



**KRİPTO PARA FİYAT DEĞİŞİMLERİNİN
BÜYÜK VERİ ARAÇLARI İLE ANALİZİ**

Yüksek Lisans Tezi

Furgan BAYRAKTAR

Eskişehir 2020

**KRİPTO PARA FİYAT DEĞİŞİMLERİNİN BÜYÜK VERİ ARAÇLARI
İLE ANALİZİ**

Furgan BAYRAKTAR

YÜKSEK LİSANS TEZİ

**Endüstri Mühendisliği Anabilim Dalı
Danışman: Doç. Dr. Zehra KAMIŞLI ÖZTÜRK**

**Eskişehir
Eskişehir Teknik Üniversitesi
Lisansüstü Eğitim Enstitüsü
Haziran 2020**

ÖZET

KRİPTO PARA FİYAT DEĞİŞİMLERİNİN BÜYÜK VERİ ARAÇLARI İLE ANALİZİ

Furgan BAYRAKTAR

Endüstri Mühendisliği Anabilim Dalı

Eskişehir Teknik Üniversitesi, Lisansüstü Eğitim Enstitüsü, Haziran 2020

Danışman: Doç. Dr. Zehra KAMIŞLI ÖZTÜRK

Tarih boyunca çoğu insanın en büyük problemlerinden biri, gelirlerinin giderlerini karşılamada yetersiz kalması olmuştur. Günümüzde insanlar daha fazla gelir elde etmek için farklı borsalarda alım-satım işlemleri yapmaktadırlar. Kripto paralar, alım-satım işlemleri için kullanılan metallerden bir tanesidir. Borsalarda alım-satım stratejileri belirlenirken temel ve teknik analiz yöntemleri kullanılmaktadır. Teknik analiz, geçmiş fiyat hareketleri incelenerek, gelecekteki fiyat değişimlerini tahmin etmeye çalışmak olarak açıklanabilir. Teknik analiz ile alım-satım stratejileri belirlenirken çok büyük verilerle karşılaşmaktadır. Bu durum, verilerin analizini zorlaştırmakta ve uygulanacak stratejilerin hatalı olma olasılığını arttırmaktadır. Bu büyük verileri doğru analiz edemeyenler kripto para yatırımlarından büyük zararlarla karşılaşabilmektedirler.

Kripto paraların yakın zamanda ortaya çıkması sebebiyle, kripto para fiyat değişimlerinin analizi hakkında çok fazla çalışma mevcut değildir. Bu çalışmada, en büyük piyasa değerine sahip 3 kripto paranın verileri kullanılmıştır. Bu kripto paralar sırasıyla Bitcoin, Ethereum ve XRP'dir. Bu kripto paraların fiyat değişimlerinin analizi için büyük veri analiz yöntemlerinden K En Yakın Komşu (KNN), Destek Vektör Regresyonu (DVR) ve Rastgele Orman (RO) algoritmaları kullanılmıştır. Çalışma sonucunda, tahmini verilerin gerçek verilerle karşılaştırılması yapılmıştır. Karşılaştırma sonucunda tüm veri setlerinde en başarılı sonucu Destek Vektör Regresyonunun verdiği görülmüştür.

Anahtar Kelimeler: Kripto Para, Bitcoin, Ethereum, XRP, Büyük veri, KNN, DVR, RO, Finans

ABSTRACT

ANALYSIS OF CRYPTOCURRENCY PRICE CHANGES WITH BIG DATA TOOLS

Furgan BAYRAKTAR

Department of Industrial Engineering

Eskişehir Technical University, Institute of Graduate Programs, June 2020

Supervisor: Doç. Dr. Zehra KAMIŞLI ÖZTÜRK

Throughout history, one of the biggest problems of most people has been the inadequacy of their income. Nowadays, people are making trading on different exchanges to generate more revenue. Cryptocurrencies is one of the commodities used for trading transactions. Basic and technical analysis methods are used in the determination of trading strategies in stock exchanges. Technical analysis can be explained by examining past price movements and trying to predict future price changes. While determining the trading strategies with technical analysis, very large data are encountered. This situation makes it difficult to analyze the data and increases the likelihood of incorrect strategies. People who are unable to analyze these big data correctly may face huge losses from cryptocurrency investments. Due to the recent emergence of cryptocurrencies, there is not much study on the analysis of changes in cryptocurrency prices. In this study, data of the 3 cryptocurrencies with the largest market value are used. These cryptocurrencies are Bitcoin, Ethereum and Xrp. For the analysis of price changes of these cryptocurrencies, K Nearest Neighbors (KNN), Support Vector Regression (SVR) and Random Forest (RF) algorithms were used. As a result of the study, the estimated data is compared with the actual data. As a result of the comparison, it was seen that the Support Vector Regression gave the most successful result in all data sets.

Keywords: Cryptocurrency, Bitcoin, Ethereum, XRP, Big data, KNN, SVR, RF, Finance

TEŞEKKÜR

Yüksek lisans öğrenimim ve tüm tez çalışması aşaması boyunca benden bilgisini ve desteğini esirgemeyen, en zor durumlarda pozitif yaklaşımı ve deneyimi ile beni daha iyiye yönlendiren ve gelişmeye yardımcı olan tez danışmanım Doç. Dr. Zehra KAMIŞLI ÖZTÜRK'e sonsuz teşekkürlerimi sunarım.

Yüksek lisans tez çalışması süresince ve hayatımın her aşamasında bana güvenen ve destek olan aileme, yüksek lisansa başlamam konusunda beni teşvik eden ve yüksek lisans boyunca bana destek olan Uzm. Dr. Ahmet Yaşar Zengin, Dursun Ali Zengin, Prof.Dr. Akın TURGUT'a, değerli dostlarım Tuğçe YAVUZ, Hatice AY, Ekrem KÖSE ve Erdem KÖSE'ye teşekkürlerimi sunarım.

ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ

Bu tezin bana ait, özgün bir çalışma olduğunu; çalışmamın hazırlık, veri toplama, analiz ve bilgilerin sunumu olmak üzere tüm aşamalarında bilimsel etik ilke ve kurallara uygun davrandığımı; bu çalışma kapsamında elde edilen tüm veri ve bilgiler için kaynak gösterdiğimi ve bu kaynaklara kaynakçada yer verdiğimi; bu çalışmanın Eskişehir Teknik Üniversitesi tarafından kullanılan “ bilimsel intihal tespit programı”yla tarandığını ve hiçbir şekilde “ intihal içermediğini” beyan ederim. Herhangi bir zamanda, çalışmamla ilgili yaptığım bu beyana aykırı bir durumun saptanması durumunda, ortaya çıkacak tüm ahlaki ve hukuki sonuçları kabul ettiğimi bildiririm.

.....

Furgan BAYRAKTAR

İÇİNDEKİLER

	<u>Sayfa</u>
BAŞLIK SAYFASI.....	i
JÜRİ VE ENSTİTÜ ONAYI	ii
ÖZET	iii
ABSTRACT	iv
TEŞEKKÜR	v
ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ.....	vi
İÇİNDEKİLER.....	vii
TABLolar DİZİNİ.....	viii
ŞEKİLLER DİZİNİ.....	x
SEMBOLLER DİZİNİ.....	xi
1. GİRİŞ.....	1
2. LİTERATÜR TARAMASI.....	4
3. BÜYÜK VERİ.....	11
3.1. Büyük Veri Bileşenleri.....	13
3.2. Büyük Verinin Sınıflandırılması	15
4. MATERYAL VE YÖNTEMLER	18
4.1. Veri Seti	18
4.2. Çalışmada Kullanılan Yöntemler.....	22
4.2.1. K En Yakın Komşu Algoritması.....	22
4.2.2. Destek Vektör Regresyonu.....	23
4.3.3. Rastgele Orman Algoritması	25
5. UYGULAMA	27
5.1. Bitcoin Fiyat Tahmin Analizi.....	27
5.2. Ethereum Fiyat Tahmin Analizi.....	33
5.3. XRP Fiyat Tahmin Analizi.....	40
6. SONUÇ VE ÖNERİLER	47
KAYNAKÇA.....	49
EKLER	
ÖZGEÇMİŞ	

TABLÖLAR DİZİNİ

	<u>Sayfa</u>
Tablo 4.1. Veri setleriyle ilgili açıklamalar	19
Tablo 4.2. Veri setini oluşturan öznitelikler.....	20
Tablo 5.1. Bitcoin KNN modellerinin k değeri ve k katlamalı çapraz doğrulama sonuçları.....	27
Tablo 5.2. KNN yöntemiyle yapılan Bitcoin fiyat tahmin analiz sonuçları.....	28
Tablo 5.3. Bitcoin DVR modellerinin C katsayısı ve k katlamalı çapraz doğrulama sonuçları.....	29
Tablo 5.4. DVR yöntemiyle yapılan Bitcoin fiyat tahmin analiz sonuçları.....	29
Tablo 5.4. (Devam) DVR yöntemiyle yapılan Bitcoin fiyat tahmin analiz sonuçları....	30
Tablo 5.5. Bitcoin RO modellerinin k katlamalı çapraz doğrulama sonuçları.....	30
Tablo 5.6. RO yöntemiyle yapılan Bitcoin fiyat tahmin analiz sonuçları.....	31
Tablo 5.7. Ethereum KNN modellerinin k değeri ve k katlamalı çapraz doğrulama sonuçları.....	34
Tablo 5.8. KNN yöntemiyle yapılan Ethereum fiyat tahmin analiz sonuçları.....	34
Tablo 5.8. (Devam) KNN yöntemiyle yapılan Ethereum fiyat tahmin analiz sonuçları.....	35
Tablo 5.9. Ethereum DVR modellerinin C katsayısı ve k katlamalı çapraz doğrulama sonuçları.....	35
Tablo 5.10. DVR yöntemiyle yapılan Ethereum fiyat tahmin analiz sonuçları	36
Tablo 5.11. Ethereum RO modellerinin k katlamalı çapraz doğrulama sonuçları.....	37
Tablo 5.12. RO yöntemiyle yapılan Ethereum fiyat tahmin analiz sonuçları	37
Tablo 5.12. (Devam) RO yöntemiyle yapılan Ethereum fiyat tahmin analiz sonuçları.	38
Tablo 5.13. XRP KNN modellerinin k değeri ve k katlamalı çapraz doğrulama sonuçları.....	40
Tablo 5.14. KNN yöntemiyle yapılan XRP fiyat tahmin analiz sonuçları.....	41
Tablo 5.15. XRP DVR modellerinin C katsayısı ve k katlamalı çapraz doğrulama sonuçları.....	42
Tablo 5.16. DVR yöntemiyle yapılan XRP fiyat tahmin analiz sonuçları.....	42
Tablo 5.16. (Devam) DVR yöntemiyle yapılan XRP fiyat tahmin analiz sonuçları....	43
Tablo 5.17. XRP RO modellerinin k katlamalı çapraz doğrulama sonuçları.....	44
Tablo 5.18. RO yöntemiyle yapılan XRP fiyat tahmin analiz sonuçları.....	44

Tablo 5.18. (Devam) RO yöntemiyle yapılan XRP fiyat tahmin analiz sonuçları 45



ŞEKİLLER DİZİNİ

	<u>Sayfa</u>
Şekil 3.1. Toplam veri büyüklüğünün temsili değişimi.....	11
Şekil 3.2. Büyük Veri Bileşenleri	14
Şekil 3.3. Büyük Veri sınıflandırılması	15
Şekil 4.1. BTC veri setinden temizlenen örnek veriler.....	18
Şekil 4.2. XRP veri setinden temizlenen örnek veriler.....	19
Şekil 4.3. K kat çapraz doğrulama yönteminin çalışma modeli	21
Şekil 4.4. KNN algoritmasının çalışma adımları.....	23
Şekil 4.5. Doğrusal DVR yumuşak marjın kayıp fonksiyon gösterimi	25
Şekil 4.6. RO algoritması çalışma adımları	26
Şekil 5.1. KNN yöntemiyle elde edilen Bitcoin verilerinin karşılaştırılması	32
Şekil 5.2. DVR yöntemiyle elde edilen Bitcoin verilerinin karşılaştırılması	32
Şekil 5.3. RO yöntemiyle elde edilen Bitcoin verilerinin karşılaştırılması	33
Şekil 5.4. KNN yöntemiyle elde edilen Ethereum verilerinin karşılaştırılması	39
Şekil 5.5. DVR yöntemiyle elde edilen Ethereum verilerinin karşılaştırılması	39
Şekil 5.6. RO yöntemiyle elde edilen Ethereum verilerinin karşılaştırılması	39
Şekil 5.7. KNN yöntemiyle elde edilen XRP verilerinin karşılaştırılması	45
Şekil 5.8. DVR yöntemiyle elde edilen XRP verilerinin karşılaştırılması	46
Şekil 5.9. RO yöntemiyle elde edilen XRP verilerinin karşılaştırılması	46

SEMBOLLER DİZİNİ

BTC	: Bitcoin
ETH	: Ethereum
LTC	: Litecoin
KNN	: K en yakın komşu
RO	: Rastgele orman
DVR	: Destek vektör regresyonu
KA	: Karar ağaçları
USD	: Amerikan doları
OMH	: Ortalama mutlak hata
GAF	: Günlük açılış fiyatı
GKF	: Günlük kapanış fiyatı
GEYF	: Günlük en yüksek fiyat
GEDF	: Günlük en düşük fiyat
GH	: Günlük hacim
GTPD	: Günlük toplam piyasa değeri
TFB	: Tahmini fiyat bilgisi
ε	: Hata sapma genişliği
C	: Ceza katsayısı
w	: Ağırlık katsayısı
p	: Toplam öznitelik sayısı
m	: Tahmin değişkeni sayısı

1. GİRİŞ

İnsanlığın ilk zamanlarında mal alım satım işlemleri günümüzdeki kadar yaygın değildi ve bahsi konu edilen dönemlerde mübadele (trampa) denilen bir malın başka bir malla takasından ibaret bir ekonomi mevcuttu. Aristoteles bu takas sisteminin tarihteki en eski ticaret sistemi olarak tanımlamıştır (Tekin, 2008; Aristoteles, 1975). Mübadele ekonomisinde birbirinin malına ihtiyaç duyan iki kişinin karşılaşması ve bu iki kişinin malları arasındaki takas oranında anlaşmış olmaları gerekmektedir. Bu ticaret sistemi büyük işlemler gerektiren, yani maliyetli, işlemleri ve hesaplamaları zor ve zaman alan bir süreçti (Kamalak, 1980). Bu gibi mübadele ekonomisindeki zorluklar paranın icadını gerekli kılmaktaydı. Paranın icadından önce insanların değerli gördüğü mallar takas için kullanılmıştır (Carradice vd. 2001). Bu yüzden insanlığın ilk dönemlerinde temel gıda maddeleri takas işlemlerinde kullanılmıştır. İnsanların göçebe hayattan yerleşik düzene geçmesiyle birlikte değer atfedilen mallarda taşınabilme, bölünebilme ve saklanabilme gibi fonksiyonlar zorunlu hale gelmişti. Bu nedenle insanoğlu doğada nadir bulunan madenleri ticarete kullanmaya başlamıştır (Keş vd. 2015). İlk paralar bu nedenle doğada bulunması zor ve değerli madenlerden yani altın, gümüş veya bakır madenleri kullanılarak üretilmişlerdir. Böylelikle insanlığın tümünde değer kavramı ortaya çıkmıştır. Ticaretin gelişmesi ve globalleşmesi ile birlikte başka madenlerden ve kağıttan üretilen paralar ortaya çıkmıştır. Günümüzde kullanılan para, devletler tarafından tedavüle çıkarılan herkesin kabul ettiği kağıt ve madenden yapılmış bir değişim, ödeme ve yatırım aracıdır (http-1). Günümüz teknolojik gelişmeleri, yatırım araçlarını geliştirmekte ve farklılaştırmaktadır. Bu gelişmeler sonucu ortaya çıkan blokzincir teknolojisi, kripto paraların altyapısını oluşturmaktadır. Kripto para, kriptoloji bilimi kullanılarak şifrelenen, ödeme ve para transfer işlemlerinde kullanılabilen dijital ve sanal paradır. Bu çalışmada kullanılacak ve kripto paralardan en bilineni olan Bitcoin (BTC), 2009 yılında Satoshi Nakamoto ismini kullanan bir geliştirici tarafından oluşturulmuştur. Bitcoin, noktadan-noktaya dağıtık ağ özelliği göstermektedir. Bu ağda gerçekleşen tüm işlemler kayıt altına alınır ve toplanan bu kayıtlar blok adı verilen yapıların içinde yer alırlar. Bu blokların her biri üzerinde çok büyük işlemler gerektiren hash algoritmaları uygulamak suretiyle belirli bir ifade bulunması gerekmektedir. Bu ifadeyi ilk gerçekleştiren kullanıcı belirli bir BTC ile ödüllendirilir. Ödül miktarı her 210.000 blokta yarıya düşürülmektedir. Bu yöntemle BTC üretme biçimine madencilik denmektedir. Bu yöntemle BTC'ler normal para birimlerinin aksine, bir merkeze bağlı olmadan, bilgisayar

donanımlarının işlemci güçleri kullanılarak üretilmektedir. Bu madencilik yöntemiyle elde edilecek toplam arzı 21.000.000 adet ile sınırlandırılmıştır. Belli bir merkeze bağlı olmaması, arzının sabit olması, kullanıcılar arası para transferi için ucuz, hızlı ve takibinin zor olması sebebiyle insanlar tarafından tercih edilmeye başlanmıştır. BTC, 22 Mayıs 2010 tarihinde bir geliştiricinin 2 büyük boy pizza karşılığında başka bir kişiye 10000 adet Bitcoin transferi yapması, BTC'nin değişim ve ödeme aracı olarak kullanılmasına örnek olarak verilebilir. Her geçen gün BTC'yi ödeme ve değişim aracı olarak kabul eden işletme sayısı artmaktadır. BTC insanlar arasında yaygınlaştıkça bir ödeme aracı olarak kullanılmasının yanı sıra, ayrıca ileride fiyatının artacağı düşüncesiyle bir yatırım aracı olarak da görülmeye başlanmıştır.

Bitcoin haricinde birçok kripto para bulunmaktadır. Bu kripto paralardan bir tanesi de Ethereum (ETH)dur. Bu çalışmada kullanılacak olan ETH piyasa değeri bakımından kripto para piyasasında ikinci sıradadır. 2013 yılında kurucusu Vitalik Buterin tarafından tanıtılan bu kripto para, 30 Temmuz 2015 tarihinde ilk kez piyasaya sürülmüştür. Ethereum da Bitcoin gibi blokzincir altyapısını kullanan, kullanıcıların merkeziyetsiz uygulamalar oluşturmaya imkan veren bir platformdur. ETH blokzincirini besleyen madenciler bu hizmetlerinin karşılığında ETH kazanmaktadır. Ethereum'un bilinen herhangi bir maksimum arzı bulunmamaktadır.

Bu çalışmada analizi yapılacak bir başka kripto para ise XRP'dir. XRP ilk olarak 2012 yılında bir ödeme ağı sistemi olarak ortaya çıkmıştır. XRP, Bitcoin ve Ethereum'un aksine merkezi bir kripto paradır ve piyasadaki maksimum arzı 100 milyar adettir. XRP, Bitcoin ve Ethereum gibi madencilik yoluyla üretilmezler.

Bu kripto para da blokzincir üzerinde işlem görmekte ancak bütün işlemler Ripple adlı şirket tarafından yönetilmektedir. XRP bir ödeme ağı sistemi olması, para transferler maliyetlerinin çok düşük olması, bu transferlerin çok kısa zamanda gerçekleşmesi ve XRP'nin kurumsal yapısı sebebiyle ödeme sistemleri sağlayıcıları ile bankaların tercih ettiği bir kripto para olmuştur. BTC, ETH ve XRP hem Türkiye'de hem de dünyanın birçok ülkesinde, 7 gün 24 saat işlem yapılabilen kripto para borsalarında alınıp satılabilmektedir.

Günümüzde insanların en büyük problemlerinden biri, sınırlı kaynaklarla elde ettikleri kazançları sınırsız ihtiyaçlarını karşılamada yetersiz kalması olmuştur. İnsanlar bu sınırlı kaynaklarını çeşitlendirmek ve elde ettikleri kazancı arttırmak için farklı borsalarda işlem gören altın, döviz, devlet tahvili, mevduat, hisse senedi, repo vb. yatırım

araçlarını kullanmaktadırlar. Kripto paralar da insanlar tarafından bir çeşit para birimi ve yatırım aracı olarak görülmeye başlandığından, alım-satım işlemleri yapılabilmesi amacıyla, yurt içinde ve yurt dışında özel şirketler tarafından yönetilen borsalar kurulmuştur. Kripto paralar, bu borsalarda alım-satım işlemleri için kullanılmaktadır. Tüm borsalarda olduğu gibi kripto para borsalarında da alım-satım stratejileri belirlenirken temel ve teknik analiz yöntemleri kullanılmaktadır. Temel analiz, bir metanın gelecekteki fiyat hareketlerini tahmin etmek amacıyla, arz ve talep istatistikleri, ekonomik, siyasi, çevresel ve diğer faktörlerle ilgili bilgi ve haberlere dayalı olarak kullanılan bir analiz çeşididir. Teknik analiz ise, geçmiş fiyat hareketleri incelenerek, gelecekteki fiyat değişimlerini tahmin etmeye çalışmak olarak açıklanabilir. Teknik analiz yapılırken çok büyük verilerle karşılaşmaktadır. Bu durumda yatırımcı bu verilerin, hangilerinin anlamlı ya da hangilerinin anlamsız olduğunu bilmeden alım-satım stratejilerini belirlemek durumunda kalmaktadırlar. Bunun sonucunda, uygulanacak stratejilerin hatalı olma olasılığı artmakta ve yatırımcı bundan zarar görebilmektedir.

Çalışmanın ikinci bölümünde, kripto paralar ve büyük veri analizi hakkında kapsamlı bir literatür çalışması sonucu günümüze kadar yapılmış benzer çalışmalarda kullanılan yöntemler ve bunun sonucunda elde edilen bulgulardan bahsedilmiştir.

Çalışmanın üçüncü bölümünde büyük verinin geçmişten günümüze değişimi, büyük veri analizinin bize sağladığı yararlar, analizde karşılaşılan zorluklar ve büyük veri için gerekli bileşenler verilmiştir.

Dördüncü bölümünde ise, kripto para fiyat değişimlerinin büyük veri araçlarıyla analizi için kullanılacak materyaller ve bu çalışmada kullanılacak K En Yakın Komşu (KNN), Rastgele Orman (RO) ve Destek Vektör Regresyonu (DVR) yöntemleri tanıtılmıştır.

Çalışmanın beşinci kısımda ise BTC, ETH ve XRP kripto para fiyatlarının K En Yakın Komşu, Rastgele Orman ve Destek Vektör Regresyonu yöntemleri ile analiz uygulaması WEKA 3.8 (Hall vd. 2009) programı kullanılarak gerçekleştirilmiş ve sonuçlar gerçek verilerle karşılaştırılmıştır. Analiz sonrasında elde edilen bulguların “Sonuç ve Öneriler” kısmında değerlendirmeleri yapılmıştır.

2. LİTERATÜR TARAMASI

Geçmişten günümüze yapılan literatür taraması sonucunda, kripto para üzerine yapılan çalışmalar, büyük veri analizlerini konu alan çalışmalar ve bu çalışmada kullanılacak yöntemlerini konu alan çalışmalar olarak üç kısımda incelenmiştir.

Kripto para, teknolojik gelişmeler sonucu yakın geçmişte ortaya çıkmıştır. Yakın geçmişte ortaya çıkmasına rağmen insanların ilgisini çekebildiğinden dolayı kripto para konusu üzerinde birçok farklı çalışma bulunmaktadır.

Aktepe (2018), kripto para piyasalarında işlem gören önemli kripto paraların hem kendi fiyat verilerini, hem de çapraz kur fiyat verilerini makine öğrenmesi tekniklerini kullanarak algoritmik işlemler yapılması amacıyla doğrusal ve karar ağacı modelleri oluşturmuştur. Bu yöntemle 22 kat getiri elde edildiği görülmüştür.

Gültekin, Özdemir ve Varıcı (2019) çalışmalarında, kripto para piyasasındaki en büyük on kripto paranın günlük fiyat değişimlerinin, ilk yüz kripto paranın fiyat değişimleriyle korelasyon içerisinde olup olmadığını incelemiştir. Çalışmada kripto paraların fiyat değişimlerinin belirli seviyelerde de olsa birbirinden bağımsız olduğu sonucuna varılmıştır.

Liu ve Serletis (2019), kripto paraların getiri ve volatiliteleri arasında ilişkiyi ve kripto para piyasalarının geleneksel piyasalara olan getiri ve volatiliteler bakımından yayılımı analiz edilmiştir. Çalışma sonucunda kripto paraların birçok yönü ile finansal varlıklara benzediği gösterilmiştir.

Bitcoin fiyat tahmini için Ji, Kim ve Im (2019) derin öğrenme yöntemlerinden Derin Sinir Ağı ve Uzun Kısa Süreli Bellek yöntemlerini kullanmıştır. Bu iki yöntemin performansları ölçüldüğünde derin sinir ağı modellemesinin Bitcoin fiyat tahmini açısından daha başarılı olduğu görülmüştür. Atlan (2019) ise Bitcoin fiyatlandırmasında Yapay Sinir Ağları, Uyarlamalı Sinirsel Bulanık Sistemi, Eğri Uydurma ve Uzun Kısa Süreli Bellek yöntemlerini kullanmış ve Bitcoin fiyatını tahmin eden bir model önermiştir. Çalışma sonucunda Eğri Uydurma ve Uyarlamalı Sinirsel Bulanık Sistem modelinin tahminleme açısından başarılı olduğu gösterilmiştir.

Mallqui ve Fernandes (2019), Bitcoin'in Amerikan Doları (USD) bazında günlük verilerini kullanarak Yapay Sinir Ağları ve Destek Vektör Makineleri yöntemleri ile fiyat yönünün, günlük en yüksek fiyat, günlük en düşük fiyat ve günlük kapanış fiyatı tahminleri yapılmıştır. Ortalama mutlak hata, ortalama mutlak yüzde hata ve ortalama

hata kare kök sonuçlarına göre en başarılı tahminleri Destek Vektör Makineleri yöntemi ile elde edildiği gözlemlenmiştir.

Kripto paraların uzun vadeli fiyat tahmininin yapılabilmesi amacıyla Yapay Sinir Ağları yöntemini kullanan bir diğer çalışma da Ömrüuzun (2019) tarafından yapılmıştır. Çalışmada, fiyat modellemesi oluşturulurken Keras kütüphanesinden faydalanılmıştır. Yapay Sinir Ağları ile kripto paraların fiyat tahmini yapılabildiği ve Yapay Sinir Ağlarına verilen veri arttırıldıkça modelin fiyat tahmininde daha başarılı olduğu sonucuna varılmıştır.

Kripto para alanında yapılmış bir başka çalışma olan Chen, Li ve Sun (2020)'da da Bitcoin fiyat tahmini yapabilmek amacıyla makine öğrenmesi teknikleri kullanılmıştır. Beş dakikalık ve günlük veriler kullanılarak Lojistik Regresyon, Doğrusal Diskriminant Analizi, Rastgele Orman Algoritması, XGBoost, ikinci dereceden diskriminant analizi, Destek Vektör Makinesi ve Uzun Kısa Süreli Bellek yöntemleriyle Bitcoin fiyat tahminlemesi yapılmıştır. Çalışma sonucunda, günlük veriler kullanılarak oluşturulan modelde, istatistiksel yöntemlerden Lojistik Regresyon %66 doğruluk oranı ile en başarılı sonucu veren yöntem olmuştur. Beş dakikalık veriler kullanılarak oluşturulan modelde ise makine öğrenmesi yöntemlerinden Uzun Kısa Süreli Bellek %67.2 doğruluk oranı ile en başarılı performansı göstermiştir. Yöntemler istatistiksel ve makine öğrenmesi olarak karşılaştırıldığında, istatistiksel yöntemlerin daha uzun zaman aralıklı verilerde, makine öğrenmesi yöntemlerinin ise kısa zaman aralıklı verilerde daha başarılı olduğu görülmüştür. Bu alanda yapılmış literatürdeki bir diğer çalışma da Kurucan (2019) tarafından yapılmıştır. Kurucan (2019), büyük veri ve makine öğrenmesi yöntemleri kullanılarak BİST100 endeksinin maksimum fiyat tahmin modellerinin oluşturulmasını amaçlamıştır. Modellerin başarısı farklı iki veri setiyle test edilmiştir. Çalışmada, Yapay Sinir Ağları ve XGBoost yöntemleriyle modeller oluşturulmuştur. Bu modeller belirli performans ölçütleriyle karşılaştırılmış; BİST100 endeksinin en yüksek fiyat tahminlemesinde XGBoost yöntemi, Yapay Sinir Ağları yöntemine göre daha başarılı olduğu görülmüştür.

Bitcoin fiyat tahmini çalışmalarından farklı olarak, başka bir çalışma da Gagarina, Nestik ve Drobysheva (2019) tarafından yapılmıştır. Bu çalışmada gençlerin kripto paraya karşı tutumlarını belirlemek amacıyla Lineer Regresyon analizi kullanılmıştır. Çalışma sonucunda gençlerin kripto paralara karşı tutumu, finansal özerklik ve devlet kurumlarına olan güvensizlikleriyle bağlantılı olduğu ortaya konmuştur.

Al-Jaroodi ve Mohamed (2014), gerçek zamanlı büyük veri uygulamalarını ve bu uygulamaların teknik zorluklarını incelemiştir. Bu çalışmada akıllı taşımacılık, finansal piyasa ticareti ve gözetimi, kalabalık kontrolü, askeri karar verme, akıllı şebekeler ve doğal afetlerde erken uyarı gibi gerçek zamanlı büyük veri uygulama konularının zorlukları ele alınmıştır. Çalışmada kritik öneme sahip gerçek zamanlı büyük veri uygulamalarında büyük verilerin toplanmasının, aktarılmasının ve işlenmesinin uzun zaman aldığı sonucuna varılmıştır.

Enerji endüstrisinde büyük veri konusundaki bir örnek için Goel, Datta ve Mannan (2017)'nin çalışması incelenebilir. Çalışmalarında, büyük veri analitikleri ile enerji endüstrisinde süreç güvenliği ve risk yönetimindeki potansiyelini ortaya koymak için veritabanı analizi, pompa arızalarının tahmine dayalı modellemesi, tesisin dinamik risk haritalamasının yapılması ve görüntü analizi uygulamalarına sahip dört vaka üzerinde çalışılmıştır. Çalışmada büyük veri analitiği uygulamalarının politika, strateji ve operasyonel risk karar verme süreçlerinde daha güvenli olduğu sonucuna varılmıştır.

Büyük veri alanında incelenmiş literatürdeki bir diğer çalışma, Özdeş (2017) tarafından yapılmıştır. Özdeş (2017), duygu analizi işlemini gerçekleştirmek amacıyla makine öğrenmesi yöntemlerinden denetimli öğrenme algoritmalarından, Naive Bayes, Lojistik Regresyon ve Karar Ağaçları yöntemleri uygulanarak, bu yöntemlerin başarı oranlarını karşılaştırmıştır. Çalışma sonucunda, Naive Bayes algoritmasının diğer yöntemlerden daha yüksek başarı oranına sahip olduğu görülmüştür.

Bürüce (2018), büyük veri setlerinin mevcut istatistiksel analiz teknikleri ve ekonometrik analiz teknikleri yardımıyla doğru yorumlanabilmesinin mümkün olduğunu incelemiştir. Çalışmada dinamik veriye dikkat etmenin önemi vurgulanmış ve büyük veri setlerini analiz etmede korelasyonun tercih edilmesi gerektiğinden bahsedilmiştir. Çalışmada, büyük veri yöntemleri kullanılarak ekonometrik analiz yapılabildiği sonucuna varılmıştır.

Sağlık alanından da bir örnek verilecek olursa; Özer (2019) çalışmasında, uyku sırasında metabolizmanın yavaşlayıp yavaşlamadığına dair verileri gezici kan basıncı izleme cihazının yerine kişilerin tahlil verileri ile veri madenciliği yöntemleri kullanarak karar verebilme altyapısını incelemiştir. Oluşturulan altyapı ile sağlanan iki adet büyük veri setinin Naive Bayes ve Rastgele Orman algoritmaları kullanılarak tahminleme modelleri değerlendirilmiştir. Çalışmanın sonucunda iki veri setinde de Rastgele Orman algoritmasının daha başarılı tahminleme modelleri oluşturduğu görülmüştür.

Bir başka alan olan eğitim alanında da Gök (2017), ortaöğretim 6, 7 ve 8. Sınıf öğrencilerinin Matematik, Türkçe ve genel başarı ortalamalarını öğrencinin ailevi durumunun, demografik, ekonomik, sosyal ve okul çevresi gibi etmenlerin etkileyip etkilemediğini araştırmıştır. Öğrencilere sorulan 24 soruluk anketle veri seti oluşturulmuştur. Bu veri seti Lojistik Regresyon, Naive Bayes, KNN, Rastgele Orman, Doğrusal Destek Vektör Makineleri ve Radyal Tabanlı Fonsiyon Destek Vektör Makineleri yöntemleri ile analiz edilmiştir. Bu analizler sonucunda başarı tahmininde Rastgele Orman algoritmasının en başarılı tahmini yaptığı görülmüştür.

Kumar, Dogra, Utreja ve Yadav (2018), hisse senetlerinin fiyat tahmini amacıyla Destek Vektör Makineleri, Rastgele Orman, K En Yakın Komşu, Naive Bayes ve Softmax algoritmaları kullanılmıştır. Çalışmada 5 farklı firmanın hisse senet fiyatları kullanılmıştır. Bu 5 veri setinin 3'ü büyük veri seti 2 tanesi ise küçük veri seti olarak ayrılarak fiyat tahminleri gerçekleştirilmiştir. Büyük veri setlerinin fiyat tahminlerinde Rastgele Orman algoritmasının, küçük veri setinde ise Naive Bayes algoritmasının diğer yöntemlere göre daha başarılı olduğu görülmüştür.

Valencia, Gomez-Espinosa ve Valdes-Aguirre (2019)'da BTC, ETH, XRP ve Litecoin (LTC) gibi kripto paraların sosyal medya ve piyasa verileri kullanılarak makine öğrenmesi ve duygu analizi ile fiyat hareketlerinin tahmin edilmesine çalışılmıştır. Çalışmada Sinir Ağları, Destek Vektör Makineleri ve Rastgele Orman algoritmaları kullanılarak Twitter ve pazar verileri analiz edilmiştir. Sonuç olarak, sosyal medya verilerinin kripto para hareketlerini tahmin etmede yardımcı olabileceği ve Sinir Ağları'nın diğer yöntemlerden daha iyi performans gösterdiği belirlenmiştir.

Enerji ürünlerinin aylık bazda fiyatları kullanılarak uzun vadeli fiyat tahmini yapmak amacıyla yaptıkları çalışmada Herrera, Constantino, Tabak, Su ve Naranpanawa (2019), geleneksel ekonomik metodların kullanıldığı melez bir model, Yapay Sinir Ağları ve Rastgele Orman yöntemlerini kullanmıştır. Sonuçlar incelendiğinde, Yapay Sinir Ağları ve Rastgele Orman yöntemlerinin, geleneksel ekonomik yöntemlere göre daha başarılı tahminler ürettiği ve bu yöntemlerin ek olarak fiyat dönüm noktalarını tahmin etmede başarısı kullanıcıya avantaj sağladığı görülmüştür.

Rastgele Orman konusunda yapılmış bir başka çalışmada, Baba ve Sevil (2020) halka arz getirilerini tahmin edebilmek ve Lineer Regresyonun çözemediği konuları çözebilmek için makine öğrenmesi yöntemlerinden Rastgele Orman yöntemini kullanmıştır. Çalışmada iki model oluşturulmuş ve bu iki modelde de Rastgele Orman

Yönteminin, Lineer Regresyon yöntemine göre çok daha iyi sonuçlar verdiği görülmüştür.

Ince ve Trafalis (2008)'in çalışmasında, hisse senetlerinin fiyat hareketlerini Çok Katmanlı Algılayıcı, Destek Vektör Regresyonu (DVR) ve Bütünleşik Otoregresif Hareketli Ortalama yöntemleri ile tahmin edilmeye çalışılmıştır. Çalışmada iki farklı strateji belirlenerek bunlarla ilgili 2 farklı model oluşturulduktan sonra analiz işlemi gerçekleştirilmiştir. Çalışma sonucunda yöntemler ortalama hata karelerinin köküne göre karşılaştırılmış, Çok Katmanlı Algılayıcı ve DVR algoritmalarının başarısının birbirine yakın olduğu görülmüştür.

DVR algoritması ile fiyat tahmini haricinde maliyet tahmini, yakıt tüketim tahmini ve talep tahmini yapılmış çalışmalar da incelenmiştir. DVR algoritması kullanılarak yapılmış bir çalışma olan Yolasıgmaz (2015)'da, kavramsal inşaat maliyetlerini tahmin etmek amaçlanmaktadır. Çalışmada 273 proje içeren 10 adet veri seti analiz edilmiştir. Çalışmada DVR ile güçlü bir model oluşturulmuş bunun yanı sıra mevcut veri seti Yapay Sinir Ağları ve vaka bazlı çözümleme yöntemleri ile de analiz edilerek karşılaştırma yapılmıştır. Çalışma sonucunda DVR ile oluşturulan modelin maliyet tahmini açısından diğer iki yönteme göre daha başarılı olduğu sonucuna varılmıştır.

Yalçın (2016), uçakların uçuş süresince harcadıkları anlık yakıt tüketiminin DVR, Karar Ağaçları, Yapay Sinir Ağları ve Gaussian Süreç yöntemleriyle tahmin etmiştir. Elde edilmiş veriler doğrusal olmayan bir fonksiyon olarak modellenmiş ve bu model gerçek bir veri seti üzerinden test edilmiştir. Çalışmanın sonunda yöntemler karşılaştırılmış, DVR ile kurulan modelin en yüksek doğruluk oranına sahip olduğu görülmüştür.

DVR yöntemi kullanan bir başka çalışma olan Fradinata, Kesuma, Rusdina ve Zaman (2019)'da, erişte talebinin tahmin edilmesi amaçlanmıştır. Veri seti rastgele üretilen veriler üzerinden oluşturulmuştur. Çalışma sonucunda standart ortalama hata kareleri ile DVR'nin ortalama hata karesi karşılaştırıldığında, DVR'nin %11 oranında daha başarılı olduğu sonucuna varılmıştır.

Enerji santrallerinin aylık doğalgaz tüketim miktarının tahmini yapmak amacıyla Meral (2019) de DVR yöntemini kullanılmıştır. Veri seti, enerji santrallerinin tüketimi gibi bağımlı değişkenlerin yanı sıra sanayi tüketimi, şehir tüketimi, üretim miktarı, ithalat miktarı ve ihracat miktarı gibi bağımsız değişkenlerden oluşmaktadır. Bu veri seti, Destek Vektör Regresyonunda kullanılan bütün çekirdek fonksiyonlarıyla analiz edilmiştir.

Çalışma sonucunda, en küçük hata kareleri değerine sahip olan Radyal Tabanlı Fonksiyona sahip Destek Vektör Regresyon modeli seçilmiştir. Bu model yardımıyla enerji santrallerinin 2018 yılının Mayıs-Aralık ayları arasındaki tahmini doğalgaz tüketimleri tespit edilmiştir.

Alkhatib, Najadat, Hmeidi ve Shatnawi (2013) çalışmalarında, Lübnan borsasında işlem gören hisse senedi fiyatlarının KNN ve Doğrusal Olmayan Regresyon yöntemleriyle tahmin etmiştir. Çalışmada, Lübnan borsasında işlem gören 6 şirketin hisse senetleri fiyatları veri seti olarak kullanılmıştır. Yapılan analizler sonucunda KNN algoritmasının tahmin ettiği veriler gerçek verilerle karşılaştırıldığında, KNN yönteminin düşük hata oranıyla tahminler gerçekleştirdiği görülmüştür.

Borsa İstanbul (BİST)'daki hisse senetlerinin fiyat tahminini İlarıslan (2016) KNN algoritmasıyla yapmıştır. Borsada işlem gören bir hisse senedinin 2014 yılı kapanış fiyatları veri seti olarak seçilmiş, çapraz doğrulama yöntemiyle uygun k parametresi belirlenmiştir. Çalışma sonucunda analiz edilen hisse senedinin 2015 yılının ilk üç ayının fiyat tahminlemesi yapılmış, gerçek verilerle karşılaştırıldığında %97 oranında başarılı sonuçlar elde edilmiştir.

Filiz, Karaboğa ve Akogul (2017), BİST-50 endeksinin günlük zaman bazında yönünü etkileyen faktörlerin KNN, Naive Bayes, C4.5 algoritması gibi makine öğrenme yöntemleriyle ve Yapay Sinir Ağları kullanılarak sınıflandırılmasını amaçlamıştır. Çalışma sonucunda, C4.5 algoritmasının %92.71 doğruluk oranı ile yöntemler arasında en iyi performansı göstermiştir. KNN algoritması ise en düşük doğruluk oranına sahip sınıflandırmayı gerçekleştirmiştir.

Yiğiter, Sarı, Karabulut ve Başakın (2018), islam hukukuna uygun sukuk adı verilen kira sertifikalarının KNN algoritmasıyla fiyat tahminlemesini yapmıştır. Veri seti olarak kira sertifikası fiyatlarının günlük verileri kullanılarak 1, 3 ve 5 gün sonrasının fiyat tahminini yapan üç farklı model oluşturulmuştur. Oluşturulan bu üç farklı modelin başarı sonuçları incelenmiştir. Çalışma sonucunda bir gün sonrasının fiyat tahminini gerçekleştiren Model 1'in ortalama hata kare değeri 0.0063, üç gün sonrasının fiyat tahminini gerçekleştiren Model 2'nin ortalama hata kare değeri 0.0161 ve beş gün sonrasının fiyatlarını tahmin eden Model 3'ün ortalama hata kare değeri ise 0.0253 olarak gerçekleşmiştir. KNN algoritması, Model 1'de en başarılı tahmin performansını göstermiştir.

Bu alanda yapılmış bir başka çalışma olan Lahmiri ve Bekiros (2020)'da Bitcoin gibi yüksek volatiliteye sahip fiyat serilerinin makine öğrenme yöntemleriyle tahmininin karşılaştırılması amaçlanmıştır. Çalışmada elde edilen veri seti DVR, Gauss Poisson Regresyonu, Regresyon Ağaçları, KNN, Bayes Düzenleyici Sinir Ağı, İleri Beslemeli Sinir Ağı ve Radyal Temelli Fonsiyon Sinir ağları yöntemleriyle analiz edilmiştir. Uygulanan yöntemler Ortalama Hata Kareleri Köküne göre karşılaştırılmış, en düşük hata değerini Bayes Düzenleyici Sinir Ağının, en yüksek hata değerini ise KNN algoritmasının verdiği gözlemlenmiştir.

Literatürdeki birçok çalışma incelendiğinde finans, sanayi ve hizmet sektörlerinde fiyat tahmini ve diğer tahmin uygulamalarında K En Yakın Komşu, Destek Vektör Regresyonu ve Rastgele Orman algoritmaları kullanıldığı görülmektedir. Finans alanındaki bu çalışmalarda genellikle BİST ve çeşitli yabancı borsalardaki hisse senetleri kullanılarak fiyat tahminlemesi yapıldığı görülmüştür. Literatürdeki kripto para ile ilgili çalışmalarda ise farklı çalışmalar olmakla birlikte genellikle sadece BTC ile ilgili çalışmalar olduğu görülmüştür.

Tez çalışmasında ise literatürden farklı olarak BTC, ETH ve XRP adlı üç farklı kripto paraların veri setleri, kripto paraların belli borsalara ilk çıkış tarihlerinden günümüze kadar ki günlük veriler göz önüne alınarak oluşturulmuştur ve bu veri setleri KNN, DVR ve Rastgele Orman algoritmalarının tümü kullanılarak fiyat tahminlemesinin yapılması amaçlanmıştır. Bu üç yöntemin birlikte kullanılarak kripto para fiyat tahmininin yapılması ve ardından gerçek fiyat değerine ilişkin en başarılı yöntemin bulunması, bu tez çalışmasının literatüre yaptığı katkı açısından oldukça önemlidir.

3. BÜYÜK VERİ

Günümüzde bilgi toplumunun hayatımızı etkileyen unsurları ile her alanda karşılaşmaktadır. Bu unsurlardan bir tanesi de büyük veri kavramıdır. Büyük veri kavramı, büyük bir hızla büyümek ve gelişmekle birlikte kabul edilen tek bir tanımı bulunmamaktadır. Büyük veri, kısaca “bireylerin ve şirketlerin oluşturduğu veriyi analiz edilerek sınıflandırılmış anlamlı ve işlenebilir hale dönüştürülmüş hale getirilen veri yığını”na denilebilir. Bireyler ve şirketler teknolojinin gelişmesi ile büyük veri kaynaklarını sürekli beslemektedir. Bireyler ve şirketler kullandıkları sosyal medya hesapları, arama motorları, banka hesapları, e-postalar, sensörler gibi araçlarla büyük veri yığını kesintisiz bir şekilde beslemeyi sürdürmektedirler (http-2).

Büyük veri yığınları, doğru yöntemler ile analiz edilerek doğru şekilde yorumlandığında kullanıcıların risklerini iyi yönetmelerine ve risk ortamında daha doğru kararlar alabilmesine yardımcı olabilmektedir. Şirketler elde ettikleri büyük veri yığınlarını işleyerek müşterilerinin hareketlerini ve tercihlerini tahmin ederek bunun için farklı geliştirmeler ile karlılıklarını arttırabilmektedirler.

Günümüzde her an oluşan veri büyüklüğü mevcut veri tabanlarının yetmeyeceği büyüklükte olduğu gibi, bu veri yığınının oluşma hızları da veri tabanları ve depolama cihazlarının kapasitesini aşmaktadır.

Geçmiş yıllarda veriler disk, CD, DVD gibi depolama araçlarında depolanırken, teknolojik gelişmelerin katkısıyla petabyte’lık veriler yüksek kapasiteli veri tabanlarına ve depolama birimlerine yüklenebilmektedir. International Data Corporation şirketi, 2009 yılına kadar oluşturulan 0.8 Zettabyte olan toplam verinin, 2020 yılına gelindiğinde 44 kat büyüyerek 35.2 Zettabyte düzeyine geleceğini söylemiştir. Şekil 3.1’de toplam veri büyüklüğünün değişimi tasvir edilmiştir (http-3).



Şekil 3.1. Toplam veri büyüklüğünün temsili değişimi

Verilerin bu kadar hızlı büyümesi verilerin saklanması, istenilen verilere ulaşma hızları, veri güvenliği ve veri analizi gibi bir çok sorunu da beraberinde getirmektedir. Bu yüzden büyük verileri işleyebilmek ve analiz edebilmek günümüzdeki şirketler için çok önemli bir konu haline gelmektedir. Büyük veri analizinin kullanılmasının bireylere ve şirketlere birçok yararı bulunmaktadır. Bunlar;

- Eldeki verinin düzgün işlenmesiyle iş anlayışının daha iyi yönetilmesi
- Müşteri davranışlarını daha önceden tanımlamak ve buna göre hareket etmek
- Pazarların risklerini öğrenmek ve pazar stratejisini buna göre belirlemek
- Müşteriyi tanımlama sonrası, müşteriye göre ürün farklılaştırması yapabilmek
- Çok daha iyi tahmin ve planlama yapabilmek
- Hedef pazara dair verilerin işlenerek hedef pazara yönelik daha iyi pazarlama stratejileri geliştirebilmek
- Ürün üretiminde verimi arttırmak
- Hedef pazarın belirlenerek bu pazardaki satış tahminini yapmak
- Yatırım stratejilerini belirleyerek daha fazla kazanç elde etmek olarak sıralayabiliriz.

Aynı şekilde büyük veriyi kullanırken birçok zorlukla da karşılaşılır. Bu zorluklar;

- Dış veriye erişimin zor olması ve mevcut özel veri ile uyumun zor olması
- Veri politikası gereği telif hakları, kişisel gizlilik ve veri güvenliği konusunun getirdiği maliyet ve problemler
- Veri formatı uyumsuzlukları
- Mevcut sistemlerle olan entegrasyon sorunları olarak sıralanabilir (Demirci, 2018).

Teknolojinin gelişmesi, bireylerin ve şirketlerin almak istediği birçok hizmetin İnternet ortamına taşınmasına neden olmuştur. Bu gelişme sonrasında hizmet sektöründe çalışan şirketler kişiye özel ürün üretme, müşteri memnuniyeti gibi konularda başarılı olabilmek için, müşterilerine ait bilgileri alıp saklamaya başlamıştır. İşte bu gibi sebeplerle büyük verinin birçok kullanım alanı bulunmaktadır. Büyük veri kullanım alanlarına ait farklı örnekleri izleyen şekilde sıralayabiliriz.

- Devlet kuruluşları, vatandaşlarına sunduğu hizmetlerini daha hızlı ve iyi şekilde gerçekleştirebilmek için devlet vatandaşlarının verilerini saklamak zorundadır.

- Hastaneler, kendilerine başvuran hastalarına iyi bir hizmet sunabilmek için hastaların geçmiş tıbbi verileri ile kişisel tüm verilerini saklayarak, buna göre bir tıbbi hizmet sunmaya çalışmaktadır.
- Bankalar müşterilerinin davranışlarını öğrenerek, müşterilerine özel farklı argümanlar sunmak ve diğer banka hizmetleri için müşterilerinin hareketlerini izleyerek kayıt altına almaktadır.
- Lojistik sektöründe araç güzergahları, harcanan yakıt miktarları ve sürücü bazlı hatalar gibi bilgiler saklanarak farklı analiz teknikleri ile ulaştırma operasyonlarında maliyetler düşürülmeye çalışılır.
- Arama motorları, müşterilerinin istediği bilgileri en hızlı ve en güvenilir biçimde müşterilerine sunmak için büyük veriyi kullanmaktadır.
- Sosyal medya ağları, kişilerin kendi uygulamalarında daha fazla zaman geçirmelerini sağlamak için müşterilerinin ilgisini çekecek bilgileri kişiye özel şekilde göstermek, müşterilerinin davranışlarını öğrenmek gibi durumlar için büyük veriden yararlanmaktadır.

3.1. Büyük Veri Bileşenleri

Büyük veriyi oluşturan beş etmen bulunmaktadır. Bu beş etmen 5V olarak bilinen Velocity (Hız), Variety (Çeşitlilik), Volume (Büyüklik), Verification (Doğrulama) ve Value (Değer)'dir.

Velocity (Hız): Büyük veri, teknolojik gelişmelerin etkisiyle doğru orantılı bir şekilde artmaktadır. Büyük verinin hızla büyümesi bu veri için gerekli işlem sayısını ve çeşitliliğinin de aynı oranda artmasını gerekli kılmaktadır.

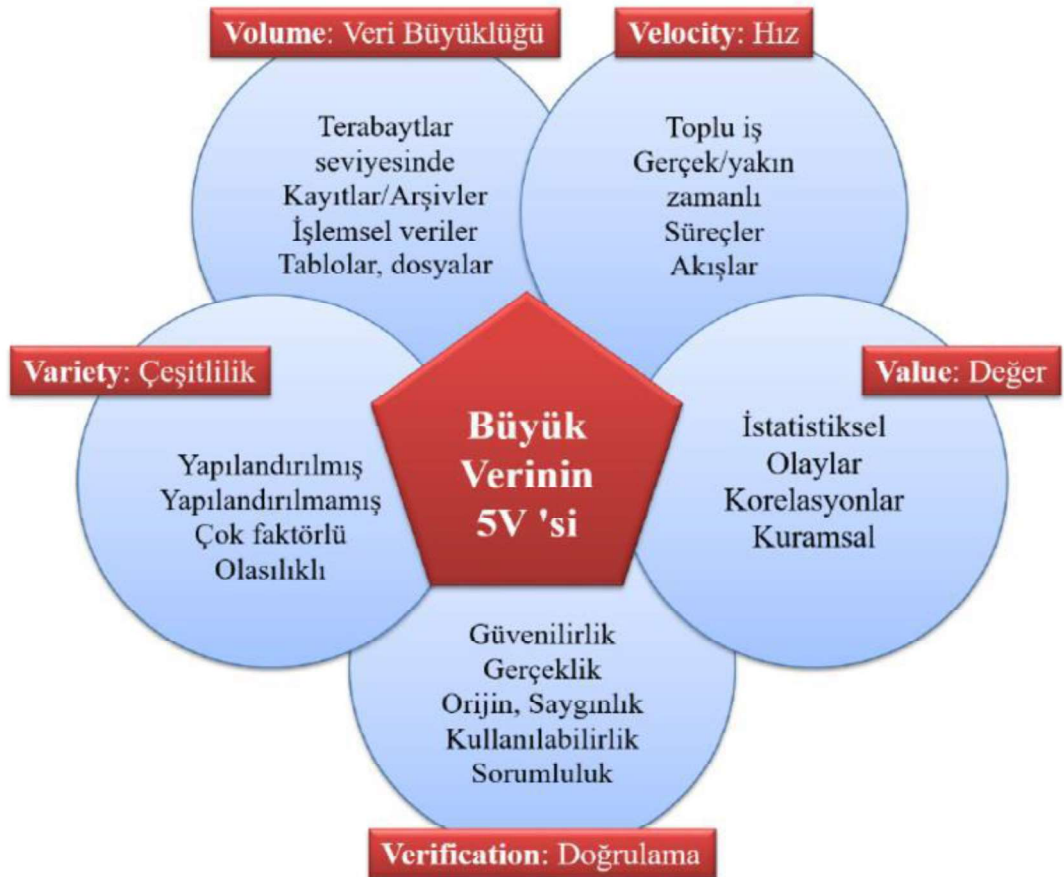
Variety (Çeşitlilik): Teknolojinin gelişmesiyle günümüzde üretilen her veri farklı kaynaklar tarafından üretilmektedir. Bu veriler bilgisayarlar, cep telefonları, tabletler ve benzer diğer tipteki cihazlar tarafından üretilmektedir. Bu cihazlarda üretilen veriler hem farklı dillerde hem de farklı formatta karşımıza çıkmaktadır. Çok yüksek hızlarda ve farklı kaynaklardan önümüze gelen bu verilerin geleneksel sistemler yardımıyla işlenebilmesi ve analiz edilebilmesi mümkün görünmemektedir. Karşılaştığımız bu farklılıkları anlamlı bir şekilde inceleyebilmemiz için aynı formata dönüştürülmeleri gerekir.

Volume (Büyüklik): International Data Corporation şirketinin istatistiklerini göz önüne aldığımızda 2009 yılında üretilen veri büyüklüğünün 2020 yılında 44 kat artacağı öngörülmüştür. Buradan yola çıkarak, kurumların elde ettikleri büyük veriyi arşivleme, işleme vb. işlemleri için nasıl bir yol izleneceği şimdiden planlanması gerekmektedir.

Verification (Doğrulama): Farklı kaynaklardan elde ettiğimiz büyük verinin doğru ve güvenilir olması gerekmektedir. Veri hacmi arttıkça bu verinin doğrulanma zorluğu da benzer şekilde artmaktadır. Büyük verinin akışı sırasında gerekli olan güvenlik seviyelerinde izlenmesi, sadece yetkili kişiler tarafından görülebilmesi gerekir ve yetkisi olmayan tüm diğer kişilere karşı da gizli kalmalıdır.

Value (Değer): En önemli büyük veri bileşenidir. Veriler, üretilmesi ve analiz aşamalarından sonra karar vermede etkili olmalıdır. Bu verilere istenilen herhangi bir durumda hemen ulaşılmalıdır. Verinin üretimi ve analizi sonrasında bireyler ya da kurumlar için fayda sağlaması gerekir.

Şekil 3.2’de büyük verinin beş bileşeni görülmektedir.

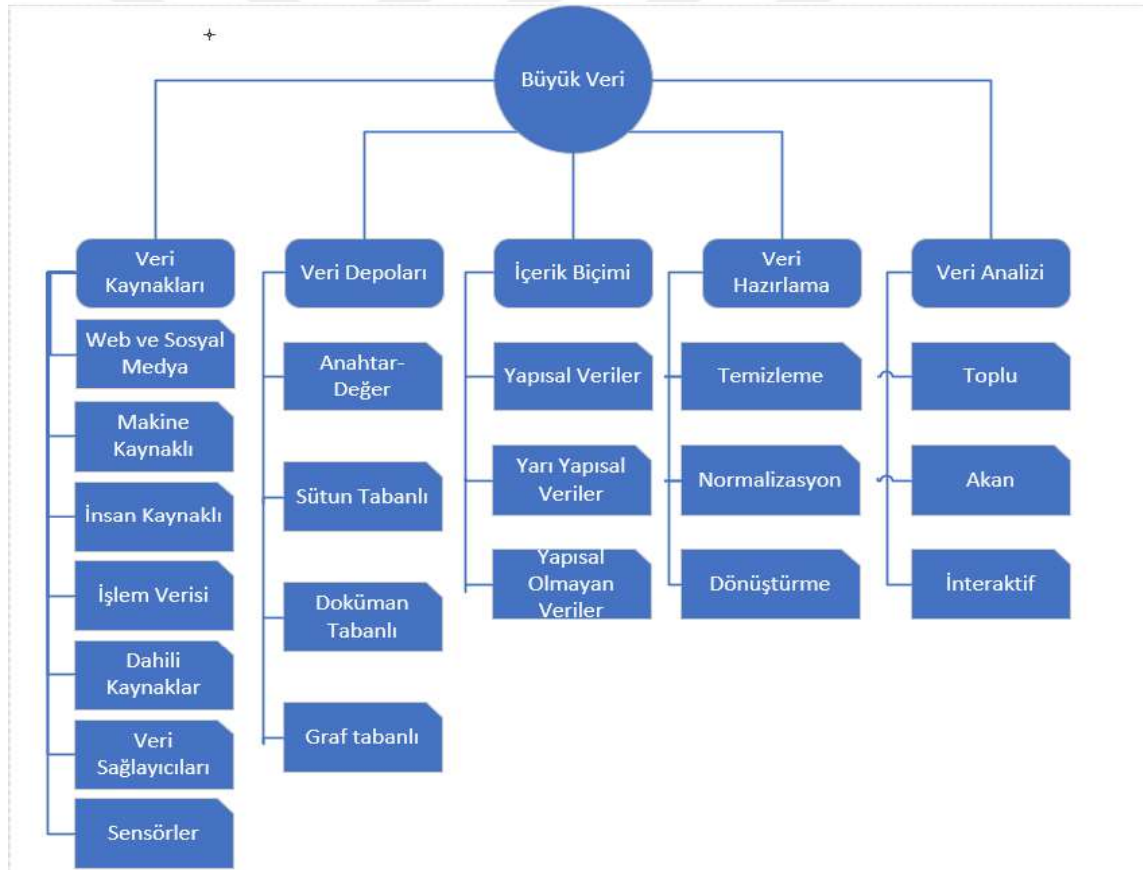


Şekil 3.2. Büyük Veri Bileşenleri (Özdeş, 2017)

Çalışmada kullanılan veri setindeki veri sayısının kesintisiz bir şekilde her gün artması, Genel Ağ ortamından elde edilen verilerin yapılandırılmamış olması, veri büyüklüğünün her geçen gün üretilen veriye bağlı artması, elde edilen veri setinin kripto paralarla ilgili güvenilir bir kaynak olan coinmarketcap.com adresinden elde edilmesi ve bu verilerin analizi sonucu elde edilecek finansal kazanç göz önüne alındığında bu veri seti büyük verinin beş etmenini karşılamaktadır.

3.2. Büyük Verinin Sınıflandırılması

Büyük veriyi daha iyi kavrayabilmek için, büyük veri farklı kategoriler altında sınıflandırılmıştır. Şekil 3.3'te büyük verinin kategoriler halinde sınıflandırılması verilmiştir.



Şekil 3.3. Büyük Veri sınıflandırılması (Özdeş, 2017)

Veri Kaynakları: Günümüzde ham veri elde etmek için birçok kaynaktan yararlanılmaktadır. Web ve Sosyal medya ağlarında, bloglardan, sosyal medya topluluklarından, bu topluluklarda yapılan paylaşım, tartışma ve bilgi alışverişiyle üretilen veriler büyük veri için ham veri kaynağıdır. Aynı şekilde, insan müdahalesi olmadan işlem yapan bilgisayarlar, cihazlar, robotlar da makine kaynaklı verilerdir. Fabrikalarda ve hizmet sektöründe kullanılan sensörler birçok farklı işlemler için kullanılmakta bu sensörler sürekli olarak veri toplamaktadır. İnsan kaynaklı veriler, işlem verileri, dahili kaynak verileri vb. kaynaklar da büyük veri için ham veri kaynağı görevi görmektedir.

Veri Depoları: Büyük veri için kullanılan veri depoları, anahtar-değer, sütun tabanlı, belge tabanlı ve graf tabanlı veri depoları olarak ayrılmaktadır. Anahtar-değer veri depoları, çok büyük veri yığınlarını ölçeklendirmek için tasarlanan, veriyi depolayan ve bu verilere erişen veri tabanı sistemleridir. Redis, Flare, Sclaris ve Valdemort bu veri tabanı sistemlerine örnek olarak gösterilebilir. Sütun tabanlı veri depoları ise içerdiği bilgileri hem satırlarda hem de sütunlarda depolayan veri tabanı sistemleridir. HBase, HadoopDB, Hypertable sütun tabanlı veri depolarına örnek olarak gösterilebilir. Doküman tabanlı veri depoları, farklı formatlardaki veri dosyalarını ve bilgileri almak, depolamak ve farklı formatlardaki belgeleri standart formatlarda destekleyebilen veri tabanı sistemleridir. SimpleDB, Terrastore doküman tabanlı veri depolarından bazılarıdır. Grafik tabanlı veri depoları, grafik modellerine sahip verileri depolamaya yaramaktadır. InfoGrid, Infinite Graph, Neo4j gibi veri depoları, graf tabanlı veri depolarına örnek olarak gösterilebilir.

İçerik Biçimi: Büyük veride yapısal veri, yarı yapısal veri ve yapısal olmayan veri olarak üç çeşit veri tipi bulunmaktadır. Yapısal veriler belirli tür ve boyutlarda organize edilmiş saklanması, modellenmesi, işlenmesi gibi işlemlerin kolayca yapılabilirdiği verilerdir. Yapısal olmayan veriler, yapısal verilerin aksine belirli tür ve boyutlarda organize edilmemiş, anlamlandırma ve işleme hazır hale getirilmesi zor olan veri tipidir. Yarı yapısal veriler ise, kolayca işlenmesi bakımından yapısal veriye benzeyebilmekle birlikte bu veri tipindeki veriler herhangi bir sistem üzerinde değildirler. Yarı yapısal verilere XML ve JSON dilleri örnek olarak verilebilir.

Veri Hazırlama: Büyük veri analiz işlemi yapılmadan önce verinin bazı aşamalardan geçmesi gerekir. Bu aşamalardan temizleme aşaması, verinin anlamsız, eksik kısımlarının belirlenerek bu verilerin yığından kaldırılması işlemidir. Dönüşüm

işlemi, verinin analiz edilmesi için uygun formatlara çevrilmesi işlemidir. Normalizasyon ise veride ortaya çıkan anlamsız ilişkileri en az seviyede tutmak için veri tabanını yapılandırma işlemidir.

Veri Analizi: Toplu veri, akan veri ve interaktif verinin farklı tabanlı sistemler yardımıyla işlemek suretiyle faydalı bilgilere ulaşma ve katma değer sağlama amacıyla işleme sürecidir.



4. MATERYAL VE YÖNTEMLER

Bu çalışmada, en büyük piyasa değerine sahip ve en bilinen üç kripto para olan Bitcoin (BTC), Ethereum (ETH) ve XRP'nin USD bazında fiyat tahmininin yapılması amaçlanmıştır. Çalışmada altı farklı öznitelik kullanılarak oluşturulan üç farklı veri seti için üç farklı yöntem ile WEKA 3.8 (Hall vd. 2009) programı kullanılarak fiyat tahminlemesi yapılmıştır. EK-1'de KNN, DVR ve RO algoritmalarının program üzerinde çalıştırılma adımları gösterilmiştir.

4.1. Veri Seti

Bu çalışmada kullanılacak olan üç veri seti de <https://coinmarketcap.com/> adresinden günlük fiyat değişimleri baz alınarak oluşturulmuştur. Veri setleri, günlük açılış fiyatı (GAF), günlük kapanış fiyatı (GKF), günlük en yüksek fiyat değeri (GEYF), günlük en düşük fiyat değeri (GEDF), günlük işlem hacmi (GH) ve kripto paranın günlük toplam piyasa değeri (GTPD) olmak üzere altı öznitelikten oluşmaktadır. Hedef sınıfı ise tahmini fiyat bilgisi (TFB) dir. Bu öznitelikler teknik analiz yapılırken kullanılan veriler olduğundan bu çalışma kapsamında öznitelik olarak belirlenmiştir. Öznitelikler kendi içerisinde ağırlıklandırılmamış, eşit önem derecesinde oldukları varsayılmıştır. Geçmiş fiyatların oluşumunda manipülasyon, borsa dalgalanmaları ve petrol fiyatları vb. faktörlerin tahmin modelleri üzerinde bir etkisinin olmadığı varsayılmıştır. EK-2'de kullanılan veri setlerinin bir bölümü verilmiştir. BTC ve XRP veri setinde günlük işlem hacmi verisi bulunmadığından dolayı bu veri seti hazırlanırken temizleme işlemi yapılmıştır. Şekil 4.1'de BTC veri setinde temizlenen verilerden bir kısmı gösterilmiştir.

1	BITCOİN						
2	GAF	GEYF	GEDF	GKF	GH	GTPD	TFB
2312	763.28	777.51	713.60	735.07	46,862,700	8,955,394,564	727.83
2313	683.94	777.75	683.94	761.98	0	9,279,735,331	735.07
2314	666.31	682.70	649.48	682.21	0	8,305,173,374	761.98
2315	672.36	684.39	645.71	665.58	0	8,100,042,042	682.21
2316	613.06	680.91	611.04	673.41	0	8,192,487,202	665.58
2317	601.78	666.74	585.64	617.18	0	7,505,757,423	673.41
2318	619.90	654.27	579.17	605.66	0	7,363,114,611	617.18
2319	694.22	729.16	595.33	625.32	0	7,599,576,492	605.66
2320	519.06	707.23	502.89	691.96	0	8,406,397,153	625.32
2321	678.20	679.32	420.51	522.70	0	6,347,159,168	691.96
2322	706.37	754.83	630.88	682.12	0	8,279,623,719	522.70
2323	880.33	882.25	668.25	705.97	0	8,565,534,010	682.12

Şekil 4.1. BTC veri setinden temizlenen örnek veriler

Şekil 4.2’de XRP veri setinde temizlenen verilerden bir kısmı gösterilmiştir.

1	XRP						
2	GAF	GEYF	GEDF	GAF	GH	GTPD	TFB
2307	0.027365	0.027810	0.026814	0.026944	58,533.40	210,649,132	0.028726
2308	0.022098	0.024891	0.022098	0.024385	0	190,638,461	0.027076
2309	0.022599	0.022599	0.021434	0.022042	0	172,318,800	0.024385
2310	0.023698	0.024124	0.022305	0.022575	0	176,487,298	0.022042
2311	0.021571	0.023965	0.021485	0.023702	0	185,301,187	0.022575
2312	0.021571	0.023839	0.020959	0.021716	0	169,773,295	0.023702
2313	0.020979	0.022572	0.020305	0.021684	0	169,521,559	0.021716
2314	0.021141	0.023279	0.019852	0.020851	0	163,010,820	0.021684
2315	0.015783	0.021546	0.015313	0.021073	0	164,747,173	0.020851
2316	0.017971	0.018571	0.012325	0.015852	0	123,933,098	0.021073

Şekil 4.2. XRP veri setinden temizlenen örnek veriler

BTC veri seti 2310 günden ve bu günlere ait özniteliklerle, toplamda 16170 adet veriden oluşmaktadır. ETH veri seti 1722 günden ve bu günlere ait belirtilen altı öznitelikle, toplamda 12054 adet veriden oluşmaktadır. XRP veri seti ise 2305 günden ve bu günlere ait özniteliklerle, toplamda 16135 adet veriden oluşmaktadır. Veri setindeki hedef sınıfı yani tahmini fiyat bilgisi, öznitelik değerleri kullanılarak elde edilen 24 saat sonrasının fiyat tahminini ifade etmektedir. Bir başka deyişle bugünün kapanışında elde edilen öznitelik ile yarının kapanış fiyatı tahmin edilmektedir. Veri setleriyle ilgili hangi tarihler arasındaki verilerin kullanıldığına ilişkin açıklamalar Tablo 4.1’de, veri setindeki öznitelikler ile ilgili bilgiler ise Tablo 4.2’de özetlenmiştir.

Tablo 4.1. Veri setleriyle ilgili açıklamalar

VERİ SETİ	AÇIKLAMA	BAŞLANGIÇ TARİHİ	BİTİŞ TARİHİ
BTC	Günlük bazda veri	27.12.2013	23.04.2020
ETH	Günlük bazda veri	07.08.2015	23.04.2020
XRP	Günlük bazda veri	01.01.2014	23.04.2020

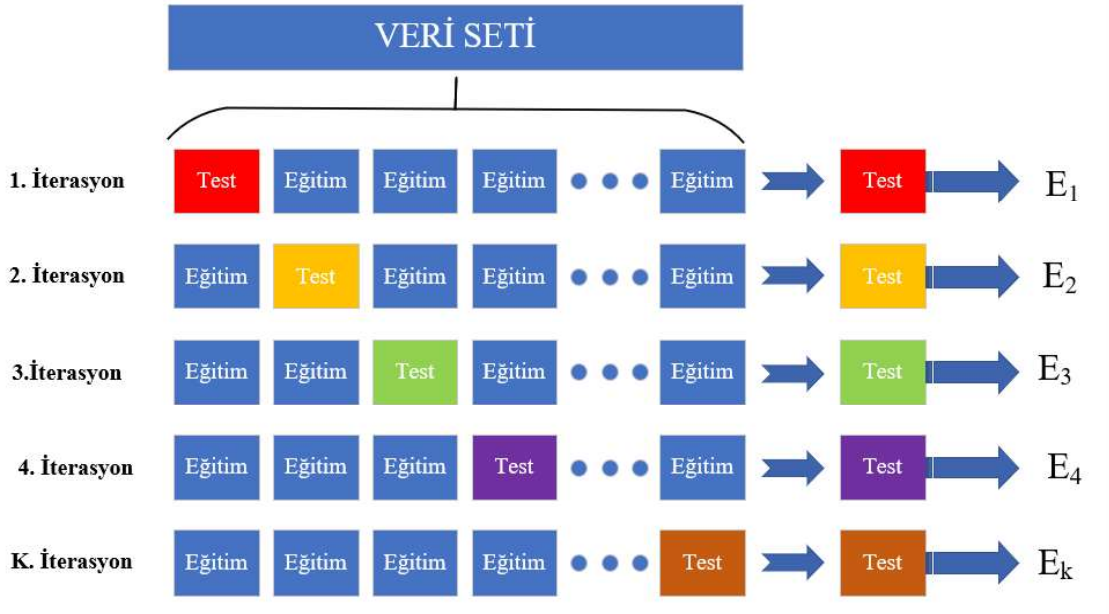
Tablo 4.2. *Veri setini oluşturan öznitelikler*

Öznitelik İsmi	Değişken Türü	Açıklama
Gün Açılış Fiyatı	Sürekli Değişken	Gün başlangıcındaki ilk gerçekleşen fiyat bilgisidir.
Gün En Yüksek Fiyatı	Sürekli Değişken	İlgili kripto paranın gün içerisinde ulaştığı en yüksek fiyatı bilgisidir.
Gün En Düşük Fiyatı	Sürekli Değişken	İlgili kripto paranın gün içerisinde ulaştığı en düşük fiyatı bilgisidir.
Gün Kapanış Fiyatı	Sürekli Değişken	Gün sonunda en son gerçekleşen fiyat bilgisidir.
Günlük Hacim	Sürekli Değişken	Günlük gerçekleşmiş toplam alım satım işleminin USD bazında değerini temsil eder.
Günlük Toplam Piyasa Değeri	Sürekli Değişken	İlgili kripto paranın USD bazında toplam piyasa değerini ifade eder.
Tahmini Fiyat Bilgisi	Sürekli Değişken	İlgili kripto paranın belirlenen tarihteki tahmini fiyat bilgisidir.

Kullanılacak veriler K kat çapraz doğrulama yöntemiyle sınanmıştır. K kat çapraz doğrulama yönteminde öncelikle bir k değeri belirlenir. Veri seti k adet eşit boyutlu alt veri setine ayrılır. Bu alt veri setlerinden bir tanesi test veri seti ve geri kalan $k - 1$ adet alt veri seti ise eğitim veri seti olarak nitelendirilmektedir.

Eğitim veri seti, kullanılan algoritmanın öğrenmesi amacıyla sunulan verilerden oluşan veri yığınlarıdır. Test verisi ise algoritmanın öğrenmesi bittikten sonra oluşturulan modele göre ve modelin gerçek verilere ne kadar yakın tahminler verdiğini test etmek için kullanılan veri yığınlarıdır.

K kat çapraz doğrulamada da tüm veri seti k adet parçaya ayrıldıktan sonra bir adet test veri seti ve $k - 1$ adet eğitim veri seti kullanılarak model eğitilmektedir. k adet iterasyon boyunca, her iterasyonda test verisi değiştirilerek modelin eğitilme işlemi gerçekleştirilir. Şekil 4.3'te k kat çapraz doğrulama yönteminin çalışma şekli modellenerek gösterilmiştir.



Şekil 4.3. K kat çapraz doğrulama yönteminin çalışma modeli

Algoritma sonucunda test veri setlerinin ortalama mutlak hata (OMH) kullanılarak en düşük OMH değerini veren model en ideal k değeri olarak kullanılır. Ortalama mutlak hata değeri Denklem (4.1) kullanılarak hesaplanmaktadır.

$$OMH = \frac{1}{k} \sum_{j=1}^k |e_j| \quad (4.1)$$

Denklem (4.1)'de belirtilen e_j değeri hedeflenen değer ile tahmin edilen değerler arasındaki fark olarak ifade edilmektedir. OMH, gerçek ve tahmini olmak üzere iki sürekli değişken arasındaki mesafedir. OMH, kolay yorumlanabilir bir değer olduğundan regresyon problemlerinde çokça kullanılmaktadır ve hesaplanırken hata değerlerinin mutlak değeri alındığından daima pozitif değerler almaktadır. OMH değeri sıfıra ne kadar yakınsa modelin tahmin performansı o kadar iyi olduğu söylenebilir.

4.2. Çalışmada Kullanılan Yöntemler

4.2.1. K En Yakın Komşu Algoritması

K En Yakın Komşu Algoritması (KNN), denetimli öğrenme algoritmalarında sınıflandırma ve regresyon çalışmalarında kullanılan yöntemlerden biridir. KNN algoritması, uygulanması çok karmaşık olmayan bir denetimli öğrenme algoritmasıdır. KNN, Cover ve Hart (1967) tarafından önerilmiş bir algoritmadır. Diğer denetimli öğrenme algoritmalarından farklı olarak, eğitim aşamasına sahip değildir. Eğitim aşaması ile test aşaması neredeyse aynı işlevi görmektedir. Tembel bir öğrenme türü olduğundan büyük veri setlerinde uygulanması ideal olmayan bir algoritmadır. KNN örnek veri setine eklenen verilerin sınıfı, veri setindeki mevcut verilere olan uzaklıkları hesaplanıp, belirlenen k sayıdaki komşularına bakılarak belirlenir. Sürekli değişkenler için uzaklık hesaplamaları genellikle Öklid, Manhattan ve Minkowski mesafe fonksiyonlarından birisi kullanılarak verilerin birbirine olan uzaklıkları hesaplanır. Öklid uzaklık fonksiyonu Denklem (4.2)'de, Manhattan uzaklık fonksiyonu Denklem (4.3)'te, Minkowski uzaklık fonksiyonu ise Denklem (4.4)'te verilmiştir.

$$\sqrt{\sum_{i=1}^k (x_i - y_i)^2} \quad (4.2)$$

$$\sum_{i=1}^k |x_i - y_i| \quad (4.3)$$

$$\left(\sum_{i=1}^k (|x_i - y_i|)^q \right)^{1/q} \quad (4.4)$$

KNN algoritması, basit ve gürültülü veri setlerine karşı dirençli olmasına rağmen verilerin uzaklık hesapları yapılırken mevcut tüm durumlar sakladığından dolayı, büyük veri setlerinde kullanıldığında fazla bellek alanı ihtiyacı oluşturmaktadır (<http-3>).

KNN algoritmasının adımları kısaca şöyledir:

Adım 1: Örnek veri seti analizi sonucunda en az hatayı veren k değeri bulunur.

Adım 2: Mevcut veri setine eklenecek yeni verilerin uzaklık fonksiyonu kullanılarak mevcut verilere uzaklıkları hesaplanır.

Adım 3: Bu uzaklıkların hesaplanması sonrasında bu veriye en yakın k kadar komşu veri ele alınır.

Adım 4: Öznitelik değerlerine göre en yakın k komşu ortalamasına göre tahmini değer elde edilir.

KNN algoritmasının çalışma adımları Şekil 4.4 (http-4)'te gösterilmiştir.



Şekil 4.4. KNN algoritmasının çalışma adımları

4.2.2. Destek Vektör Regresyonu

Destek Vektör Makineleri algoritması, sınıflandırma ve regresyon problemlerinin çözümünde kullanılan en iyi kernel tabanlı yöntemlerden birisidir. Makine öğrenme algoritma ilkelerine dayalı ve çok güçlü olan bu öğrenme stratejisi Vapnik (1998) tarafından geliştirilmiştir. DVM algoritmasının regresyon problemleri üzerine en kapsamlı çalışması Smola ve Schölkopf (2004) tarafından yapılmıştır. Bu çalışmada doğrusal DVR kullanılmış ve farklı bir kernel kullanılmamıştır.

Eğitim verisi $\{(x_i, y_i)\}_{i=1}^n \subset \mathbb{R}$ şeklinde verilen ifade de X girdi verilerinin alanını göstermektedir. Bu X değişkeni kripto paralarda günlük fiyat verileri olabilir. Y değişkeni ise regresyon problemlerinde tahmin değişkeni olarak tanımlanmaktadır.

DVR de amaç, tüm eğitim verileri için elde edilen hedeflerden en fazla ε sapmasına sahip, aynı zamanda mümkün olduğunca doğrusal $f(x)$ fonksiyonunu bulmaktır. Bu ifadeden anlaşılacağı üzere, hatalar ε değerinden küçük olduğu sürece yöntem tarafından

önemsenmeyecek ancak ε değerinden büyük herhangi bir sapma da kabul edilmeyecektir. İlk olarak doğrusal fonksiyonların durumu tanımlanır. w parametresi x_i girdilerinin ağırlık katsayısını belirtmekte ve b parametresi ise herhangi bir reel sayısını ifade etmektedir. Denklem (4.5)'te verilen f fonksiyonu tanımlanır.

$$f(x) = \langle w, x \rangle + b \quad w \in X, b \in R \quad (4.5)$$

Belirtilen bu fonksiyonun düz olmasının istenmesi, w değerinin küçük olması anlamına gelmektedir. Bu durumu sağlamanın bir yolu da, $\|w\|^2$ ifadesinin normalize edilmesidir (Smola vd. 2004). Bu problem, dışbükey bir optimizasyon problemi olarak Denklem (4.6)'da ifade edilmiştir.

$$\begin{aligned} \min \quad & \frac{1}{2} \|w\|^2 \\ & y_i - \langle w, x_i \rangle - b \leq \varepsilon \quad \forall i = 1, 2, \dots, n \\ & \langle w, x_i \rangle + b - y_i \leq \varepsilon \quad \forall i = 1, 2, \dots, n \end{aligned} \quad (4.6)$$

Bazı varsayımlara göre, tüm (x_i, y_i) çiftlerinde ε değeri kesinliğiyle yaklaştıran f fonksiyonunun var olduğu düşünülmektedir. Ancak bazen durum böyle olmayabilir ve bazı farklı hatalara izin vermek istenilebilir. Bu durumu oluşturabilmek amacıyla yumuşak marjın kayıp fonksiyonuna benzer şekilde, optimizasyon problemiyle başa çıkabilmek için modele ξ_i ve ξ_i^* gibi gevşek değişkenler ilave edilebilir (Cortes vd. 1995). Bu durumda model Denklem (4.7)'deki gibi yazılır (Vapnik, 1995).

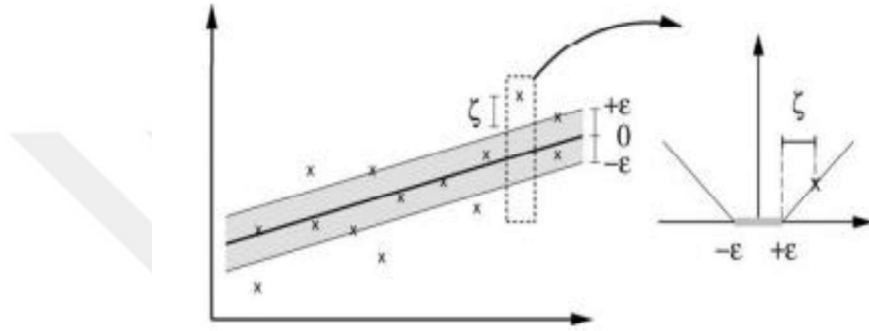
$$\begin{aligned} \min \quad & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \\ & y_i - \langle w, x_i \rangle - b \leq \varepsilon + \xi_i \quad \forall i = 1, 2, \dots, n \\ & \langle w, x_i \rangle + b - y_i \leq \varepsilon + \xi_i^* \quad \forall i = 1, 2, \dots, n \\ & \xi_i, \xi_i^* \geq 0 \end{aligned} \quad (4.7)$$

$C > 0$ sabit değeri, f fonksiyonun düzlüğü ile tolere edilebilecek ε değeri arasında dengeyi sağlamak için kullanılmaktadır. Bu durum ise ε duyarsız kayıp fonksiyonu ile

uğraşmak anlamı gelmektedir. ε duyarsız kayıp fonksiyonu, Denklem (4.8)'deki gibi tanımlanmaktadır.

$$|\xi|_{\varepsilon} = \begin{cases} 0, & |\xi| \leq \varepsilon \\ |\xi| - \varepsilon, & \text{diğer durumlar} \end{cases} \quad (4.8)$$

Sadece ε değerinden büyük değerlerin doğrusal olarak cezalandırıldığı Şekil 4.5'te gösterilmiştir.



Şekil 4.5. Doğrusal DVR yumuşak marjin kayıp fonksiyon gösterimi (Smola vd. 2004)

4.3.3. Rastgele Orman Algoritması

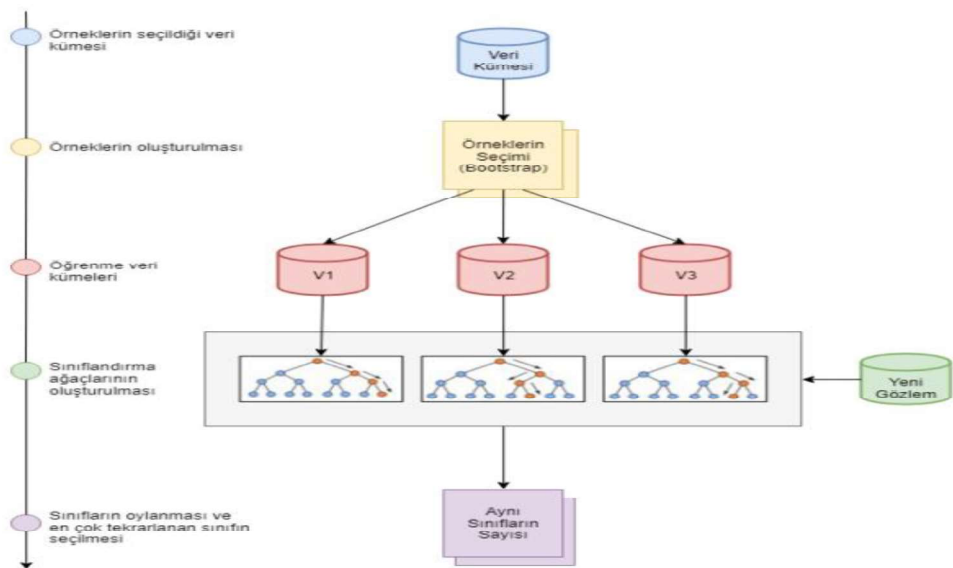
Veri madenciliği ve makine öğrenmesi yöntemleri arasında en çok kullanılan algoritmalarından bir tanesi de Karar Ağaçlarıdır (KA). KA algoritmasında ağacın doğru kurulamaması karşılaşılan en büyük problemlerin başında gelmektedir. Doğru kurulamayan ağaç, yüksek başarı performans gösterememektedir. KA algoritmasının bu dezavantajına karşılık Rastgele Orman adlı yeni bir ağaç tabanlı yöntem geliştirilmiştir. Rastgele Orman (RO) algoritması, torbalama (bagging) temelli bir makine öğrenmesi yöntemidir (Breiman, 2001). RO algoritması, birden fazla karar ağacı oluşturarak, çok daha doğru ve kararlı modeller oluşturmak için bunları birleştirir. RO algoritması, ağaçları büyütürken modele ilave olarak rastgelelik katmaktadır. Dğümleri ayırırken ise en önemli özellik yerine, rastgele bir özellik alt kümesi içindeki en iyi özelliği aramaktadır. Bu durum geniş bir çeşitlilik yaratarak daha iyi modellerin elde edilmesini sağlamaktadır (http-5). RO algoritması regresyon analizi yapmak için de kullanılmaktadır. Regresyon analizinde kullanıldığında, hedef değeri sürekli veriden oluşmaktadır. Bundan dolayı, RO algoritmasıyla oluşturulan her ağaç bize sürekli bir

tahmin modeli oluşturur. Bireysel ağaçlardan elde edilen tahminlerin ortalaması alınarak tüm modelin tahmini elde edilir (Cutler vd. 2011).

RO algoritmasının adımları aşağıdaki gibi sıralanabilir (Akman vd. 2011):

1. Veri setinden N adet veri seçilir ve seçilen bu verilerin belirli bir kısmı öğrenme verisi olarak kullanılarak ağaçlar oluşturulur.
2. Öğrenme veri seti içerisinde m adet öznelik seçilir ve en çok bilgi kazancı elde edecek olan belirlenir. Regresyon ağacı için ideal değer $m = p/3$ olarak önerilmektedir (Breiman, 2001). p , toplam öznelik sayısını belirtmektedir.
3. Seçilen tahmin değişkeninin en iyi dallanma kriterini entropi ve Gini indeksi kullanarak hesaplanır. Bu değere göre veri seti her düğümde iki alt dala ayrılır.
4. Buraya kadarki işlemler yaprak düğüm elde edilinceye kadar tekrar edilir ve N adet ağacın oluşturmak için önceki adımlar N defa tekrarlanır.
5. Bu ağaçların oluşturduğu kurallar kullanılarak test verileri ile model test edilir.
6. Sınıflandırma problemleri için en fazla oyu alanı ağacın tahminini tüm ormanın tahmini olarak seçilir. Regresyon problemleri için ise yapılan tüm oylamaların ortalamaları hesaplanır ve buna göre tahmin yapılır.

RO algoritmasının çalışma adımları Şekil 4.6'da gösterilmiştir.



Şekil 4.6. RO algoritması çalışma adımları (Fidan, 2020)

5. UYGULAMA

Çalışmada coinmarketcap.com adresinden elde edilen veriler online şekilde çekildikten sonra veri setleri üzerinde eksik veriler tespit edildiği için veri setleri üzerinde temizleme işlemi yapılmıştır. BTC, ETH ve XRP veri setleri temizleme işlemleri yapıldıktan sonra x_i , i .nci satırdaki öznitelik değeri olmak üzere Denklem (5.1) kullanılarak min-max normalizasyonu yapılmıştır.

$$\frac{x_i - \min(x_i)}{\max(x_i) - \min(x_i)} \quad (5.1)$$

Normalizasyon işlemi uygulandıktan sonra, WEKA 3.8 (Hall vd. 2009) programı kullanılarak tüm veri setlerine KNN, DVR ve RO yöntemleri uygulanmıştır.

5.1. Bitcoin Fiyat Tahmin Analizi

Bitcoin veri setine K En Yakın Komşu algoritmasının uygulanması için öncelikli olarak, ayrılan eğitim veri seti kullanılarak en düşük OMH değerine sahip k değeri ve k katlamalı çapraz doğrulama parametresi bulunmuştur. Tablo 5.1’de Bitcoin eğitim veri setinin değişen k parametresi ve k katlamalı çapraz doğrulama değerine göre elde edilen OMH değerleri gösterilmiştir.

Tablo 5.1. Bitcoin KNN modellerinin k değeri ve k katlamalı çapraz doğrulama sonuçları

		K kat çapraz doğrulama değeri					
		5	6	7	8	9	10
K parametre değeri	1	167,9031	169,6786	167,1244	167,2375	167,9599	167,4398
	3	137,4052	139,9591	136,6819	138,0851	137,8709	137,6872
	5	131,9727	132,8269	130,6864	131,5973	131,6963	131,3264
	7	127,4385	129,8612	128,0758	128,7563	128,7736	127,7193
	9	125,9282	127,1355	125,579	124,8323	126,7831	126,7391
	11	126,1409	126,6652	125,1089	124,8286	125,0342	124,8712

Bitcoin için KNN algoritmasının k parametresi 11 ve k katlamalı çapraz doğrulama değeri ise 8 olduğunda en düşük OMH sonucu elde edildiği görülmektedir. Bu parametreler kullanılarak elde edilen modellerle fiyat tahminlemesi yapılmıştır. KNN ile yapılan 23 günlük fiyat tahmin değerleri, hata değerleri ve OMH sonucu Tablo 5.2’de gösterilmiştir.

Tablo 5.2. KNN yöntemiyle yapılan Bitcoin fiyat tahmin analiz sonuçları

Gün	Gerçek Değer	Tahmini Değer	Hata Değeri
1	6793,62	6479,698	-313,922
2	6733,39	6749,171	15,781
3	6867,53	6587,405	-280,125
4	6791,13	6874,925	83,795
5	7271,78	7061,659	-210,121
6	7176,41	6615,73	-560,68
7	7334,1	6734,787	-599,313
8	7302,09	7034,512	-267,578
9	6865,49	6998,456	132,966
10	6859,08	6750,616	-108,464
11	6971,09	7077,93	106,84
12	6845,04	6808,015	-37,025
13	6842,43	6587,405	-255,025
14	6642,11	6874,925	232,815
15	7116,8	6874,925	-241,875
16	7096,18	6749,171	-347,009
17	7257,66	7077,93	-179,73
18	7189,42	7204,079	14,659
19	6881,96	7450,625	568,665
20	6880,32	6575,62	-304,7
21	7117,21	6950,613	-166,597
22	7429,72	7139,011	-290,709
23	7550,9	6887,646	-663,254
OMH değeri			260,0715

KNN yönteminin sonuçları incelendiğinde bazı noktalarda elde edilen tahmin başarısının genele yayılmadığı görülmüştür. Tahmini değerlerin ortalama mutlak hata değeri incelendiğinde %4’e yakın hata oranı görülmüştür.

Çalışmada kullanılan ikinci analiz yöntemi olan Destek Vektör Regresyonu yöntemi için Bitcoin eğitim veri seti C katsayısı ve k katlamalı çapraz doğrulama değerlerine göre modeller oluşturulmuş, en düşük OMH değerine sahip model belirlenmeye çalışılmıştır. Tablo 5.3'te C katsayısı ve k katlamalı çapraz doğrulama değerine göre elde edilen OMH değerleri verilmiştir.

Tablo 5.3. Bitcoin DVR modellerinin C katsayısı ve k katlamalı çapraz doğrulama sonuçları

		K kat çapraz doğrulama değeri					
C katsayısı		5	6	7	8	9	10
	16	111,5852	112,0149	112,0487	111,8246	112,1473	111,9077
	17	112,1624	111,9298	111,8022	111,8157	112,4048	111,9224
	18	111,898	111,8856	111,7877	112,086	112,3533	111,935
	19	111,7888	111,8868	112,2211	112,0916	112,3781	112,0042
	20	111,8313	112,1986	111,9965	112,0613	112,069	111,8505

DVR yöntemiyle eğitim seti analiz edildiğinde C katsayısı 16, k kat çapraz doğrulama değeri ise 5 seçildiğinde en düşük OMH değeri elde edilmiştir. Bu değerler kullanılarak elde edilen modelle fiyat tahminlemesi yapılmıştır. Yapılan 23 günlük fiyat tahmininde elde edilen sonuçlar Tablo 5.4'te gösterilmiştir.

Tablo 5.4. DVR yöntemiyle yapılan Bitcoin fiyat tahmin analiz sonuçları

Gün	Gerçek Değer	Tahmini Değer	Hata Değeri
1	6793,62	6572,693	-220,927
2	6733,39	6815,943	82,553
3	6867,53	6748,372	-119,158
4	6791,13	6846,874	55,744
5	7271,78	6779,577	-492,203
6	7176,41	7246,495	70,085
7	7334,1	7183,665	-150,435
8	7302,09	7313,286	11,196

Tablo 5.4. (Devam) *DVR yöntemiyle yapılan Bitcoin fiyat tahmin analiz sonuçları*

Gün	Gerçek Değer	Tahmini Değer	Hata Değeri
9	6865,49	7277,519	412,029
10	6859,08	6885,013	25,933
11	6971,09	6846,004	-125,086
12	6845,04	6972,312	127,272
13	6842,43	6825,422	-17,008
14	6642,11	6837,624	195,514
15	7116,8	6658,311	-458,489
16	7096,18	7091,024	-5,156
17	7257,66	7083,185	-174,475
18	7189,42	7239,421	50,001
19	6881,96	7178,65	296,69
20	6880,32	6899,536	19,216
21	7117,21	6865,808	-251,402
22	7429,72	7101,938	-327,782
23	7550,9	7413,428	-137,472
OMH değeri			166,3401

DVR yöntemiyle Bitcoin fiyat tahmin işleminde hata oranının %2-%2.3'lere kadar düştüğü gözlenmiştir. OMH değeri baz alınarak incelendiğinde ise KNN yönteminden daha başarılı tahmin sonuçları elde edilmiştir.

Rastgele Orman yöntemi için oluşturulacak modeli belirlemek amacıyla ayrılan Bitcoin eğitim veri seti k katlamalı çapraz doğrulama değerleri değiştirilerek modeller oluşturulmuş, en düşük OMH değerine sahip model tahmin modeli olarak seçilmiştir. RO algoritması ile eğitilen modellere ait k katlamalı çapraz değerler ve bu değerlere ait OMH değerleri Tablo 5.5'te verilmiştir.

Tablo 5.5. *Bitcoin RO modellerinin k katlamalı çapraz doğrulama sonuçları*

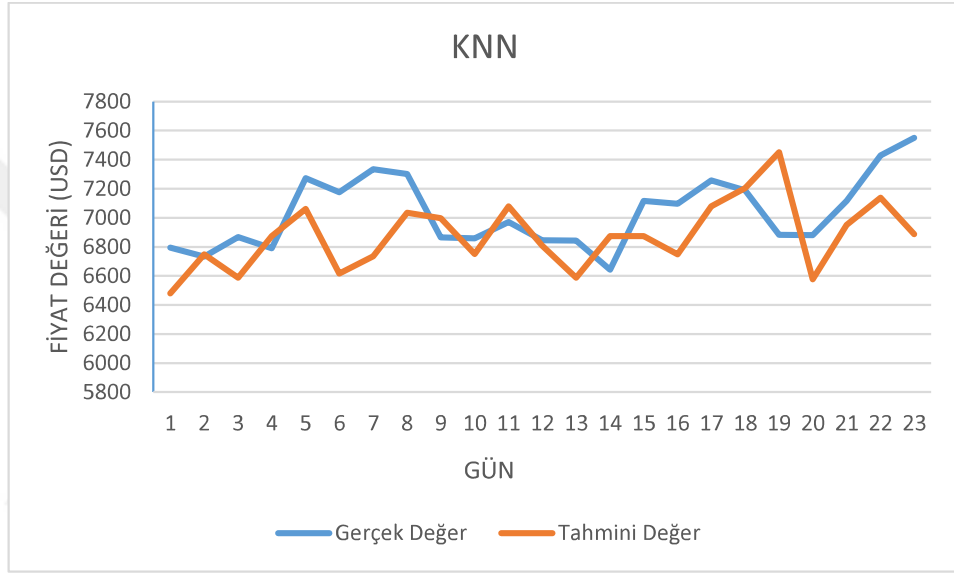
K	Hata değeri	K	Hata değeri
3	153,2518	7	148,689
4	150,775	8	149,646
5	148,6133	9	147,0638
6	148,4178	10	149,2364

Tablo incelendiğinde k katlamalı çapraz doğrulama değeri 9 seçildiğinde modeller arasında en düşük OMH değeri elde edilmiştir. RO algoritmasıyla oluşturulan bu model kullanılarak 23 günlük Bitcoin fiyat tahminlemesi gerçekleştirilmiştir. Rastgele Orman algoritmasıyla elde edilen veriler Tablo 5.6’da gösterilmiştir.

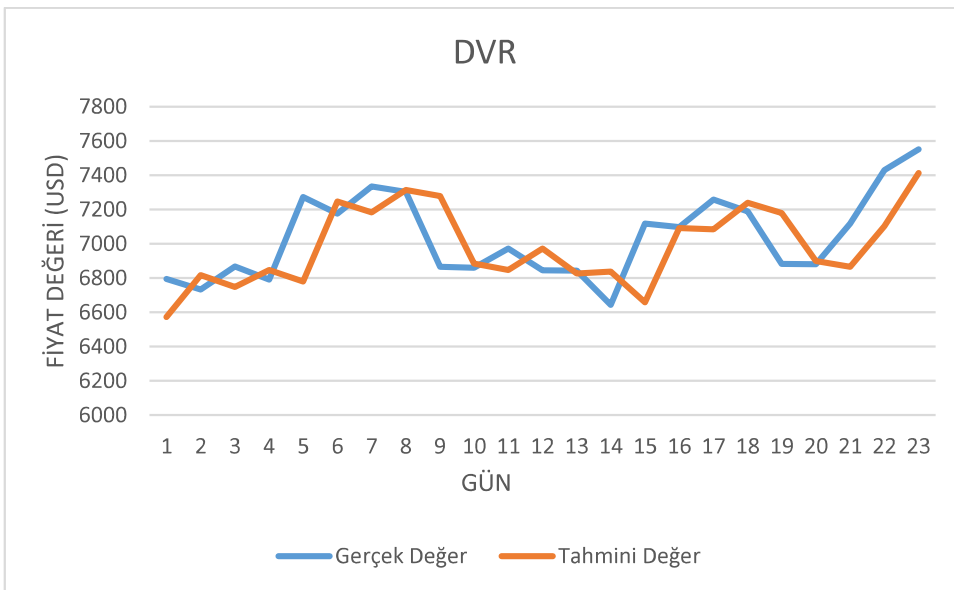
Tablo 5.6. RO yöntemiyle yapılan Bitcoin fiyat tahmin analiz sonuçları

Gün	Gerçek Değer	Tahmini Değer	Hata Değeri
1	6793,62	6595,482	-198,138
2	6733,39	6931,618	198,228
3	6867,53	6875,443	7,913
4	6791,13	6871,49	80,36
5	7271,78	6758,375	-513,405
6	7176,41	7223,788	47,378
7	7334,1	7153,88	-180,22
8	7302,09	7102,177	-199,913
9	6865,49	7262,753	397,263
10	6859,08	6888,912	29,832
11	6971,09	6834,281	-136,809
12	6845,04	7281,132	436,092
13	6842,43	6982,989	140,559
14	6642,11	6920,834	278,724
15	7116,8	6856,116	-260,684
16	7096,18	7178,216	82,036
17	7257,66	7136,045	-121,615
18	7189,42	7258,978	69,558
19	6881,96	7238,854	356,894
20	6880,32	6909,697	29,377
21	7117,21	6921,328	-195,882
22	7429,72	7139,658	-290,062
23	7550,9	7147,65	-403,25
OMH değeri			202,3562

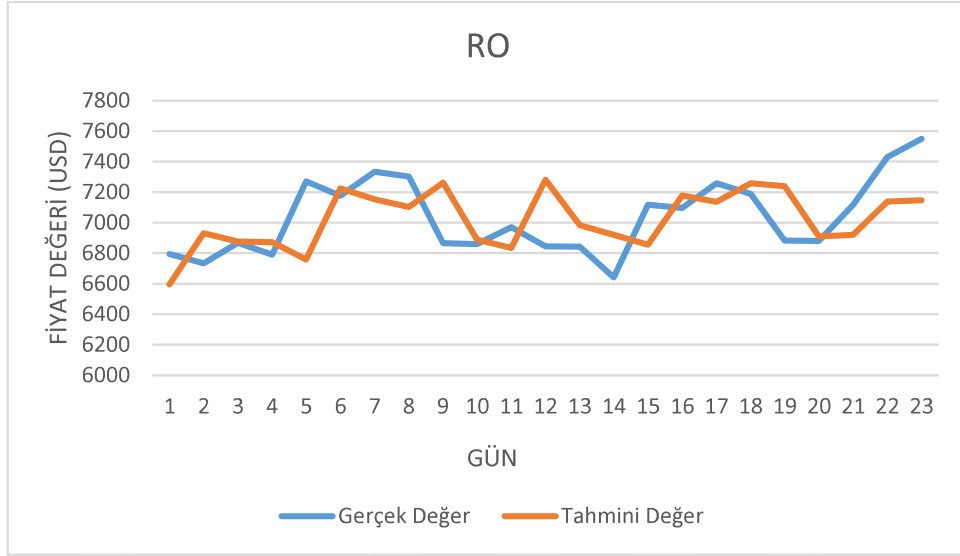
Rastgele Orman algoritmasıyla elde edilen tahmini değerler incelendiğinde elde edilen sonuçların, KNN yönteminde elde edilen sonuçlara göre daha başarılı olduğu görülmüştür. Ancak DVR yöntemiyle kıyaslandığında daha hatalı tahmini veriler üretmiştir. Bitcoin veri setinin KNN, DVR ve RO yöntemleriyle analiz edilmesi sonucunda bu veri setinde en başarılı sonuç veren yöntemin Destek Vektör Regresyonu yöntemi olduğu sonucuna varılmıştır. Şekil 5.1, Şekil 5.2 ve Şekil 5.3'te yöntemlerde elde edilen tahmini veriler gerçek verilerle karşılaştırılmıştır.



Şekil 5.1. KNN yöntemiyle elde edilen Bitcoin verilerinin karşılaştırılması



Şekil 5.2. DVR yöntemiyle elde edilen Bitcoin verilerinin karşılaştırılması



Şekil 5.3. RO yöntemiyle elde edilen Bitcoin verilerinin karşılaştırılması

Yöntemlere ait grafikler incelendiğinde en başarılı sonuçlar veren DVR algoritmasının değerleri gerçek değerlerle paralel olarak hareket ederek diğer yöntemlere göre fiyat değişimlerine daha hızlı tepki verdiği görülmüştür. RO algoritmasında ise bazı noktadaki ciddi sapmalar olmuştur ve fiyat değişimlerine verdiği tepkiler daha yavaş gerçekleşmiştir. KNN grafiğinde diğer iki algoritmaya göre hem oluşturduğu veriler açısından hem de değişime verdiği tepki açısından başarısız olduğu söylenebilir.

5.2. Ethereum Fiyat Tahmin Analizi

Ethereum fiyat tahmini amacıyla K En Yakın Komşu algoritması kullanılması için ilk olarak veri seti eğitim ve test setleri olarak iki ayrılmıştır. Eğitim veri seti K parametresi ve k katlamalı çapraz doğrulama parametre değerleri değiştirilerek en düşük OMH puanına sahip olan parametreler, fiyat tahminlemesi amacıyla kullanılmıştır. ETH eğitim veri setinin KNN algoritmasına ait k ve k katlamalı çapraz doğrulama parametrelerine göre elde edilen OMH değerleri Tablo 5.7’de verilmiştir.

Tablo 5.7. *Ethereum KNN modellerinin k değeri ve k katlamalı çapraz doğrulama sonuçları*

		K kat çapraz doğrulama değeri					
K parametre değeri		5	6	7	8	9	10
	1	13,0159	13,2628	13,1559	13,1512	13,3136	13,2327
	3	10,8292	10,8088	10,8353	10,8153	10,749	10,7878
	5	10,2658	10,3125	10,2009	10,3753	10,4389	10,1677
	7	9,9672	10,0324	10,0655	10,0421	10,1298	9,9247
	9	9,8676	10,0128	10,0549	9,8955	9,9435	9,8559
	11	9,9171	9,9974	10,0006	9,9409	9,9566	9,8729
	13	10,0075	9,9736	9,9868	9,9976	9,976	9,858

ETH eğitim veri setinin KNN yöntemiyle analizleri sonucunda, modellerden k parametresi 9 ve k katlamalı çapraz doğrulama parametresi ise 10 olarak seçildiğinde en düşük OMH sonucu elde edildiği görülmüştür. ETH fiyat tahmini için bu parametreler ile eğitilen model kullanılmıştır. KNN ile yapılan 23 günlük fiyat tahmin analizinin gerçek fiyatlarla karşılaştırılması, hata değeri ve OMH sonucu Tablo 5.8’de gösterilmiştir.

Tablo 5.8. *KNN yöntemiyle yapılan Ethereum fiyat tahmin analiz sonuçları*

Gün	Gerçek Değer	Tahmini Değer	Hata Değeri
1	142,03	129,538	-12,492
2	142,09	146,436	4,346
3	145,22	135,661	-9,559
4	143,55	136,261	-7,289
5	169,14	140,246	-28,894
6	165,1	196,706	31,606
7	172,64	196,706	24,066
8	170,81	173,2	2,39
9	158,41	174,142	15,732
10	158,22	175,074	16,854
11	161,14	158,757	-2,383
12	156,28	169,019	12,739
13	157,6	154,547	-3,053

Tablo 5.8. (Devam) *KNN yöntemiyle yapılan Ethereum fiyat tahmin analiz sonuçları*

Gün	Gerçek Değer	Tahmini Değer	Hata Değeri
14	153,29	165,967	12,677
15	172,16	153,474	-18,686
16	171,64	198,897	27,257
17	186,91	173,2	-13,71
18	181,61	207,418	25,808
19	172,3	202,218	29,918
20	172,74	196,706	23,966
21	182,6	175,141	-7,459
22	185,03	193,913	8,883
23	189,24	204,247	15,007
OMH değeri			15,4249

ETH fiyat tahmininin KNN yöntemi kullanılarak elde edilen tahmini değerler incelendiğinde gerçek değerden sapmaların çok yüksek değerler aldığı gözlemlenmiştir. Modelin elde ettiği tahmini değerlerin hata değerleri incelendiğinde %8'den fazla ortalama hata oranı görülmüştür.

Destek Vektör Regresyonu yöntemi kullanılarak Ethereum fiyat tahmini öncesinde en başarılı parametre değerleri elde edilmeye çalışılmıştır. Bunun için ETH eğitim veri seti C katsayısı ve k katlamalı çapraz doğrulama parametre değerlerine göre modeller oluşturularak, en düşük ortalama mutlak hata değerine sahip model belirlenmiştir. Modellerin C katsayısı ve k katlamalı çapraz doğrulama değerlerine göre değişen ortalama mutlak hata değerleri Tablo 5.9'da gösterilmiştir.

Tablo 5.9. *Ethereum DVR modellerinin C katsayısı ve k katlamalı çapraz doğrulama sonuçları*

		K kat çapraz doğrulama değeri					
		5	6	7	8	9	10
C katsayısı	5	8,4533	8,4837	8,4958	8,4657	8,454	8,4586
	6	8,4727	8,4744	8,4714	8,4634	8,4471	8,4608
	7	8,4482	8,4761	8,4797	8,4573	8,4521	8,4582
	8	8,4446	8,4707	8,4699	8,4528	8,4543	8,4509
	9	8,4472	8,4422	8,4679	8,4505	8,4593	8,4489
	10	8,4585	8,4686	8,4651	8,4439	8,4538	8,4495

Modellere ait ortalama mutlak hata değerlerine göre değerlendirildiğinde C katsayısının 9, k kat çapraz doğrulama parametresinin değeri 6 seçildiğinde en düşük puanı aldığı görülmüştür. Bu modelle elde edilen 23 günlük tahmini fiyat değeri, gerçek fiyat değeri ile karşılaştırılmış, gerçek fiyat ile tahmini fiyat arasındaki hata oranı belirlenmiş ve tüm tahminin ortalama mutlak hata puanı hesaplanarak Tablo 5.10’da verilmiştir.

Tablo 5.10. *DVR yöntemiyle yapılan Ethereum fiyat tahmin analiz sonuçları*

Gün	Gerçek Değer	Tahmini Değer	Hata Değeri
1	142,03	135,337	-6,693
2	142,09	141,828	-0,262
3	145,22	142,113	-3,107
4	143,55	144,776	1,226
5	169,14	143,439	-25,701
6	165,1	168,197	3,097
7	172,64	165,508	-7,132
8	170,81	171,884	1,074
9	158,41	170,539	12,129
10	158,22	159,92	1,7
11	161,14	158,022	-3,118
12	156,28	160,84	4,56
13	157,6	156,587	-1,013
14	153,29	157,23	3,94
15	172,16	153,644	-18,516
16	171,64	171,46	-0,18
17	186,91	171,229	-15,681
18	181,61	185,996	4,386
19	172,3	181,757	9,457
20	172,74	173,459	0,719
21	182,6	172,235	-10,365
22	185,03	181,83	-3,2
23	189,24	184,57	-4,67
OMH değeri			6,1707

Yukarıdaki tablo incelendiğinde, Destek Vektör Regresyonu algoritması ile Ethereum fiyat tahmininin ortalama hata oranı %3 olarak gerçekleştiği görülmektedir. Aynı yöntemin Bitcoin fiyat tahmininin aksine Ethereum fiyat tahmininden elde edilen

OMH değeri, eğitim setinin OMH değerinden daha düşük puan almıştır. Bu açıdan bakıldığında DVR yöntemi Ethereum fiyat tahmini için tatmin edici düzeyde başarılı olmuştur.

Ethereum fiyat tahmini için kullanılacak Rastgele Orman yönteminin en ideal modelini belirlemek amacıyla ayrılan ETH eğitim veri seti k katlamalı çapraz doğrulama değerleri değiştirilerek modellerin OMH değeri incelenmiştir. Ortalama mutlak hata değeri en düşük puana sahip model tahmin modeli olarak seçilmiştir. Tablo 5.11’de RO algoritması ile eğitilen modellere ait k katlamalı çapraz değerler ve bu değerlere ait OMH değerleri verilmiştir.

Tablo 5.11. *Ethereum RO modellerinin k katlamalı çapraz doğrulama sonuçları*

K	Hata değeri	K	Hata değeri
3	10,9729	7	10,8775
4	10,9648	8	10,883
5	10,7245	9	10,9903
6	10,8411	10	10,9258

RO algoritması ile fiyat tahminlemesi yapılması amacıyla modeller oluşturularak k katlamalı çapraz doğrulama parametresi 5 olarak seçilen modelde en düşük OMH değeri elde edilmiştir. Ethereum fiyat tahmini için RO algoritmasıyla oluşturulan bu model kullanılmıştır. Rastgele Orman algoritmasıyla yapılan 23 günlük fiyat tahmini ile ilgili veriler Tablo 5.12’de gösterilmiştir.

Tablo 5.12. *RO yöntemiyle yapılan Ethereum fiyat tahmin analiz sonuçları*

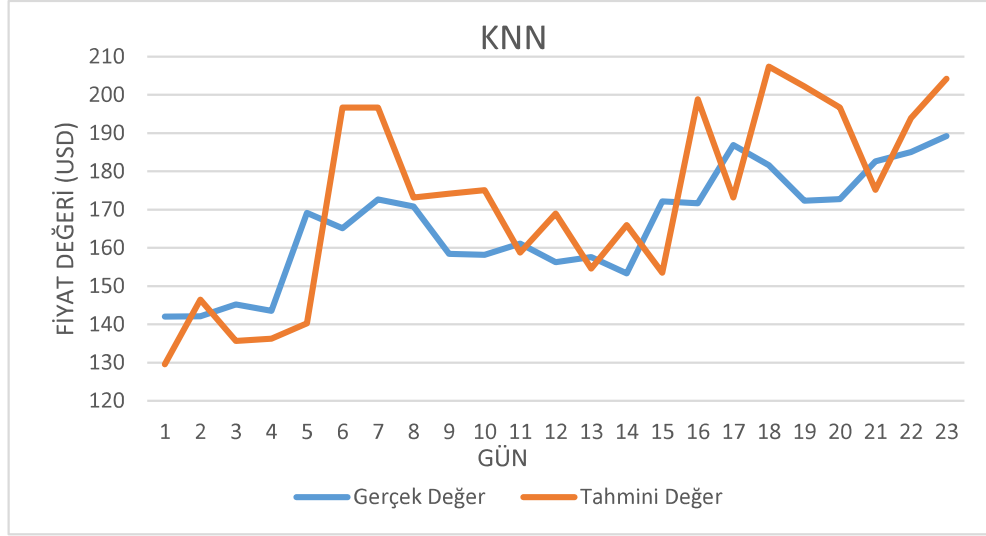
Gün	Gerçek Değer	Tahmini Değer	Hata Değeri
1	142,03	135,21	-6,82
2	142,09	143,006	0,916
3	145,22	143,239	-1,981
4	143,55	145,454	1,904
5	169,14	146,52	-22,62
6	165,1	164,318	-0,782
7	172,64	166,165	-6,475
8	170,81	169,312	-1,498
9	158,41	173,132	14,722

Tablo 5.12. (Devam) *RO yöntemiyle yapılan Ethereum fiyat tahmin analiz sonuçları*

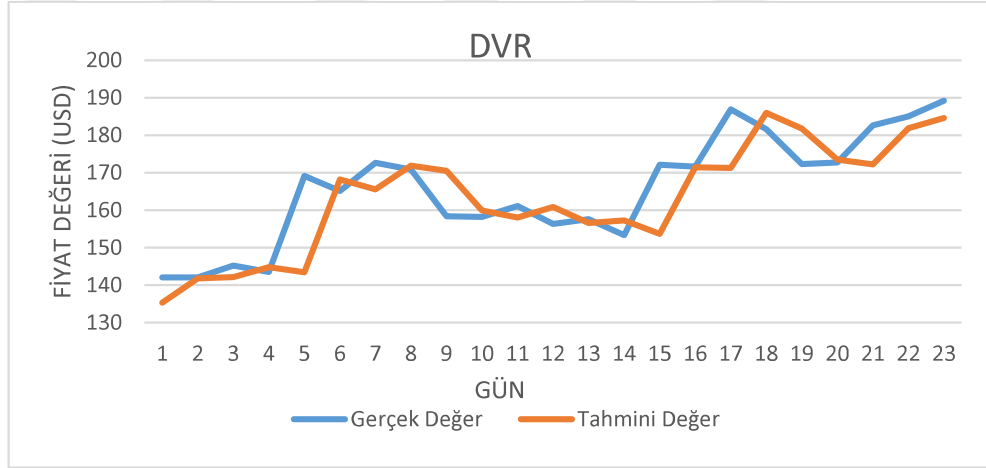
Gün	Gerçek Değer	Tahmini Değer	Hata Değeri
10	158,22	162,395	4,175
11	161,14	160,784	-0,356
12	156,28	163,818	7,538
13	157,6	158,764	1,164
14	153,29	160,421	7,131
15	172,16	155,948	-16,212
16	171,64	167,002	-4,638
17	186,91	172,026	-14,884
18	181,61	165,232	-16,378
19	172,3	166,107	-6,193
20	172,74	169,028	-3,712
21	182,6	172,026	-10,574
22	185,03	165,561	-19,469
23	189,24	164,957	-24,283
OMH değeri			8,4533

Rastgele Orman algoritmasıyla elde edilen tahmin değerleri incelendiğinde ortalama hata oranının %4 civarlarında olduğu hesaplanmıştır. DVR yöntemine göre başarısı daha düşüktür. KNN algoritmasına göre ise daha başarılı sonuçlar vermiştir. DVR yönteminde olduğu gibi Rastgele Orman algoritmasında da eğitim verisinden elde edilen OMH değerinden daha düşük bir ortalama mutlak hata değeri elde edilmiştir.

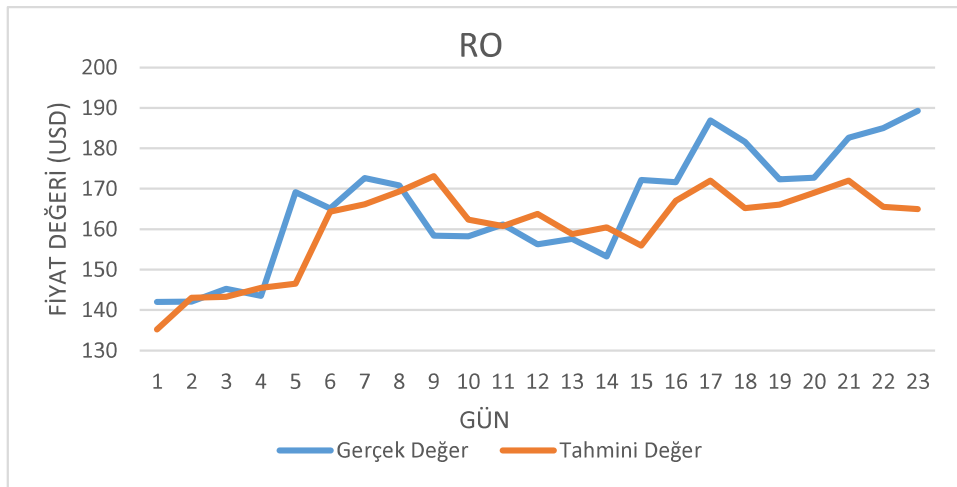
Tüm bu sonuçlardan yola çıkarak Ethereum fiyat tahmini analizinde en başarılı sonuç veren yöntemin Destek Vektör Regresyonu algoritması olduğu söylenebilir. KNN, DVR ve RO algoritmalarıyla elde edilen verilerin gerçek verilerle karşılaştırılması Şekil 5.4, Şekil 5.5 ve Şekil 5.6'deki grafiklerde gösterilmiştir.



Şekil 5.4. KNN yöntemiyle elde edilen Ethereum verilerinin karşılaştırılması



Şekil 5.5. DVR yöntemiyle elde edilen Ethereum verilerinin karşılaştırılması



Şekil 5.6. RO yöntemiyle elde edilen Ethereum verilerinin karşılaştırılması

Ethereum fiyat analizinin yapıldığı üç yönteme ait grafikler incelendiğinde en başarılı sonuçlar veren yöntemin DVR olduğu görülmektedir. Destek Vektör Regresyonu yöntemi diğer yöntemlere kıyasla gerçek fiyatla benzer şekilde hareket etmiş hata değerleri de diğer algoritmalara göre makul olduğu görülmektedir. RO algoritmasında ise bazı noktalardaki ciddi sapmalarla birlikte hata değerleri de DVR'ye göre ciddi oranda artmıştır. RO algoritması DVR ile kıyaslandığında fiyat değişimlerine verdiği tepkiler çok daha yavaş gerçekleşmiştir. KNN yönteminin grafiğinde ise diğer iki yönteme göre her açıdan başarısız sonuçlar verdiği söylenebilir.

5.3. XRP Fiyat Tahmin Analizi

K En Yakın Komşu algoritması kullanılması için ilk olarak XRP veri seti eğitim ve test setleri olarak iki ayrılmıştır. Ayrılan XRP eğitim veri seti kullanılarak k değeri ve k katlamalı çapraz doğrulama parametrelerinin değerleri değiştirilerek yöntem için en ideal parametre değerleri bulunmuştur. En düşük OMH puanına sahip olan parametre değerleri, fiyat tahmini için kullanılmıştır. Tablo 5.13'de XRP eğitim veri setine ait k ve k katlamalı çapraz doğrulama parametrelerine göre elde edilen OMH değerleri verilmiştir.

Tablo 5.13. XRP KNN modellerinin k değeri ve k katlamalı çapraz doğrulama sonuçları

		K kat çapraz doğrulama değeri					
		5	6	7	8	9	10
K parametre değeri	1	0,0153	0,0158	0,0159	0,0157	0,0161	0,0156
	3	0,0136	0,0136	0,0135	0,0136	0,0138	0,0135
	5	0,0126	0,0129	0,0127	0,0128	0,0129	0,0127
	7	0,0126	0,0129	0,0129	0,0129	0,013	0,0127
	9	0,0125	0,0128	0,0125	0,0127	0,0128	0,0125
	11	0,0124	0,0126	0,0126	0,0124	0,0126	0,0123
	13	0,0128	0,0129	0,0131	0,0127	0,0126	0,0125

XRP veri setinde en düşük ortalama mutlak hata puanına sahip parametre ikilisi k parametresi 11 ve k katlamalı çapraz doğrulama parametresi ise 10 olarak seçildiğinde en ideal model elde edilmiştir. Bu modelle 23 günlük fiyat tahminlemesi gerçekleştirilmiştir. Fiyat tahmin analizinin gerçek fiyatlarla karşılaştırılması, gerçekleşen hata değeri ve OMH sonucu Tablo 5.14’te gösterilmiştir.

Tablo 5.14. *KNN yöntemiyle yapılan XRP fiyat tahmin analiz sonuçları*

Gün	Gerçek Değer	Tahmini Değer	Hata Değeri
1	0,179	0,17	-0,009
2	0,179	0,18	0,001
3	0,182	0,17	-0,012
4	0,179	0,192	0,013
5	0,196	0,201	0,005
6	0,192	0,211	0,019
7	0,201	0,211	0,01
8	0,198	0,19	-0,008
9	0,187	0,207	0,02
10	0,188	0,175	-0,013
11	0,19	0,194	0,004
12	0,188	0,175	-0,013
13	0,185	0,184	-0,001
14	0,181	0,182	0,001
15	0,19	0,172	-0,018
16	0,19	0,2	0,01
17	0,195	0,189	-0,006
18	0,191	0,181	-0,01
19	0,183	0,185	0,002
20	0,184	0,177	-0,007
21	0,188	0,178	-0,01
22	0,192	0,203	0,011
23	0,194	0,184	-0,01
OMH değeri			0,0093

XRP fiyat tahmininin KNN algoritması ile elde edilen tahmini değerler incelendiğinde gerçek değerlerinden ortalama sapmaları %5 civarlarında gerçekleştiği görülmüştür. Üç veri setinin KNN yöntemiyle ortalama hata oranlarına bakıldığında bu yöntem, XRP veri setinde en başarılı sonuçları vermiştir.

XRP fiyat tahminlemesi için kullanılacak diğer bir yöntem olan Destek Vektör Regresyonu yöntemi için ilk olarak en ideal parametre değerleri elde edilmeye çalışılmıştır. XRP eğitim veri seti C katsayısı ve k katlamalı çapraz doğrulama parametre değerlerine göre modeller oluşturularak, en düşük OMH değerine sahip model belirlenmiştir. Tablo 5.15’de modellere ait C katsayısı ve k katlamalı çapraz doğrulama değerlerine göre değişen ortalama mutlak hata değerleri verilmiştir.

Tablo 5.15. XRP DVR modellerinin C katsayısı ve k katlamalı çapraz doğrulama sonuçları

		K kat çapraz doğrulama değeri					
		5	6	7	8	9	10
C katsayısı	11	0,0111	0,0112	0,0113	0,0111	0,0112	0,01127
	12	0,0111	0,0113	0,0111	0,0112	0,0111	0,01116
	13	0,0112	0,0113	0,0113	0,0112	0,0111	0,01126
	14	0,0112	0,0111	0,0112	0,0111	0,0112	0,01137
	15	0,0111	0,0112	0,0112	0,0110	0,0115	0,01135

Modellere ait C ceza katsayısının 15, k kat çapraz doğrulama parametresinin değeri 8 olduğunda ortalama mutlak hata değerlerine göre değerlendirildiğinde en düşük puanı aldığı görülmüştür. Tablo 5.16’da 23 günlük tahmini fiyat değeri, gerçek fiyat ile tahmini fiyat arasındaki hata değeri ve ortalama mutlak hata puanı verilmiştir.

Tablo 5.16. DVR yöntemiyle yapılan XRP fiyat tahmin analiz sonuçları

Gün	Gerçek Değer	Tahmini Değer	Hata Değeri
1	0,179	0,176	-0,003
2	0,179	0,178	-0,001
3	0,182	0,18	-0,002
4	0,179	0,182	0,003

Tablo 5.16. (Devam) *DVR yöntemiyle yapılan XRP fiyat tahmin analiz sonuçları*

Gün	Gerçek Değer	Tahmini Değer	Hata Değeri
5	0,196	0,181	-0,015
6	0,192	0,195	0,003
7	0,201	0,192	-0,009
8	0,198	0,201	0,003
9	0,187	0,199	0,012
10	0,188	0,189	0,001
11	0,19	0,189	-0,001
12	0,188	0,19	0,002
13	0,185	0,189	0,004
14	0,181	0,186	0,005
15	0,19	0,182	-0,008
16	0,19	0,19	0
17	0,195	0,191	-0,004
18	0,191	0,195	0,004
19	0,183	0,192	0,009
20	0,184	0,184	0
21	0,188	0,185	-0,003
22	0,192	0,189	-0,003
23	0,194	0,192	-0,002
OMH değeri			0,0042

DVR algoritması kullanılarak XRP fiyat tahmininin ortalama hata oranı %2,2 olarak gerçekleştiği görülmektedir. DVR yöntemiyle tahmin edilen verilerin ortalama mutlak hata değeri, eğitim setinin ortalama mutlak hata değerinden %60 daha az puan almıştır. Böylelikle bu yöntem en başarılı fiyat tahminini XRP veri setinde göstermiştir.

Rastgele Orman yöntemi için en ideal modeli belirlemek amacıyla ayrılan XRP eğitim veri seti k katlamalı çapraz doğrulama değerleri değiştirilerek modellerin ortalama mutlak hata değerleri hesaplanmıştır. Bu modellerden en düşük ortalama mutlak hata değerine sahip model XRP tahmin modeli olarak seçilmiştir.

Tablo 5.17’de RO algoritması ile eğitilen modellere ait k katlamalı çapraz doğrulama değerleri ve bu değerlere ait OMH değerleri verilmiştir.

Tablo 5.17. *XRP RO modellerinin k katlamalı çapraz doğrulama sonuçları*

K	Hata değeri	K	Hata değeri
3	0,0134	7	0,0135
4	0,0136	8	0,0135
5	0,0135	9	0,0137
6	0,0135	10	0,0135

Fiyat tahminlemesi için modeller oluşturularak en düşük ortalama mutlak hata değerini veren k katlamalı çapraz doğrulama parametresi 3 olarak belirlenmiştir. Bu parametre değeri kullanılarak elde edilen modelle 23 günlük XRP fiyat tahminlemesi yapılmıştır. Rastgele Orman algoritmasıyla elde edilen tahmin verileri, bu verilere ait hata değerleri ve OMH değeri Tablo 5.18’de verilmiştir.

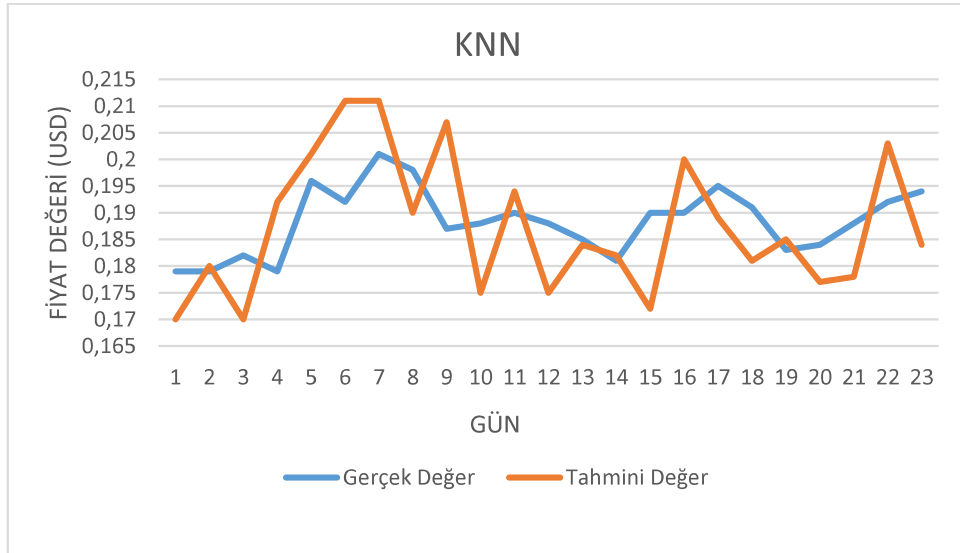
Tablo 5.18. *RO yöntemiyle yapılan XRP fiyat tahmin analiz sonuçları*

Gün	Gerçek Değer	Tahmini Değer	Hata Değeri
1	0,179	0,175	-0,004
2	0,179	0,18	0,001
3	0,182	0,181	-0,001
4	0,179	0,182	0,003
5	0,196	0,182	-0,014
6	0,192	0,188	-0,004
7	0,201	0,193	-0,008
8	0,198	0,192	-0,006
9	0,187	0,198	0,011
10	0,188	0,189	0,001
11	0,19	0,191	0,001
12	0,188	0,19	0,002
13	0,185	0,187	0,002
14	0,181	0,189	0,008
15	0,19	0,184	-0,006
16	0,19	0,187	-0,003
17	0,195	0,191	-0,004

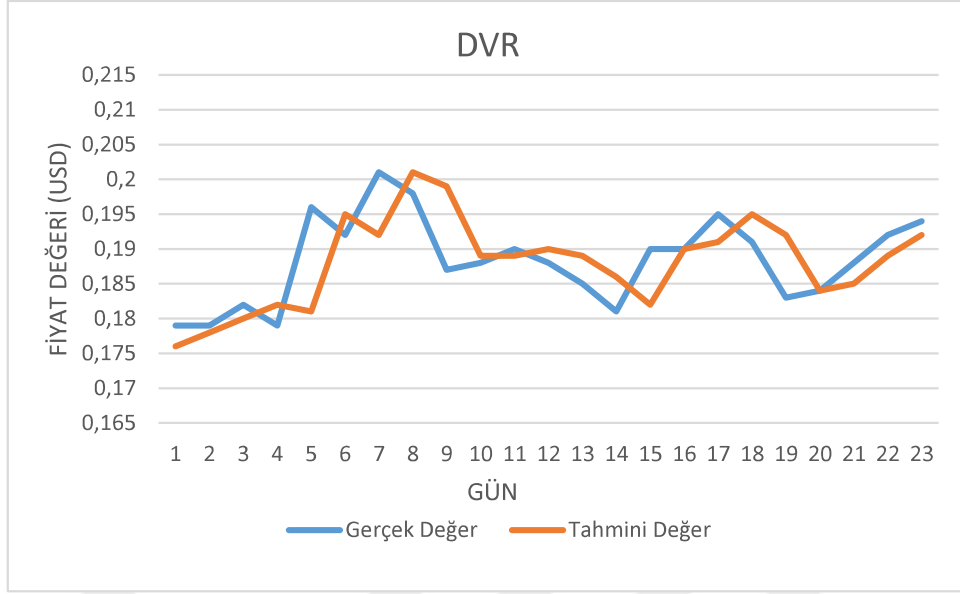
Tablo 5.18. (Devam) *RO yöntemiyle yapılan XRP fiyat tahmin analiz sonuçları*

Gün	Gerçek Değer	Tahmini Değer	Hata Değeri
18	0,191	0,194	0,003
19	0,183	0,193	0,01
20	0,184	0,189	0,005
21	0,188	0,188	0
22	0,192	0,189	-0,003
23	0,194	0,191	-0,003
OMH değeri			0,0044

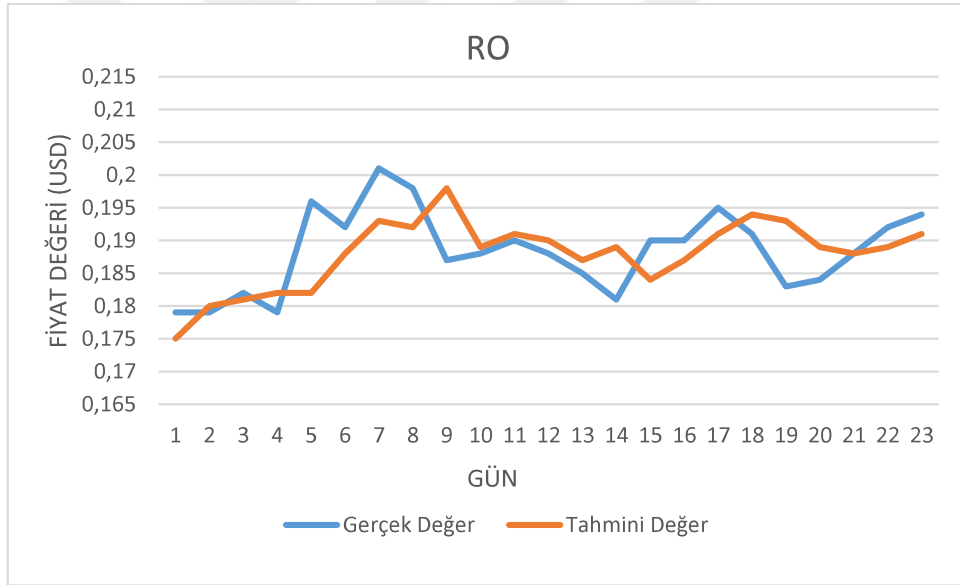
XRP tahmin verileri incelendiğinde ortalama hata oranının %2.3 olarak hesaplanmıştır. DVR ile RO algoritmaları XRP veri seti bazında kıyaslandığında birbirine çok yakın sonuçlar verdiği görülmüştür. KNN algoritması ise bu yöntemlere kıyasla daha başarısız sonuçlar vermiştir. DVR yönteminde olduğu gibi Rastgele Orman algoritmasında da eğitim verisinden elde edilen OMH değerinden daha düşük bir ortalama mutlak hata değeri elde edilmiştir. Şekil 5.7, Şekil 5.8 ve Şekil 5.9'daki grafiklerde KNN, DVR ve RO algoritmalarıyla elde edilen verilerin gerçek verilerle karşılaştırılması gösterilmiştir.



Şekil 5.7. *KNN yöntemiyle elde edilen XRP verilerinin karşılaştırılması*



Şekil 5.8. DVR yöntemiyle elde edilen XRP verilerinin karşılaştırılması



Şekil 5.9. RO yöntemiyle elde edilen XRP verilerinin karşılaştırılması

Üç yönteme ait grafikler incelendiğinde XRP fiyat analizinde en başarılı sonuçlar veren yöntemin DVR olduğu görülmüştür. DVR ile RO yöntemleri birbirine çok yakın OMH değerleri vermiş olmasına rağmen, RO yöntemiyle elde edilen tahmini verilerde ciddi sapmalar görülmüştür. Rastgele Orman algoritması, DVR ile kıyaslandığında fiyat değişimlerine verdiği tepkiler çok daha yavaş gerçekleşmiştir. KNN yönteminin grafiği incelendiğinde ise diğer iki yönteme göre her açıdan başarısız sonuçlar verdiği söylenebilir.

6. SONUÇ VE ÖNERİLER

Kripto paralar, farklı amaçları gerçekleştirebilmek için farklı özelliklere sahip, birçoğunun toplam arzı belirli dijital varlıklardır. Kripto paralar ortaya çıktıkları 2009 yılından günümüze kadar geçen sürede her türlü yatırımcının dikkatini çeken yatırım araçlarının başında gelmektedir. Kripto paraların asıl amacı ödeme ve para transfer işlemlerinde kullanılması olsa da, büyük bir kitlenin yatırım aracı olarakta kullandığı bir emtia olmuştur. Teknolojik gelişmelerin finans sistemini değiştirmeye başladığı bu günlerde farklı birçok olumsuz değerlendirme olmasına rağmen, kripto paraların değer saklama aracı ve potansiyeli göz önüne alındığında gelecekte de yatırım aracı olarak kullanılacaktır. Öyle ki, hiçbir devlet ve kurum tarafından desteklenmemesine rağmen kripto paraların ilki ve en popüler olan Bitcoin sürekli olarak yeni yatırımcıların dikkatini çekebilmektedir. Bu nedenle kripto paraların gelecekteki fiyat tahmininin yapılması, yatırımcıların vazgeçilmezleri arasında yerini almaktadır.

Bu çalışmada, kripto para dünyasının en değerli üç kripto parası olan BTC, ETH ve XRP'nin günlük gerçekleşen açılış, kapanış, en yüksek ve en düşük fiyat, günlük hacim ve günlük toplam piyasa değeri öznitelikleri kullanılarak KNN, DVR ve RO algoritmalarıyla fiyat tahmini yapılmıştır. Fiyat tahmininin kur fiyatlarından etkilenmemesi için USD baz alınarak gerçekleştirilmiştir. Kullanılacak veri seti, ilgili kripto paranın borsalarda işlem gördüğü ilk tarihten günümüze kadar geçen zamandaki tüm veriler online ortamdan çekilerek oluşturulmuştur.

Çalışmada veri setlerindeki bazı özniteliklerin eksik verilerden oluştuğu görülmüş ve eksik olan tarihteki tüm veriler silinerek veri setleri kullanıma hazır hale getirilmiştir. Tüm veri setleri kendi içerisinde normalize edilmiş, sonrasında eğitim ve tahmin veri seti olarak ayrılarak analize hazır hale getirilmiştir. WEKA 3.8 (Hall vd. 2009) programı ile eğitim veri setleri K En Yakın Komşu, Destek Vektör Regresyonu ve Rastgele Orman algoritmalarına özgün parametreler ile k kat çapraz doğrulama değerleri değiştirilerek en ideal parametre değerleriyle eğitilmiştir. Oluşturulan modeller kullanılarak 23 günlük fiyat tahmini yapılmıştır. Fiyat tahminleri gerçekleşen gerçek verilerle karşılaştırılmıştır. Karşılaştırma sonucunda yöntemlerin başarı sıralaması sırasıyla DVR, RO ve KNN olmuştur. Destek Vektör Regresyonu algoritması ile Bitcoin, Ethereum ve XRP fiyat tahmininde en düşük ortalama mutlak hata sonuçları elde edilmiştir. K En Yakın Komşu algoritması ise en yüksek mutlak hata sonuçları verdiği için üç yöntem arasından en başarısız yöntem olarak belirlenmiştir.

Sonuç olarak DVR algoritmasının tahmini verileri gerçek verilerle karşılaştırıldığında, tahmini fiyat verilerinin gerçek verilerle uyduğu ve fiyat değişimlerine diğer yöntemlere kıyasla çok iyi tepki verdiği görülmüştür. Destek Vektör Regresyonu yönteminin, bu çalışmadaki başarısı göz önüne alındığında BTC, ETH ve XRP'nin işlem gördüğü borsalardaki diğer tüm kripto paralara da uygulanabileceği sonucuna varılmıştır.

Bu çalışmada büyük veri araçlarıyla kripto para fiyat değişimlerinin analizinde 16000 adet veri ile, büyük veri setlerine oranla nispeten daha küçük veri seti ile çalışılmıştır. Ancak çalışılan veri setinin veri satırı günlük veriler yerine 5 dakikalık gibi kısa zaman dilimlerine odaklanıldığında doğrudan artacaktır. Aynı zamanda, veri setinin öznelilik sayısı teknik analizde kullanılan indikatör değerlerinin eklenmesi ile arttırıldığında çalışma boyutu daha da arttırılabilir. Tüm bu durumlar göz önüne alındığında, bu çalışma tam anlamı ile bir büyük veri çalışması olmasa da belirtilen bu durumların göz önüne alınması ile yine aynı araçların kullanılacağı bir büyük veri çalışması olacaktır.

İleriki çalışmalarda, bu çalışmada kullanılan özneliliklerin yanı sıra teknik analizlerde kullanılan bazı indikatör değerleri de eklenerek daha fazla özneliliğe sahip bir veri seti kullanılabilir. Yine ilerideki başka bir çalışmada ise bu çalışmada kullanılan özneliliklerin farklı niteliklerle tahmininin yapıldıktan sonra, tahmini öznelilik değerleriyle fiyat tahmini yapılması şeklinde tahmin modeli iki aşamalı olacak şekilde analizi gerçekleştirilebilir.

KAYNAKÇA

- Akman, M., Genç, Y. ve Ankaralı, H. (2011). Random Forests Yöntemi ve Sağlık Alanında Bir Uygulama. *Türkiye Klinikleri Biyoistatistik Dergisi*, 3(1), 36–48.
- Aktepe, Ç. (2018). Algorithmic trading on cryptocurrency markets using machine learning techniques. Yüksek Lisans Tezi. *İstanbul: Boğaziçi University, Graduate Program in Industrial Engineering*.
- Alkhatib, K., Najadat, H., Hmeidi, I. ve Shatnawi, M.K.A. (2013). Stock price prediction using k-nearest neighbor (kNN) algorithm. *International Journal of Business, Humanities and Technology*, 3(3), 32-44.
- Al-Jaroodi, J. ve Mohamed, N. (2014). Real-time big data analytics: applications and challenges. *2014 International Conference on High Performance Computing and Simulation*, 305-310.
- Aristoteles. (1975). Politika, Çev. Mete Tuncay, *İstanbul: Remzi Kitabevi Yayınları*
- Atlan, F. (2019). Kripto para değerlerinin yapay zeka teknikleri ile tahmini. Yüksek Lisans Tezi. *Burdur: Burdur Mehmet Akif Ersoy Üniversitesi, Sosyal Bilimler Enstitüsü*.
- Baba, B. ve Sevil, G. (2020). Predicting IPO initial returns using random forest. *Borsa İstanbul Review*, 9(1), 13-23.
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32.
- Bürüce, B. (2018). Ekonomik ve finansal araştırmalarda büyük veri. Yüksek Lisans tezi. *Sakarya: Sakarya Üniversitesi, Sosyal Bilimler Enstitüsü*.
- Carradice, I. ve Price, M. (2001). Hellen Dünyasında Sikke, Çev. Oğuz Tekin, *İstanbul: Homer Kitabevi ve Yayıncılık*.
- Chen, Z., Li, C. ve Sun, W. (2020). Bitcoin price prediction using machine learning: An approach to sample dimension engineering. *Journal of Computational and Applied Mathematics*, 365.
- Cover, T. M. ve Hart, P.E. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*. 13(1), 21–27.
- Cortes, C. ve Vapnik, V. (1995). Support-Vector Networks, *Machine Learning*, 20, 273-297.
- Cutler, A., Cutler, D. R. ve Stevens, J. R. (2011). Random Forests, *Machine Learning* , 157- 176.

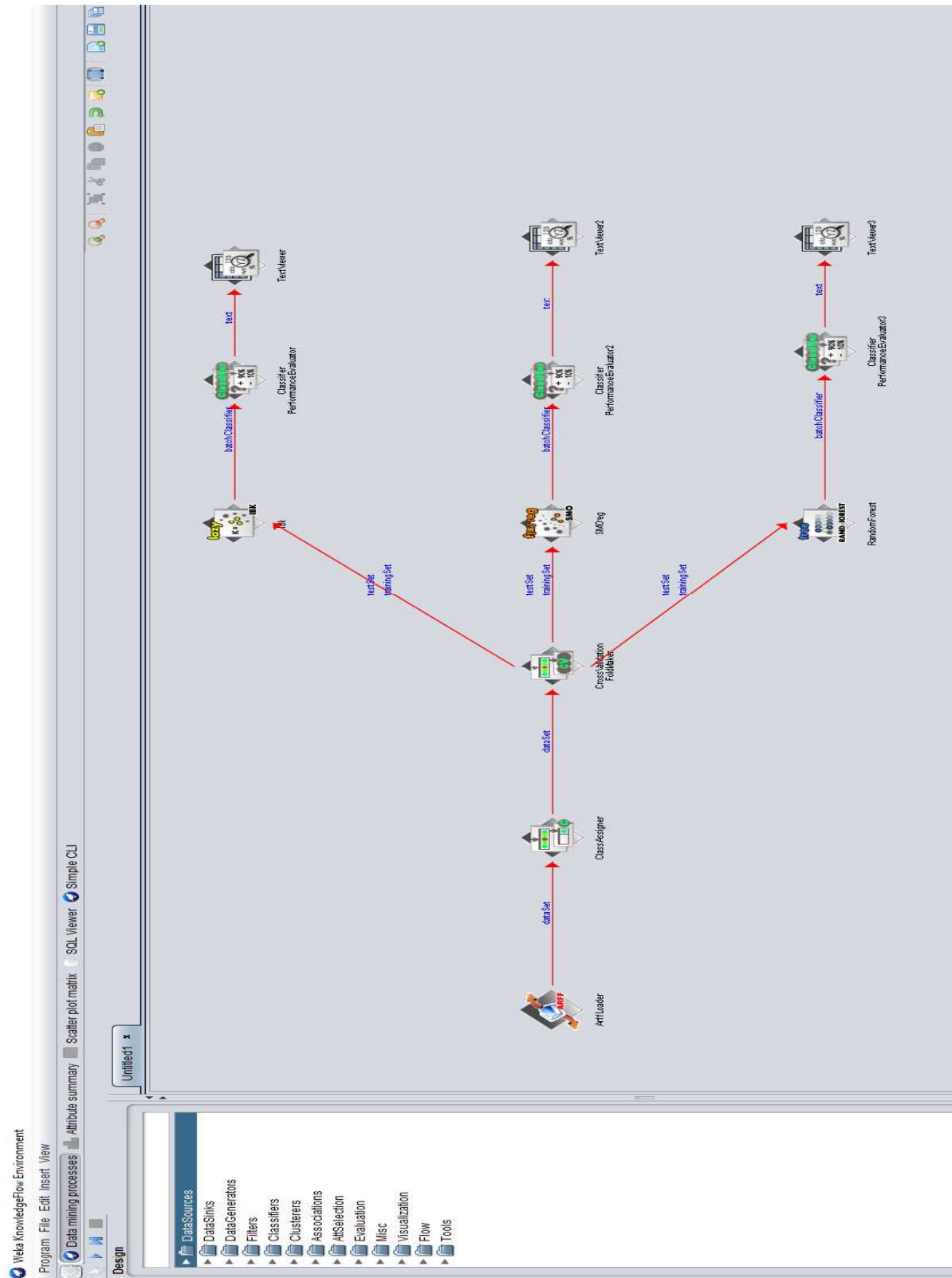
- Demirci, C. (2018). Büyük veri platformlarından hadoop ile örnek veri analizi. Mezuniyet Projesi. *İstanbul: Maltepe Üniversitesi, Fen Bilimleri Enstitüsü.*
- Fidan, H. (2020). Random forest (Rastgele Orman) algoritması temelli süreç izleme yönteminin ambulatuvar kan basıncı izlemede hipertansiyon erlen tanısı için kullanımı. Yüksek Lisans Tezi. *Muğla: Muğla Sıtkı Koçman Üniversitesi, Fen Bilimleri Enstitüsü.*
- Fradinata, E., Kesuma, Z.M., Rusdina, S. ve Zaman, N. (2019). Forecast analysis of instant noodle demand using support vector regression(SVR). *1ST SOUTH ACEH INTERNATIONAL CONFERENCE ON ENGINEERING AND TECHNOLOGY (SAICOET): IOP Conference Series-Materials Science and Engineering*, 506.
- Filiz, E., Karaboğa, H.A. ve Akogul, S. (2017). Bist-50 endeksi değişim değerlerinin sınıflandırılmasında makine öğrenmesi yöntemleri ve yapay sinir ağları kullanımı. *Çukurova Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 26(1), 231-241.
- Gagarina, M., Nestik, T. ve Drobysheva, T. (2019). Social and psychological predictors of youths' attitudes to cryptocurrency. *Behavioral Sciences*, 9, 118.
- Goel, P., Datta, A. ve Mannan, M. (2017). Application of big data analytics in process safety and risk management. *2017 IEEE International Conference on Big Data(BIGDATA)*, 1143-1152.
- Gök, M. (2017). Makine öğrenmesi yöntemleri ile akademik başarının tahmin edilmesi. *Gazi Üniversitesi Fen Bilimleri Dergisi*, 5(3), 139-148.
- Gültekin, Y., Özdemir, F.S. ve Varıcı, İ. (2019). Kripto para birimlerinin piyasa değerlerindeki değişimlerinin analizi. *Yönetim ve Ekonomi*, 26(2), 677-688.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. ve Witten, I. H. (2009). The WEKA data mining software: An update. *SIGKDD Explorations*, 11(1), 10-18.
- Herrera, G.P., Constantino, M., Tabak, B.M., Pistori, H., Su, J.J., ve Naranpanawa, A. (2019). Long-term forecast of energy commodities price using machine learning. *Energy*, 179, 214-221.
- Ince, H. ve Trafalis, T.B. (2008). Short term forecasting with support vector machines and application to stock price prediction. *International Journal of General Systems*, 37(6), 677-687.

- İlarslan, K. (2016). K-en yakın komşu (KNN) algoritması ile hisse senedi fiyatlarının tahmin edilmesi: Bist'den örnek bir uygulama. *Akademik Sosyal Araştırmalar Dergisi*, 30, 375-392.
- Ji, S., Kim, J. ve Im, H. (2019). A comparative study of bitcoin price prediction using deep learning. *Mathematics*, 7, 898.
- Kamalak, M. (1980). Para ve Para Sistemleri, *Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 4(3-4), 71-97.
- Keş, Y. ve Turgut, B. F. (2015). Kültür ve Sanatın Yansıması Para Tasarımı (Osmanlı Son Dönem Kaimeleri ve Cumhuriyetin İlk Emisyonları, *Medeniyet Sanat, İMÜ Sanat Ve Tasarım Fakültesi Dergisi*, 1(2), 31-47.
- Kumar, I., Dogra, K., Utreja, C. ve Yadav, P. (2018). A comparative study of supervised machine learning algorithms for stock market trend prediction. *International Conference on Inventive Communication and Computational Technologies (ICICCT)*, 1003-1007.
- Kurucan, F. (2019). Büyük veri ve finansal piyasalarda istatistiki makine öğrenmesi metodlarının yatırım kararlarında kullanılması. Yüksek Lisans Tezi. *İstanbul: İstanbul Üniversitesi, Sosyal Bilimler Enstitüsü*.
- Lahmiri, S. ve Bekiros, S. (2020). Intelligent forecasting with machine learning trading systems in chaotic intraday Bitcoin market. *Chaos Solitions&Fractals*, 133.
- Liu, J. ve Serletis, A. (2019). Volatility in the cryptocurrency market, *Open Economies Review*, 30, 779-811.
- Mallqui, D.C.A. ve Fernandes R.A.S. (2019). Predicting the direction, maximum, minimum, and closing prices of daily bitcoin exchange rate using machine learning techniques. *Applied Soft Computing Journal*, 75, 596-606.
- Meral, G. (2019). Türkiye'deki enerji santrallerinde doğal gaz tüketiminin destek vektör regresyon ile tahmini. Yüksek Lisans Tezi. *Afyon: Afyon Kocatepe Üniversitesi, Fen Bilimleri Enstitüsü*.
- Ömrüuzun, B. (2019). Yapay sinir ağları ile kripto paraların fiyat modellemesi. Yüksek Lisans Tezi. *İstanbul: İstanbul Üniversitesi, Sosyal Bilimler Enstitüsü*.
- Özdeş, M. (2017). Büyük veri araçlarını kullanarak duygu analizi gerçekleştirimi. Yüksek Lisans Tezi. *Denizli: Pamukkale Üniversitesi, Fen Bilimleri Enstitüsü*.

- Özer, S. (2019). Büyük veri teknolojileri ve veri madenciliği yöntemleri ile medikal veri analizi. Yüksek Lisans Tezi. *İstanbul: Marmara Üniversitesi, Fen Bilimleri Enstitüsü.*
- Smola, A. J. ve Schölkopf, B. (2004). A Tutorial on Support Vector Regression, *Statistics and Computing*, 14, 199-222.
- Tekin, O. (2008). Antik Nümismatik ve Anadolu (Arkaik ve Klasik Çağlar), *İstanbul: Arkeoloji ve Sanat Yayınları.*
- Valencia, F., Gomez-Espinosa, A., ve Valdes-Aguirre B. (2019). Price movement prediction of cryptocurrencies using sentiment analysis and machine learning. *Entropy*, 21(6), 589.
- Vapnik, V.N. (1998). Statistical learning theory. *John Wiley and Sons*, New York.
- Yalçın, C. (2016). Destek vektör regresyon ile uçaklarda anlık yakıt tüketiminin analizi. Doktora Tezi. *İstanbul: Mimar Sinan Güzel Sanatlar Üniversitesi, Fen Bilimleri Enstitüsü.*
- Yiğiter, Ş.Y., Sarı, S.S., Karabulut, T. ve Başakın, E.E. (2018). Kira sertifikası fiyat değerlerinin makine öğrenmesi metodu ile tahmini. *International Journal of Islamic Economics and Finance Studies*, 4(3), 74-82.
- Yolasığmaz, İ.B. (2015). A support vector regression method for conceptual cost estimate of construction projects. Yüksek Lisans Tezi. Ankara: *Middle East Technical University, Fen Bilimleri Enstitüsü.*
- http-1, <https://ekonomihukuk.com/para-banka/para-nedir/>
- http-2, <https://medium.com/dusunenbeyinler/big-data-b%C3%BCy%C3%BCk-veri-analizi-d53d8f8ab52b>, Big Data(Büyük Veri) nedir? , 2019.
- http-3, <https://kodcu.com/2011/12/2012nin-modasi-big-data>, 2012'nin modası Big Data, 2011.
- http-4, <https://medium.com/@k.ulgen90/makine-%C3%B6%C4%9Frenimi-b%C3%B6l%C3%BCm-2-6d6d120a18e1>, Makine Öğrenimi Bölüm-2 (k-En Yakın Komşuluk), 2017.
- http-5, <https://tr.sciencewal.com/77885-the-random-forest-algorithm-d457d499ffcd-33>

EKLER

EK-1



EK-2

Bitcoin Veri Seti

1	BITCOİN						
2	GAF	GEYF	GEDF	GKF	GH	GTPD	TFB
3	7,121.31	7,491.78	7,081.59	7,429.72	43,500,782,316	136,275,638,031	7,550.90
4	6,879.44	7,145.87	6,867.78	7,117.21	33,249,153,866	130,530,916,077	7,429.72
5	6,879.78	6,934.55	6,834.44	6,880.32	32,589,741,511	126,175,238,603	7,117.21
6	7,186.87	7,240.29	6,835.50	6,881.96	37,747,113,936	126,192,239,912	6,880.32
7	7,260.92	7,280.52	7,167.05	7,189.42	31,311,210,215	131,815,853,850	6,881.96
8	7,092.29	7,269.96	7,089.25	7,257.66	32,447,188,386	133,053,136,876	7,189.42
9	7,116.55	7,167.18	7,050.33	7,096.18	32,513,423,567	130,078,472,088	7,257.66
10	6,640.45	7,134.45	6,555.50	7,116.80	46,783,242,377	130,445,413,270	7,096.18
11	6,845.56	6,928.66	6,633.40	6,642.11	32,288,311,031	121,731,940,639	7,116.80
12	6,843.28	6,958.56	6,793.82	6,842.43	34,110,434,052	125,388,687,977	6,642.11
13	6,965.62	6,965.62	6,668.26	6,845.04	38,619,308,647	125,422,221,237	6,842.43
14	6,858.07	7,119.95	6,811.08	6,971.09	35,759,567,632	127,716,846,208	6,845.04
15	6,867.44	6,926.07	6,789.92	6,859.08	31,222,085,946	125,651,623,116	6,971.09
16	7,303.82	7,303.82	6,802.47	6,865.49	43,622,840,992	125,755,496,117	6,859.08
17	7,337.97	7,341.45	7,179.09	7,302.09	34,815,139,178	133,737,857,010	6,865.49
18	7,179.28	7,356.22	7,153.31	7,334.10	37,563,249,549	134,309,895,036	7,302.09
19	7,273.64	7,427.94	7,136.71	7,176.41	44,243,482,668	131,409,299,432	7,334.10
20	6,788.05	7,271.78	6,782.89	7,271.78	46,896,904,615	133,140,400,948	7,176.41
21	6,862.54	6,883.41	6,715.93	6,791.13	29,510,409,856	124,328,853,512	7,271.78
22	6,738.38	6,878.95	6,696.48	6,867.53	33,185,988,584	125,712,834,530	6,791.13
23	6,797.40	7,003.22	6,673.34	6,733.39	38,976,504,903	123,244,051,100	6,867.53
24	6,606.78	7,088.25	6,595.92	6,793.62	47,660,646,124	124,335,130,640	6,733.39
25	6,437.32	6,612.57	6,202.37	6,606.78	40,346,426,266	120,903,014,693	6,793.62
26	6,430.61	6,504.52	6,374.16	6,438.64	32,786,468,812	117,814,077,267	6,606.78
27	5,925.54	6,517.20	5,903.23	6,429.84	37,101,651,525	117,640,143,688	6,438.64
28	6,245.62	6,250.47	5,920.09	5,922.04	28,373,690,931	108,338,520,156	6,429.84
29	6,467.25	6,467.50	6,117.84	6,242.19	34,885,225,901	114,183,689,641	5,922.04
30	6,719.39	6,793.84	6,466.70	6,469.80	34,585,598,367	118,336,082,172	6,242.19
31	6,675.17	6,735.46	6,590.96	6,716.44	35,319,797,642	122,834,375,216	6,469.80
32	6,738.72	6,892.51	6,536.93	6,681.06	44,590,107,888	122,174,009,143	6,716.44
33	6,436.64	6,789.02	6,411.07	6,734.80	48,221,910,672	123,148,917,787	6,681.06
.							
.							
.							
.							
2306	773.44	820.31	767.21	802.39	38,489,500	9,781,073,921	818.72
2307	754.97	775.35	754.97	771.40	22,489,400	9,403,308,145	802.39
2308	760.32	760.58	738.17	754.01	20,897,300	9,191,325,349	771.40
2309	741.35	766.60	740.24	756.13	20,707,700	9,217,167,990	754.01
2310	728.05	748.61	714.44	745.05	19,011,300	9,082,103,621	756.13
2311	737.98	747.06	705.35	727.83	32,505,800	8,869,918,644	745.05
2312	763.28	777.51	713.60	735.07	46,862,700	8,955,394,564	727.83

EK-2 (Devam)

ETH Veri Seti

1				ETHEREUM			
2	GAF	GEYF	GEDF	GKF	GH	GTPD	TFB
3	182.62	189.09	180.34	185.03	21,275,740,032	20,472,236,139	189.24
4	172.67	183.75	171.83	182.60	17,994,666,395	20,200,984,352	185.03
5	172.02	175.18	170.55	172.74	16,458,767,984	19,107,591,182	182.60
6	181.48	186.08	170.32	172.30	21,266,681,335	19,056,520,718	172.74
7	186.86	188.10	180.12	181.61	19,146,038,381	20,084,614,446	172.30
8	171.62	187.94	171.62	186.91	20,160,323,443	20,668,075,857	181.61
9	172.28	174.28	169.75	171.64	16,714,684,266	18,976,660,194	186.91
10	153.20	173.16	150.36	172.16	22,910,469,236	19,031,693,317	171.64
11	157.57	160.71	152.80	153.29	14,171,753,737	16,943,501,886	172.16
12	156.36	160.94	155.87	157.60	14,723,156,630	17,417,697,024	153.29
13	160.72	160.75	151.61	156.28	16,465,282,133	17,270,031,992	157.60
14	158.23	164.52	156.32	161.14	15,123,721,386	17,805,206,568	156.28
15	158.54	161.17	155.30	158.22	13,555,089,447	17,479,687,023	161.14
16	170.83	170.95	154.91	158.41	17,980,944,616	17,499,250,378	158.22
17	172.76	172.90	167.99	170.81	14,901,696,210	18,866,106,818	158.41
18	165.24	173.21	164.49	172.64	17,063,110,836	19,066,388,908	170.81
19	169.15	175.20	164.20	165.10	21,659,346,091	18,231,455,422	172.64
20	143.61	169.27	143.54	169.14	21,636,382,525	18,674,571,701	165.10
21	145.26	146.13	142.10	143.55	11,280,993,591	15,847,283,514	169.14
22	142.22	145.38	140.12	145.22	11,946,658,257	16,029,973,618	143.55
23	142.11	146.90	139.62	142.09	13,184,603,402	15,682,757,538	145.22
24	135.73	146.79	135.73	142.03	15,322,861,686	15,673,960,169	142.09
25	133.61	135.63	129.73	135.63	12,456,564,435	14,966,424,710	142.03
26	132.82	134.27	131.65	133.59	11,065,246,317	14,739,376,438	135.63
27	125.58	133.91	125.29	132.90	11,841,123,306	14,661,550,831	133.59
28	131.02	131.17	125.45	125.58	9,938,133,669	13,852,241,607	132.90
29	134.03	134.03	126.96	130.99	12,162,403,960	14,446,379,743	125.58
30	138.37	139.90	133.94	133.94	11,396,286,629	14,770,071,750	130.99
31	136.19	138.83	134.89	138.36	11,367,261,176	15,255,985,586	133.94
32	138.91	141.40	134.30	136.20	13,433,092,920	15,015,344,150	138.36
33	135.19	141.95	133.17	138.76	14,609,068,673	15,296,295,171	136.20

.

.

.

1717	1.81	2.26	1.75	1.83	4,637,030	110,672,322	1.69
1718	1.22	1.97	1.17	1.83	4,068,680	110,607,192	1.83
1719	1.06	1.29	0.883608	1.22	2,150,620	73,645,011	1.83
1720	0.708087	1.13	0.663235	1.07	1,463,100	64,569,288	1.22
1721	0.713989	0.729854	0.636546	0.708448	405,283	42,818,364	1.07
1722	0.706136	0.879810	0.629191	0.701897	532,170	42,399,573	0.708448
1723	2.79	2.80	0.714725	0.753325	674,188	45,486,894	0.701897
1724	2.83	3.54	2.52	2.77	164,329	166,610,555	0.753325

EK-2 (Devam)

XRP Veri Seti

1	XRP						
2	GAF	GEYF	GEDF	GAF	GH	GTPD	TFB
3	0.187889	0.194790	0.186766	0.191711	2,271,111,354	8,452,461,961	0.193665
4	0.184175	0.189676	0.183120	0.187970	1,783,803,231	8,287,541,721	0.191711
5	0.183337	0.185644	0.182762	0.184328	1,912,728,846	8,126,957,533	0.187970
6	0.190821	0.194900	0.181831	0.183490	2,364,973,860	8,090,006,361	0.184328
7	0.195004	0.196344	0.190793	0.190859	1,904,866,741	8,414,904,361	0.183490
8	0.189911	0.195356	0.189820	0.194925	1,918,245,143	8,594,155,313	0.190859
9	0.190252	0.192006	0.187614	0.190202	1,887,808,776	8,385,919,739	0.194925
10	0.181254	0.191235	0.177414	0.190104	2,610,251,456	8,381,628,981	0.190202
11	0.185360	0.189266	0.181155	0.181270	1,963,736,221	7,989,127,042	0.190104
12	0.187808	0.188985	0.184262	0.185335	1,889,687,587	8,168,316,007	0.181270
13	0.190533	0.190533	0.181808	0.187732	2,253,497,429	8,273,953,308	0.185335
14	0.188229	0.193848	0.187617	0.190135	2,087,857,417	8,361,929,949	0.187732
15	0.187451	0.192394	0.186113	0.188063	1,847,482,831	8,270,834,076	0.190135
16	0.198156	0.198783	0.183706	0.187464	2,105,282,079	8,244,475,832	0.188063
17	0.200973	0.201336	0.195926	0.198257	1,878,628,668	8,719,129,423	0.187464
18	0.192137	0.201275	0.191408	0.200869	2,091,042,647	8,834,009,937	0.198257
19	0.195686	0.203868	0.191207	0.192093	2,626,151,910	8,448,042,799	0.200869
20	0.179628	0.195865	0.179359	0.195506	2,529,832,372	8,598,152,596	0.192093
21	0.181764	0.181996	0.177277	0.179406	1,660,971,367	7,882,313,124	0.195506
22	0.179612	0.182300	0.178305	0.181732	1,793,826,153	7,984,528,291	0.179406
23	0.178769	0.182728	0.176886	0.179467	2,106,869,711	7,884,998,557	0.181732
24	0.175180	0.183321	0.174310	0.178640	2,352,651,038	7,848,687,168	0.179467
25	0.174315	0.175323	0.169134	0.175323	2,053,108,172	7,702,917,454	0.178640
26	0.172134	0.177883	0.171549	0.174563	2,101,862,050	7,669,550,526	0.175323
27	0.163762	0.174170	0.163176	0.172171	2,131,319,206	7,564,440,755	0.174563
28	0.175680	0.176066	0.163225	0.163829	1,996,124,443	7,193,114,648	0.172171
29	0.175932	0.179475	0.167561	0.175783	2,689,925,753	7,717,955,718	0.163829
30	0.175003	0.186385	0.170564	0.176195	2,849,306,896	7,736,067,263	0.175783
31	0.161680	0.175916	0.160422	0.175019	2,236,102,811	7,684,422,876	0.176195
32	0.162290	0.163891	0.158634	0.161633	2,096,308,696	7,096,688,846	0.175019
33	0.157574	0.163493	0.156824	0.162255	2,155,981,658	7,123,997,865	0.161633

2300	0.022825	0.024723	0.022257	0.023227	82,940.90	181,589,253	0.022368
2301	0.025799	0.026409	0.022986	0.022986	73,760.70	179,703,578	0.023227
2302	0.025298	0.027385	0.024276	0.025700	184,478	200,922,113	0.022986
2303	0.028098	0.029668	0.024890	0.025230	200,438	197,241,451	0.025700
2304	0.027781	0.028371	0.026794	0.028129	60,973.40	219,911,767	0.025230
2305	0.028788	0.029018	0.027320	0.027627	89,779.80	215,987,968	0.028129
2306	0.027014	0.029430	0.026439	0.028726	128,914	224,573,575	0.027627
2307	0.027365	0.027810	0.026814	0.026944	58,533.40	210,649,132	0.028726

ÖZGEÇMİŞ

Adı Soyadı : Furgan BAYRAKTAR
Yabancı Dil : İngilizce
Doğum Yeri ve Yılı : Bursa/1992
E-Posta : furganbayraktar@gmail.com

Eğitim Geçmişi:

- 2017, Karadeniz Teknik Üniversitesi, Mühendislik Fakültesi, Endüstri Mühendisliği, Endüstri Mühendisliği Anabilim Dalı.