makine çevirisi ve diyalog sistemlerine kadar her düzeyde çeşitli görevlerin gerçekleştirilmesine olanak sağlamaktadır (Young v.d., 2018: 55).

Yıllardır DDİ problemlerini hedefleyen makine öğrenmesi yaklaşımları çok yüksek boyutlu ve seyrek özellikler üzerine eğitilmiş sığ modellere dayanmaktaydı. Son yıllarda yoğun vektör temsillerine dayanan sinir ağları, çeşitli DDİ görevlerinde üstün sonuçlar elde edilmesinde önemli rol oynamaktadır (Young v.d., 2018: 55).

Doğal dil işleme yöntemi kullanılarak insan benzeri dil işleme gerçekleştirilmek istenmiştir. Bunu başarabilmek amacıyla doğal dil işleme süreci şu şekilde gerçekleştirilmektedir (Liddy, 2001).

- Kullanıcı istediği metni bilgisayara girer.
- Cümlemin Türkçe dilbilgisi kurallarına göre çözümlenmesi yapılır.
- Bilgi erişim sistemleri ile hazırlanmış olan uygulamalar tarafından bilgi çıkarım analizi gerçekleştirilir.

- Elde edilen analizler yorumlanır.

DDİ alanında istatistiksel yöntemlerin kullanılması ile başarılı çalışmalar elde edilmiştir. Sözcük etiketlemesi, sözcük türlerinin sözcüklere atanması ile bu süreç gerçekleştirilmektedir. Sözcük etiketlemesi sayesinde makinenin sözcük ile ifade edileni anlaması kolaylaştırılmaktadır. Fakat DDİ çalışmalarında bazı problemler ile karşılaşılabilir. Bu problemler dilin yapısından dolayı kaynaklanan bazı belirsiz durumlarda ortaya çıkmaktadır. Örneğin ‘yüz’ sözcüğü birden fazla anlam ifade etmektedir. Kullanıldığı cümleye göre isim veya fiil etiketi alabilir. Bu nedenle her sözcüğe etiketleme işlemi yapılamayabilir. Etiket işlemi yapılamayan sözcüğün öncesinde ve sonrasında kullanılan sözcükler bu bağlamda önem kazanmaktadır. Bu kelimelere bakarak sözcüğün doğru etiket sınıfına eklenmesi gerekmektedir (Babüroğlu v.d., 2019).

DDİ sürecinde karşılaşılan diğer zorluklar ise kuralsız ve anlaşılabilir konuşmalar, yazım hataları olarak ifade edilebilmektedir. Bu hataların tespit edilip düzeltilmesi ya da veri setinden çıkarılması analizin sağlıklı bir şekilde sonuçlanmasını sağlayacaktır.

2.1.2. Metin Madenciliği Kullanım Alanları

Kamu ve özel kuruluşları, endüstrilerin ve birçok kurumun hemen her bilgisi elektronik ortamda saklanmaktadır. Bu bilgiler karmaşık, düzensiz bir yapıda olabilir. Günümüzde metin madenciliği, yapılandırılmamış yapıda farklı yerlerde depolanan çok sayıda bilgi yığını olması sebebiyle önemli bir araştırma alanı haline gelmiştir. Var olan bilgilerin neredeyse %80'i gibi büyük bir kısmı yapılandırılmamış metinlerden oluşur. Bu bilgi yığını bilgisayar programları kolayca kullanamaz. Dolayısıyla yapılandırılmamış metinlerden bazı yararlı bilgiler elde edebilmek için birtakım işlemler yapılması gerekmektedir (Dang, 2014: 22).

Metin madenciliği müşteri ilişkileri yönetimi (CRM), sahtekarlık tespiti, bilimsel araştırmalar, medikal araştırmalar, güvenlik ve istihbarat, pazar araştırması, bankacılık, reklamcılık gibi birçok farklı alanda bilgi çıkarımı, tahminde bulunma, sınıflandırma, özetleme, kümeleme, görselleştirme gibi amaçlar doğrultusunda kullanılabilir.

2.1.3. Metin Madenciliği Aşamaları

Metin madenciliği metnin analizi, bilginin çıkarılması, sınıflandırılması, kümeleme, görselleştirme, makine öğrenmesi ile ilgili çok disiplinli bir alandır.

Metin madenciliği uygulaması temelde beş adımda gerçekleşmektedir. Bu adımlar; yapılandırılmamış verilerden bilgi toplama, alınan bu bilgileri yapılandırılmış veriye dönüştürme işlemi olan ön işleme, özellik seçme, sınıflandırma ve sonuçların değerlendirmesi olarak özetlenebilir (Dang, Ahmad,2014:22).

2.1.3.1. Veri Setinin Oluşturulması

Metin madenciliğinin ilk aşaması veri setinin oluşturulmasıdır. Belirlenen araştırma konusuna dair mevcut olan bilgi erişim sistemlerini kullanabilme sürecidir. Veri toplama aşamasında birçok farklı kaynaktan yararlanılabilir. Metin madenciliği uygulamaları için günümüzde çoğunlukla Google, Yahoo, Yandex gibi bilgi erişim sistemleri ve Twitter, Facebook gibi sosyal ağlar en temel kaynaklar olarak kullanılır. Veri toplama sürecinde veriler toplanırken elde edilen verilerin ihtiyaçları karşılayıp karşılayamayacağı değerlendirilmelidir. Veri seti içerisindeki eksiklikler ve yanlış bilgilerin tespit edilip bu sorunun giderilmesi gerekir (Göker, Tekedere, 2017: 293).

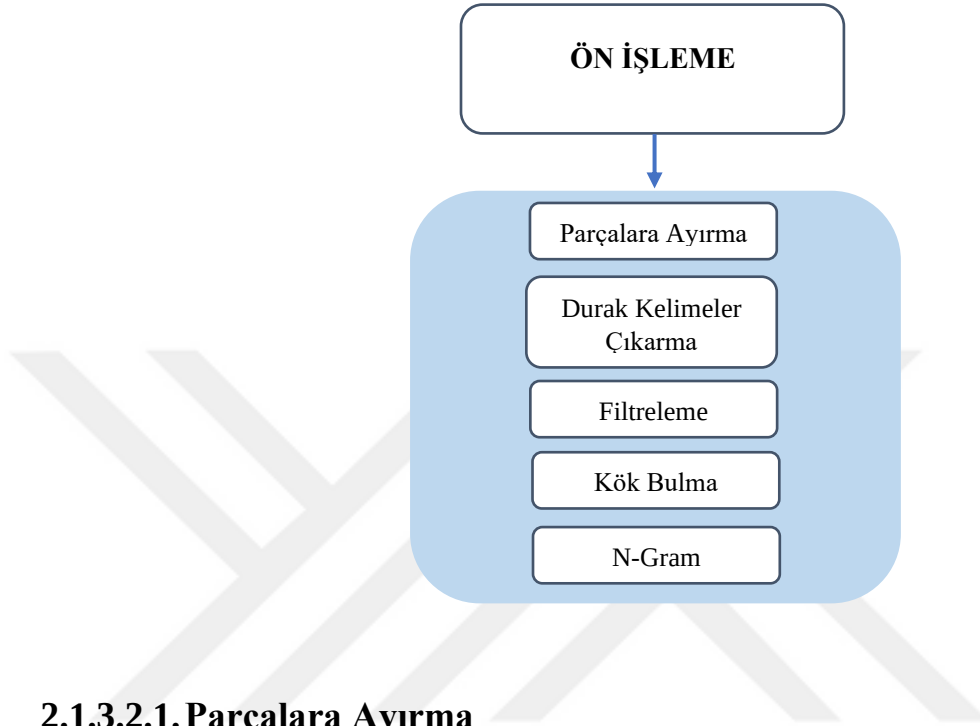
2.1.3.2. Ön İşleme

Ön işleme aşamasında öncelikle metinler simge haline getirilir ve cümlelere bölünür. Çünkü elde edilen veriler yapılandırılmış veri değildir ve metin yığınınından oluşur. Bu veriler karmaşıktır ve gereksiz bilgiler içerir. Bu nedenle veriler üzerinde analiz yapabilmek için birtakım işlemler yapılarak yapılandırılmış veriye dönüştürülmesi gerekir.

Her cümle bir özelliğin oluşum sıklığını ya da ağırlığını özelliğin oluşup oluşmaması şeklinde temsil eden veya sayısal olarak ifade edilen bir özellik vektörü ile temsil edilir. Veri setindeki verinin kelimelere bölünmesi, kelime köklerinin bulunması, gereksiz kelimelerin silinmesi, kelimelerin anlamsal değerlerinin saptanması, imla hatalarının ortadan kaldırılması ve yazım yanlışlarının düzeltilmesi

gibi işlemlerin bütünü metin ön işleme aşamasında gerçekleşir (Kuzucu, 2015:13). Bu işlemler metin ön işleme olarak adlandırılır. Birçok ön işleme tekniği vardır. Bu teknikler Şekil 2-1’de gösterilmektedir.

Şekil 2-1: Metin Ön İşleme Süreci



2.1.3.2.1. Parçalara Ayırma

Parçalara ayırma (tokenization), bir metin akışını ifadelere, kelimelere, simgelere veya belirteç adı verilen diğer anlamlı öğelere bölme işlemidir. Parçalara ayırma işleminin amacı, metin içerisindeki kelimelerin araştırılmasıdır. Bilgiye erişimde veri kümesinin kelimeleri gereklidir. Bu nedenle belgelerin parçalara ayrılmasını (tokenleştirilmesini) işleyen bir ayrıştırıcıya ihtiyaç bulunmaktadır. Bu işlemler yapılırken virgül, tire gibi noktalama işaretleri, boşluk vb. karakterler dikkate alınarak yapılmaktadır (Verma v.d.,2014:16).

2.1.3.2.2. Durak Kelimleri Çıkarma

Durak kelimeleri (Stop Words Removal), içinde bulundukları cümlede önemli bir anlamsal ilişkisi olmayan, bir metinde sıklıkla kullanılan sözcüklerdir. Araştırmacının ihtiyacına uyan belgelerin seçilmesine yardımcı olmada değeri yok denecek kadar az, fakat yaygın kullanılan kelimelerdir. Durak kelimelerin kaldırılması

analizin en önemli süreçlerinden birisidir. Durak kelimeleri olarak herhangi bir kelime grubu seçilebilir. Bu kelime grubunu belirlemek araştırmacının inisiyatifi doğrultusunda gerçekleşir. Genellikle edat, bağlaç veya ek olarak kullanılan “ve”, “veya”, “ya da”, “ancak” gibi kelimeler bu gruba dahil edilir. Durak kelimelerinin çıkarılması terim sayısının azaltılmasına yardımcı olacaktır (Alajmi v.d., 2012:8).

2.1.3.2.3. Filtreleme

Filtreleme, kapsamlı bir veri kaynağı görünümüne dayalı olarak bir araştırma yapısı oluşturmamızı sağlamaktadır. Filtreleme ile eğitim ve test veri setlerinin yalnızca istenen bir bölümünü kullanma imkanı elde edilmektedir. Bu sayede her veri alt kümesi için farklı bir yapı oluşturmaya gerek kalmamaktadır. İçerik, dil, bölge, karakter sayısı gibi farklı filtreler kullanılabilir. Böylece veri seti daha sadeleştirilmiş ve anlaşılır hale getirilebilir (Verma v.d., 2014: 16).

Filtreleme işlemi, “ama, fakat, ve, bu yüzden” gibi bağlaç, edat ve durak kelimelerini tek başına anlamı olmadığından filtreleyerek gereksiz veri kalabalığından kurtulmamızı sağlar. Metinlerde sık sık kullanılan bu kelimelerin analiz aşamasında hiçbir katkısı olmayacaktır. Aksine kelime kalabalığına neden olacaktır. Veri seti bu şekilde yapılandırılarak araştırmaya uygun hale getirilmelidir (Kotu ve Deshpande, 2014:283).

Araştırmanın amacına uygun olarak durak kelimesi olmasa bile çok sık veya çok seyrek olarak kullanılan kelimeler de filtrelenebilir.

2.1.3.2.4. Kök Bulma

Bir kelimenin kökünü bulabilmek için kök bulma (stemming) teknikleri kullanılır. Kök bulma işlemi, kelimelerin yalın hale getirilerek ortak bir kelime dahilinde frekans hesaplamaları yapılabilmesi amacıyla gerçekleştirilir. Kelimeleri dile bağlı dilbilimsel bilgi içeren köklerine dönüştürür. Bunun avantajı toplam terim sayısının ve boyutunun azaltılmış olmasıdır (Srividhya, Anita, 2010:50). Kök bulma işlemi yapılırken dikkat edilmesi gereken iki nokta vardır. İlk olarak bir kelimenin morfolojik formlarının aynı temel anlama sahip olması ve dolayısıyla aynı gövdeye

eşlenmesi, ikincisi ise aynı anlamı olmayan kelimelerin ayrı tutulmasıdır (Jivani, 2011: 1931).

2.1.3.2.5. N-Gram

Uzun metin dizilerinin ön işleme aşamasında kullanılan yöntemlerden bir diğeri N-gram tekniğidir. Bu teknik kelime tabanlı özelliklere ihtiyaç duymadığından küçük eğitim verileri ve kısa test örneklerini kapsayan dil tanımlama işlemleri için uygundur. Karakter tabanlı n-gram modelleri dil tanımlamada yaygın olarak kullanılır (Vatanen v.d., 2010: 3423).

N-gram tekniği, kelimeleri tek tek değil ‘n’ kadar kelimeyi birlikte ele alır. Örneğin “Bitcoin oldukça yükseldi” cümlesi parçalara ayrıldığında “bitcoin”, “oldukça”, “yükseldi” şeklinde parçalanır. 2li tamlama hali ise “bitcoin oldukça”, “oldukça yükseldi” şeklinde olacaktır. Metnin yapısına göre veya yapılan araştırma doğrultusunda kelimelerin ne şekilde ayrılacağına karar verilebilir.

2.1.3.3. Öznitelik Seçme

Metin madenciliğinde kullanılan veri setleri öznitelik seçimi işlemi yapılmadan önce, içerisinde nitelikli verilerin yanında niteliksiz verileri de barındıran çok yüksek boyutlu bir yapıdadır. Öznitelik seçme aşaması aslında çok yüksek boyutlardaki kalabalık veri setinin ayıklanması işlemidir. Bu tekniğin amacı, belirli bir öznitelik vektöründen alakasız ve fazlalık olan öznitelikleri çıkarmaktır. Öznitelik seçme işlemi ile birlikte daha küçük boyutta temsil gücü yüksek alt kümeler oluşturulmaktadır. Bu işlem sonunda yapılandırılmamış veriler yapılandırılmış veriye dönüştürülmektedir. Bu süreç zaman tasarrufu sağlamakla birlikte sınıflandırma işleminin başarılı bir şekilde yapılmasını da sağlamaktadır (Halta v.d., 2015: 420). Fakat hangi özniteliklerin kullanılacağını doğru bir şekilde seçebilmek önem arz etmektedir. Metinsel verilere uygulanabilecek farklı öznitelik seçme yaklaşımları bulunmaktadır. Bunların çoğu bir puanlama şemasına dayanmaktadır. Kategori için lokal bir sözlük oluşturmak için her kategorinin öznitelik kümesi seçilmek istenmektedir. Bunun için ele alınan kategorinin özniteliklerinin tahmin kapasitelerine göre bağımsız olarak

sıralanmaktadır. Bu yaklaşıma göre sıralamanın en üstünde yer alan öznitelik seçilmektedir. Bir niteliğin tahmin kapasitesi, özniteliğin sınıf etiketi ile arasındaki korelasyon ilişkisini gösteren bir değer ile ölçülmektedir. Örneğin n adet dokümanın içerisinde X_j , j niteliğinin X dokümanında bulunup bulunmadığını göstermektedir. Dokümanın etiketini göstermek için bazı sıralama skorları kullanılmaktadır. Özniteliklerin puanı, belge veri kümesindeki terimlerin kalitesini temsil etmektedir. Yüksek puana sahip bir terim veri setiyle ilgili veya önemli olduğu anlamına gelmektedir. Denetimli yaklaşımlarda dönem puanları, sınıf bilgisi olan etiketli eğitim setine dayanmaktadır. Popüler denetimli öznitelik seçimi yaklaşımlarından bazıları bilgi kazancı, karşılıklı bilgi ve ki-kare istatistikleridir (Do v.d., 2006: 62).

En yaygın kullanılan sıralama skoru ise bilgi kazancıdır. Bilgi kazancı, $IG = L_{etiket} - L_j$ formülü ile ifade edilmektedir (Oğuzlar & Kızılkaya, 2019: 37).

Denetimsiz öznitelik seçimi yaklaşımları ise bir veri setindeki terimlerin kalitesini tahmin etmek için sezgisel yöntemlere dayanmaktadır. Metinsel belgelerin bir veri seti için buluşsal yöntemler genellikle veri seti arasındaki terim dağılımına odaklanmaktadır. Popüler denetimsiz öznitelik seçimi yaklaşımları belge frekansını ve terim gücünü içermektedir. Belge frekansı, öznitelik seçimi için basit ama etkili bir ölçüdür. Öznitelik seçimi yaklaşımı, her terim için belge frekansını hesaplamakta ve belge frekansı önceden tanımlanmış bir eşikten daha az olan terimleri kaldırmaktadır. Temel varsayım, sık kullanılan terimlerin sık olmayanlara kıyasla daha önemli olduğu ve veri setini daha doğru bir şekilde temsil ettiği yönündedir (Do v.d., 2006: 62).

2.1.3.4. Duygu Sınıflandırma Yöntemleri

Veri madenciliğinde kullanılan tanımlayıcı ve tahmin edici teknikler vardır. Tahmin edici yöntemlerden olan sınıflandırma yöntemleri literatürde geniş yer kaplamaktadır. Tıp, biyoloji, sigorta, sosyal medya gibi farklı alanlarda tahmin edici yöntem olarak sınıflandırma yöntemleri sıklıkla kullanılır. Sınıflandırma algoritmaları öncelikle geçmişe yönelik verileri baz alarak kendisini eğitir. Eğitimi tamamlanmış olan algoritmalar ise geçmiş veriler ışığında verilerin hangi sınıfa ait olduğuna karar verir. Yani verilerin hangi sınıfa ait olduğuna dair tahmin yapılmış olur (Ayhan, Erdoğan, 2014: 176). Sınıflama işlemi, verilerin eğitilmesi ve modelin test edilmesi

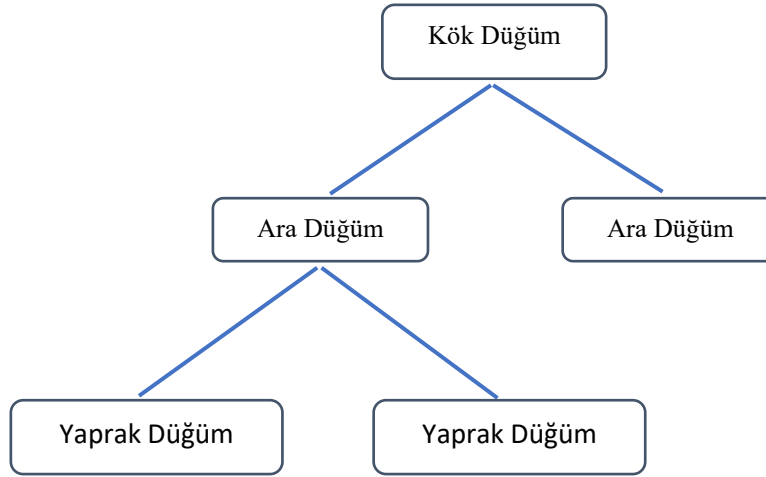
ile gerekleřir. Eđitim kmesi aracılıđıyla model oluřturulur ve test kmesi kullanılarak da modelin dođruluđu kontrol edilir. Sınıflandırma problemlerinin özm için birok makine đrenimi yaklařımı geliřtirilmiřtir. Bařlıca karar ađaları, Naive Bayes, destek vektr makineleri, K en yakın komřu ve yapay sinir ađları olarak sıralanabilir.

Metin madenciliđi uygulamaları eřitli yntemlerden oluřur. Farklı yntemler farklı amalara hizmet eder ve her yntemin avantaj ve dezavantajları vardır (Gupta v.d., 2011: 191). Bu yntemlerden olan sınıflandırma iřlemi en nemli grevlerden biridir. Sınıflandırma, verileri nceden tanımlanmıř hedeflere eřler. Sınıflandırma iřleminin amacı, nesneleri tanımlamak için bazı z niteliklere veya nesnelerin grubunu tanımlamak için bir z niteliđe sahip bazı durumlara dayalı bir sınıflandırıcı oluřturmaktır. Daha sonra sınıflandırıcı, diđer z niteliklerin deđerlerine dayalı olarak veri setindeki her bir verinin grup z niteliklerini tahmin etmek için kullanılır (Krishnaiah v.d., 2013: 41). Metin madenciliđi sınıflandırma grevleri için yaygın olarak kullanılan yntemlerden ařađıda bahsedilmiřtir.

2.1.3.4.1. Karar Ađaları

En ok kullanılan sınıflandırma aralarından birisi olan karar ađalarının basit bir đrenme algoritması vardır. Karar ađaları algoritması birtakım testler gerekleřtirerek en dođru hedef tahminini yapmaya alıřmaktadır. Yapılan her test karar ađacının dallarını oluřturur. Bu dallar sayesinde yeni sınıflar oluřmuř olur. Yapılan testleri ifade eden bu dallar bir yaprak dđmnde sonlanana kadar devam eder. Yani her karar ađacının kk dđm, ara dđm ve yaprak dđm olmak zere  dđm noktası bulunur. Karar ađacı algoritmasında, bulunulan dđmdeki tm nesneler eđer aynı kategoriye ait ise dđmlerinin blnmesi sreci sona erer. Her dalın bir olasılık deđerı vardır. Yaprak dđmne ulařıldıđında her bir dalın olasılık deđerleri ile toplam olasılık bulunur. Karar ađacının yapısı řekil 2-2'de gsterilmektedir.

Şekil 2-2: Karar Ağacı Yapısı



Kaynak: Akpınar, 2017

Karar ağacının oluşturulmasında ağaçtaki dallanmanın hangi kriterlere göre yapılacağı oldukça önemlidir. Tablo 2-1’de gösterilen araba alımı örnek verisinde görüldüğü üzere araba alıp almama kararı yaş, gelir ve kişinin kredi değerlendirilmesi göz önünde bulundurularak yedi kişiye ait veriler incelendiğinde “Eğer 50 yaşından büyük ise ve gelir düzeyi yüksek ise ve kredi değerlendirmesi yüksek ise araba alacak” olduğu görülmektedir. İşletmelerin elde ettiği veri setlerine dayanarak hangi müşteri portföyünün hangi ürünü alabileceği tahmini yapılabilir (Akpınar, 2017: 242-250).

Tablo 2-1: Araba Alımı Örnek Dizisi

Yaş	Gelir	Kredi Değerlendirmesi	Alır - Almaz
<= 30	Yüksek	Yüksek	Alır
<=30	Düşük	Orta	Almaz
>50	Yüksek	Orta	Almaz
>50	Yüksek	Yüksek	Alır
30<x<50	Düşük	Orta	Almaz
>50	Yüksek	Yüksek	Alır
30<x<50	Yüksek	Orta	Alır

Karar ağaçları algoritması, anlaşılması ve yorumlaması kolaydır, güvenilirliği yüksektir. Ayrıca algoritmanın görselleştirilebilmesi anlaşılmasını daha da kolaylaştırmaktadır. Güvenilirlik bazı istatistikî teknikler kullanılarak test edilebilir.

Fakat karar ağaçlarında dikkat edilmesi gereken bazı hususlar vardır. Sayısal veriler karar ağacında kategorik veri olarak değerlendirilmektedir. Verilerin karmaşık hale gelmesi oluşan dal sayısını arttıracaktır. Bu da ağaç modelinin anlaşılmasını ve yorumlanmasını zorlaştıracaktır. Bu nedenle sayısal verileri veri setinin yapısına uygun bir şekilde kategorilere ayırarak ağaç oluşturulur ise daha sağlıklı sonuç elde edilmektedir.

Karar ağacı algoritmasında veri setindeki hangi özelliğe göre dallanmanın yapılacağına entropi kavramını içeren bilgi teorisini kullanarak karar verilir. Entropi, bir sistemdeki belirsizliği ifade eden bir ölçüdür. Belirsizliğin olmadığı bir sistemin entropisi sıfırdır. Belirsizlik arttıkça entropi de doğru orantılı olarak artacaktır. Entropi değeri, her öz niteliğin kategorilerinin gerçekleşme sayıları ile bulunur. Entropi değerini bulmak için kullanılan formül denklem 1’de gösterilmiştir (Zizka v.d., 2020: 175).

$$\text{Entropi (T)} = - \sum_{i=1}^n p_i \log_2(p_i) \quad (1)$$

Entropi hesabında öncelikle özniteliklerin her bir kategorisinin sınıflara göre toplam gerçekleşme sayısına olan oranı bulunur. Her sınıfa ait ilgili kategorinin entropisi hesaplanır ve her kategorinin entropi değeri öznitelikteki orana göre ağırlıklandırılır. Bu değerlerin toplanması ile öznitelik entropi değeri elde edilir.

Her bir sınıfın özniteliği için hesaplanan entropi değeri ile öznitelik için hesaplanan entropi değeri arasındaki fark ile bilgi kazancı bulunur. B özniteliğine göre T sınıf değerleri $T_1, T_2, T_3 \dots T_n$ şeklinde alt kümelere ayrıldığında B öznitelik değerleri kullanılarak T sınıf değerlerinin bölünmesi ile elde edilecek bilgi kazancı denklem 2’de gösterilen formül ile bulunur.

$$\text{Kazanç (B,T)} = \text{Entropi (T)} - \sum_{i=1}^n [|T_i| \div |T|] \text{Entropi (T}_i) \quad (2)$$

T kümesi için B özniteliğinin değerini belirlenmesinde bölünme bilgisi kullanılır. Bunun için aşağıdaki denklem kullanılır.

$$\text{Bölünme Bilgisi (B)} = - \sum_{i=1}^k \frac{|T_i|}{|T|} \log \frac{|T_i|}{|T|} \quad (3)$$

Bu işlemler sonucunda ise kazanç oranı 4. denklemdeki gibi hesaplanmaktadır.

$$\text{Kazanç Oranı} = \text{Kazanç (B,T)} / \text{Bölünme Bilgisi (B)} \quad (4)$$

(Zizka v.d., 2020: 176-177)

Karar ağacının her düğümünde kazanç oranı maksimum olacak şekilde T eğitim kümesinin tekrarlı bir şekilde ayrılması yukarıdaki işlemler gerçekleştirilerek sağlanır. İşlem, yaprak düğümüne kadar devam eder (Kavzoğlu & Çölkesen, 2010: 40-41).

1950’li yıllardan itibaren karar ağacı öğrenimi için birçok algoritma geliştirilmiştir. Bu algoritmaların bir kısmı sınıflandırma, bir kısmı geleceği tahmin amaçlı geliştirilmiştir. Bazı algoritmalar ise her iki görevi de yerine getirebilir. Kök düğümünden yaprak düğümlere doğru ilerleyen karar ağacı homojen düğümler elde ederek bütünü en iyi bölen özniteliği bularak kendini tekrarlayan bir süreçtir.

Geliştirilen her algoritma bir öncekinin eksikliklerini tamamlar niteliktedir. En sık kullanılanlar ID3, C4.5, CHAID ve CART gibi algoritmalarlardır. ID3 algoritmasında entropi ve gini endeksi kullanılarak bölen özniteliğinin seçimi gerçekleştirilir. Daha önce bölen olarak seçilmemiş olan öznitelikleri seçerek yaprak düğümlere doğru ağacın homojenleşmesini sağlar. C4.5 algoritması, ID3 algoritmasının devamı niteliğindedir ve iyileştirilmiş halidir. En iyi bölen öznitelik seçiminde entropi ve enformasyon kazancı tekniklerini kullanır. Özniteliklerin hiçbirinin enformasyon kazancına sahip olmaması ve daha önce olmayan bir sınıfın meydana gelmesi durumunda, sınıfın beklenen değeri kullanılarak bir karar düğüm oluşturulur. ID3 algoritması ile kıyaslandığında hem kategorik hem de sürekli öznitelik değerleri işlenebilir. Ayrıca veri setinde eksik veri bulunuyorsa bu değerler “?” ile işaretlenir ve entropi ve enformasyon kazancı hesaplamalarına dahil edilmez. CHAID algoritması çoğunlukla doğrudan pazarlama için tüketici gruplarının seçiminde kullanılır. Her seviyede ikiden daha fazla dala ayrılabilmesi, tüm ölçü birimleri ile çalışabiliyor olması, eksik verilerin ayrı bir kategoride analiz edilebiliyor olması önemli avantajlarındandır. CHAID algoritması parametrik bir algoritma olmadığından veri seti normal dağılım göstermesede analiz gerçekleştirilebilir. Bağımlı değişkenin sürekli olması durumunda F, kategorik olması durumunda χ^2 testleri kullanılır. CART algoritması her seviyede daha homojen düğümler elde etmek amacıyla kök düğümünden başlayarak her seviyede iki alt diziye ayrılarak büyür. Kategorik ve sürekli değerlerle

çalışabilen CART hem sınıflandırma hem de regresyon ağacı olarak analiz gerçekleştirebilir (Akpınar, 2017: 260-267).

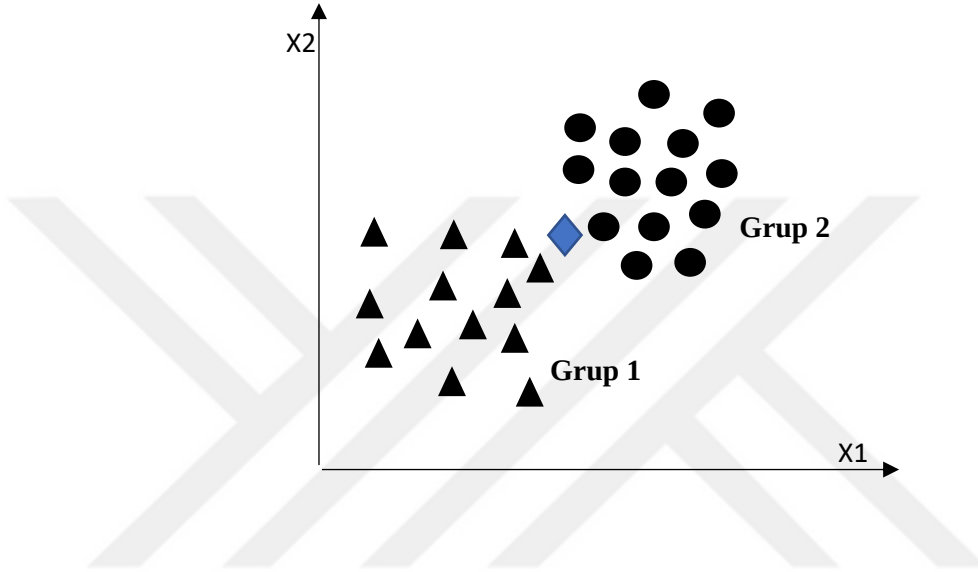
2.1.3.4.2. K En Yakın Komşu

K en yakın komşu (KNN) algoritması makine öğrenmesi yöntemlerinden birisidir. K en yakın komşu algoritması ile sınıflandırılmak istenen bir nesne öznitelik değerlerine göre kendisine en yakın öznitelik değerlerine sahip olan komşu veya komşuların sınıfına atanarak gerçekleştirilmektedir. En yakın k komşu bulunur ve en yakın komşunun özellikleri nesneye atanmış olur. Sınıflandırma işleminin yapılmasında komşu sayıları 'k' harfi ile gösterildiğinden k-en yakın komşu algoritması olarak ifade edilmektedir. KNN algoritması ile sınıflandırma işlemi yapılırken eğitim veri setindeki bir test örneğinin en yakın 'k' komşusu belirlenmektedir. Daha sonra yapılacak olan tahminler de 'k' en yakın komşunun sınıf etiketlerine göre yapılmaktadır (Sun & Huang, 2010; 91). Eğitim veri setindeki her sınıfın örnek dağılımları eşit değildir. Bazı sınıflarda daha fazla örnek bulunabilmektedir. Bu nedenle sınıflandırma işleminin performansı 'k' parametresinin seçimine oldukça duyarlıdır. Sabit bir 'k' değerinin daha büyük sınıflar üzerinde bir yanlılıkla sonuçlanması muhtemeldir. KNN ile ilk olarak n sayıda nesneden 'k' kadar nesne için küme merkezi veya kümenin orta noktası belirlenmektedir. 'k' değeri kadar en yakın komşunun sınıfına sınıflandırma işlemi yapılmaktadır. Kalan diğer nesneler ise kendisine en yakın olan küme merkezine göre sınıflandırılmaktadır. Oluşan her bir sınıf için ortalama değer hesaplanmaktadır. Hesaplanan ortalama değer kümenin merkezi konumundadır. Genellikle KNN algoritması, vektör girdileri olarak gösterilen örnekler arasındaki farklılıkları ölçmek amacıyla basit Öklid uzaklığı metriğini kullanmaktadır. Verilen 'k' değeri için Öklid uzaklık hesaplama formülü ile tüm noktalar arasındaki uzaklıklar tek tek hesaplanmaktadır. Öklid uzaklığı aşağıda verilen formül ile hesaplanmaktadır (Li vd., 2003).

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^n (x_{ir} - x_{jr})^2}$$

KNN algoritması, etiketlenmemiş her bir örneği eğitim veri kümesindeki en yakın komşuları arasında çoğunluk etiketini göz önünde bulundurarak sınıflandırmaktadır. Bu nedenle algoritmanın performansı en yakın komşuları belirlemek için kullanılan mesafe metriğine bağlıdır.

Şekil 2-3: K- En Yakın Komşu Algoritması



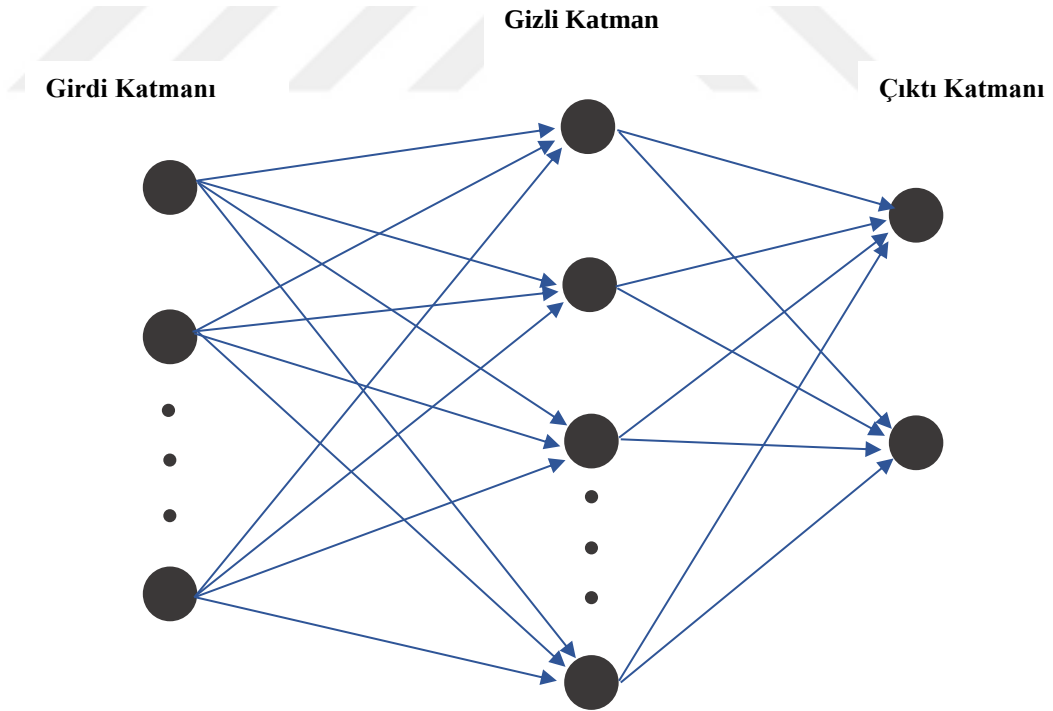
Şekil 2-3'ün yaşam memnuniyetine göre iki gruba sınıflandırılmış bir veri seti olduğu varsayılırsa, k değerine göre $\blacklozenge$ simgesi ile gösterilen nesnenin hangi sınıfa atanabileceği Öklid uzaklığı ile hesaplanabilmektedir. Dikkate alınacak olan komşu sayısı $k=1$ olduğunda en yakın komşu grup 2 olduğundan grup 2 olarak sınıflandırılması söz konusu olacaktır. Komşu sayısı $k=3$ olarak belirlendiğinde ise 2. Gruptan 1, 1. Gruptan 2 gözlem ile komşu olduğundan grup 1 olarak sınıflandırması yapılacaktır. Komşular uzaklaştıkça sınıflandırmaya olan etkilerin azaltmak için komşu etkileri uzaklıklarına göre ağırlıklandırılabilir. Bu işlem için uzaklık (d) ve ağırlık (w) değeri olmak üzere $w = 1/d^2$ eşitliği kullanılmaktadır (Akpınar, 2017; 274).

2.1.3.4.3. Yapay Sinir Ağları

İnsan beyninin örüntü tanıma problemlerini çözme konusundaki etkinliğinden esinlenerek bu yapının matematiksel olarak modellenmesi yapay sinir ağları olarak adlandırılmaktadır. Basitleştirilmiş bir insan sinir ağı modeline dayanan bilgisayar

sistemlerinin istatistiksel sınıflandırma yöntemlerinden daha etkili olup olmayacağı konusundaki merak yapay sinir ağlarının geliştirilmesine katkı sağlamıştır. Yapay sinir ağı modelinin tasarlanmasında biyolojik sinir ağları temel alınmaktadır (Mather & Tso, 2016; 77). Bir yapay sinir ağı modeli Şekil 2-4'te görüldüğü üzere n adet katman ve her katmanda biyolojik sinir hücrelerine benzer şekilde işlevi yerine getiren katmanlar boyunca hesaplama elemanları arasındaki bağlantılardan meydana gelmektedir. Biyolojik bir sinir hücresinde olduğu gibi faaliyet gösteren her bir düğümün birden fazla girdisi bulunmaktadır. Fakat sinir ağında yer alan diğer düğümlere gönderilebilecek bir tek çıktı bulunmaktadır. Şekil 2-4'te görüldüğü üzere bir yapay sinir ağı içerisinde A düğümünden önce işlemde bulunan düğümler kendi çıktılarını A düğümüne girdi olarak göndermektedir. Önceki düğümlerden gönderilen girdilerin her biri A düğümünün bu girdiler için belirlediği ve w_i , w_j ile gösterilen ağırlık düğümleri ile çarpılmaktadır. Elde edilen çarpım sonuçları toplanarak A düğümünün net girdi değeri saptanmaktadır (Akpınar, 2017; 307).

Şekil 2-4: Yapay Sinir Ağı Modeli



Kaynak: Akpınar, 2017; 307

Yapay sinir ağıları, çok büyük veri setlerinin gerçek zamanlı olarak işlenmesine izin veren, yoğun ara bağlantı ve basit işlemci (nöron) düzenlenmesinden ortaya çıkan büyük paralellikleriyle elde edilen yüksek hesaplama hızı sayesinde avantaj sağlamaktadır.

Yapay sinir ağıları parametrik olmayan algoritma olarak tanımlanmaktadır. Yani bir sinir ağının kullanımı verilerin normal dağılması gibi istatistiksel olarak dağılımı hakkında herhangi bir varsayım gerektirmemektedir. Bir sinir ağının performansı istatistiksel dağılımına dair varsayımların yeterli olup olmamasına değil, veri setinin ne ölçüde eğitildiğine bağlıdır. Eğitim aşamasında eğitim verilerinde bulunan düzenlilikleri öğrenir ve öğrenilen düzenliliklere bağlı olarak bilinmeyen veriler için kurallar oluşturulmaktadır. Kullanıcı tarafından sinir ağının mimarisini belirlemeli ve eğitim süresi, performansı, öğrenme hızı gibi parametreler tanımlanmalıdır (Yang & Liu, 1999: 45).

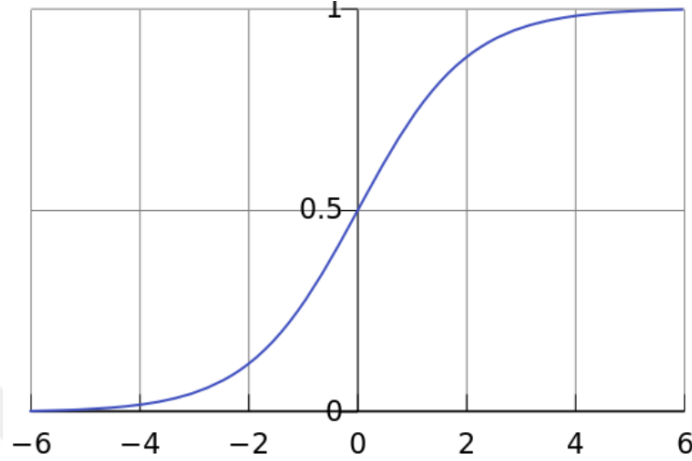
İşletmelerin finansal durumlarının derecelendirilmesi, ülkelerin politik durumlarının derecelendirilmesi, dolandırıcılık tespiti, yüz tanıma, duygu sınıflandırma gibi çeşitli uygulamalarda etkin bir şekilde kullanılmaktadır.

2.1.3.4.4. Lojistik Regresyon

Lojistik regresyon, istatistiksel bir sınıflandırma algoritmasıdır. Bağımlı ve bağımsız değişkenler arasındaki ilişkinin anlaşılmasına yardımcı olan, doğrusal olmayan bir regresyon modelidir. Lojistik regresyon, tahmini değişkenlerin kategorik sonuçlar üzerindeki etkilerinin incelemek için kullanılmaktadır. Lojistik regresyonda bağımlı değişken bir durumun varlığı veya yokluğu gibi nitel, kesikli değerler alabilmektedir. Örneğin sigara kullanıp/kullanmamak, bir hastalığın varlığı/yokluğu gibi ikili (binary) olarak ifade edilmektedir. Binary lojistik regresyon 1 veya 0 olarak kodlanan verileri içermektedir. Tahmin edilmek istenen değişkenler kategorik değişken olduğu durumlarda kullanılmaktadır. Lojistik regresyon algoritmasını doğrusal regresyondan ayıran temel özellik verilerin doğrusal değil logaritmik olarak eğri üzerinde ifade edilebilmesidir (Nick & Campbell, 2007: 274).

Şekil 2-5'te lojistik regresyon eğrisi gösterilmektedir. S şekli 0 ile 1 arasında sınırlandırılmaktadır. Bağımsız değişkende meydana gelebilecek bir birimlik değişimin alt ve üst kısımda etkisi az olmasına rağmen grafiğin ortalarında bu etkinin arttığı görülmektedir.

Şekil 2-5: Lojistik Regresyon Modeli



Kaynak: Nick & Campbell, 2007

Lojistik regresyon analizinde, doğrusal regresyonda olduğu gibi bağımsız değişkenlerin dağılımına ilişkin herhangi bir varsayımda bulunulmamaktadır. Fakat doğrusal regresyon modellerinde olduğu gibi lojistik regresyonda da çoklu doğrusal bağlantı problemi ortaya çıkabilmektedir. Lojistik regresyon da bağımsız değişkenler arasındaki yüksek korelasyona oldukça duyarlıdır. Bu durum ortaya çıktığında bir ya da birden fazla değişkenin modelden çıkarılması önerilmektedir. Lojistik regresyon algoritması kullanılırken veri setindeki bağımlı değişkene ait uç değerler incelenmeli ve gereken düzenlemeler yapılmalıdır. Çünkü lojistik model uç değerlere oldukça duyarlıdır. Ayrıca gözlem sayısı yeterli düzeyde olmalıdır. Gözlem sayısı azaldıkça modelin güvenilirliği de azalmaktadır. Anlamlı sonuçların elde edilebilmesi için yüksek miktarda verilere ihtiyaç duyulmaktadır (Çokluk, 2010: 1368).

2.1.3.4.5. Naive Bayes

Naive Bayes sınıflandırma algoritması, Bayes teoremine dayanan, olayların gerçekleşme sıklığının hesaplanmasına dayanan istatistiksel bir sınıflandırma yöntemidir. Metinsel analizlerde oldukça popüler bir sınıflandırma algoritmasıdır.

Eğitim veri seti küçük boyutta olsada sınıflandırma için gerekli parametreleri tahmin etmek için Naive Bayes algoritması oldukça başarılı sonuçlar verebilmektedir. Sağladığı bu avantaj sayesinde oldukça fazla tercih edilen bir sınıflandırma algoritmasıdır. Naive Bayes algoritması, kelimelerin ve kategorilerin ortak olasılıklarını kullanarak bir metin belgesinde verilen kategorilerin olasılıklarının bulunması temeline dayanmaktadır (Dey v.d., 2016: 2). Daha önce sınıfı belirlenmiş olan verilerin istatistik sonuçlarını kullanarak yeni verilerin hangi sınıfta etiketlenmesi gerektiğinin olasılığını hesaplamak amaçlanmaktadır. Eğitim veri seti ile her bir öznitelik ve sınıf arasındaki ilişkinin olasılık değeri öğrenilerek test veri setindeki herhangi bir sınıf etiketi olmayan veriler için sınıflandırma işlemi yapılmaktadır. Öncelikle sınıfların olasılıkları ve özniteliklerinin tüm durumları için sınıf değerine bağlı koşullu olasılıkları bulunmaktadır. Test edilecek olan veriler ise ilgili olan sınıflar için tüm olasılıklar çarpılmaktadır. Yapılan işlem sonucunda test edilen veri en yüksek sonucu veren sınıfa atanmaktadır.

Tablo 2-2: Naive Bayes Algoritması Örnek Veri Seti

	J	Sınıf Değeri		Koşullu Olasılık	
		Evet	Hayır	P (evet/j)	P(hayır/j)
Eğitim Durumu	Lisans	a	f	$a/(a+b+c)$	$f/(f+g+h)$
	Yüksek Lisans	b	g	$b/(a+b+c)$	$g/(f+g+h)$
	Doktora	c	h	$c/(a+b+c)$	$h/(f+g+h)$
Medeni Durum	Evli	d	ı	$d/(d+e)$	$ı/(ı+k)$
	Bekar	e	k	$e/(d+e)$	$k/(ı+k)$

Naive Bayes sınıflandırma algoritmasında sınıf değerinin evet veya hayır, özniteliklerin ise eğitim durumu ve medeni durum olduğu bir veri seti Tablo3'te gösterilmektedir. Tablo 3 'te gösterilen veri setinde öncelikle P(evet/j) ve P(hayır/j) sütunlarında gösterildiği şekilde sınıflara ait koşullu olasılıklar hesaplanmalıdır. Tabloda gösterilen a, b, c, ... ,k değerleri ise her özneliğin durumunun ilgili sınıfa ait frekansını ifade etmektedir. Bu bilgilere göre test edilen verilerde

- Eğitim durumu “lisans”, medeni durumu “evli”

Şeklinde ifade edilen bir verinin hangi sınıfa ait olduğunu belirlemek için her sınıfa ait koşullu olasılık değerleri çarpılarak karşılaştırılmalıdır. Buna göre test edilecek olan yeni veri;

- $P(\text{evet}/\text{lisans}) * P(\text{evet}/\text{evli}) * P(\text{evet}) > P(\text{hayır}/\text{lisans}) * P(\text{hayır}/\text{evli}) * P(\text{hayır})$
ise evet sınıfına
- $P(\text{evet}/\text{lisans}) * P(\text{evet}/\text{evli}) * P(\text{evet}) < P(\text{hayır}/\text{lisans}) * P(\text{hayır}/\text{evli}) * P(\text{hayır})$
ise hayır sınıfına atanmalıdır.

Naive bayes algoritması, büyük hacimli veri setlerinde oldukça hızlı ve başarılı sonuçlar verdiği için literatürde araştırmacılar tarafından çok fazla tercih edilmiştir. Ayrıca diğer veri madenciliği algoritmalarına göre uygulanması ve anlaşılabilir olması da tercih sebebi olmuştur. Metin belgelerinin sınıflandırılması ve duygu analizi uygulamalarında sıklıkla kullanılan bir algoritmadır (Dean, 2014).

2.1.3.4.6. Destek Vektör Makineleri

Temelleri ilk olarak Vladimir Vapnik ve Alexey Chervonenkis tarafından 1960’lı yıllarda atılan destek vektör makineleri, istatistiksel öğrenme teorisini temel alan, sınıflandırma ve regresyon analizinde kullanılan bir denetimli öğrenme algoritmasıdır. DVM, iki sınıfı birbirinden en uygun şekilde ayırabilen karar fonksiyonunun tahmin edilmesi sürecidir (Ayhan, Erdoğan, 2014: 176). İstatistiksel öğrenme teorisindeki yapısal risk minimizasyonu prensibi ile çalışır. DVM, iki veya daha fazla sınıflı problemlerin çözümü amacıyla optimal hiper düzlemin elde edilmesini, farklı sınıflara ait olan destek vektörleri arasındaki uzaklığın maksimum olmasını sağlar (Yakut v.d., 2014: 143).

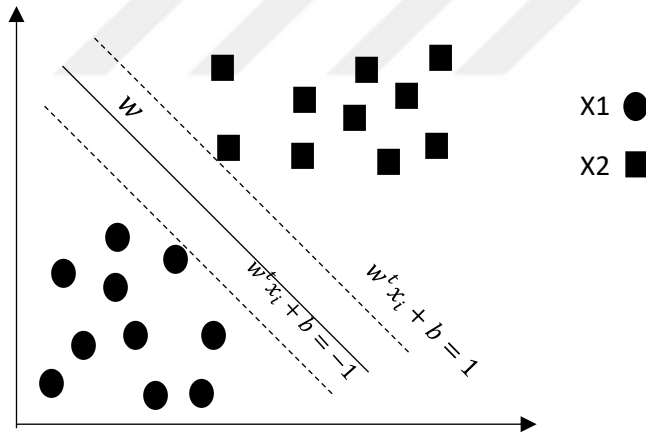
DVM algoritması, bir sınıflandırma probleminde öğrenme sürecinde yapılan işlem sayısını en aza indirgeyerek modelin daha hızlı çözüme ulaşmasını sağlayan kareli optimizasyon yöntemine dönüştürür. Bu avantajının yanı sıra yüksek güvenilirliği, lineer olmayan sınıflandırmadaki başarısı da tercih edilen bir yöntem olmasını sağlamıştır (Ayhan, Erdoğan, 2014: 176).

DVM algoritması, eğitim veri setindeki örneklerin normal dağıldığı ve birbirinden bağımsız olduğu varsayımlarına dayanır. Bu varsayımlar doğrultusunda DVM, lineer ve lineer olmayan sınıflandırma yöntemleri olarak ikiye ayrılabilir. Gerçekte karşılaştığımız problemler genellikle lineer olmayan sınıflandırma ile çözümü gerçekleştirilir.

Lineer olarak ayrılabilen iki boyutlu veri dizisi karar doğrusu ile sınıflandırılır. Ancak veri dizisini ikiye ayırabilen sonsuz sayıda karar doğrusu çizmek mümkündür. Veri setine yeni nesneler ilave edildikçe karar doğrusunun değişikliklere karşı direnç göstermesi ile hangi karar doğrusunun optimal karar doğrusu olduğuna karar verilir. İki sınıfın sınır çizgisine en yakın noktaları destek vektörleri, noktaların sınır çizgisine olan uzaklıkları ise genişlik olarak adlandırılır (Akpınar, 2017: 291).

DVM algoritmasında hedef, optimal karar doğrusunun bulunabilmesi için;

$y_i(w^t x_i + b) \geq 1 \forall i$ kısıtını sağlayacak, $\min \frac{1}{2} w^t w$ amaç fonksiyonunun minimize edilmesidir.



İki boyutlu bir veri seti için amaç fonksiyonu ve kısıtı şu şekilde yazılabilir;

$$\text{Amaç fonksiyonu: } Z_{\min} = (w_1 \cdot w_1 + w_2 \cdot w_2) / 2$$

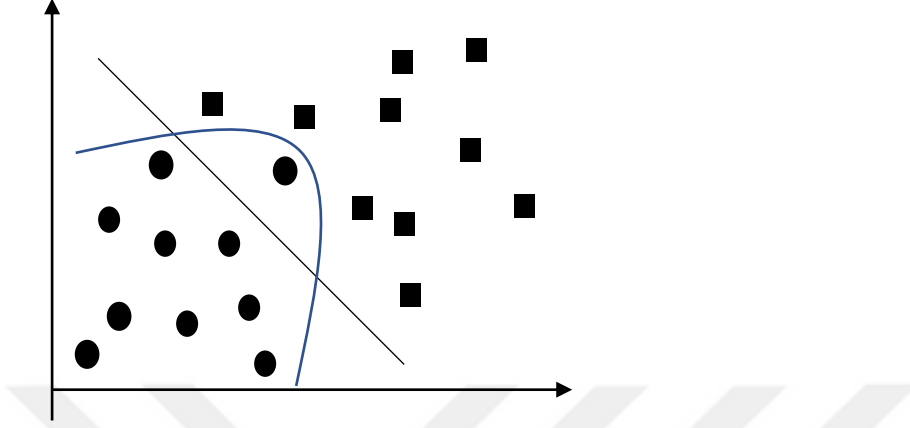
$$\text{Kısıt: } y_i(w_1 \cdot x_1 + w_2 \cdot x_2 + b)$$

Karar doğrusu $w_1 \cdot x_1 + w_2 \cdot x_2 + b = 0$ eşitliğini,

$y_1 = +1$ karar doğrusu eşleniği $w_1 \cdot x_1 + w_2 \cdot x_2 + b > 0$ eşitliğini,

$y_2 = -1$ karar doğrusu eşleniği $w_1 \cdot x_1 + w_2 \cdot x_2 + b < 0$ eşitliğini sağlamalıdır.

Lineer ayrılabilirlik durumunda veri setleri doğrusal bir düzlem aracılığıyla iki sınıfa ayrılabilir. Fakat lineer ayrılmayan durumlarda veriler birbirinden doğrusal bir düzlem ile ayrılamaz.



Veriler doğrusal olarak ayrılamadığı durumlarda lineer olmayan sınıflandırma algoritması kullanılır. Bu bağlamda lineer olmayan özellik uzayı $x \in R^n$ gözlem vektörünü daha yüksek dereceden bir uzayda z vektörüne dönüştürür. Ancak oluşturulan yeni uzay ile doğrusal sınıflandırıcıları elde etmek mümkündür. z vektörünün yer aldığı özellik uzayı F ile gösterilsin. Bu durumda ϕ ifadesi $R^n \rightarrow R^F$ eşlemesini yapmak için $z = \phi(x)$ şeklinde ifade edilir. (Yakut v.d., 2014: 144)

2.1.3.5. Sınıflandırma Algoritmalarının Karşılaştırılması

Sınıflandırma çalışmalarında eğitim veri setinin özelliklerine bağlı bir model oluşturulmaktadır. Oluşturulan bu model test verisi üzerinde sınıflandırma işlemi yapılması amacıyla kullanılmaktadır. Sınıflandırma modelinin performansını ölçebilmek için etiketlenmiş veri setinin bir kısmı rastgele ayrılarak test verisi olarak kullanılmaktadır. Eğitim verisi ile oluşturulmuş olan sınıflandırma modeli test veri setine uygulanmaktadır. Sınıflandırma algoritmaları karşılaştırılırken Tablo 2-3'te gösterilen kontenjans tablosundan hareketle hesaplanan kriterlerden yararlanılmaktadır.

Tablo 2-3: Kontenjans Tablosu

Tahmini Değerler \ Gerçek Değerler	Pozitif	Negatif
Pozitif	Doğru Pozitif (TP)	Yanlış Pozitif (FP)
Negatif	Yanlış Negatif (FN)	Doğru Negatif (TN)

Sınıflandırma işlemi sonucunda performans değerlendirme amaçlı kullanılan bazı değerler hesaplanmaktadır. Bu değerler doğruluk, kesinlik, duyarlılık ve F ölçütü değerleridir (Perikos & Hatzilygeroudis, 2016; 195).

Doğruluk: Doğruluk ölçütü, sınıflandırıcının tüm örnekleri sınıflandırma konusunda ne kadar başarılı olduğunu göstermektedir. Doğru olarak bilinen pozitif ve negatif değerlerin tüm veriye oranlanması ile elde edilmektedir.

$$doğruluk = \frac{TP + TN}{TP + TN + FP + FN}$$

Kesinlik: Kesinlik kavramı tutarlılık olarak da ifade edilmektedir. Sınıflandırıcının bir sınıf için yaptığı sınıflandırmada ne kadar hassas olduğunu göstermektedir.

$$kesinlik = \frac{TP}{TP + FP}$$

Duyarlılık: Duyarlılık kavramı bir sınıfta bulunan örneklerin ne kadarının doğru sınıflandırıldığını göstermektedir. Pozitif olarak etiketlenen örneklerin gerçekte pozitif olan örneklerin toplamına oranı olarak ifade edilmektedir.

$$duyarlılık = \frac{TP}{TP + FN}$$

F-Ölçütü: Kesinlik ve duyarlılık ölçütleri tek başına yeterli ölçütler değildir. Hem kesinlik hem de duyarlılık ölçütlerini bir arada kullanıp sistemi kesinlik ve duyarlılık ölçütlerine göre optimize eden F ölçütü kullanılmaktadır.

$$F - \text{Ölçütü} = 2 \frac{\text{Kesinlik} \times \text{Duyarlılık}}{\text{Kesinlik} + \text{Duyarlılık}}$$

2.3. Duygu Analizi

İnsanlar genellikle ürün ve hizmetler, kuruluşlar, bireyler, etkinlikler hakkında fikirlerini tartışırlar. Duygu analizi, insanların bu varlıklar ile ilgili ve yazılı metinlerdeki ifadelerle yönelik fikirlerini, değerlendirmelerini, tutumlarını analiz edebilen bir çalışma alanıdır. Duygu analizinin amacı, yapılandırılmış ve işlenmeye hazır bilgi oluşturmak için öznel bilgileri çıkarabilen otomatik araçları tanımlamaktır. Yani yapılandırılmamış metni yapılandırılmış verilere dönüştürür. Daha sonra ise veri tabanı yönetim sistemleri ve analitik işleme araçları yardımıyla nitel ve nicel analizler gerçekleştirilir (Liu, 2020).

Veri madenciliği ve web madenciliği alanlarında çalışan birçok araştırmacı metin verileri ile ilgilendiğinden duygu analizi geniş çapta üzerinde çalışılan bir alan olmuştur. Genel olarak olumlu ve olumsuz duyguları ifade eden, günlük dilde olumlu ve olumsuz olarak adlandırılan görüşlere odaklanır. Literatürde duygu analizi, fikir madenciliği, fikir analizi, duygu madenciliği, etki analizi gibi pek çok farklı kavram kullanılmıştır. Duygu analizi ve fikir madenciliği olarak adlandırılıp adlandırılmaması konusunda bazı karışıklıklar ve tartışmalar yaşanmıştır. Merriam- Webster sözlüğünde duygu, hislerin yol açtığı bir tutum, düşünce veya yargı olarak tanımlanır. Fikir ise bir konu ile ilgili zihnimizde oluşan bir görüş, yargı veya değerlendirme olarak tanımlanır. Örneğin, “Ekonominin mevcut durumu hakkında endişeliyim.” cümlesi duyguyu ifade

ederken “Ekonominin iyi gitmediğini düşünüyorum.” cümlesi bir fikri ifade etmektedir. Yani aradaki fark çok incedir. Sonuçta duygu analizi kavramı daha çok araştırmacı tarafından kabul görmüştür. Duygu analizi kavramını ilk olarak Nasukawa ve Yi 2003 yılında yaptıkları web sayfaları ve haber metinlerine üzerine gerçekleştirdikleri bir çalışmada kullanmışlardır (Pozzi v.d., 2016: 2).

Yaklaşık son yirmi yılda internet ve sosyal medyanın hızla büyümesi ve geniş kitleler tarafından kullanılmasıyla dijital formlarda kaydedilmiş bir fikir akışına ulaşılabilir. Web’deki bu hızlı büyüme ile web madenciliği ve duygu analizi çalışmaları da hızla gelişme göstermektedir. Web’deki sosyal medya uygulamaları insanların katılımının sağlanması için oldukça kullanışlı ve etkili platformlardır.

Sosyal medya temel olarak insanların duygu, düşünce ve fikirlerini ifade etmeleri için kullandıkları bir araç olduğundan duygu analizi, sosyal medyanın araştırma ve uygulamasının merkezindedir. Sosyal medya içeriğinden çıkarmak istenilen bilgi insanların ne hakkında konuştuğu ve ne düşündüğüdür. Duygu analizi de bu amaç doğrultusunda gelişmektedir. Sosyal medyadaki metin yığınından anlamlı bilgileri çıkarmak ve bu bilgileri kullanmak için duygu analizi gerekli bir alan haline gelmiştir. Bu işlemi gerçekleştirebilmek için de duygu analizi, metin madenciliği ve doğal dil işleme tekniklerini kullanmaktadır (Oğuzlar & Kızılkaya, 2019: 58).

Fikirler, duygu ve düşünceler öznel yapıda olan cümleler iken olguları ifade eden cümleler nesnel cümlelerdir. Duygu analizi öznel cümlelerin yanı sıra nesnel cümleleri de inceler. Örneğin “Dün X markasından bir fincan aldım ama bugün kırıldı.” cümlesi nesnel bir cümledir ve istenmeyen bir gerçeği ifade etmektedir. X markası ile ilgili olumsuz bir ifade içeren bu gibi nesnel cümleler de duygu analizinin inceleme alanını oluşturmaktadır.

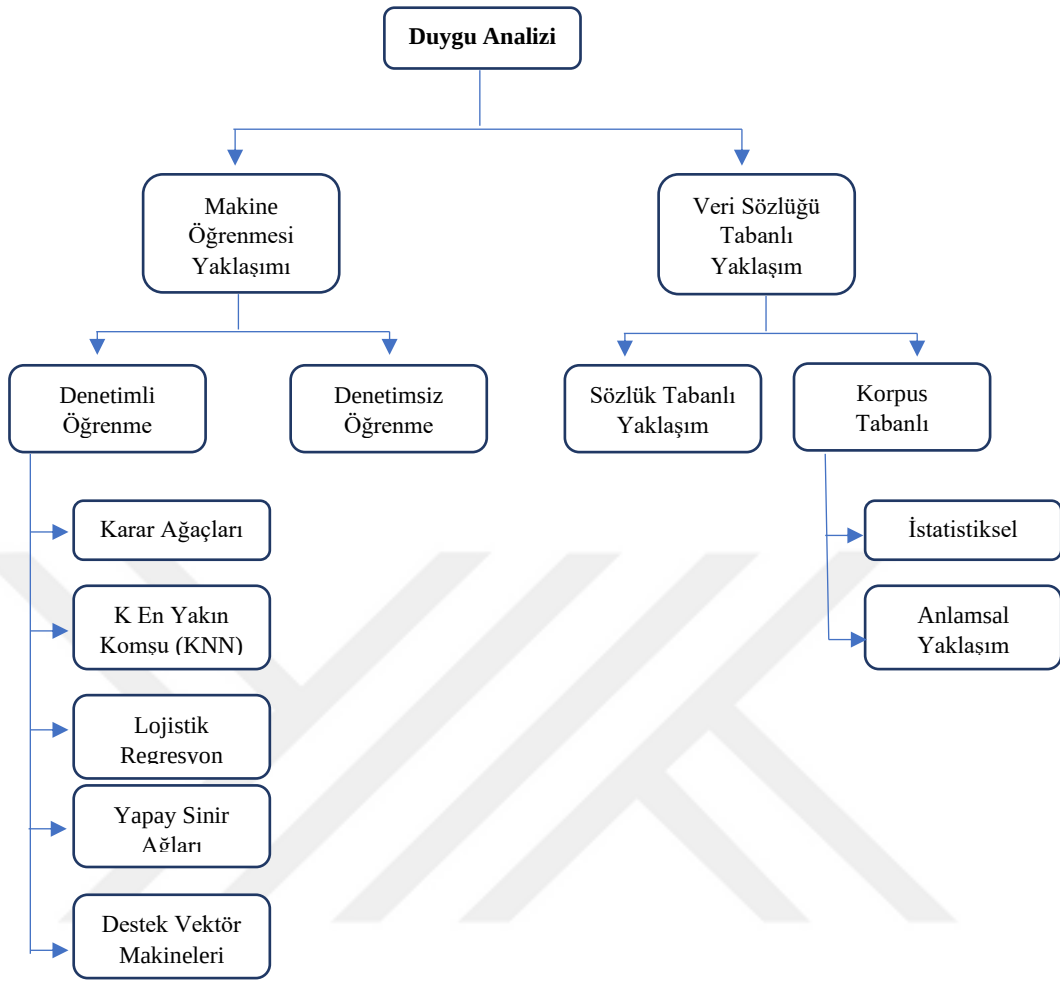
Duygu analizi bilgisayar bilimi sayesinde gelişse de son yıllarda işletmeler, toplum, sanat vb. için de önemi nedeniyle yönetim ve sosyal bilimlerde de kendisine önemli bir yer edinmektedir. Böylece yalnızca doğal dil işleme alanına katkı sağlamakla kalmayıp yönetim bilimi, siyaset bilimi, ekonomi gibi alanlardaki araştırmaların da ilerlemesine katkıda bulunmaktadır (Liu, 2020).

Sosyal bilimlerde duygu analizi işletmeler ve toplum için hemen her alanda anlamlı bilgilere ulaşmak için başvurulan bir yöntemdir.

Günümüzde Facebook, Twitter, Instagram gibi sosyal ağ platformlarının çok fazla kullanılmaya başlanmıştır. Sosyal medya aracılığı ile insanlar belirli bir konu ile ilgili duygularını, görüşlerini ve tutumlarını ifade etmektedirler. Sosyal medya platformlarında aktif kullanıcılar arasında olan yoğun bağlantı sayesinde insanları ortak hedeflere bağlayan, daha büyük kitlelere ulaşabilen, motive edebilen bir tartışma alanı oluşturmuştur. Bununla birlikte duygu analizi daha da değerli hale gelmektedir. Sosyal medya platformlarından elde edilen veriler doğru bir şekilde toplanır ve analiz edilirse birçok karmaşık sosyal olgu anlaşılabilir hale gelebilir ve aynı zamanda doğru tahminlerde bulunulmasına olanak sağlanabilmektedir.

Duygu analizi yöntemleri Şekil 2-6'da görüldüğü üzere genellikle makine öğrenmesi temelli yaklaşımlar ve sözlük tabanlı yaklaşımlar olmak üzere kategorize edilmektedir. Makine öğrenmesi yaklaşımı daha büyük veriler ile çalışırken veri sözlüğü tabanlı yaklaşımlar küçük veri setleri için kullanımı daha uygundur (Medhat v.d., 2014; 1095).

Şekil 2-6: Duygu Analizi Yöntemleri



Kaynak: Medhat v.d., 2014; 1095

2.3.1. Duygu Analizi İçin Makine Öğrenmesi Yaklaşımı

Makine öğrenmesi yaklaşımı denetimli öğrenme ve denetimsiz öğrenme tekniklerini kapsamaktadır. Denetimsiz öğrenme sınıflandırma ve regresyon tekniklerini kullanarak önceden bilinmeyen ilişkileri modellemek amacıyla kullanılmaktadır. Bu öğrenme yöntemi, veri setindeki benzer örnekleri tespit ederek veriler ile ilgili bilgi sağlamakla birlikte verilerin dağılımının modellenmesi için kullanılmaktadır (Zhao & Liu, 2007; 1151). Ancak öğrenme süreçleri denetimli öğrenmeye göre çok daha yavaş olabilmektedir.

Denetimli öğrenme ise veri setlerindeki kelimeleri girdi olarak kullanmaktadır. Veriler elde edilip ön işleme aşamalarından geçtikten sonra verinin bir kısmı eğitim

veri seti olarak bir kısmı da modeli test edebilmek amacıyla test kümesi olarak belirlenmektedir. (Akay, 2018: 46) Test kümesini kullanarak kurulan model kelime çantası (Bags of Words) tekniği ile karar ağaçları, Naive Bayes, destek vektör makineleri gibi sınıflandırma algoritmalarını kullanarak duyguları sınıflandırabilmektedir. Fakat kelime çantası genellikle büyük bir özellik kümesine yol açıp aşırı uyuma neden olduğundan bu sorunu hafifletmek için çeşitli öznelik seçim yöntemleri kullanılmaktadır (Deng v.d., 2017: 66-67).

2.3.2. Duygu Analizi için Sözlük Temelli Yaklaşım

Sözlük tabanlı duygu analizi tekniklerinde öncelikle kavram düzeyinde analize odaklanılmaktadır. Kavramlar, “kar”, “boğa sezonu”, “yükseliş trendi” gibi tek veya birden çok kelimeli kavramsal öğeler ile oluşmaktadır. İlk adım olarak kavramlar, cümlelerin söz dizimsel yapısına uygun örüntü tabanlı yöntemler ile kavram çıkarma işlemi gerçekleştirilmektedir. Kavram çıkarma adımından sonra cümlelerin duygu değerini elde edebilmek amacıyla sözlük tabanlı yaklaşımlar kullanılmaktadır.

Sözlük tabanlı yaklaşım ile karşılaşılan bazı zorluklar vardır. Örneğin karmaşık cümle yapısı, cümlelerin hem olumlu hem olumsuz ifadeleri içermesi gibi karmaşık durumlar ile karşılaşılabilmektedir. “Sizin sevdiğiniz filmi ben sevmiyorum.” cümlesinde “sevmek” ve “sevmemek” eylemleri birlikte kullanılmıştır. Fakat cümlelerin anlamına baktığımızda olumsuz bir anlam ifade etmektedir. Çok büyük ölçeklerde işlenmesi mümkün olmayan ya da çok zaman alacak olan metinlerde sözlük tabanlı yaklaşımlar yerine makine öğrenmesi temelli yaklaşımların kullanılması daha sağlıklı olacaktır. Web tabanlı kaynaklar çok fazla yapılandırılmamış ve gürültülü veri içermektedir. Metinlerin ve tweetlerin içinde kısaltılmış kelimeler ile duygu analizi yapılsa bile anlamlı ya da doğru sonuçlar vermeyebilir. Diğer bir örnek ise cümlelerin içinde kullanılan ironi, metafor içerek kelime ve kelime gruplarıdır. İroni yapılmış bir cümlelerin tespit edilmesi oldukça güçtür (Poria v.d., 2018: 4-5).

2.3.3. Sosyal Medya Verileri ile Duygu Analizi

Sosyal medya platformları, insanların herhangi bir konu ile ilgili görüşlerini coğrafi ve mekânsal sınırların ötesinde özgürce ve rahatça dile getirmelerine olanak sağlıyor. Kullanıcılar başkaları ile kolayca bağlantı kurabiliyor ve bilgilerini paylaşabiliyor. Bu sayede kullanıcılar tarafından birkaç saniye içinde binlerce mesaj oluşabilir. Bu mesajlar artık bir bilgi kaynağı haline geldi. Örneğin insanların sosyal ağlarda kullandıkları ürünler ile ilgili ifade ettikleri olumlu ya da olumsuz duygular, şirketlerin ürünleri hakkında oluşan tutum ile ilgili bilgi edinmek için kullandıkları bir araç haline geldi. Oluşan genel duygu durumlarını yorumlayabilmek için ise öncelikle duygu analizi teknolojilerinden faydalanılır.

Sosyal ağlarda kullanıcıların kullandığı dilin sürekli değişimi ve gelişimi dikkate alındığında duygu analizinin karşılaştığı zorlukları da tahmin etmek güç değildir. Sosyal ağlarda iletişim kurarken kullandığımız dil resmi yazı diline göre daha esnek ve informal yapıdadır. Bu dil değişimine adapte olmak ve sağlıklı analizler yapabilmek için güçlü doğal dil işleme ve dilbilim yeteneğine ihtiyacımız vardır (Agarwal v.d., 2011: 30).

Sosyal medya metinleri genellikle kısa ve heterojen metinleri içerir. İçeriğin kısa ve düzensiz yapısı analiz aşamasında karşımıza bazı zorluklar çıkarabilir. Twitter gibi mikroblog verileri ile yapılan duygu analizi literatüründe iki ana araştırma yönü olduğunu söyleyebilir. Bunlardan ilki, Twitter takipçi grafiklerinde duygu etiketi gerçekleştirmek ve kullanıcı düzeyinde duyarlılık analizi gerçekleştirmek. Bir diğeri ise ifadelerin, hashtag'lerin, karakter tekrarlarının eğitilmiş modele duygu tanımlaması için eklenecek özellik kümelerini belirlemeye odaklanmıştır (Saif v.d., 2012: 508-509).

ÜÇÜNCÜ BÖLÜM

UYGULAMA

3.1. Araştırmanın Amacı

Kripto para birimleri, internet aracılığıyla gerçekleştirilen finansal işlemlerin güvence altına alınması, ek kripto para birimlerinin oluşturulmasını kontrol edebilmek ve varlık transfer işlemlerinin doğrulanabilmesi için güçlü bir kriptografi alt yapısına sahiptir. Kripto paralar, farklı bilgisayar ağları tarafından uygulanabilen dağınık ağ yapısına sahip blok zinciri teknolojisi üzerine kuruludur. İlk merkezi olmayan kripto para birimi olan Bitcoin, 2009 yılında açık kaynaklı yazılım olarak piyasaya sürülmüştür. Günümüzde birçok farklı kripto para var olmasına rağmen Bitcoin hala en yüksek piyasa değerine sahip olan kripto paradır.

Alışılmış yatırım araçlarının aksine, kripto paraların gelecek değerini tahmin etmek daha zordur. Fakat özellikle son yıllarda yatırımcılar için oldukça tercih edilir hale gelmiştir. Kripto paraların piyasa hacimlerine duyarlılığı oldukça yüksektir. Bu nedenle zaman zaman kripto para değerlerinin çok fazla yükseliş trendine girdiği görülse de bazı dönemlerde tam tersi yönde azalış trendine girmektedir. Piyasadaki aşırı yükselişlerin yaşandığı döneme boğa sezonu (bullish market), önemli ölçüde düşüşlerin olduğu döneme ise ayı sezonu (bear market) denilmektedir. Boğa piyasalar sonsuza kadar süremeyeceğinden bir noktada yatırımcının güveni azalabilir veya son yıllarda yaşanan Covid-19 pandemisi gibi öngörülemez durumlardan dolayı aşağı yönlü bir fiyat hareketi gerçekleşebilmektedir. Bu gibi durumlarda yatırımcı daha fazla kaybı önlemek adına satış yapmayı tercih etmektedir. Ayı sezonu bu şekilde başlayabilir. Ayı piyasasında arz talepten daha fazla olduğu ve piyasaya güvenin düşük olduğu dönem olarak tanımlanabilmektedir. Bu sürecin ne zaman sona ereceği ve en düşük fiyata ne zaman ulaşıldığını tahmin etmek oldukça güçtür. Tekrar yukarı yönlü hareketin ekonomik büyüme, yatırımcı psikolojisi, dünya gündemi, siyasi sebepler gibi birçok dış etkenden etkilenmektedir. Bu nedenle öngörülemez bir süreç olarak görülmektedir.

Kripto para piyasasında yüksek miktarda düşüşlerin ve yükselişlerin olduğu dönemlerde yatırımcıların davranışlarını, piyasaya bakış açılarını incelemek amacıyla

bir sosyal medya analizi yapılmıştır. Bu çalışmada Türk yatırımcıların kripto para piyasasına bakış açısı ve en çok hangi sezonlarda yatırım yapma isteği olduğu belirlenmek istenmiştir. Ayrıca merkezi olmayan yapıya sahip kripto paraların aksine bazı kurum ve kuruluşlar kendi kripto para birimlerini çıkarmaya başlamışlardır. Çalışmanın bir diğer amacı da piyasaya yeni bir kripto para birimi sürülmesinin yatırımcılar tarafından ilgi görüp görmeyeceğini öngörebilmektir. Yatırımcıların her koşulda kripto para piyasasında yatırım fırsatlarını değerlendirip değerlendirmeyeceklerini öngörebilmek adına iki farklı dönem seçilmiştir. Kripto paraların yükseliş ya da düşüş trendinde olmasının yatırımcının piyasaya olan bakış açısının değişip değişmediği test edilmek istenmiştir.

Çalışma kapsamında kripto para piyasasını domine eden ve diğer kripto paralarında durumunu etkileyen Bitcoin esas alınmıştır. Bitcoin değerindeki artış ve azalışa göre yatırımcının davranışı test edilmiştir. Bu bağlamda yatırımcıların Twitter paylaşımları baz alınmıştır.

3.2. Araştırmanın Örnekleme

Bir sosyal medya platformu olan Twitter aracılığıyla kullanıcılar duygularını, görüşlerini diğer kullanıcılar ile paylaşabilmektedir. Twitter ile ilgili bazı terimler kullanılmaktadır. Kullanıcıların paylaştığı metinlere ‘tweet’ denilmektedir. Tweetler toplamda 280 karakter içermektedir. Kullanıcıların paylaştığı tweetler başka kullanıcılar tarafından tekrar tweetlenebilir. Başka bir kullanıcıya ait tweetin kaynak gösterilerek bir başkası tarafından paylaşmasına retweet (rt) denilmektedir. Paylaşılan tweetlerin altında bulunan kalp simgesine tıklandığında ise tweet beğeni listesine eklenmektedir. Kullanıcılar bir konu hakkında paylaşılan içeriklerin kolayca bulunabilmesi için tweetlerini paylaşırken hashtag (#) kullanmaktadır.

Yatırımcıların Twitter istatistikleri ile ilgili analizleri gerçekleştirmek için örneklem dahilinde seçilen kripto para birimlerinden Bitcoin’in en yüksek ve en düşük değere sahip olduğu iki hafta seçilmiştir. Boğa sezonu olarak adlandırılan 2021 yılı 5-12 Kasım tarih aralığı ve ayı sezonu olarak adlandırılan 2022 yılı 4-11 Mayıs tarih aralığında #bitcoin ve #btc hashtagi ile paylaşılan tweetler ele alınmıştır. Filtreleme işlemi yapılırken sadece Türkiye lokasyonundan Türkçe tweetler elde edilmiştir. İki

farklı dönemin seçilmesinin amacı, farklı koşullarda yatırımcının piyasaya bakış açısını ortaya koyabilmektir.

3.3. Araştırmanın Kısıtları

Uygulama kapsamında R Studio aracılığı ile Twitter sosyal medya platformundan veriler elde edilmiştir. Twitter API aracılığı ile elde edilebilecek veri setinin bazı kısıtları bulunmaktadır. Bu kısıtlar Twitter tarafından uygulanmaktadır. Twitter API ile her bir istekte en fazla 18000 adet tweet elde edilebilmektedir. Ayrıca Twitter sadece son yedi günlük süre içerisinde veri çekişmesine izin vermektedir. Yedi gün öncesine ait tweetler elde edilememektedir. Bu nedenle daha fazla veri çekilememiştir. Veri çekilen her bir gün #btc için 18000 ve #bitcoin için 18000 adet tweet elde edilmiştir. Bazı tweetlerde her iki hashtag (#) kullanıldığından aynı tweetler veri setinden çıkarılmıştır. Ayrıca sadece Türkiye’den Türkçe paylaşılan tweetlerin kullanılması da araştırmanın bir diğer kısıtını oluşturmaktadır.

Kripto para son dönemlerde fazlasıyla popüler olduğundan ve sosyal medyada çok fazla konuşulan bir konu olduğundan her bilgisayarın işlemcisi tüm verileri elde edebilmek için yeterli olmayabilir ya da verileri elde etmek uzun sürebilir. Bu da çalışmanın zaman ve maliyet açısından kısıtını ifade etmektedir.

3.4. Veri Toplama

Duygu analizi yapılması amacıyla toplanan tweetler, Twitter hesabı ile bağlantılı olan Twitter API aracılığıyla toplanmıştır. Twitter’dan veri elde edebilmek için birçok paket program kullanılmaktadır. Duygu analizinde yaygın bir şekilde kullanılan, istatistiksel veri analizi ve veri görselleştirmede oldukça başarılı olan R Studio 4.2.0 verilerin toplanması aşamasında da kullanılmıştır. Çalışma süresince Bitcoin fiyat trendinin negatif yönlü olduğu 05 Mayıs – 11 Mayıs 2022 tarih aralığında toplamda 257.786 adet veri toplanmıştır. Bitcoin fiyatının en yüksek olduğu dönem olan 06 Kasım – 12 Kasım 2021 tarih aralığında ise 265.432 adet veri toplanmıştır. Kasım ayına ait oluşturulan 7 günlük veri seti ‘veril’, Mayıs ayına ait oluşturulan 7

günlük veri seti ise ‘veri2’ olarak adlandırılmıştır. Çalışmanın geri kalan kısmında elde edilen veri setleri ‘veri1’ ve ‘veri2’ olarak ifade edilecektir.

Tweetlerin çekilmesi aşamasında ve işlenmesi aşamasında karşılaşılan en önemli kısıt bilgisayarın kapasitesidir. R programı bilgisayarın RAM olarak adlandırılan belleğini kullanmaktadır. Bu nedenle kullanılan bilgisayarın RAM kapasitesi verinin toplanması ve işlenecek olan verinin hacmini belirlemekte oldukça belirleyicidir.

3.4.1. Aplikasyon Oluşturma

Sosyal medya madenciliğinin ilk aşaması olan verilerin elde edilmesi Twitter API desteği ile R Studio 4.2.0 aracı kullanılarak gerçekleştirilmiştir. Twitter sosyal medya platformundan veri elde edebilmek için Twitter hesabına sahip olmak gerekmektedir. Twitter hesabı oluşturduktan sonra <https://developer.twitter.com> adresinden “Create New App” butonu tıklandıktan sonra açılan sayfada istenilen detaylar doldurulur. Tüm detaylar doldurulduktan sonra yaklaşık 24 saat içerisinde Twitter’dan geri dönüş sağlanmaktadır. Eğer girilen bilgiler yetersiz veya eksik ise Twitter Developer hesabı tarafından E-mail ile geri dönüş sağlanmaktadır. Eksik olan bilgiler gelen E-mail’e cevap olarak iletildiğinde onay maili yine 24 saat içerisinde gelmektedir. Onay maili geldikten sonra Twitter Developer hesabı oluşturulmuş olur. Developer Portal üzerinden giriş yapıldığında “Keys and Token” tıklanır, “API Key and Secret” ve “Access Token and Secret” tıklanarak API ile veri elde edebilmek için gerekli aplikasyon anahtarları elde edilmiş olur.

Twitter, bilgilere erişim sağlamak için Açık Kimlik Doğrulama (OAuth) özelliğini kullanmaktadır. Twitter’dan yetki alabilmek için öncelikle ROAuth paketi *install.packages(ROAuth)* komutu ile kurulmalıdır. Daha sonra *library(ROAuth)* komutu ile kütüphaneden paket çağırılır. Aplikasyon oluştururken elde edilen API Key, API Secret, Access Token ve Access Token Secret anahtarları tanımlandıktan sonra *setup_twitter_oauth()* işlevi kullanılarak kimlik doğrulama işlemi gerçekleştirilmektedir. Kimlik doğrulama işleminden sonra *searchTwitter()* komutu ile istenilen anahtar kelime, elde edilmek istenen veri sayısı, dil ve tarih gibi bilgiler girilerek veri çekme işlemi gerçekleştirilmektedir. Twitter API her bir istekte 18000

adet veri elde etmeye izin vermektedir. Bu nedenle her bir tarih için 18000 adet ile sınırlandırılmıştır.

Resim 3-1: Veri Çekebilmek İçin Gerekli R Kodları

```
53 options(RCurlOptions = list(cainfo= system.file("CurSSL", "cacert.pem", package = "RCurl1")))  
54  
55  
56 api_key <- "  
57 api_secret <- "  
58 access_token <- "  
59 access_token_secret <- "  
60  
61  
62 setup_twitter_oauth(api_key, api_secret, access_token, access_token_secret)  
63  
64  
65  
66  
67  
68 btc <- searchTwitter("#bitcoin",n= 18000, lang= "tr", since = '2021-11-07' , until = '2021-11-08')  
69  
70 class(btc)  
71  
72 df_btc <- twListToDF(btc)  
73  
74 view(df_btc)  
75
```

3.5. Araştırmanın Yöntemi ve Araçlar

Bu tez çalışmasında kripto para yatırımcılarının Twitter kullanımına ait değerlendirmeler sunulmuştur. Kripto para piyasalarındaki dalgalanmalardan yatırımcıların ne derecede etkilendikleri, endişe, umut gibi duygu durumları ele alınmıştır. Bu kapsamda yapay sinir ağları algoritması kullanılarak duygu analizi gerçekleştirilmiştir. Duygu analizi gerçekleştirmek için R Studio paket programı kullanılmıştır.

Duygu analizini gerçekleştirebilmek adına R Studio programında gerekli olan paketler yüklenmelidir. Bu paketlerden bazıları *twitteR*, *tm*, *wordcloud*, *ggplot2*, *RColorBrewer*, *string*, *plyr*, *SnowballC* paketleridir. Bu paketler *install.packages()* fonksiyonu kullanılarak yüklenmektedir. Bu paketlerin her birinin farklı bir işlevi vardır. “*twitteR*” paketi, kullanıcıların Twitter web API’sine bağlanabilmek için bir arayüz sağlamaktadır. “*tm*” paketi, R Studio programı içindeki metin madenciliği uygulamaları için bir çerçeve oluşturmaktadır. Tweetler ilk olarak bir veri çerçevesi (data frame) halinde elde edilmektedir. Bu veri çerçevesi korpus olarak adlandırılan metin belgeleri koleksiyonuna dönüştürülmektedir. Oluşturulan korpus “*tm*” paketi ile sağlanan işlevlerle işlenebilir hale getirilmektedir. Metin madenciliği kapsamında kelime bulutları oluşturularak belgeler arasındaki benzerlik ve farklılıkları ortaya koyabilmek için “*wordcloud*” paketi kullanılmaktadır. Kullanıcıların sağladığı veriler

üzerindeki değişkenlerin nasıl eşleştirileceği, hangi grafiklerin kullanılacağı gibi fonksiyonlar “*ggplot2*” paketi ile gerçekleştirilmektedir. “*stringr*” paketi birçok veri temizleme, hazırlama aşamasında önemli rol oynamaktadır. İşlevler arasındaki tutarlılığı sağlayarak metin işleme işlemi kolaylaştırmaktadır. “*plyr*” paketi, çalışmayı parçalara ayırarak her bir parça üzerinde çalışma ve sonrasında bu parçaları tekrar bir araya getirme işlevini sağlamaktadır. “*SnowballC*” paketi ise kelimeleri ortak bir köke indirgemek için kullanılan kelime kökleme algoritmasını uygulayan C’libstemmer’ kütüphanesinin R Studio içerisindeki ara yüzüdür. Bu paket dahilinde Flemenkçe, Danca, İngilizce, Fince, Fransızca, Almanca, Macarca, İtalyanca, Norveççe, Portekizce, Romence, Rusça, İspanyolca, İsveççe ve Türkçe dilleri desteklenmektedir.

Bahsedilen paketler *library()* fonksiyonu ile çağrılmaktadır.

Resim 3-2: R Kütüphanesinden Gerekli Olan R Paketlerinin Çağırılması

```
38  
39 library(twitterR)  
40 library(ROAuth)  
41 library(tm)  
42 library(wordcloud)  
43 library(ggplot2)  
44 library(plyr)  
45 library(SnowballC)  
46 library(stringr)  
47
```

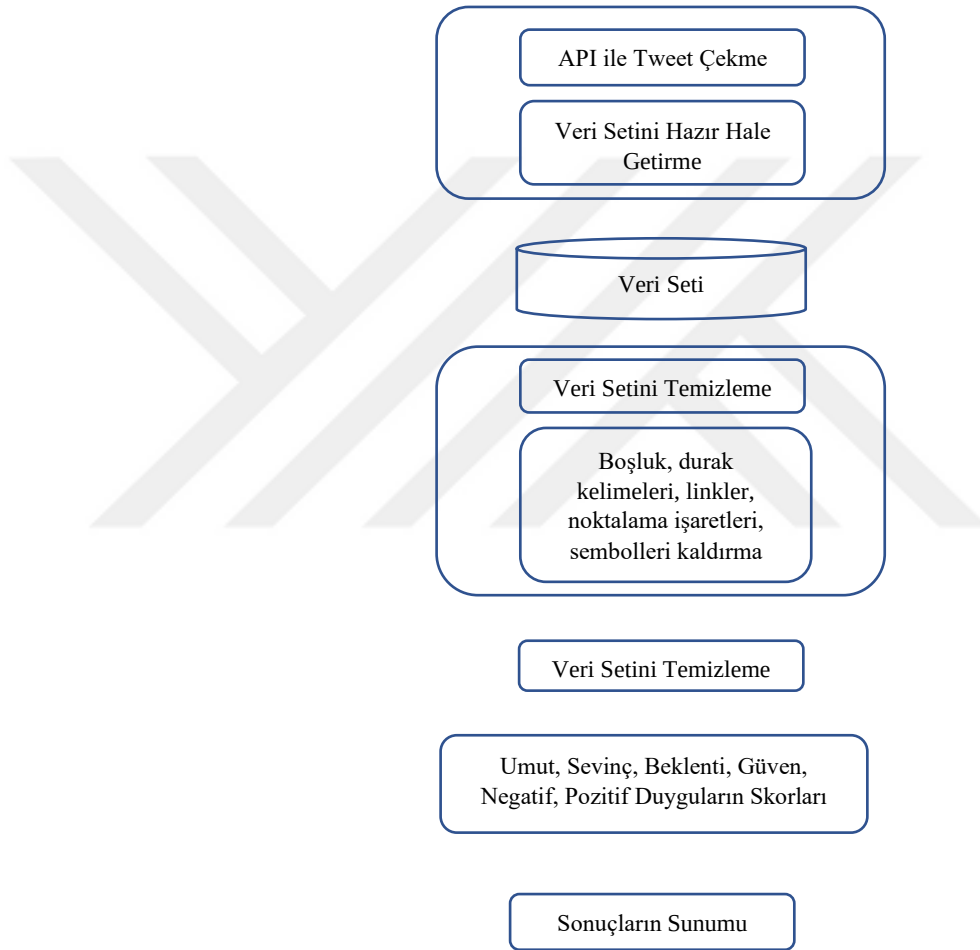
3.6. Veri Ön İşleme Süreci

Günlük konuşma, yazma ve okuma için kullanılan dil programlama dillerinden farklıdır. Twitter paylaşımları, günlük konuşma dili kullanılarak yazıldığından karmaşık ve gürültülü bir yapıya sahiptir. Kelimelerin kısaltılması, uzatılması, yanlış yazılması, sosyal medya dilinin kullanılması gibi sebeplerden dolayı tweetler karmaşık görülebilmektedir. Bu nedenle metinler üzerinde sınıflandırma işlemi gerçekleştirebilmek için öncelikle metinlerin ön işleme sürecinden geçirilmesi gerekmektedir. Bu ön işleme sürecinde kısaltılarak veya yanlış yazılan kelimeler üzerinde normleştirme işlemi yapılmaktadır. Ayrıca metinler İngilizce karakterler kullanılarak yazılmış olabilir. Örneğin ‘Bitcoin çıkıyor’ yerine ‘Bitcoin cikiyor’ şeklinde ifade edilmiş olabilir. İngilizce karakterler ile yazılan metinlerde Türkçeleştirme yapılmalıdır. Veriler ön işleme aşamasında temizlenip sınıflandırma işlemine hazır hale getirilmelidir. Ön işleme aşamasında metinlerde parçalama, tüm harfleri küçük harf yapma, durak kelimelerinin çıkarılması, kelimelerin köklerine

ayrılması, Terim-Doküman matrisinin oluşturulması işlemleri gerçekleştirilmektedir (Sreeja ve ark., 2020: 2). İngilizce dili gibi kelimelerin fazla ek almadığı dillerde sözlük tabanlı duygu analizini gerçekleştirmek Türkçe diline göre daha kolaydır. Türkçe gibi sondan eklemeli bir dil ile duygu analizi gerçekleştirebilmek için çok ciddi bir önışleme aşaması gerekmektedir.

Şekil 3.1’de araştırmanın metodolojisi gösterilmektedir.

Şekil 3-1: Twitter Verilerinin Elde Edilmesi ve İşlenmesi Süreci



Kaynak: Sreeja ve ark., 2020: 2

Twitter’dan elde edilen veri setinde #btc ve #bitcoin hashtag ile paylaşılan tweetler ayrı ayrı elde edilip daha sonradan bir tek Excel dosyası olarak birleştirilmiştir. Bu işlem tweet çekilen iki dönem için ayrı ayrı yapılarak iki veri seti

oluşturulmuştur. Bazı tweetler hem #btc hem de #bitcoin hashtag kullanılarak paylaşıldığından yinelenen tweetler veri setinden çıkarılmıştır.

Kripto para konusu günümüzde popüler bir konu olduğundan, sosyal medya gündeminde oldukça büyük bir hacim kaplamakta ve etkileşim almaktadır. Bu nedenle #bitcoin, #kriptopara, #crypto gibi hashtag içeren tweetler her an sosyal medyada karşımıza çıkmaktadır. Kripto para piyasası ile ilgili olmasa da etkileşim alabilmesi adına farklı konu ile alakalı içerikler de bu hashtag'ler kullanılarak paylaşıldığı görülmektedir. Özellikle SMA hastası çocuklar için paylaşılan tweetlerde görünürlüğünün artması ve daha fazla etkileşim alabilmesi adına bu hashtaglerin kullanıldığı görülmüştür. Kripto para piyasası ile ilgili olmadığından bu tweetler de veri setinden çıkarılmıştır.

Sosyal medyada bazı hesaplar yeni çıkan kripto paraların piyasada etkileşimini arttırmak adına çekilişler düzenlemektedir. Bu çekilişler ile ilgili reklam içerikli tweetler de veri setinden çıkarılmıştır. Reklam içerikli olan tweetlerin birçoğu aynıdır. Farklı zamanlarda paylaşılmış fakat aynı olan metinler de veri setinden çıkarılmıştır.

Excel üzerinden sütunlara filtreleme yapılmış ve 'çekiliş var', 'çekiliş zamanı', 'SMA' gibi kelime ve kelime grubu içeren tweetlere kolayca ulaşılmıştır. Bu tweetler toplu olarak silinerek veri setinden çıkarılmıştır. Bu işlem sonucunda Kasım 2021 veri setinden veri sayısı 265.432'den 63.076 adet veriye inmiştir. Mayıs 2022 veri setinde ise veri sayısı 257.786'den 59.159 adet veriye inmiştir.

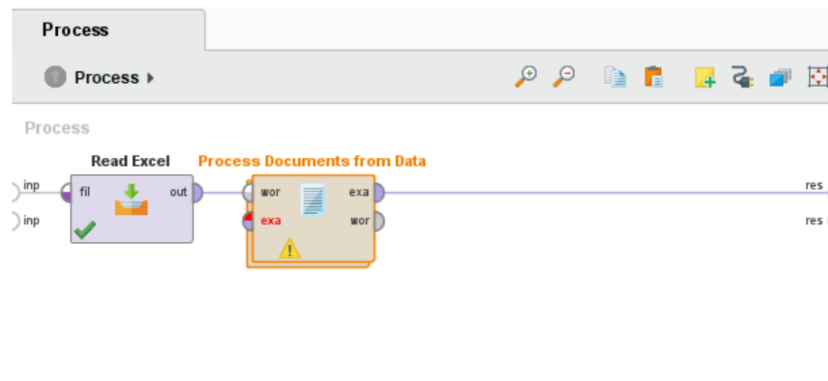
Silinmesi gereken veriler veri setinden çıkarıldıktan sonra kalan veri setinin temizlenmesi işlemi gerçekleştirilmiştir. Elde edilen tweetlerde anlamsız birçok karakter bulunmaktadır. Analizi gerçekleştirirmeden önce anlamsız karakterlerin ve fazlalıkların temizlenmesi gerekmektedir. Veri setini temizlemek için R programı içerisinde yer alan "tm" kütüphanesinden yararlanılmaktadır. Bu aşamada semboller, sayılar, linkler gibi analize dahil olmayacak tüm karakterler ortadan kaldırılmıştır. Ayrıca cümle içinde kullanılan, fakat analiz için herhangi bir anlamı olmayan edat ve bağlaçlarında tweetlerden çıkarılması gerekmektedir. Bu kelimelere durak kelimeleri denir. R programı içerisinde durak kelimeleri içeren İngilizce dilinde hazır veri setleri bulunmaktadır. Fakat Türkçe durak kelimeleri sözlüğü henüz geliştirilmemiştir. Türkçe dili ile hazırlanan durak kelimeleri veri tabanı olmadığından 650 adet kelimedenden oluşan veri tabanı oluşturulmuştur.

3.6.1. Verilerde Karakter Düzenlemesi

R Studio aracılığı ile elde edilen veri setinden silinmesi gereken veriler silindikten sonra metinler üzerinde karakter düzenlemesi yapılmak için RapidMiner kullanılmıştır. Kullanışlı ve anlaşılır bir arayüze sahip olduğundan bu program tercih edilmiştir. RapidMiner, Yale Üniversitesi tarafından geliştirilmiş Java programlama dili ile yazılmış açık kaynak kodlu bir yazılımdır. Veri ve metin madenciliği uygulamalarında oldukça kullanışlı bir programdır. RapidMiner farklı dosya uzantıları ile çalışmayı desteklemesi, içerisinde çok fazla operatör bulundurması, açık kaynak kodlu olması, kurulumunun ve kullanımının kolay olması, grafik ve görsellerin etkili bir şekilde kullanılabilmesi gibi avantajları sayesinde veri madenciliği uygulamalarında oldukça fazla tercih edilmektedir.

R Studio ile elde edilen twitter veri seti toplamda on dört sütundan oluşmaktadır. Bu sütunlar; 'text', 'favorited', 'favoriteCount', 'replyToSN', 'created', 'truncated', 'replyToSID', 'id', 'statusSource', 'screenName', 'retweetCount', 'isRetweet', 'retweeted', 'longitude', 'latitude' sütunlarıdır. Ancak analizde sadece 'text', 'favoriteCount' ve 'retweetCount' sütunları kullanılarak analiz gerçekleştirileceğinden diğer sütunlar veri setinden çıkarılmıştır. Her iki veri setinde de aynı işlem gerçekleştirilmiştir. Text sütununda yer alan metinler üzerinde karakter düzenlemesi yapabilmek için 'Read Excel' fonksiyonu kullanılmıştır. Read Excel fonksiyonu ile veri seti RapidMiner programına tanıtılmıştır. Daha sonra metin içerisindeki karakterler üzerinde ön işleme aşaması devam ettirilmiştir.

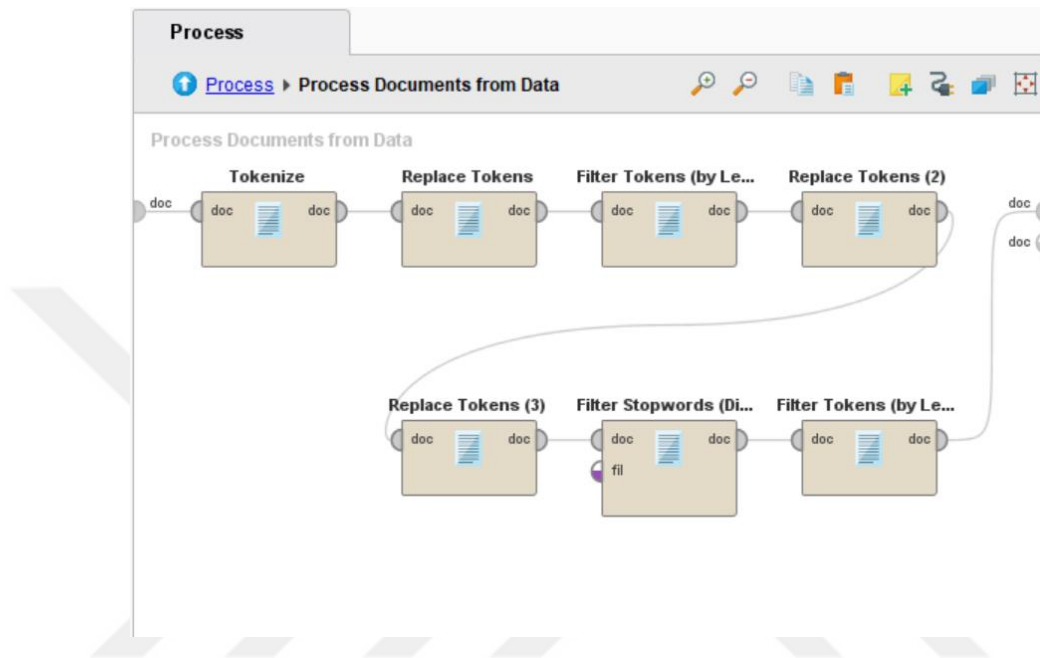
Resim 3-3: RapidMiner Kullanılarak Ön İşleme Sürecinin Gerçekleştirilmesi



Excel dosyasında kayıtlı olan veri seti 'Read Excel' fonksiyonu kullanılarak RapidMiner programına yüklenmiştir. Veriyi üzerinde parçalara ayırma, durak

kelimelerinin çıkarılması, kelime filtreleme ve metin içerisinde bazı parçaları değiştirebilmek için ‘Process Document From Data’ fonksiyonu kullanılmıştır. Bu fonksiyon içerisindeki alt fonksiyonlar aracılığıyla bu işlemleri gerçekleştirmektedir. Bu fonksiyon içerisinde kullanılan alt fonksiyonlar şekilde gösterilmiştir.

Resim 3-4: Verilerde Karakter Düzenlemesi



‘**Tokenize**’ fonksiyonu kelimelere ayırma işlemi için kullanılmaktadır. Metin içerisinde bulunan her kelime boşluk ya da noktalama işaretleri tanınarak ayrılmaktadır.

‘**Replace Tokens**’ fonksiyonu aracılığı ile metin içerisinde yer alan aynı anlama sahip kelimelerin aynı şekilde ifade edilebilmesi için kullanılmaktadır.

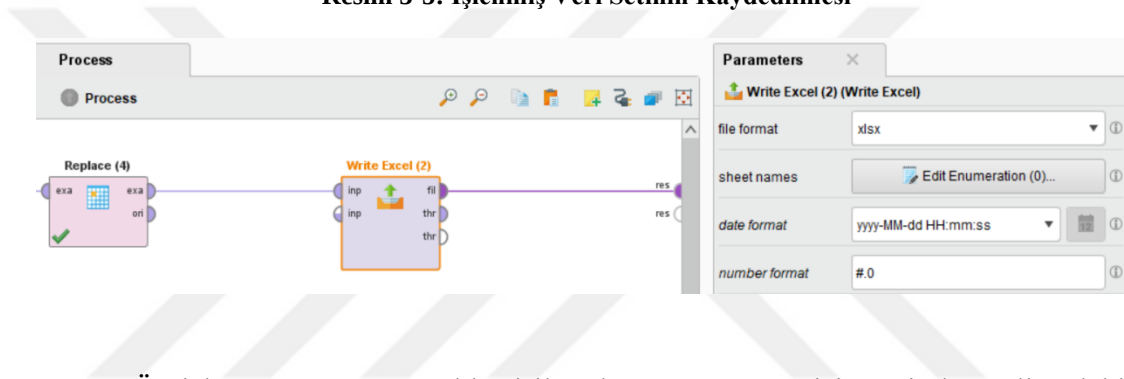
‘Filter Tokens – by length’ fonksiyonu karakter temelinde sınıflandırma yapmaktadır. Bu fonksiyon aracılığı ile en az 3 karakter bulunma şartı konularak daha az karaktere sahip kelimeler silinebilmektedir.

‘**Filter Stopwords – Dictionary**’ fonksiyonu aracılığı ile tek başına bir anlam ifade etmeyen edatlar, bağlaçlar, zamirler, bazı fiiller temizlenebilmektedir. Türkçe metinlerde bu işlem için öncelikle sözlük oluşturmak gerekmektedir. Oluşturulan sözlük tanımlanarak durak sözcükleri silinmektedir.

Ön işleme gerçekleştirilmiş olan veri seti ‘Write Excel’ fonksiyonu kullanılarak tekrar Excel formatında kaydedilebilmektedir.

Şekil 3.5’te görüldüğü üzere işlem sonrasında ‘Write Excel’ fonksiyonu kullanılarak işlenmiş veri seti Excel dosyası olarak bilgisayara kaydedilmiştir.

Resim 3-5: İşlenmiş Veri Setinin Kaydedilmesi



Ön işleme aşaması gerçekleştirilmeden önce text verisi üzerinde analize dahil olmayacak, hiçbir anlam ifade etmeyen semboller bulunmaktaydı. Analizin daha başarılı bir şekilde gerçekleştirilmesi için veri ön işleme aşamasının sağlıklı bir şekilde yürütülmesi gerekmektedir. Veri seti gereksiz ifadelerden arındırılarak bu ifadelerin analize girmesi önlenmelidir.

Verilere ön işleme yapılmadan önceki verinin ham haline ilişkin örnek veri seti aşağıda gösterilmektedir. R programında *head()* fonksiyonu kullanılarak veri setinden ilk altı satır gösterilmektedir. Aşağıda gösterilen tweetler Bitcoin’in düşüş trendinde olduğu hafta elde edilen “veri2” veri setinden örnek olarak alınmıştır.

[1] #Luna'da giriş dediğim yerden girenler bugün 1 saatte %660 a kadar karı gördü, karını alanda yaşadı Yeni alım ye... <https://t.co/aIRUGfy0Am>

[2] RT @OnurSuBrk: #BTC \$Usdt 1 W Tam olarak tamam mı devam mı beklenen 28.8 K güçlü desteği bu seferde çalışacak mı

[3] RT @GrafikAnaliz3: #Bitcoin #BTC \$BTC

Bu grafik teknik anlamda çizilmedi. #Btc'nin bir kaç ay akümüle olması lazım, hem altcoinleri rahatla...

[4] RT @M49814133: Alım satım yapan yapay zeka mı müthiş birşey 1 gğnde gelen kar ilerde buralar dutluk olur şimdiden alın. #dxgm #dexgame #oxr...

[5] " 4\$ beri shortluyorum sonunda hak ettiği yeri buldu "
sukur 1.4\$ <https://t.co/E8XfqC09Ji>

[6] RT @Kriptocum4: #btc anlık dalgalanmalar çok oldu 30 31 k arası baya alımlar geldi ama satıcılı bölge durmadı sürekli satış short yedi bunu...

İşlenmemiş hali verilen tweetler önışleme aşamasından sonra analiz edilebilir hale getirilmiştir. Noktalama işaretleri, hastagler, kullanıcı adları silinip tüm harfler küçük harf olarak düzenlenmiştir. Tweetlerin analiz edilmeye hazır hali örnek olarak aşağıda gösterilmiştir.

1 luna da giriş dediğim yerden girenler bugün 1 saatte 660 a kadar karı gördü karını alanda yaşadı yeni alım ye

2 btc \$ustd 1 w tam olarak tamam mı devam mı beklenen 28.8 k güçlü desteği bu seferde çalışacak mı

3 bitcoin btc btc bu grafik teknik anlamda çizilmedi btc nin bir kaç ay akümüle olması lazım hem altcoinleri rahatla

4 alım satım yapan yapay zeka mı müthiş bir şey 1 günde gelen kar ilerde buralar dutluk olur şimdiden alın dxgm dexgame oxr

5 4\$ beri shortluyorum sonunda hak ettiği yeri buldu şükür 1.4\$

6 btc anlık dalgalanmalar çok oldu 30 31 k arası baya alımlar geldi ama satıcılı bölge durmadı sürekli satış short yedi bunu

3.6.2. Sözlük Oluşturma

Literatürde yer alan çalışmalara bakıldığında duygu analizi çalışmaları sözlük temelli ve makine öğrenmesi temelli algoritmalar ile gerçekleştirilmiştir. Sözlük temelli duygu analizi yapabilmenin en kritik noktası sözlük oluşturma aşamasıdır. Her veri setinin kendisine özel terimleri, dili, jargonu olduğundan her birisinin kendisine özel sözlüğü olmalıdır. İngilizce içeriklerde pozitif ve negatif kelime veri setleri hazır olduğundan web ortamında kolayca ulaşılabilir. Bu alanda geliştirilmiş olan üç farklı veri seti mevcuttur. Bunlar Finn Arup Nielsen tarafından geliştirilen AFINN, Bing Liu tarafından geliştirilen Bing ve Saif Mohammad ile Peter Turney tarafından geliştirilmiş olan NRC isimli kütüphanelerdir (Oğuzlar & Kızılkaya, 2019: 100). Türkçede pozitif ve negatif anlam içeren kelimeler için geliştirilmiş etkili bir sözlük bulunmamaktadır. Elde edilen veri setinin içeriği kapsamında tweetler taranarak Türkçe pozitif ve negatif duygu ifade eden kelimelerden oluşan birer sözlük oluşturulmuştur. Pozitif sözlük toplamda 1392 adet ifade içerirken, negatif duygu sözlüğü toplamda 1517 adet ifade içermektedir.

3.7. Duygu Skoru

R programı kullanılarak “veri1” ve “veri2” veri setleri için ayrı ayrı duygu skorları hesaplanmıştır. Duygu skorları hesaplanabilmesi için elde edilen negatif, pozitif ve durak kelimeleri sözlükleri R programına yüklenmelidir. Yüklenen sözlükler *sapply()* fonksiyonu kullanılarak veri seti liste formatına dönüştürülmelidir. Daha sonra *score.sentiment()* fonksiyonu kullanılarak her bir tweete ait duygu skorları hesaplanmaktadır. Her bir tweetin duygu skoru paylaşılamayacağından duygu skorlarının ortalaması *mean(analysis\$score)* fonksiyonu kullanılarak elde edilmiştir. Duygu skorlarının ortalaması tek başına bir fikir vermeyeceğinden duygu skorlarının frekans dağılımları da elde edilmiştir.

Duygu skorları hesaplanırken ön işleme sonucunda elde edilen veri setlerinde bulunan tüm tweetler için duygu puanları hesaplanmıştır. Hesaplanan duygu skorlarının ortalamaları alınmıştır. Oluşturulan pozitif ve negatif duygu sözlüğü yardımı ile duygu skorları hesaplanmıştır. Pozitif duygu sözlüğünde bulunan her kelime için +1 puan ve negatif duygu sözlüğünde bulunan her kelime için -1 puan

verilmiştir. Tweetlerde bulunan pozitif ve negatif kelime değerlerinin toplanması ile duygu skorları elde edilmiştir. Bir tweet içerisinde pozitif ve negatif kelime sayısı eşit olduğunda ya da oluşturulan duygu sözlüğünde bulunan kelimeleri içermeyen bir tweet olduğunda bu tweetin duygu skoru 0 olduğundan nötr bir tweet olduğunu göstermektedir.

Tablo 3.1’de her iki veri seti için elde edilen duygu skorlarının ortalaması gösterilmektedir.

Tablo 3-1: Duygu Skorlarının Aritmetik Ortalaması

Veri Seti	Duygu Skorları Ortalaması
‘veri1’	0,35
‘veri2’	0,14

İki veri seti arasında en yüksek duygu skoruna sahip olan veri seti ‘veri1’dir. Veri1 kripto para piyasasının yükseliş trendinde olduğu dönemde elde edildiğinden daha fazla olumlu tweet olduğu düşünülmektedir. Bu nedenle duygu skoru veri2 veri setine göre daha yüksektir.

Bitcoin’in yükseliş trendinde olduğu zaman diliminde elde edilen tweetlerin tweet puanları ve frekansları ile ilgili değerler tablo 3.2’de gösterilmektedir.

Tablo 3-2: Bitcoin Yükseliş Trendi Gösterdiği Zamana İlişkin Tweet Puanları ve Frekansları

Tweet Puanı	Frekansı	Nisbi Frekans (%)
-3	32	0,05
-2	104	0,1
-1	3.298	5
0	38.846	65
1	15.598	26
2	1.056	1
3	178	0,3
4	47	0,07
Toplam Frekans	59.159	100

Yukarıda ‘veri1’ veri seti için tweetlerin duygu skorları ve frekans dağılımları gösterilmektedir. Tabloda görüldüğü üzere 06-12 Kasım 2021 tarih aralığında elde edilen tweelerin duygu skorları -3 ile 4 puan arasında değer almaktadır. 0 puan alan nötr tweetler veri setinin %65’in oluşturmaktadır. -3 ve 4 puan alan tweetler oldukça düşük frekans göstermektedir. En yüksek ikinci frekansa sahip duygu puanı ise 1’dir. 1 puana sahip tweetler ise veri setinin %26’sını oluşturmaktadır. Genel olarak negatif tweetler veri setinin %5’ini oluştururken, pozitif tweetler %26’sını oluşturmaktadır. Tabloda görüldüğü üzere nötr tweetler duygu skoru ortalamasına en fazla katkı sağlayan tweetler olmuştur.

Tablo 3.3’te ise Bitcoin’in düşüş trendinde olduğu zaman diliminde elde edilen tweetlerin tweet puanları ve frekansları gösterilmektedir.

Tablo 3-3: Bitcoin Düşüş Trendi Gösterdiği Zamana İlişkin Tweet Puanları ve Frekansları

Tweet Puanı	Frekansı	Nisbi Frekans (%)
-4	7	0,01
-3	116	0,1
-2	3.727	5
-1	14.761	23
0	32.846	55
1	11.513	18
2	98	0,1
3	8	0,01
Toplam Frekans	63.076	100

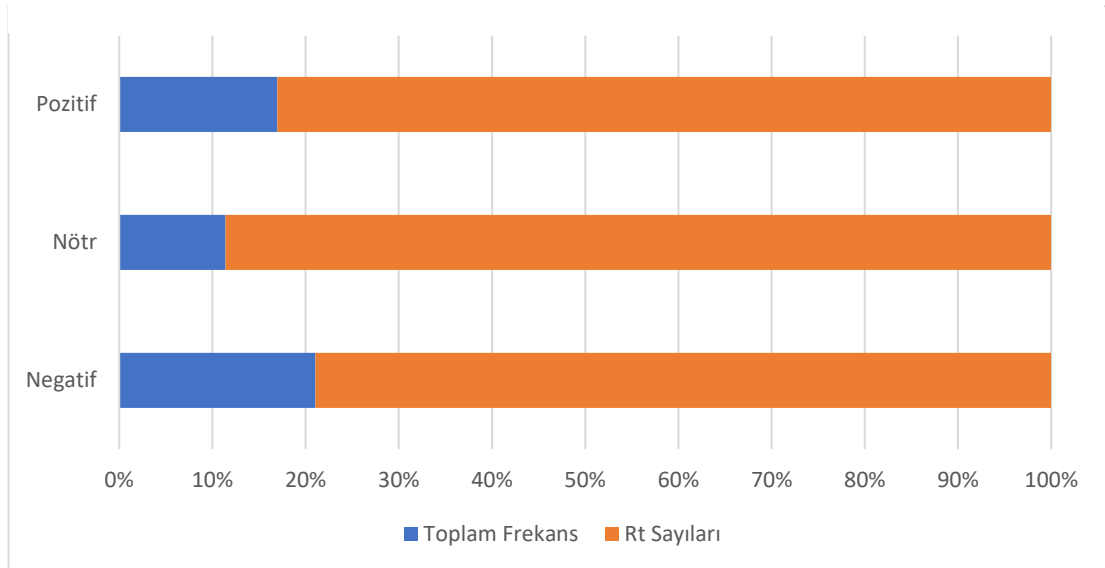
Yukarıdaki tabloda görüldüğü üzere ‘veri2’ veri seti için elde edilen tweetlerin %55’i nötr tweetlerden oluşmaktadır. Genel olarak bakıldığında tweetler nötr duygu ifade etse de en yüksek frekansa sahip ikinci tweet puanı -1’dir. Negatif duygulu tweetler veri setinin %31’ini oluştururken pozitif duygulu tweetler veri setinin %19’unu oluşturmaktadır. Bitcoin’in yükseliş trendinde olduğu dönemin aksine Bitcoin’in düşüş trendinde olduğu dönemde negatif duygulu tweetlerin oranı artış göstermektedir.

Veri setlerinin elde edildiği dönemlerde Bitcoin'in sahip olduğu değerin ve piyasa hacminin de çalışma için önemli olduğu düşünülmektedir. Şekil 3.5'te ilgili tarihlerde Bitcoin'in sahip olduğu piyasa değeri grafiği gösterilmektedir.

3.7.1. Betimsel İstatistikler

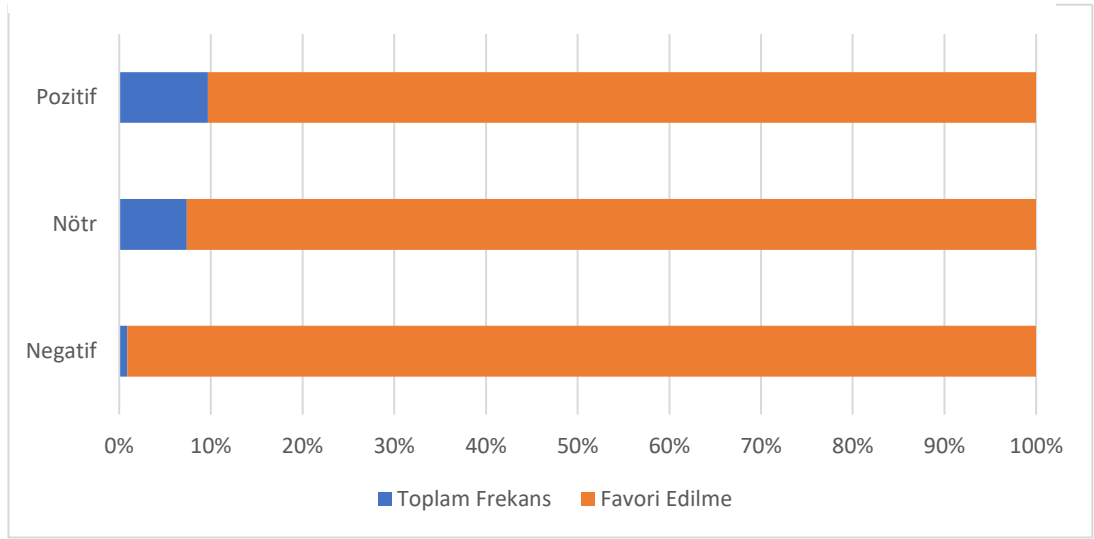
Çalışmanın bu bölümünde paylaşılan Twitter içeriklerinin aldıkları etkileşimler ifade edilmek istenmiştir. Bu kapsamda her iki veri seti için de pozitif, negatif ve nötr duygu içeren tweetlerin aldıkları retweet ve favori edilme sayıları gösterilmiştir. Şekil 3-2'de kripto para piyasasının yükseliş trendinde olduğu dönem elde edilen tweetlerin retweet edilme sayıları gösterilmektedir. Grafikte görüldüğü üzere en yüksek etkileşim alan tweetler nötr duygu içeren tweetler olmuştur. İkinci olarak pozitif tweetler en fazla retweet edilme oranına sahiptir. Negatif duygu içeren tweetler ise en az retweet edilme sayısına sahip tweetler olmuştur.

Şekil 3-2: 'veri1' Tweetlerin Retweet Edilme Oranları



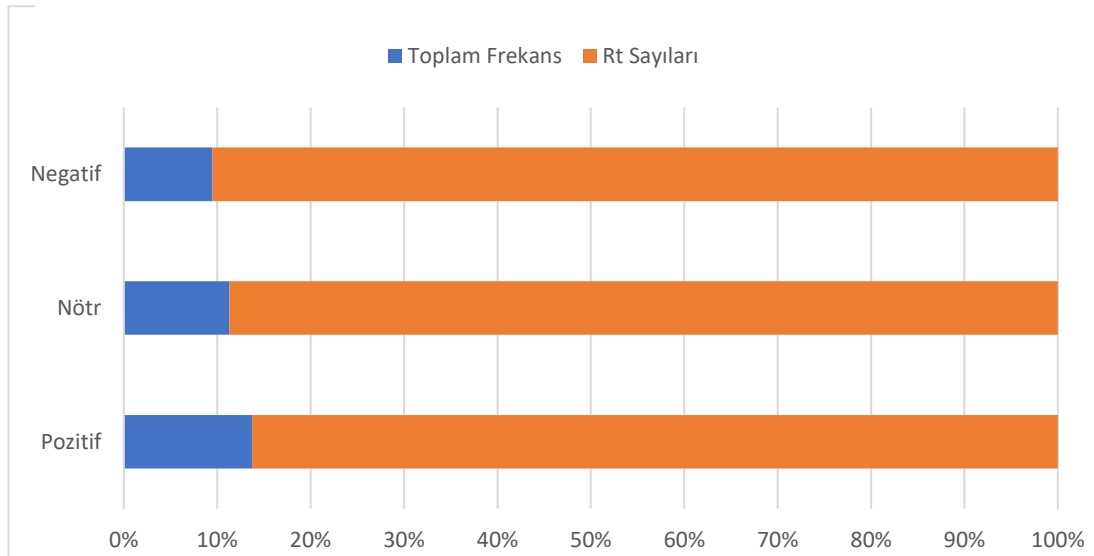
Şekil 3-3'te ise yine Bitcoin'in yükseliş trendinde olduğu dönem elde edilen tweetlerin favori edilme sayıları gösterilmiştir. Retweet edilme sayılarında olduğunun aksine en yüksek favori edilmesine sahip olan tweetler negatif duygu içeren tweetler olduğu görülmektedir. En az favori sayısına sahip olan tweetler ise pozitif duygu içeren tweetler olmuştur.

Şekil 3-3: ‘veri1’ Tweetlerin Favori Edilme Oranları



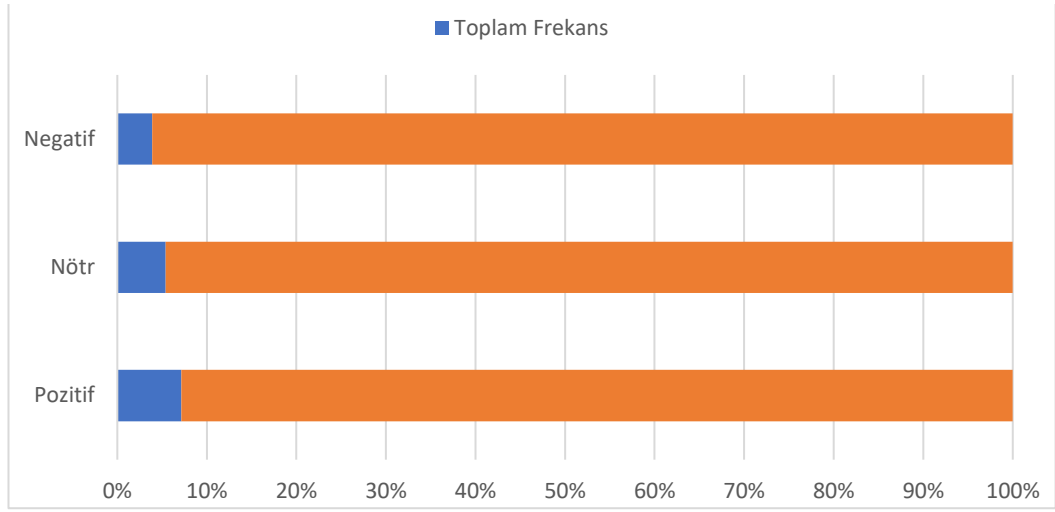
Şekil 3-4’te kripto para piyasasının düşüş trendinde olduğu dönem elde edilen tweetlerin retweet edilme sayıları gösterilmektedir. Grafikte görüldüğü üzere en yüksek etkileşim alan tweetler negatif duygu içeren tweetler olmuştur. İkinci olarak nötr tweetler en fazla retweet edilme oranına sahiptir. Pozitif duygu içeren tweetler ise en az retweet edilme sayısına sahip tweetler olmuştur.

Şekil 3-4: ‘veri2’ Tweetlerin Retweet Edilme Oranları



Şekil 3-5’te ise Bitcoin’in düşüş trendinde olduğu dönem elde edilen tweetlerin favori edilme sayıları gösterilmektedir. Retweet edilme sayılarında olduğu gibi en yüksek favori edilmesine sahip olan tweetler negatif duygu içeren tweetler olduğu görülmektedir. En az favori sayısına sahip olan tweetler ise pozitif duygu içeren tweetler olmuştur.

Şekil 3-5: ‘veri2’ Tweetlerin Favori Edilme Oranları



Yukarıdaki şekillerde görüldüğü üzere her iki dönemde elde edilen tweetler oldukça yüksek miktarlarda etkileşim almışlardır. Kripto para piyasasında değer yükselmesi veya değer düşmesi yaşansa da kamuoyu tarafından ilgi gösterilen bir alan olduğu sonucuna ulaşılabilmektedir.

Yatırımcılar, piyasa koşulları nasıl olursa olsun her zaman sahip olduğu maddi kaynakların değerini arttırmak istemektedir. Piyasanın yükseliş trendinde olduğu dönemlerde kazanç sağlamak daha kolay olarak görülmektedir. Fakat önemli olan piyasa düşüş trendine geçtiği dönemlerde maddi varlıkların değerini koruyabilmek ya da varlıkların değerini arttırabilmektir. Bu nedenle yatırımcılar hemen her koşulda kripto para piyasalarına karşı ilgili olmuşlardır. Günümüzde fikirleri paylaşmanın en kolay ve en hızlı yolu sosyal medya olduğundan yatırımcılar kripto para piyasasına olan ilgilerini de sosyal medyada bilgi alışverişi yaparak, fikirlerini paylaşarak, uzman yatırımcıların tavsiyelerini takip ederek desteklemişlerdir.

Kripto para günümüzde oldukça popüler bir konu haline gelmesiyle birlikte sosyal medyada kripto para içeriklerinin aldığı etkileşimde paralel olarak artmaktadır.

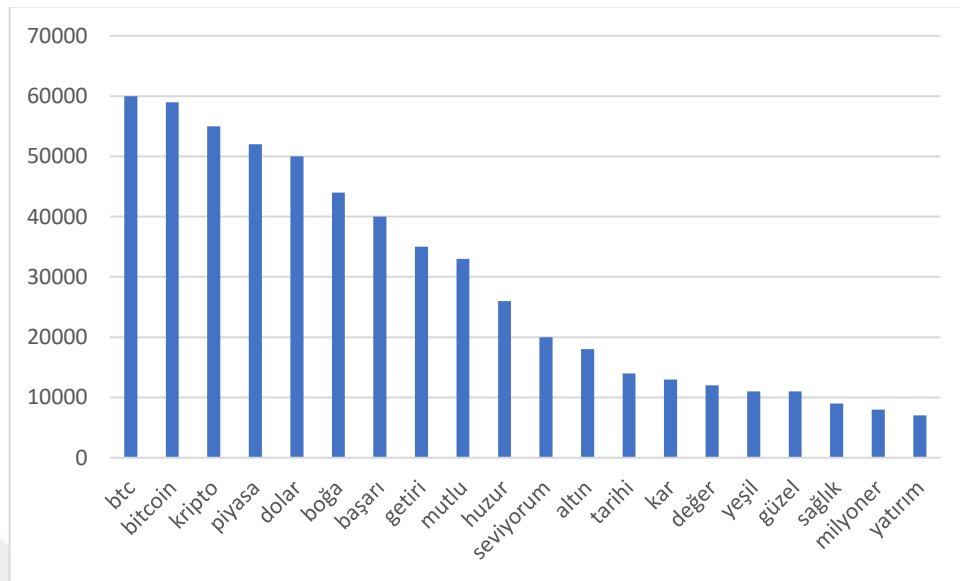
3.7.2. Kelime Frekanslarının Değerlendirilmesi

Bir metin içerisinde sık kullanılan kelimeler liste halinde gösterilmenin yanı sıra kelime bulutu kullanılarak farklı şekillerde de görselleştirilerek gösterilebilmektedir. Kelime bulutu, metinsel verilerin görselleştirilebilmesi için

Şekil 3-6’da kripto para piyasasının yükseliş trendinde olduğu ve Bitcoin’in tarihindeki en yüksek değere ulaştığı döneme ilişkin kullanılan veri setine ait kelime bulutu gösterilmektedir. Elde edilen kelime bulutunda veri setinin en yüksek frekansa sahip olan kelimelerinin btc, bitcoin, kripto, dolar ve piyasa kelimeleri olduğu gözlenmektedir. Bunun yanı sıra mutlu, huzur, güzel, başarı, kar gibi olumlu ifade belirten kelimelerin de sıklıkla kullanıldığı görülmektedir.

79

Şekil 3-7: 06-12 Kasım 2021 Tarih Aralığında Elde Edilen Tweetlerde En Sık Kullanılan Kelimelerin Frekans Dağılımı



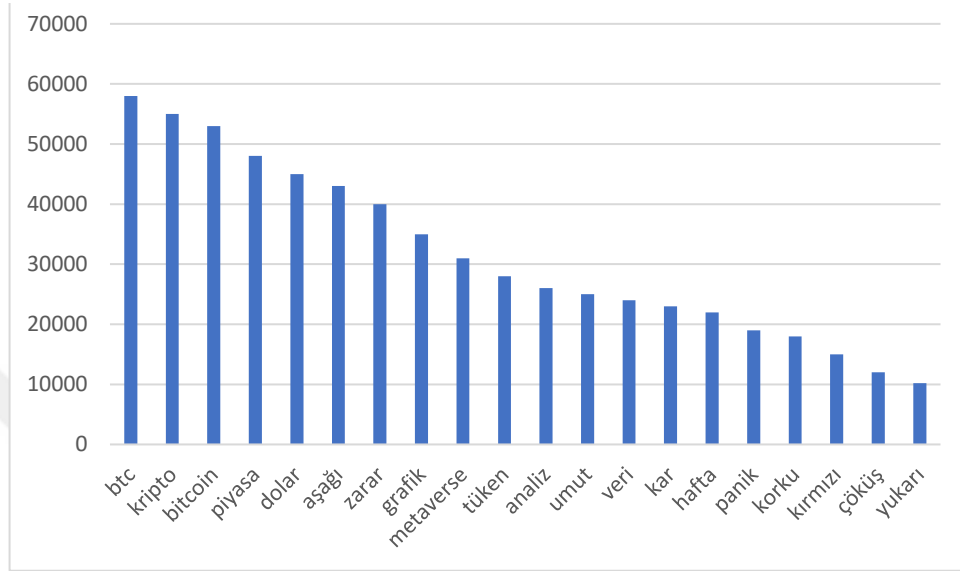
Şekil 3-8’de kripto para piyasasının düşüş trendinde olduğu döneme ilişkin kullanılan veri setine ait kelime bulutu gösterilmektedir. Elde edilen kelime bulutunda veri setinin en yüksek frekansa sahip olan kelimelerinin ‘veri1’ veri setinde olduğu gibi btc, bitcoin, kripto, dolar ve piyasa kelimeleri olduğu gözlenmektedir. Bunun yanı sıra korku, zor, kayıp, çöküş gibi olumsuz ifade belirten kelimelerin de sıklıkla kullanıldığı görülmektedir.

Şekil 3-8: 05-11 Mayıs 2022 Tarih Aralığında Elde Edilen Tweetlerde En Sık Kullanılan Kelimelerin Kelime Bulutu



Şekil 3-9’da ise ilgili dönemde en yüksek frekans gösteren 20 kelime grafik üzerinde gösterilmektedir.

Şekil 3-9: 05-11 Mayıs 2022 Tarih Aralığında Elde Edilen Tweetlerde En Sık Kullanılan Kelimelerin Frekans Dağılımı

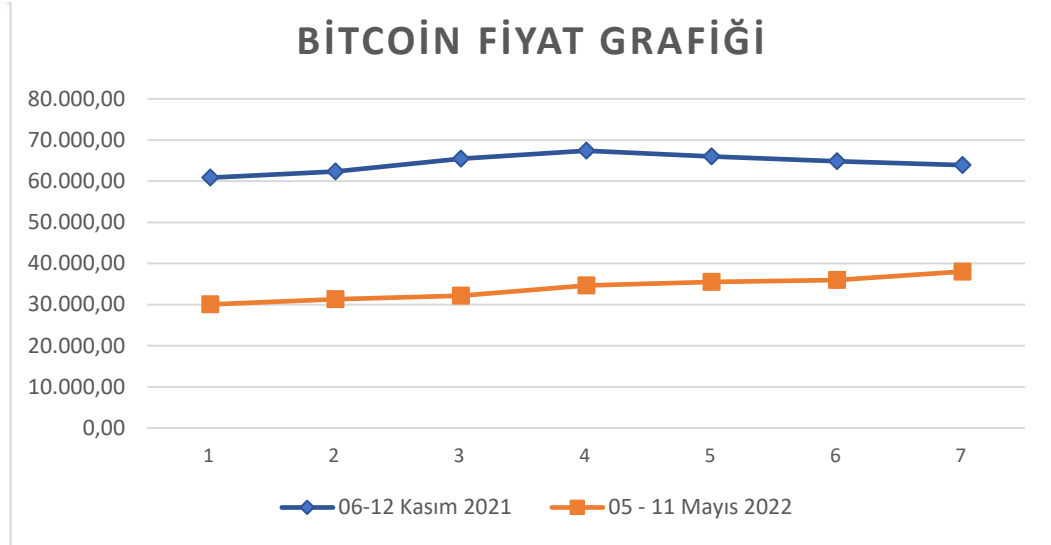


Oluşturulan ikinci kelime bulutunda olumsuz ifadelerin yanı sıra grafik, analiz, yatırım, oran gibi kelimelerde sıklıkla kullanılmıştır. Kripto para piyasasının düşüş trendinde olmasıyla birlikte yatırımcılar zararlarını en aza indirmek veya kar edebilmek amacıyla daha fazla grafik ve analiz yorumlarına ilişkin tweet paylaştığı söylenebilir. Kripto para piyasalarının yükseliş trendinde olduğu dönemler de yatırımcıların hemen hepsi kar edebilmektedir. Düşüş trendinde olan, ayı sezonu olarak nitelendirilen dönemlerde ise kar elde etmek daha zordur. Bu nedenle yatırımcılar daha fazla kripto piyasasına ilişkin grafiksel analizleri takip ettikleri söylenebilir. Ayrıca ‘veri1’ veri setinde gün ve saat kelimeleri hafta kelimesine göre daha fazla kullanılmışken, ‘veri2’ veri setinde hafta kelimesi gün ve saat kelimelerine göre daha fazla kullanılmıştır. Kripto para piyasalarında boğa sezonu olarak adlandırılan yükseliş trendinin yoğun olduğu dönemler düşüş trendinin yaşandığı dönemlere göre daha kısa sürmektedir. Bu nedenle günlük ve saatlik veriler daha fazla takip edilmektedir. Çünkü bazı veriler günlük olarak çok fazla değişkenlik gösterebilir. Fakat düşüş trendinin yaşandığı sezonlarda bazı fiyat verileri haftalarca yaklaşık aynı değer etrafında seyredebilmektedir. Bu nedenle saatlik veriler kripto para piyasasının düşüş sezonlarında önemini yitirebilmektedir.

3.7.3. Bitcoin Fiyat ve Hacim Verilerinin Değerlendirilmesi

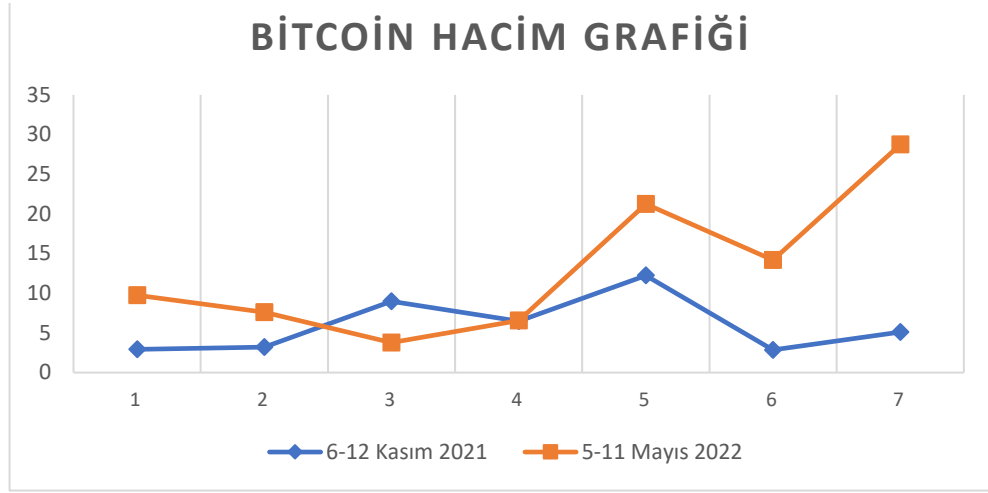
2022 yılı itibariyle farklı kripto para piyasalarında 9000 civarı kripto para işlem görmektedir. Her geçen gün piyasalarda işlem gören kripto para sayısı artmaktadır. Kripto para piyasalarının yaklaşık %40 hacmine sahip olan Bitcoin ile ilgili fiyat ve piyasa hacmi verileri grafikler üzerinde gösterilmektedir. Bitcoin fiyat grafiğinde görüldüğü üzere 06-12 Kasım 2021 tarih aralığında Bitcoin değeri 60.000 Dolar bandında iken 5-11 Mayıs 2022 tarih aralığında %50 civarında değer kaybı gerçekleşmiş ve 30.000 Dolar bandına gerilemiştir.

Şekil 3-10: İlgili Tarihlerde Bitcoin Fiyat Grafiği



Şekil 3-11'de ise yine aynı dönemlere ait Bitcoin'in sahip olduğu günlük hacim grafiği gösterilmektedir. Genel itibarıyla Bitcoin'in düşüş trendinde olduğu dönem piyasa hacmi yükselmiştir. Yatırımcı tarafından, Bitcoin'de yaşanan değer kaybı bir yatırım fırsatı olarak görülmüş olabilir.

Şekil 3-11: İlgili Tarihlerde Bitcoin Hacim Grafiği



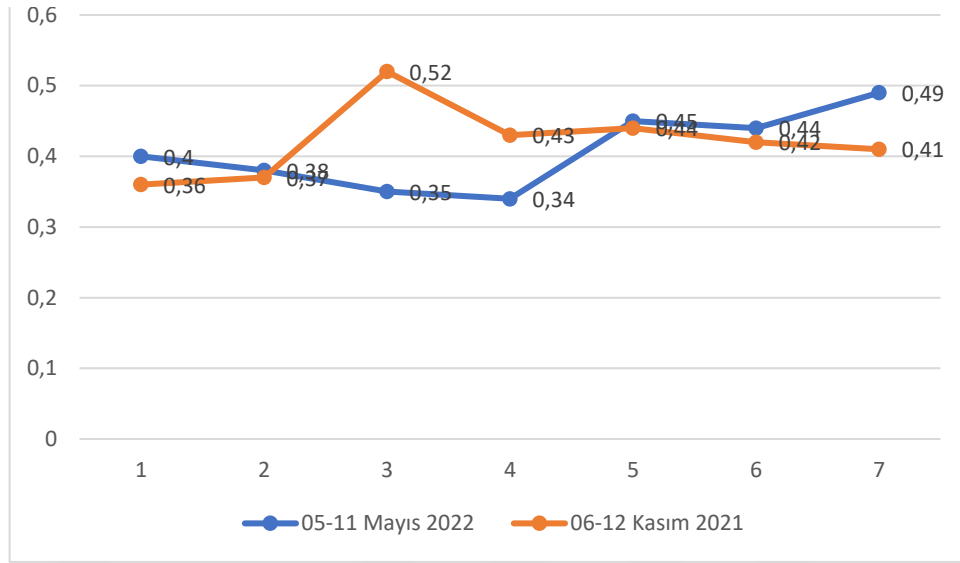
Duygu skorları ve Bitcoin'e ilişkin fiyat ve hacim grafikleri birlikte değerlendirildiğinde duygu skorlarının daha çok fiyattan etkilendiği söylenebilmektedir.

3.7.3.1. Türk Kripto Para Borsalarında Bitcoin

Kripto paralar tüm dünyada olduğu gibi ülkemizde de yoğun bir ilgi görmektedir. Coinmarketcap verilerine göre 2022 yılı itibariyle dünya genelinde 294 adet farklı kripto para piyasası mevcuttur. Türkiye'de ise 9 farklı Türk kripto para borsasında işlem yapılabilir. Bunlar BtcTurk, Paribu, Bitexen, Bitlo, Bitci, Icrypex, Bitturk, Felixo, Bitubu borsalarıdır. Bu borsalar arasında en çok kullanılanlar BtcTurk ve Paribu platformlarıdır. En iyi Türk kripto para borsalarına bakıldığında en çok kullanılan kripto para çiftlerini sistemlerine entegre ettikleri görülmektedir. Bu borsalar talep durumlarını göz önünde bulundurarak sistemlerine yeni kripto para birimleri eklemeye devam etmektedirler.

Aşağıdaki grafiklerde Paribu ve BtcTurk borsalarında ilgili dönemlere ait Bitcoin hacim grafikleri gösterilmektedir.

Şekil 3-122: İlgili Tarihlerde Paribu Platformunda Gerçekleşen Bitcoin Hacim Grafiği

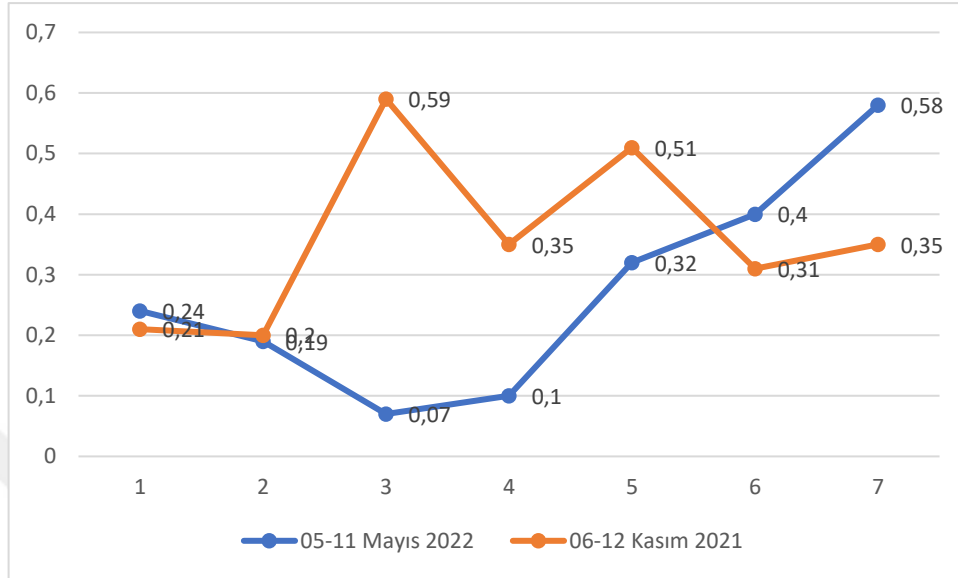


Paribu, İstanbul merkezli olarak 2017 yılının Şubat ayında faaliyetlerine başlamıştır. Yaklaşık 5 milyon kullanıcısı ile Türkiye’de en fazla kullanıcıya sahip kripto para borsalarından birisidir. 2022 yılı itibariyle borsada 97 farklı kripto para birimi ile işlem yapılabilmektedir. Paribu’nun entegre olduğu bankalardan herhangi biri aracılığıyla para yatırılabilir ve çekilebilmektedir.

Şekil 3-12’de 6-12 Kasım 2021 tarih aralığına ve 5-11 Mayıs 2022 tarih aralığına ilişkin Paribu platformu üzerinde gerçekleşen Bitcoin hacminin artış ve azalış grafiği görülmektedir. Paribu üzerinden gerçekleştirilen işlemlerde Bitcoin’in en yüksek hacme sahip olduğu tarih 8 Kasım 2021 tarihine aittir. Bu tarihten sonra Bitcoin fiyatının düşüş trendine girmesiyle hacim miktarı da azalmıştır. Kullanıcılar bu süreçte zararını en aza indirmek için sahip oldukları Bitcoin’i satıp yeni bir alım fırsatı beklemiş olabilirler. Bitcoin hacminin en düşük olduğu tarih ise 8 Mayıs 2022 tarihine aittir. Fakat hemen sonrasında bir hacim artışı gözlenmektedir. Bu durumda yatırımcılar için alım fırsatı doğduğu söylenebilir. Genel itibariyle grafiğe bakıldığında hacim miktarlarında artış ve azalışlar yaşansa da her iki dönemde de

yatırımcılar için alım fırsatlarının olduğu ve yatırımcıların bu fırsatları değerlendirdiği söylenebilmektedir.

Şekil 3-13: İlgili Tarihlerde BtcTurk Platformunda Gerçekleşen Bitcoin Hacim Grafiği



BtcTurk, 2013 yılında faaliyetlerine başlayan İstanbul merkezli bir kripto para borsasıdır. Türkiye’de faaliyete başlayan ilk kripto para borsasıdır. BtcTurk 5 milyondan fazla kullanıcıya hizmet vermektedir. Paribu’da olduğu gibi BtcTurk’te de anlaşmalı bankalar aracılığıyla para yatırılıp çekilebilmektedir. BtcTurk, kullanıcılarına herhangi bir transfer ücreti ödemediği Bitcoin çekebilme hizmeti sunmaktadır.

Şekil 3-13’te 6-12 Kasım 2021 tarih aralığına ve 5-11 Mayıs 2022 tarih aralığına ilişkin BtcTurk platformu üzerinde gerçekleşen Bitcoin hacminin artış ve azalış grafiği görülmektedir. BtcTurk üzerinden gerçekleştirilen işlemlerde Bitcoin’in en yüksek hacme sahip olduğu tarih 8 Kasım 2021 tarihine aittir. Bu tarihten sonra Bitcoin fiyatının düşüş trendine girmesiyle hacim miktarı da 8 Kasım 2021 tarihine göre azalmıştır. Kullanıcılar bu süreçte zararını en aza indirmek için sahip oldukları Bitcoin’i satıp yeni bir alım fırsatı beklemiş olabilirler. Bitcoin hacminin en düşük olduğu tarih ise 7 Mayıs 2022 tarihine aittir. Fakat hemen sonrasında ciddi oranda bir hacim artışı gözlenmektedir. Bu durumda yatırımcılar için alım fırsatı doğduğu söylenebilir. Paribu platformuna ait grafik ile karşılaştırıldığında hacim miktarının daha düşük olduğu gözlenmiştir. Yatırımcıların Bitcoin alım satım işlemlerini

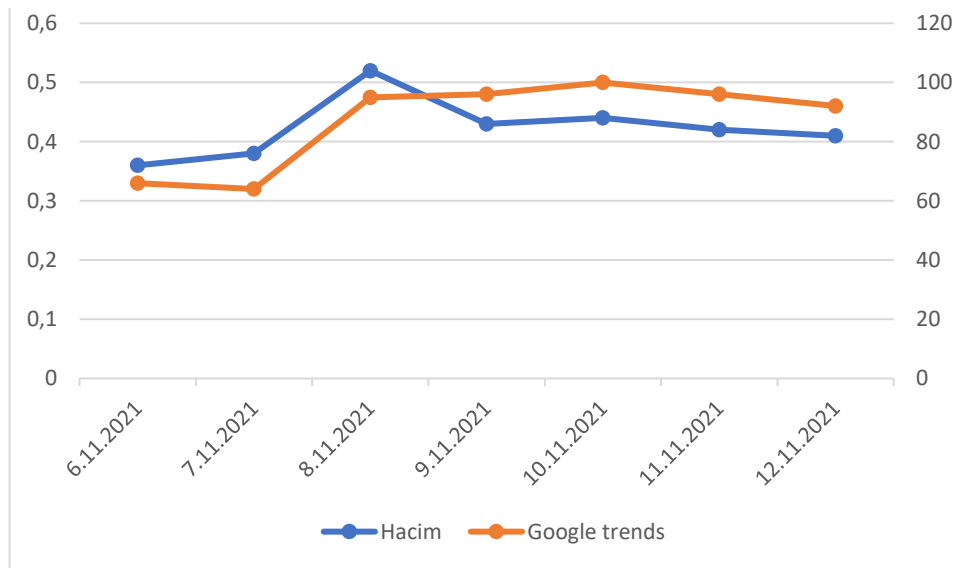
gerçekleştirmek için genellikle Paribu platformunu kullandığı söylenebilir. Genel itibariyle grafiğe bakıldığında hacim miktarlarında artış ve azalışlar yaşansa da her iki dönemde de yatırımcılar için alım fırsatlarının olduğu ve yatırımcıların bu fırsatları değerlendirdiği söylenebilmektedir.

3.7.3.2. Bitcoin Hacmi ile Google Trends Verileri Arasındaki İlişki

Google Trends, Google arama motoru üzerinde aranan kelimelerin ne sıklıkla aratıldığını gösteren ücretsiz veri sunan bir servistir. Belirli bir tarihte aranan popüler kelimelerin yanı sıra istenilen tarih içerisindeki istenilen anahtar kelimeler de istatistiksel olarak görülebilmektedir. Google’da aranan kelimelerin popülerlik düzeyi tarih aralığı seçilerek bölgesel olarak hangi bölgelerde daha popüler olduğu görülebilmektedir.

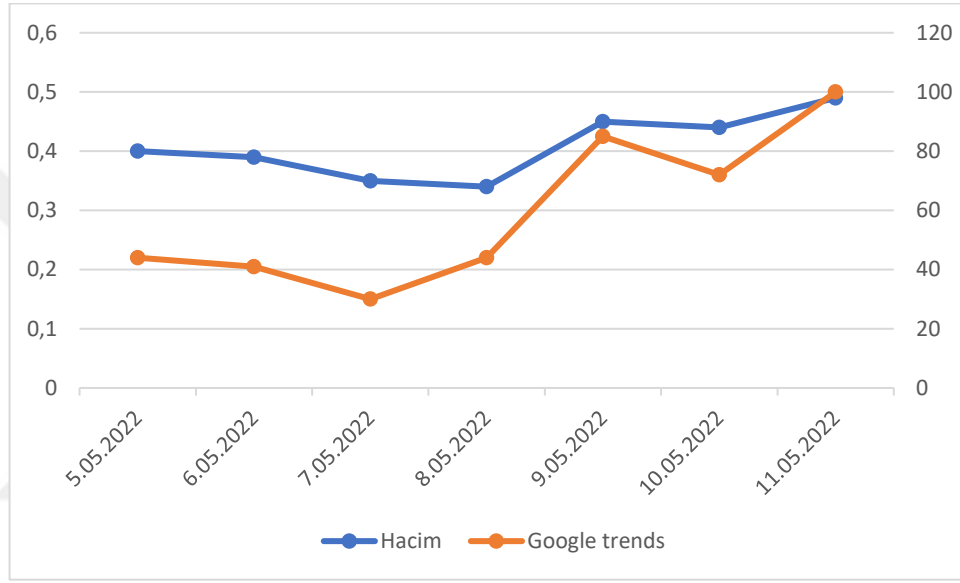
Farklı kripto para piyasalarındaki Bitcoin hacmi ile Google üzerinden tıklanma sayılarına ilişkin grafikler paylaşılmaktadır. Çalışmada kullanılan her iki tarih aralığı için Paribu ve BtcTurk kripto para piyasalarında Bitcoin hacmi ile Google Trends verileri karşılaştırılmıştır. Google Trends verileri elde edilirken ilgili tarih aralıkları baz alınmıştır. Ayrıca lokasyon olarak Türkiye seçilmiş ve Türkiye’de ‘Bitcoin’ kelimesinin Google’da aratılma sayıları elde edilmiştir.

Şekil 3-14: İlgili Tarihlerde Paribu Platformunda Gerçekleşen Bitcoin Hacmi ve Google Trends Verilerinin Karşılaştırılması



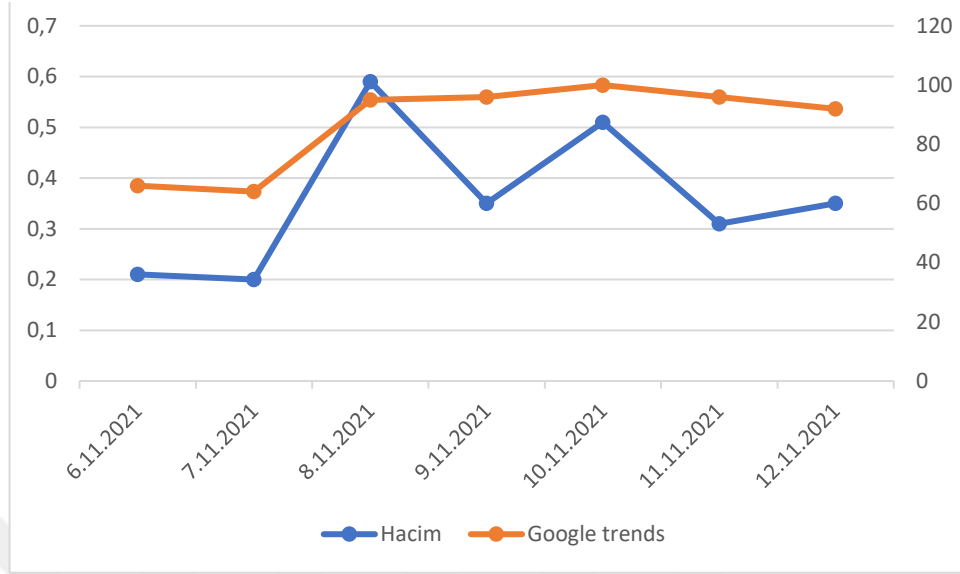
Şekil 3-14'te 'Bitcoin' kelimesinin Google'da aratılma sayıları ile Paribu kripto para piyasasındaki Bitcoin hacmi ile karşılaştırmalı grafiği gösterilmektedir. Grafikte görüldüğü üzere hacim verilerindeki artış ile 'Bitcoin' kelimesinin aratılma sayıları eş zamanlı olarak artmaya başlamıştır. 8 Kasım 2021 tarihinden sonra hacimde gerçekleşen düşüşe rağmen 'Bitcoin' kelimesinin tıklanma sayıları artış göstermiştir.

Şekil 3-15: İlgili Tarihlerde Paribu Platformunda Gerçekleşen Bitcoin Hacmi ve Google Trends Verilerinin Karşılaştırılması



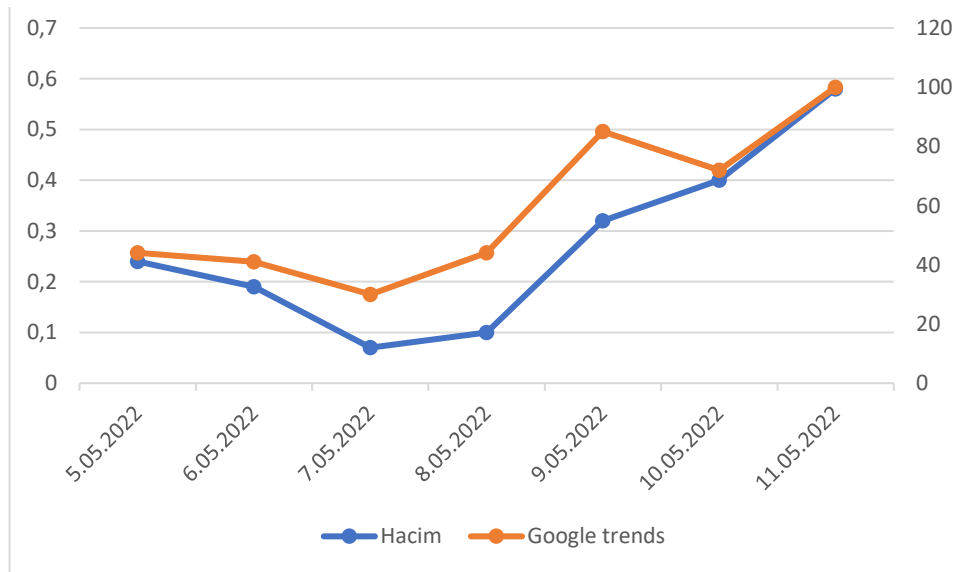
Şekil 3-15'te 'Bitcoin' kelimesinin 05-11 Mayıs 2022 tarih aralığında Google'da aratılma sayıları ile Paribu kripto para piyasasındaki Bitcoin hacminin karşılaştırmalı grafiği gösterilmektedir. Grafikte görüldüğü üzere hacim verilerindeki azalış ile 'Bitcoin' kelimesinin aranma sayıları da azalış göstermiştir. 9 Mayıs 2022 tarihinden sonra hacimde gerçekleşen artış ile 'Bitcoin' kelimesinin tıklanma sayıları artış göstermiştir.

Şekil 3-16: İlgili Tarihlerde BtcTurk Platformunda Gerçekleşen Bitcoin Hacmi ve Google Trends Verilerinin Karşılaştırılması



Şekil 3-16’da ‘Bitcoin’ kelimesinin 06-12 Kasım 2021 tarih aralığında Google’da aratılma sayıları ile BtcTurk kripto para piyasasındaki Bitcoin hacminin karşılaştırmalı grafiği gösterilmektedir. Grafikte görüldüğü üzere genel itibariyle Google üzerinden ‘Bitcoin’ kelimesinin aratılma sayısı hacim verisi eş zamanlı artış azalış göstermemiştir.

Şekil 3-17: İlgili Tarihlerde BtcTurk Platformunda Gerçekleşen Bitcoin Hacmi ve Google Trends Verilerinin Karşılaştırılması



Şekil 3-17’de ‘Bitcoin’ kelimesinin 05-11 Mayıs 2021 tarih aralığında Google’da aratılma sayıları ile BtcTurk kripto para piyasasındaki Bitcoin hacminin karşılaştırmalı grafiği gösterilmektedir. Grafikte görüldüğü üzere Google üzerinden ‘Bitcoin’ kelimesinin aratılma sayısı hacim verisi eş zamanlı artış azalış göstermiştir.

Paribu ve BtcTurk kripto para piyasaları için Bitcoin’in düşüş trendinde olduğu 05-11 Mayıs 2022 tarih aralığında hacim verileri ile eş zamanlı artış ve azalış göstermektedir. 06-12 Kasım 2021 tarih aralığında ise ‘Bitcoin’ kelimesi görece daha fazla aratılmıştır. Bitcoin’in yükseliş trendinde olduğu dönem yatırımcılar piyasayı takip edebilmek adına daha fazla konu ile ilgili bilgi edinme ihtiyacı hissetmiş olabilir. Kripto para piyasasının yükseliş trendinde olması ve tarihinin en yüksek seviyelerini görmesi nedeniyle yatırımcıların ilgi odağı haline gelmiştir.

3.7.4. Makine Öğrenmesi İle Duygu Analizi

Duygu analizi yapılırken makine öğrenmesi algoritmaları kullanılarak her iki veri setine ilişkin analizler gerçekleştirilmiştir. Literatür incelendiğinde en çok kullanılan ve en etkili şekilde sonuçlar veren algoritmalarından üç tanesi seçilmiştir. Bu amaçla karar ağacı, Random Forest ve Naive Bayes algoritmaları kullanılmıştır.

Analiz sürecinde öncelikle terim doküman matrisleri oluşturulmuştur. Algoritmaların performansları kıyaslanabilmesi amacıyla kesinlik, duyarlılık, doğruluk ve F-ölçütü hesaplanarak karşılaştırma yapılmıştır.

Analiz yapılma aşamasında veri setinin %70’i eğitim %30’u ise test veri seti olarak kullanılmıştır.

Tablo 3-4: Makine Öğrenmesi Sonuçları

Veri Seti	Algoritma	P	R	A	F
'veri1'	Karar Ağacı	0,25	0,72	0,84	0,37
	Naive Bayes	0,86	0,35	0,35	0,49
	Random Forest	0,34	0,85	0,85	0,48
'veri2'	Karar Ağacı	0,57	0,78	0,83	0,65
	Naive Bayes	0,38	0,67	0,72	0,48
	Random Forest	0,75	0,86	0,81	0,80

Tablo 3-4'te her iki veri seti için algoritma sonuçları gösterilmektedir. Kesinlik (P), duyarlılık (R), doğruluk (A), ve F-ölçütü (F) hesaplanarak gösterilmektedir. 'veri1' için yapılan analizde kesinlik ölçütü ve F-ölçütü için en iyi sonucu Naive Bayes algoritması verirken duyarlılık ve doğruluk için en iyi sonucu Random Forest algoritması vermektedir.

Kesinlik oranı gerçekte pozitif olup pozitif olarak etiketlenmiş tweetlerin (TP), gerçekte pozitif olup pozitif olarak etiketlenmiş tweetler ve gerçekte negatif olup pozitif olarak etiketlenmiş (FP) olan tweetlerin toplamına oranı olarak ifade edilmektedir. Kesinlik ölçütü için en iyi sonucu veren algoritma 'veri1' için Naive Bayes algoritması olmuştur. Gerçekte negatif olan tweetleri en az pozitif olarak etiketleyen algoritma Naive Bayes algoritması olmuştur denilebilir. 'veri2' için ise en iyi sonucu karar ağacı algoritması vermiştir.

Duyarlılık gerçekte pozitif olup pozitif olarak etiketlenen tweetlerin yine gerçekte pozitif olup pozitif olarak etiketlenen ve gerçekte pozitif olup negatif olarak etiketlenen tweetlerin toplamına oranı olarak ifade edilmektedir. Duyarlılık değerinde en iyi sonucu veren algoritma 'veri1' için de 'veri2' için de Random Forest algoritması olmuştur.

Doğruluk oranı, doğru olarak etiketlenmiş pozitif ve negatif tweetlerin toplamının, tüm tweetlerin toplamına oranı olarak ifade edilmektedir. Bu oranı ise en

yüksek başarıda veren algoritma ‘veri1’ için Random Forest algoritması iken ‘veri2’ için karar ağacı algoritması olmuştur.

F-ölçütü hesaplanabilmesi için kesinlik ve duyarlılık oranlarına ihtiyaç duyulmaktadır. F-ölçütü hesaplanarak sistem kesinlik ve duyarlılık ölçütlerine göre optimize edilmektedir. Bu optimizasyon sonucunda ‘veri1’ için en iyi sonucu Naive Bayes algoritması verirken, ‘veri2’ için en iyi sonucu Random Forest algoritması vermiştir.



SONUÇ

İnternetin yaygınlaşması ve gelişmesiyle birlikte insanlar hemen her konuda interneti kullanır hale gelmiştir. İnsanlar tatilde, işe ara verdiklerinde, uyumadan önce, toplu taşımada hemen hemen her anlarında bilgisayar, akıllı telefonlar gibi teknolojik aletleri kullanmaktadır. Teknolojinin günlük hayatta önemli bir parça haline gelmesiyle birlikte veri kavramı bilinen anlamdaki halinden çok daha farklı bir noktaya ulaşmıştır. Veri, kişi veya kurumların soru cevap yoluyla topladığı anket verilerinden çok daha ötesine ulaşmış durumdadır. Günümüzde herhangi bir yerde paylaşılan fotoğraflar, yazılan mesajlar, konum bilgileri, ziyaret edilen internet siteleri ve internette aranan içerikler gibi kişisel teknolojik aletler aracılığıyla yapılan tüm aktiviteler kayıt altına alınmaktadır. Özellikle Web 2.0 teknolojisi ile birlikte hayatımıza giren sosyal medya internet kullanımını çok daha çekici hale getirmiştir. İnsanlar sosyal medyanın gücünü keşfederek kendilerini ifade etme biçimi olarak sosyal medya kullanmaya başlamışlardır. Bu nedenle sosyal medya, son yıllarda en çok dikkat çeken araştırma konuları arasında yer almaktadır. Sosyal medyanın popülerliği ve büyük veri elde edebilmek için oldukça etkili bir araç olması nedeniyle metin madenciliği için önemli bir veri kaynağı olarak görülmektedir.

Veri boyutunda gerçekleşen büyüme nedeniyle yapılandırılmamış ya da yarı yapılandırılmış veri tiplerinin ortaya çıkmasına neden olmuştur. Veri oluşturmak günümüzde oldukça hızlı olmasıyla birlikte büyük veri kavramı da hayatımıza girmiştir. İşletmelerin kullandığı teknolojik alt yapıların gelişmesi ve akıllı cihazların kullanılması da veri elde edebilmek için kolaylık sağlamaktadır. Çok büyük miktarda verilerin elde edilmesiyle birlikte bu büyük hacimli veri yığınlarını anlamlandırma ihtiyacı doğmuştur. Yapılandırılmamış veya yarı yapılandırılmış halde olan karmaşık ve anlaşılmaz durumda olan veri yığınlarını anlamlandırma ihtiyacı doğmuştur. Büyük veri, gelişmiş bilgisayar teknolojileri aracılığıyla hızlı bir şekilde oluşturulur ve depolanır. Depolanan büyük veri, veriye uygun olan yöntem veya yöntemler kullanılarak analiz edilmektedir. Anlamsız bilgi yığını olarak görünen büyük veri, veri analiz yöntemleri kullanılarak değerli bilginin oluşturulmasını sağlamaktadır.

Yapılandırılmamış veriye en uygun örnek metin verileridir. Büyük metin yığınlarından oluşan metin verilerinin işlenmesi ve anlamlı sonuçlar elde edilmesi

süreci olan metin madenciliği, istatistik bilimi için oldukça önemlidir. Metin madenciliği süreçleri uygulanarak metinler işlenip önemli ve anlamlı bilgilere ulaşılabilmektedir. Yapılan bu tez çalışmasında da yapılandırılmamış, karmaşık yapıda olan metin verilerine uygulanan ve son yıllarda oldukça popüler olan duygu analizi uygulanmıştır. Duyguların ölçülebilir olması ve analiz edilebilmesi sayesinde insan davranışları ölçülebilir ve anlaşılabilir hale gelmektedir.

Literatürde duygu analizi kullanılan çalışmaların birçoğu İngilizce metinler üzerinde gerçekleştirilmiştir. İngilizce'nin dil yapısının sadeliği ve hazır İngilizce sözlüklerin bulunması bunun en önemli nedeni olarak görülmektedir.

Bu tez çalışması kapsamında ise Türkçe metinler kullanılarak duygu analizi gerçekleştirilmiştir. Sosyal medya platformu olan Twitter üzerinden elde edilen yorumlar ile veri setleri oluşturulmuştur. Türkçe metinler için oluşturulan hazır bir sözlük olmadığından üç farklı sözlük oluşturulmuştur. Bu sözlükler hem Türkçe sözlük hem de veri seti taranarak oluşturulmuştur. Sözlükler pozitif kelimeler, negatif kelimeler ve durak kelimeleri içermektedir. Literatürde Türkçe bir sözlük olmadığından oluşturulan sözlüklerin literatüre katkı sağlaması amaçlanmaktadır.

Duygu analizi alanında yapılan çalışmaların birçoğu bilgisayar yazılımı geliştirme üzerine fen bilimleri alanında olduğu gözlenmiştir. Sosyal bilimler alanında duygu analizi çalışmaları ise oldukça kısıtlı ve az sayıdadır. Yapılan tez çalışmasında makine öğrenmesi yöntemleri kullanılarak duygu analizi gerçekleştirilmiştir. Bu bağlamda analiz gerçekleştirilirken makine öğrenmesi yaklaşımları ile olasılığa dayalı algoritmalar kullanılmıştır.

Geleceğe yönelik tahminlerde bulunabilmek bilim dünyası için oldukça önemlidir. Sağlık, ekonomi, sanayi gibi birçok alanda önceden tahminler bulunmak kişilerin, kurum veya kuruluşların gelecek için hazırlıklı olmalarını, planlarını daha etkili bir şekilde yapmalarını sağlamaktadır. Kripto para piyasasında da geleceğe yönelik değer tahmini yapabilmek amacıyla birçok çalışma gerçekleştirilmiştir. Bu tahminler genellikle zaman serisi gibi geleneksel yöntemlere dayanarak gerçekleştirilen çalışmalarda karşımıza çıkmaktadır.

Yapılan tez çalışmasının uygulama kısmında, kripto para piyasasındaki dalgalanmaların sosyal medyada nasıl yankı uyandırdığını ve kullanıcıların kripto para piyasasına karşı tutumlarının ne yönde değiştiğini gözler önünde sermek amacıyla

duygu analizi gerçekleştirilmiştir. Kripto para piyasasının yükseliş trendinde olduğu bir dönem ve düşüş trendinde olduğu bir dönem seçilmiştir. Seçilen dönemlere ilişkin tweet verileri günlük olarak elde edilmiştir. Twitter'dan elde edilen veriler düzenli olarak birleştirilmiştir. Birleştirilen veriler sonucunda iki farklı veri seti elde edilmiştir. Oluşturulan veri setleri ön işleme aşamasından geçtikten sonra duygu analizi gerçekleştirilmiştir.

Verilerin elde edilmesi açık kaynak kodlu olan R yazılımı kullanılmıştır. Verilerin ön işleme aşamasında ise kullanım kolaylığı nedeniyle RapidMiner programı kullanılmıştır. Analiz aşamasında ise yine R yazılımı kullanılmıştır. R yazılımı birçok farklı imkan sunabilmesi, temel istatistikten makine öğrenmesine kadar kullanım alanının geniş olması nedeniyle son yıllarda oldukça popüler hale gelmiştir. Ayrıca en büyük avantajı da açık kaynak kodlu, ücretsiz bir yazılım olmasıdır.

Tez çalışmasının en büyük kısıtı R'da kullanılan '*twitteR*' paketinin kısıtlarıdır. Twitter kullanılarak yapılan tüm çalışmalarda bu paketin kullanılması zorunludur. '*twitteR*' paketi elde edilen verilerde 140 karaktere kadar analiz yapabilmektedir. Fakat günümüzde 280 karaktere kadar Twitter paylaşımı yapılabilmektedir. Ayrıca Twitter'dan veri elde ederken her bir istekte 18000 adet tweet ile sınır konulmaktadır. Tweetler günlük olarak elde edildiğinden 18000'den fazla paylaşım içeren günlerde yalnızca verilerin izin verildiği kadarına ulaşılmıştır.

Verilerin analiz edebilmesi için en önemli aşama ön işleme aşamasıdır. Ön işleme aşaması çalışmanın en uzun süren aşaması olmuştur. Ön işleme aşamasında öncelikle tekrar eden tweetler ayıklanmıştır. Hem #bitcoin hem de #btc hashtah ile paylaşılan tweetler ayrı ayrı elde edildiğinden ve her iki hashtag kullanılarak paylaşılan tweetler olduğundan tekrar eden birçok içerik olmuştur. Daha sonra reklam içerikli, konu ile ilgisi olmayan tweetler silinmiştir. Bu işlem için her iki veri seti de taranmıştır. Tekrar eden tweetlerin temizlenmesi gerçekleştirildikten sonra tweetlerin temizlenmesi aşamasına geçilmiştir. Bu aşamada tüm semboller, noktalama işaretleri, kullanıcı adları, fotoğraf ve videolar silinip tüm harfler küçük harf olarak düzenlenmiştir. Bu işlemler için RapidMiner programı kullanılmıştır.

Ön işleme aşamasında durak kelimelerin ayıklanabilmesi ve analize dahil edilmemesi için durak kelimeler sözlüğü oluşturulmuştur. Böylece bağlaç, edat gibi analize herhangi bir katkısı olmayan sözcükler ayıklanmıştır.

Gerçekleştirilen analiz iki aşamada gerçekleştirilmiştir. Öncelikle duygu skorları hesaplanmıştır. Oluşturulan Türkçe pozitif ve negatif duygu sözlükleri kullanılarak duygu skorları hesaplanabilmektedir. Her iki veri setinin içerdiği her tweetin duygu skoru hesaplanmış ve hesaplanan duygu skorlarının ortalamaları alınmıştır. Duygu skorları hesaplandıktan sonra pozitif, negatif ve nötr duygulu tweetlerin retweet ve favori edilme sayıları betimsel istatistik olarak gösterilmiştir. İkinci aşamada ise üç farklı makine öğrenmesi algoritmaları kullanılarak duygu analizi gerçekleştirilmiştir. Elde edilen sonuçlar tablolastırılarak analiz gerçekleştirilmiştir.

Son yıllarda kripto para piyasasının popüler olması ve piyasada yatırımcı olmayanların dahi ilgisini çekmesi sebebiyle sosyal medyada birçok kişi kripto para piyasası hakkında yorum paylaşmaktadır. Sadece 1 günde #btc hashtag kullanılarak 20000 civarı tweet paylaşılmaktadır. Bu nedenle yalnızca yedi günlük iki veri seti halinde çalışılabilmektedir. Verilerin elde edildiği zaman diliminde dahi oldukça büyük bir veri seti oluşturulmuştur. Verilerin analiz edilebilmesi için kullanılan bilgisayarın RAM kapasitesi oldukça önemli bir kısıt haline gelmiştir. Daha gelişmiş bilgisayarlar kullanılarak daha geniş bir zaman aralığı seçilebilir ve daha büyük veri setleri kullanılarak analiz gerçekleştirilebilir. Literatürdeki çalışmalarda Türkçe metinler kullanılarak kripto paralar üzerine duygu analizi daha önce yapılmamıştır. Yapılan bu çalışma daha sonra yapılacak olan çalışmalar için bir örnek teşkil etmektedir.

Verilerin elde edildiği ilgili dönemlere bakıldığında kripto para piyasasına her dönemde oldukça yoğun ilgi olduğu söylenebilmektedir. Bitcoin fiyatı ve duygu skorları arasında bir ilişki olup olmadığına bakıldığında, Bitcoin fiyatının yükseliş trendinde olduğu dönem daha olumlu tweetlerin daha fazla etkileşim aldığı gözlenmiştir. Fakat Bitcoin'in düşüş trendinde olduğu dönemde de olumlu ve nötr tweetlerin aldığı etkileşim yadsınamayacak kadar fazla olduğu görülmektedir. Yatırımcılar kripto para piyasasına olumlu ya da olumsuz yaklaşım sergilemektense her durumda optimist olarak yaklaşarak maddi varlıklarını koruma ve varlıklarını arttırma amacı ile hareket etmektedirler.

Son yıllarda merkeziyetsiz kripto paralar dışında kurum ve kuruluşların ürettiği, bir merkezi otoriteye bağlı olan kripto paralar da piyasalarda işlem görmektedir. Yatırımcıların her koşulda kripto para piyasasına olan ilgisi sayesinde kripto para üretmek, piyasaya sürmek isteyen kurum ve kuruluşlar daha cesur davranabilmektedir. Böylece kripto para piyasalarındaki para birimleri gün geçtikçe artmakta ve yatırımcılara yeni fırsatlar sunmaktadır.

Kripto paralar her geçen gün ülkeler tarafından yasal ödeme aracı olarak görülmeye başlamıştır. Kripto paraların blok zinciri teknolojisi sayesinde kontrol edilebilir olması ve yasallaşması ile çek, EFT, kredi kartları gibi sistemleri geride bırakacağı öngörülmektedir. Ayrıca kredi kartı veya nakit kullanımında olabilecek hırsızlık durumlarına karşı da kripto paraların saklandığı bellek cüzdanların kullanılması bir tedbir olacağı düşünülmektedir.

İleride yapılacak olan çalışmalarda daha yüksek RAM gücüne sahip bilgisayarlar ile daha geniş bir zaman aralığında elde edilen veriler ile duygu analizi gerçekleştirilerek günlük olarak kripto para hacim veya fiyat verileri ile arasındaki nedensellik ilişkisi kurulabilir. Ayrıca günümüzde Ripple, Ethereum, Avalanche gibi oldukça popüler birçok altcoin bulunmaktadır. Analize daha fazla kripto para dahil edilerek de duygu analizi çalışması gerçekleştirilebilir.

KAYNAKÇA

- Abraham, J., Higdon, D., Nelson, J., & Ibarra, J. **“Cryptocurrency price prediction using tweet volumes and sentiment analysis”**, SMU Data Science Review, 1(3), 2018, 1.
- Agarwal, A., Xie, B., Vovsha, I., Rambow, O., Passonneau, R. J. **“Sentiment analysis of twitter data”**, In Proceedings of the workshop on language in social media (LSM 2011), 2011, June, pp. 30-38.
- Akay, E. Ç. **“Ekonometride Yeni Bir Ufuk: Büyük Veri ve Makine Öğrenmesi”**, Sosyal Bilimler Araştırma Dergisi, 2018, 7(2), 41-53.
- Akpınar, H. **“Data, Veri Madenciliği – Veri Analizi”**, 2017.
- Alajmi, A., Saad, E.M., Darwish, R.R. **“Toward an Arabic Stop-Words List Generation”**, International Journal of Computer Application, Vol.46, No:8, 2012, pp. 8-13
- Alharby, M., & Van Moorsel, A. **“Blockchain-based smart contracts: A systematic mapping study”**, 2017, arXiv preprint arXiv:1710.06372.
- Ashford, K. & Schmidt, J. **“What Is An IPO?”**, Forbes Advisor, <https://www.forbes.com/advisor/investing/initial-public-offering-ipo/>
- Ayhan, S., Erdoğan, Ş. **“Destek Vektör Makineleriyle Sınıflandırma Problemlerinin Çözümü İçin Çekirdek Fonksiyonu Seçimi”**, Eskişehir Osmangazi Üniversitesi İİBF Dergisi, 9(1), 2014, ss. 175-198
- Babüroğlu, B., Tekerek, A., Tekerek, M. **“Türkçe İçin Derin Öğrenme Tabanlı Doğal Dil İşleme Modeli Geliştirilmesi”**, 2019
- Benbya, H., & McKelvey, B. **“Bitcoin and Cryptocurrency Technologies”**, Journal of Information Technology, 2006, 21(4), 284-298.

- Bondarenko, O., Kichuk, O., & Antonov, A. **“The possibilities of using investment tools based on cryptocurrency in the development of the national economy”**, Baltic Journal of Economic Studies, 2019, 5(2), 10-17.
- Bunjaku, F., Gjorgieva-Trajkovska, O., & MitevamerKacarski, E. **“Cryptocurrencies – advantages and disadvantages”**, Journal of Economics, 2017, 2(1), 31-39.
- Buterin, V. **“A Next Generation Smart Contract & Decentralized Application Platform”**. Ethereum White Paper, 2014, 1-36.
- Chris Clark **“Bitcoin Internals”**, July 31, 2013
- Colianni, S., Rosales, S., & Signorotti, M. **“Algorithmic trading of cryptocurrency based on Twitter sentiment analysis”**, 2015, CS229 Project, 1(5), 1-4.
- Coşkun, A., & Ülker, Ü. **“Ulusal Bilgi Güvenliğine Yönelik Bir Kriptografi Algoritması Geliştirilmesi ve Harf Frekans Analizine Karşı Güvenirlilik Tespiti”**, Bilişim Teknolojileri Dergisi, 2013, 6(2), 31.
- Cox, J. **“Bitcoin and Digital Currencies”**, 2013, Laissez Faire Books.
- Çabuk, H., Yüksel, Ç., Mocan, Z., Diri, B., Amasyalı, M.F. **“Metin Analizi ve Sorgulama”**, 11. Sinyal İşleme ve İletişim Uygulamaları Kurultayı, İstanbul, Haziran 2003, ss. 679-681
- Çarkacıoğlu, A. **“Kripto-Para Bitcoin”**, Sermaye Piyasası Kurulu Araştırma Dairesi Araştırma Raporu, 2016
- Çokluk, Ö. **“Lojistik regresyon analizi: Kavram ve uygulama”**, Kuram ve uygulamada eğitim bilimleri, 2010, 10(3), 1357-1407.

- Dang, S., Ahmad, P. H **“Text mining: Techniques and its application”**, International Journal of Engineering & Technology Innovations, 2014, pp: 22-25
- Dean, J. **“Big Data, Data Mining, and Machine Learning: Value Creation for Business Leaders and Practitioners”**, John Wiley & Sons, 2014.
- Deng, S., Sinha, A. P., Zhao, H **“Adapting sentiment lexicons to domain-specific social media texts”**, Decision Support Systems, 2017, 94, 65-76.
- Dey, L., Chakraborty, S., Biswas, A., Bose, B., & Tiwari, S. **“Sentiment analysis of review datasets using naive bayes and k-nn classifier”**, 2016, arXiv preprint arXiv:1610.09982.
- Do, T.D., Hui, S.C., Fong, A.C.M. **“Associative Feature Selection For Text Mining”**, International Journal of Information Technology, Vol.12, No:4, 2006, pp. 59-68
- DuPont, Q. **“Cryptocurrency And Blockchains”**, 2019
- Fan, W., Wallace, L., Rich, S., Zhang, Z. **“Tapping the power of text mining”**, Communications of the ACM, 2006, 49(9), 76-82.
- Feldman, R., & Sanger, J. **“The text mining handbook: advanced approaches in analyzing unstructured data”**, 2007, Cambridge university press.
- Göker, H. , Tekedere, H. **“FATİH Projesine Yönelik Görüşlerin Metin Madenciliği İle Otomatik Değerlendirilmesi”**, Bilişim Teknolojileri Dergisi, 2017, 10(3), 291-299
- Gupta, S., Kumar, D., Sharma, A. **“Data Mining Classification Techniques Applied For Breast Cancer Diagnosis and Prognosis”**, Indian Journal of Computer Science and Engineering (IJCSE), 2011, Vol:2, No:2, pp. 188-195

- Halta, A., Alkan, A.,
Karabulut, M. **“Metin Sınıflandırmada Sezgisel Arama Algoritmalarının Performans Analizi”**, Gazi Üniversitesi Mühendislik ve Mimarlık Fakültesi Dergisi, Vol.30, No.3, 2015, pp. 417-427
- Hassan, M. K., Hudaefi, F. A.,
& Caraka, R. E. **“Mining netizen’s opinion on cryptocurrency: sentiment analysis of Twitter data”**, 2021, Studies in Economics and Finance.
- Hotho, A., Nürnberger, A., &
Paaß, G. **“A brief survey of text mining”**, In Ldv Forum, 2005, May, Vol. 20, No. 1, pp. 19-62).
- Inamdar, A., Bhagtani, A.,
Bhatt, S., & Shetty, P. M. **“Predicting cryptocurrency value using sentiment analysis”**, In 2019 International Conference on Intelligent Computing and Control Systems (ICCS), 2019 May, pp. 932-934, IEEE.
- İlhan, S., Duru, N., Karagöz,
Ş., & Sağır, M. **“Metin madenciliği ile soru cevaplama sistemi”**, Elektronik ve Bilgisayar Mühendisliği Sempozyumu (ELECO), 2008, Bursa, 26, 30.
- İşler, B., Takaoğlu M., &
Küçükali, U. F. **“Blok Zinciri ve Kripto Paraların İnsanlığa Etkileri”**, Yeni Medya Elektronik Dergisi, (2019), 3(2), 71-83.
- J. Stark **“Making sense of blockchain smart contracts”**, Available online at: <http://www.coindesk.com/making-sense-smart-contracts/> [e.t.: 09.03.2022].
- Jerman C. Kaminski, **“Nowcasting the Bitcoin Market with Twitter Signals”**, 2016, MIT Media Lab.
- Jivani, A.G. **“A Comparative Study of Stemming Algorihms”**, Int. J. Comp. Tech. Appl, 2(6), 2011, 1930-1938.
- Karaarslan, E., Akbaş, M.F. **“Blokzinciri Tabanlı Siber Güvenlik Sistemleri”**, Uluslararası Bilgi Güvenliği Mühendisliği Dergisi, Vol.3, No.2, 2017, s.s.: 18.

- Kavzoğlu, T., Çölkesen, İ. **“Karar ağaçları ile uydu görüntülerinin sınıflandırılması: Kocaeli örneği”**, Harita Teknolojileri Elektronik Dergisi, 2010, 2(1), 36-45.
- Kılınç, D., Borandağ, E., Yücalar, F., Tunalı, V., Şimsek, M., Özçift, A. **“KNN Algoritması ve R Dili ile Metin Madenciliği Kullanılarak Bilimsel Makale Tasnifi”**, Marmara Fen Bilimleri Dergisi, 2016, pp: 89-94
- Kim, Y. B., Kim, J. G., Kim, W., Im, J. H., Kim, T. H., Kang, S. J., & Kim, C. H. **“Predicting fluctuations in cryptocurrency transactions based on user comments and replies”**. PloS one, 2016, 11(8), e0161197.
- Kotu, V., & Deshpande, B. **“Predictive analytics and data mining: concepts and practice with rapidminer”**, 2014, Morgan Kaufmann.
- Kraaijeveld, O., & De Smedt, J. **“The predictive power of public Twitter sentiment for forecasting cryptocurrency prices”**, Journal of International Financial Markets, Institutions and Money, 2020, 65, 101188.
- Krishnaiah, V., Narsimha, G., Chandra, S. **“Diagnosis of Lung Cancer Prediction System Using Data Mining Classification Techniques”**, International Journal of Computer Science and Information Technologies, Vol:4, No:1, 2013, pp.39-45
- Kumar, A., Smith, C. **“Crypto Currencies An Introduction to not so Funny Moneys”**, Journal of Economic Surveys, 2017, Vol:32, No:5, pp:1531-1559
- Kuzucu, K. **“Müşteri Memnuniyeti Belirlemek İçin Metin Madenciliği Tabanlı Bir Yazılım Aracı”**, Maltepe Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, İstanbul, 2015, ss:13
- Lamon, C., Nielsen, E., & Redondo, E. **“Cryptocurrency price prediction using news and social media sentiment”**. SMU Data Sci. Rev, 2017, 1(3), 1-22.

- Li, B., Yu, S., & Lu, Q. **“An improved k-nearest neighbor algorithm for text categorization”**, 2003, *arXiv preprint cs/0306099*.
- Li, T. R., Chamrajnagar, A. S., Fong, X. R., Rizik, N. R., & Fu, F. **“Sentiment-based prediction of alternative cryptocurrency price fluctuations using gradient boosting tree model”**. *Frontiers in Physics*, 2019, 7, 98.
- Liddy, E.D. **“Natural Language Processing”**, In *Encyclopedia of Library and Information Science*, Syracuse University, 2001
- Liu, B. **“Sentiment analysis: Mining opinions, sentiments, and emotions”**, Cambridge university press, 2020, pp. 1-8
- Mather, P., & Tso, B. **“Classification methods for remotely sensed data”**, 2016, CRC press.
- Matta, M., Lunesu, I., & Marchesi, M. **“Bitcoin Spread Prediction Using Social and Web Search Media”**, In *UMAP workshops*, 2015, June, pp. 1-10.
- Medhat, W., Hassan, A., & Korashy, H. **“Sentiment analysis algorithms and applications: A survey”**, *Ain Shams engineering journal*, 5(4), 2014, 1093-1113.
- Nakamoto, S. **“Bitcoin: A Peer-to-Peer Electronic Cash System”**, *Journal for General Philosophy of Science*, 2008
- Narayanan, A., Bonneau, J., Felten, E., Miller, A., & Goldfeder, S. **“Bitcoin and cryptocurrency Technologies”**, *Network Security*, 2016(8), 4.
- Nick, T. G., & Campbell, K. M. **“Logistic regression. Topics in biostatistics”**, 2007, 273-301.
- Oğuzlar, A., Kızılkaya, Y.M. **“Metin Madenciliğinde Duygu Analizi: R Uygulamalı”**, 2019

- Ozbay, Ö. **“Veri Madenciliği Kavramı ve Veri Madenciliği Uygulamaları”**, Uluslararası Eğitim Bilimleri Dergisi, Vol:2, No: 5 , 2015, pp: 262-272
- Perikos, I., & Hatzilygeroudis, I. **“Recognizing emotions in text using ensemble of classifiers”**. Engineering Applications of Artificial Intelligence, 2016, 51, 191-201.
- Poria, S., Majumder, N., Gelbukh, A., Akhtar, M. S., Cambria, E., & Ekbal, A. **“IARM: Inter-aspect relation modeling with memory networks in aspect-based sentiment analysis”**. In Proceedings of the 2018 conference on empirical methods in natural language processing (pp. 3402-3411).
- Pozzi, F., Fersini, E., Messina, E., & Liu, B. **“Sentiment analysis in social networks”**, 2016, Morgan Kaufmann.
- Ravindran, Sharan K., Vikram Garg: **Mastering Social Media Mining with R**, Packt Publishing Ltd., 2015.
- Rodeck, D. & Curry, B. **“What Is Blockchain?”**, 2022, Forbes Advisor, <https://www.forbes.com/advisor/investing/cryptocurrency/what-is-blockchain/>
- Rognone, L., Hyde, S., & Zhang, S. S. **“News sentiment in the cryptocurrency market: An empirical comparison with Forex. International Review of Financial Analysis”**, 2020, 69, 101462.
- Rotman, S. **“Bitcoin Versus Electronic Money”**, 2014.
- Saif, H., He, Y., Alani, H. **“Semantic sentiment analysis of twitter”**, In International semantic web conference, 2012, November, pp. 508-524.
- Sattarov, O., Jeon, H. S., Oh, R., & Lee, J. D. **“Forecasting Bitcoin Price Fluctuation by Twitter Sentiment Analysis”**, International Conference on Information Science and Communications

- Technologies (ICISCT), 2020, November, (pp. 1-4). IEEE.
- Sreeja, I., Sunny, J. V., & Jatian, L. **“Twitter sentiment analysis on airline tweets in India using R language”**, In Journal of Physics: Conference Series, 2020, (Vol. 1427, No. 1, p. 012003). IOP Publishing.
- Srividhya, V., Anita, R. **“Evaluating Preprocessing Techniques in Text Categorization”**, International Journal of Computer Science and Application Issue, 2010, pp. 49-51
- Sun, S., & Huang, R. **“An adaptive k-nearest neighbor algorithm”**. In 2010 seventh international conference on fuzzy systems and knowledge discovery, IEEE, 2010, August, Vol. 1, pp. 91-94,
- Surana, S., Jain, A. **“What Is Bitcoin and How Does It Work?”** March 2022, E.T.: 30/03/2022 <https://www.forbes.com/advisor/in/investing/what-is-bitcoin-and-how-does-it-work/>
- Tanrıverdi, M., Uysal, M., & Üstündağ, M. T. **“Blok zinciri Teknolojisi Nedir? Ne Değildir?: Alanyazın İncelemesi”**, Bilişim Teknolojileri Dergisi, 2019, 12(3), 203-217.
- Torres J. A. **“Blockchain Technology & The Cryptocurrency Revolution”**, 2021
- Ünal, G. & Uluyol, Ç. **“Blok Zinciri Teknolojisi”**, Bilişim Teknolojileri Dergisi , 2020, 13 (2) , 167-175 . DOI: 10.17671/gazibtd.516990
- Vaidhyanathan, J.& Jain, A. **“What Is Cryptocurrency And How Does It Work?”**, 2022, Forbes Advisor, Çevrimiçi: <https://www.forbes.com/advisor/in/investing/what-is-cryptocurrency-and-how-does-it-work/>

- Vatanen, T., Väyrynen, J. J., Virpioja, S. **“Language Identification of Short Text Segments with N-gram Models”**, In LREC, 2010 May, pp. 3423- 3430
- Verma, T., Renu, Gaur D. **“Tokenization And Filtering Process in Rapidminer”**, International Journal of Applied Information Systems, Vol.7, No:2, 2014, pp.16-18.
- Voican, Oona **“Using Data Mining Methods To Solve Classification Problems In Financial Banking Institutions”**, Economic Computation and Economic Cybernetics Studies and Research Issue, 2020, pp:159-175
- Wild, J. **"Blockchain believers seek to shake-up financial services."** *Financial Times* 14 (2015).
- Wright, C. **Cryptography and Bitcoin**, 2020, Çevrimiçi: <https://craigwright.net/blog/law-regulation/cryptography-and-bitcoin/>
- Yakut, Y. B. E. T. Y., YAKUT, E., & Yavuz, S **“Yapay Sinir Ağları Ve Destek Vektör Makineleri Yöntemleriyle Borsa Endeksi Tahmini”**, Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 2014, 19(1), 139-157.
- Yang, Y., & Liu, X. **“A re-examination of text categorization methods”**, In Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval, 1999 August, pp. 42-49.
- Young, T., Hazarika, D., Poria, S., Cambria, E. **“Recent Trends in Deep Learning Based Natural Language Processing”**, IEEE Computational Intelligence Magazine, 2018, pp. 55-75

Zhao, Z., & Liu, H.

“Spectral feature selection for supervised and unsupervised learning”, In Proceedings of the 24th international conference on Machine learning, 2007, June, pp. 1151-1157.

Žižka, T., Podaras, A.,
Nejedlová, D., & Kubát, D.

“A Business Intelligence Solution For Business Continuity And Safety Management In Public Universities”, 2020, Ad Alta: Journal of Interdisciplinary Research, 10(2).

<https://apps.twitter.com>

[https://developer.twitter.com/en/docs/tweets/timelines/api-reference/get-statusesuser_
timeline](https://developer.twitter.com/en/docs/tweets/timelines/api-reference/get-statusesuser_timeline)

<https://developer.twitter.com/en/docs/tweets/search/overview>

<https://help.twitter.com/en/rules-and-policies/twitter-limits>

<https://www.binance.com/tr>

<https://www.ecb.europa.eu/home/html/index.en.html>

<https://coinmarketcap.com/>