

**T.C.  
İSTANBUL AYDIN ÜNİVERSİTESİ  
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ**



**MAKİNE ÖĞRENMESİ ALGORİTMALARI İLE HİSSE SENEDİ KAPANIŞ  
TAHMİNİ**

**YÜKSEK LİSANS TEZİ**

**Salim DEMİRAY**

**Bilgisayar Mühendisliği Ana Bilim Dalı**

**Bilgisayar Mühendisliği Programı**

**Mart, 2021**

T.C.  
İSTANBUL AYDIN ÜNİVERSİTESİ  
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ



MAKİNE ÖĞRENMESİ ALGORİTMALARI İLE HİSSE SENEDİ KAPANIŞ  
TAHMİNİ

YÜKSEK LİSANS TEZİ

Salim DEMİRAY  
(Y1713.010046)

Bilgisayar Mühendisliği Ana Bilim Dalı

Bilgisayar Mühendisliği Programı

Tez Danışmanı: Dr. Öğr. Üyesi Ahmet GÜRHANLI

Mart, 2021

## **YEMİN METNİ**

Yüksek lisans tezi olarak sunduğum “MAKİNE ÖĞRENMESİ ALGORİTMALARI İLE HİSSE SENEDİ KAPANIŞ TAHMİNİ” adlı çalışmanın, tezin proje safhasından sonuçlanmasına kadarki bütün süreçlerde bilimsel ahlak ve geleneklere aykırı düşünecek bir yardıma başvurulmaksızın yazıldığını ve yararlandığım eserlerin Bibliyografya’da gösterilenlerden oluştuğunu, bunlara atıf yapılarak yararlanılmış olduğunu belirtir ve onurumla beyan ederim. (18/03/2021)

**Salim DEMİRAY**

## **ÖNSÖZ**

Önemli yatırım araçlarından olan borsa sistemlerinden hisse senetler alış, satış, miktar ve firma bilgileri hakkında yüksek miktarda veri oluşmaktadır. Bu veriler üzerinden temel ve teknik analizler ile getiri sağlanmaya çalışılmaktadır. Bu araştırmada da konu olan çalışma Microsoft hisse senedi kapanış fiyatları üzerinden zaman analizi ve makine öğrenme yöntemleri ile gelecekteki değerlerin tahminde edilmeye çalışılmaktadır.

Bu çalışmalar sırasında ve hayatımın her aşamasında bana yardımcı ve destek olan çok kıymetli eşim Nurten DEMİRAY 'a, hayatıma güzellik ve anlam katan Oğlum Ege ve kızım İrem'e bana yaşattıkları mutluluk ve desteklerinden dolayı teşekkür eder şükranlarımı sunarım.

Yüksek lisansa başlamamda ve ders ve tez süresince destek ve fikirleri ile her zaman yanımda olan iş ve yüksek lisans sınıf arkadaşlarım İlker POYRAZ ile Ahmet YILDIRIM 'a teşekkür ve şükranlarımı sunmayı borç bilirim.

Tez çalışmamın planlanmasında, araştırılmasına, yürütülmesine ve oluşumunda ilgi, bilgi ve desteğini esirgemeyen, üstün bilgi ve tecrübelerinden yararlandığım, yönlendirme ve bilgilendirmeleriyle çalışmamı bilimsel temeller ışığında şekillendiren yine çalışmamda her zaman konu, kaynak ve yöntem açısından yol gösteren ve yönlendiren bunun yanında bana sürekli yardımda bulunarak gelecekteki hayatında çok daha başarılı olup başarılarla imza atacağına inandığım kıymetli Tez Danışman hocam Assist. Prof. Dr. AHMET GÜRHANLI sonsuz teşekkürlerimi ve saygılarımı sunarım.

**Mart, 2021**

**Salim DEMİRAY**

## İÇİNDEKİLER

### Sayfa

ÖNSÖZ.....	iv
İÇİNDEKİLER .....	v
KISALTMALAR .....	vi
ÇİZELGE LİSTESİ.....	vii
ŞEKİL LİSTESİ.....	viii
ÖZET.....	ix
ABSTRACT .....	x
<b>1. GİRİŞ .....</b>	<b>1</b>
1.1 Hisse Senedi Bilgileri.....	2
1.2 Zaman Serileri Analiz Yöntemleri .....	3
1.3 Veri Madenciliği Ve Makine Öğrenmesi Farkları .....	5
1.4 Veri Madenciliği .....	6
1.5 Makine Öğrenmesi Yöntemleri.....	8
1.6 Uygulama Alanları .....	12
<b>2. YÖNTEM.....</b>	<b>14</b>
2.1 Polinomsal Regresyon.....	14
2.2 LSTM .....	15
2.3 Arima.....	16
2.4 Facebook Prophet .....	20
2.5 XGBoosting Yöntemi.....	20
2.6 Hata Ölçme Yöntemi.....	21
2.6.1 Ortalama Mutlak Hata.....	21
2.6.2 Ortalama Kare Hata .....	21
2.6.3 Kök Ortalama Kare Hata.....	22
<b>3. LİTERATÜR ÇALIŞMASI .....</b>	<b>23</b>
<b>4. ARAŞTIRMA DEĞERLENDİRİLMESİ.....</b>	<b>25</b>
4.1 Araştırma Metodu .....	25
4.2 Ölçümlerin Değerlendirilmesi.....	26
<b>5. SONUÇ.....</b>	<b>34</b>
<b>KAYNAKLAR .....</b>	<b>36</b>
<b>ÖZGEÇMİŞ.....</b>	<b>40</b>

## **KISALTMALAR**

<b>AI</b>	:Artificial Intelligence
<b>ARIMA</b>	:Autoregressive Integrated Moving Average
<b>ARMA</b>	:Autoregressive Moving Average
<b>CPU</b>	:Central Process Unit
<b>CRM</b>	:Customer Relationship Management
<b>EEG</b>	:Elektroensefalografi
<b>GPU</b>	: Graphic Process Unit
<b>IMKB</b>	:İstanbul Menkul Kıymetler Borsası
<b>KNN</b>	:K-Nearest Neighbors Algorithm
<b>LSTM</b>	:Long Short-term Memory
<b>MAE</b>	:Mean Absolute Error
<b>MSE</b>	:Mean Squared Error
<b>RAM</b>	:Random Access Memory
<b>RMSE</b>	:Root Mean Square Error
<b>RNN</b>	:Recurrent Neural Network
<b>SARMA</b>	:Seasonal Autoregressive Moving Average

## ÇİZELGE LİSTESİ

	<b><u>Sayfa</u></b>
<b>Çizelge 4.1:</b> Hata Sonuç Tablosu.....	27
<b>Çizelge 4.2:</b> 1 Yıllık Hata Tablosu .....	31

## ŞEKİL LİSTESİ

### Sayfa

Şekil 1.1: Microsoft Hissesi Kapanış Değerleri .....	4
Şekil 1.2: Zaman Serisi Bölümleri .....	5
Şekil 1.3: Veri Madenciliği İşleme Adımları .....	7
Şekil 1.4: Makine Öğrenme Sınıflandırması .....	8
Şekil 1.5: Sınıflandırma .....	9
Şekil 1.6: Kümeleme .....	10
Şekil 2.1: Polinomsal Durumu .....	14
Şekil 2.2: RNN Çalışma Adımları .....	15
Şekil 2.3: LSTM Çalışma Adımları .....	16
Şekil 2.4: Microsoft Hisse Senedi Otokorelasyon Grafiği .....	18
Şekil 2.5: Microsoft Hisse Senedi Kısmı Otokorelasyon Grafiği .....	18
Şekil 2.6: Arima Çalışma Şekli .....	19
Şekil 4.1: Yıllık Bazda Kapanış Değişimi .....	26
Şekil 4.2: Son 5 Gün Tahmin Grafiği .....	28
Şekil 4.3: Son 10 Gün Tahmin Grafiği .....	29
Şekil 4.4: Son 20 Gün Tahmin Grafiği .....	30
Şekil 4.5: 1 Yıllık Detaylı Tahmin Grafiği .....	31
Şekil 4.6: 1 Yıllık Tahminde Eğitim -Gerçek-Tahmin Grafiği .....	32
Şekil 4.7: 5 ,10,20 Günlük Hata Sonuç Grafiği .....	33
Şekil 4.8: 1Yıllık Hata Sonuç Grafiği .....	33



# MAKİNE ÖĞRENMESİ ALGORİTMALARI İLE HİSSE SENEDİ KAPANİŞ TAHMİNİ

## ÖZET

Günümüzde Finans ve Bankacılık Sektöründe teknoloji ve iletişim sistemleri ile birlikte gelişmesinden dolayı yüksek miktarda veri iletilmektedir. Özellikle borsa, döviz, kıymetli emtia gibi yatırım yapılabilen her alanda gelecekte olabileceği değerleri tahmin edilmeye çalışılmaktadır.

Zaten istatistik bilimi ile iç içe olan finans gelişen teknoloji ile analizlerde veri madenciliği yöntemlerini çok etkin bir şekilde kullanmaktadır. Yatırımcı araçları Banka gibi finansal kurumlardan başka bireysel bazda da takip edilip analize edilmektedir. Yatırımcılar, en az risk ile olabilecek en yüksek getiri kazanmayı hedeflemektedir. Bu hedefin sağlıklı bir şekilde gerçekleştirilmesi için değerlerin doğru tahmin çok önem kazanmıştır.

Bu çalışmada 2014-2019 yılları arasındaki Microsoft Firmasının hisse senetleri kapanış fiyatları Makine Öğrenmesi Yöntemi ile tahmin edilmeye çalışılmıştır. Kullanılan yöntemler Polinomial Regresyon, Arima, XGBooster, LSTM (Uzun Kısa Süreli Bellek) ve Facebook Prophet algoritmalıdır. Çalışmada hissenin alacağı değerler; kısa vadede son 5,10 ve 20 günlük değerler tahmin edilmeye iken 1 yıllık uzun vadeli değerler de tahmin edilmeye çalışılmıştır. Bu sayede algoritmaların kısa ve uzun vadeli davranışları hakkında bilgi edinme imkânı ortaya çıkmıştır.

**Anahtar Kelimeler:** *Makine Öğrenmesi, Tahmin, Zaman Serileri, Hisse Senedi Kapanış Değerleri.*

# STOCK CLOSING PREDICTION WITH MACHINE LEARNING ALGORITHMS

## ABSTRACT

Today, it transfers large amounts of data due to the development of technology and communication systems in the Finance and Banking Sector. It tries to predict future values in every field that can be invested, especially in the stock market, foreign exchange and precious commodities.

Finance, which is already intertwined with the science of statistics, uses data mining methods very effectively in analysis with developing technology. Investor instruments are followed and analyzed on an individual basis and also on the basis of financial institutions such as banks. Investor instruments are also followed and analyzed on an individual basis other than financial institutions such as the Bank. Investors aim to earn the highest possible return with the least risk. It has become very important to correctly estimate the values in order to achieve this goal in a healthy way.

In this study, the closing prices of Microsoft Company stocks between 2014 and 2019 are tried to be estimated by using the Machine Learning Method. The methods used are Polynomial Regression, Arima, XGBooster, LSTM (Long Short Term Memory) and Facebook Prophet algorithm. The values that the share will get in the study; While estimating the last 5, 10 and 20-day values in the short term, 1-year long-term values are also tried to be estimated. In this way, it has become possible to obtain information about the short and long term behavior of the algorithms..

**Keywords:** *Machine Learning, Forecasting, Time Series, Stock Closing Forecasting.*

## 1. GİRİŞ

Borsa hisse senetlerin el deęiřtirmesi ile yatırım yapıp kar saęlanabilen riskli ve önemli yatırım araçlarından biridir. Hisse senedinin ileri tarihte olabileceęi piyasa deęerlerinin doęru tahmin edilmesi yüksek kar kazanmak için önemli bir adımdır. Hisse senetlerinin fiyat tahmini ülke ve dünyadaki politik duruma, finansal beklentiler, yatırımcı davranıřları, ekonomik beklenti ve gelişim duruma, sektörel beklentiler vs. gibi birçok farklı faktöre baęlı olduęu için çok zordur. Yatırımcılar hisse senetlerine yatırım yapmak için řirketin ekonomik durumu, politik beklentiler gibi temel analizi kullandıkları gibi; sadece hisse senetlerinin deęerleri ve hareket hacimleri üzerinden gelecekteki deęer eğilimlerinin tahmin edilmesi için teknik analiz adı verilen bir çalıřma da yapmaktadır. Son yıllarda gelişen donanım ve yazılım teknolojisi ile yapay zekâ teknolojisi hisse senedi fiyatları tahmininde de kullanılması yaygınlařmaya başlamıřtır.

Makine öğrenmesi ile elimizdeki verilerle, matematiksel ve istatistiksel operasyonlar ile geçmiş verilerden yararlanarak sınıflandırma, kümeleme veya Eğitim belirleme (Regresyon) gibi tahminde bulunulan yazılım modellemesidir. Makine öğrenmesine Sınıflandırma verinin hangi sınıfa ait bulunması, kümeleme verinin özellikleri bakımından en yakın kümesinin bulunması, Eğitim belirleme (Regresyon) ise verilerin hareket yönünü belirlemektir. Deęiřkenlerin zaman akıřına göre deęiřimlerini izleyen verilere zaman serisi adı verilmektedir. Zaman serileri yıl, ay, gün gibi farklı frekans deęerlerine sahip olabilmektedir. Zaman serileri analizinde önceki deęerler üzerinden gelecek periyotlara ait deęerlerin tahmin edilmesidir. Hisse senetlerinin kapanıř deęer tahmini, zaman serisi analizi olup bir regresyon analizidir.

Yapılan çalıřmada Microsoft hisse senedinin kapanıř fiyatları 01.01.2014 ile 30.05.2020 tarihleri arasında zaman serisi makine öğrenme algoritmaları ile incelenmiřtir. Kullanılan Polinomial Regresyon, Arima, XG Booster ve Prophet algoritmalarda 5 ,10 ve 20 günlük ile 1 yıllık tahminler yapılmıř olup

algoritmalarının veri miktarına göre başarısını ölçmek için veriler hem kısa vadeli hem de uzun vadeli olmak üzere 4 farklı durum için çalıştırılmıştır.

### **1.1 Hisse Senedi Bilgileri**

Hisse Senedi borsada işlem yapan bir firmaya belirli oranda ortak olduğunuzu gösteren kıymetli evraktır [1]. Hisse senedi ile şirkete ortak olarak kar ve zarara da ortak olursunuz. Şirketin değeri arttığında hisse senedi değeri artar iken eğer iflas eğer ise sizde hisse senetleriniz değersiz olacaktır. Şirketler belirli periyotlarda kar -temettü- dağıttıklarında ortak olarak size de üzerinize düşen karı alırsınız. Hisse senedi değer artışında satarak veya temettü dağıtıldığında bir kazanç elde edebilirsiniz.

Yatırım sırasında hangi hisse senedi ile yatırım yapılacağı elde edilecek kar için çok önemlidir. Hisse senedi değerlerini incelemek için temel ve teknik olarak 2 analiz çeşidi bulunmaktadır. Temel analiz firmanın finansal durumu, mal varlıklarını, borç ve alacaklarını, firmanın mali değerleri, hisse senedi değerinin firma değerine göre oranları gibi birçok değeri inceleyerek hisse senedinin alınıp, alınmamasına karar vermekte yardımcı olan analizdir [2]. Teknik analiz hisse senedinin geçmiş zamandaki değerlerine bakarak, çeşitli istatistiksel ve matematiksel formüller ile gelecekteki değerlerini tahmin etme analizidir. Teknik analizde kullanılan incelenen bilgiler Trendler ve Konsolidasyon, Grafik, Destek direnç, işlem hacmi gibi bilgilerdir [3]. Trend hisse senedi fiyatının belli bir oranda yükselme veya azalma eğilimi iken konsolide ise belirli aralıklarda sıkışmasıdır. Hisse senetleri değerleri üzerinden çizilen grafikler üzerinden çeşitli yöntemler ile gelecekteki kritik dip veya tepe noktalarını bulmaya çalışarak hisse senedi fiyatı hakkında tahminde bulunmaktadır. Direnç ve destek hisse senetlerinin fiyat değişimleri için aşması veya düşmesi beklenen kritik noktalar. Belli bir değeri aştığında fiyatların yükselmesini beklenen noktaya destek, düşme eğilimi olduğunda o değer altına indiğinde düşmeye devam edileceğine inanılan noktaya ise direnç adı verilmektedir. İşlem hacmi ise yapılan alış-satış işlem miktarları üzerinden değişim -histogram- grafiği üzerinden piyasanın alış veya satış yönünde istekli olduğuna karar vermek için kullandığımız bir grafik. Teknik analizde grafik okuma yanında, göstergeler, hareketli ortalama, Bollinger bandı gibi çeşitli

istatistiksel yöntemler de kullanılmaktadır. Son zamanlarda gelişen yazılım, yapay zekâ ve donanımlar ile hisse senedi teknik analizlerinde yapay zekâ da kullanılmaktadır. Bu çalışmanın konusu ise kısa vadeli 5, 10 ve 20 günlük ve uzun vadeli 1 yıllık hisse senedi kapanış değerlerinin makine öğrenmesi ile tahmin edilmesidir.

Hisse senedi yatırımı da zamana göre kısa, orta ve uzun vadeli olmak üzere 3 çeşit yatırımcılık bulunmaktadır. Kısa vadeli yatırımlarda hızlıca al, sat yaparak yüksek riskler ile kazanç yapılmaya çalışılmaktadır. Kısa vadeli yatırımlar saatlik, günlük haftalık gibi periyotlarda olabilirler [4]. Orta vadeli yatırımcılar ise genellikle 1 ile 3 yıl arasında vadeler ile hisse senedi alıp bekledikleri kar oranını yaptıkları veya piyasa değerleri düşüş eğilimi gösterdiği sırada satıp, uygun fiyatlarda tekrar alarak kar elde edilmeye çalışılan yatırım yöntemidir. Uzun vadeli yatırımda ise belirli periyotlarda düzenli hisse senedi alarak biriktirerek gelecek için birikim sağlamayı amaçlayan bir yöntemdir [5]. Uzun vadeli yatırımda güvenli ve güçlü bir firma seçilerek farklı fiyatlardan sürekli alış yapıldığı için belli zaman sonunda yatırımcı elinde ciddi miktarda hisse senedi olduğu için yüksek miktarda birikim yapmış olmaktadır. Önemli yatırım araçlarından biridir. Hisse senedinin ileri tarihte olabileceği piyasa değerlerinin doğru tahmin edilmesi yüksek kar kazanmak için önemli bir adımdır.

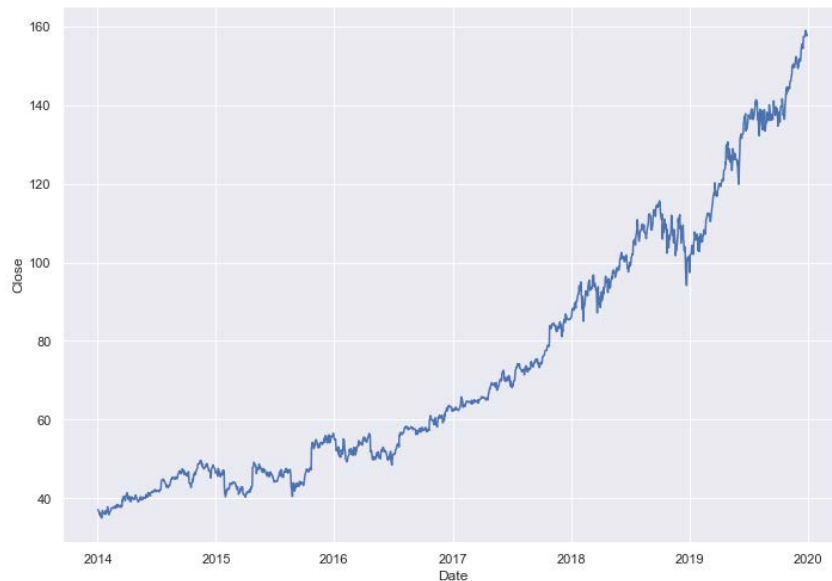
## **1.2 Zaman Serileri Analiz Yöntemleri**

Zamana bağlı olarak değişen bilgi ve verilerin, Grafik 1-1 de görüldüğü belirli periyotlara göre sıralı olarak düzenlenip incelenmesidir [6]. Zaman serileri analizinde, verilerin değişimi, sebepleri ve gelecekteki alabilecekleri değerler hakkında bilgi toplanması amaçlanmaktadır. Zaman Serilerinin önemli bileşenleri Eğilim, mevsimsellik, konjonktürel ve düzensiz hareket bileşenlerden oluşmakta olup Grafik 1-2 de grafiğinde görünmektedir [7]. Eğilim (Trend) verinin genel süreç içerisinde sürekli artış veya azalış eğiliminde olmasını göstermektedir. Veri kısa zaman aralığında küçük farklı hareketlerde bulunursa da genel görünüm de azalma veya artma eğiliminde ise zaman serisinde trend bulunmaktadır. Zaman serilerinde mevsimsellik (Seasonal) belirli dönemlerde değerlerin benzer yapıda değerler almasıdır. Mevsimsellik etkisini turizm verilerinin yaz aylarında artar iken kış aylarında azalma

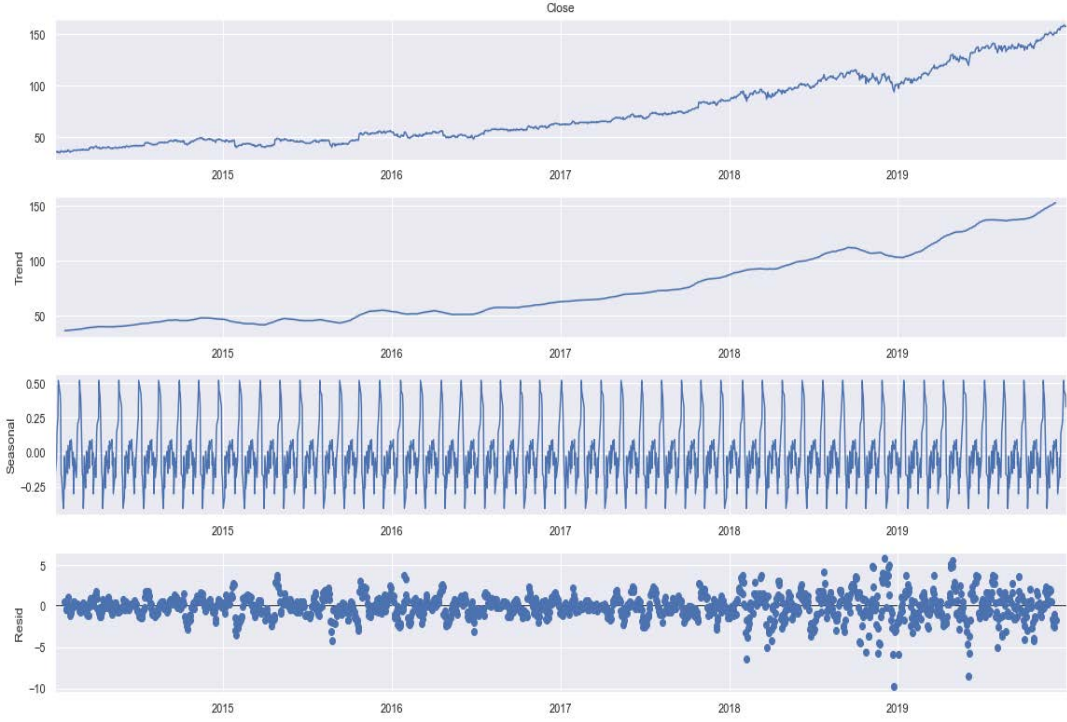
eğiliminde bulunması gibi özetlenir [8]. Zaman serilerinde mevsimsellik, trend gibi beklenen durumlar dışında belirli zamanlarda beklenen hareket yönünün dışsal etkiler ile değerlerin değişim olması konjonktürel (Cyclical) bileşenidir. Konjonktürel değişimler belli dönemler için değerleri etkileyen durumlardır. Konjonktürel için ekonomik iyileşme ve beklentilerin artması veya bir ekonomik sıkıntı gibi durumlardan dolayı olabilecek değişimlerdir. Düzensiz hareket (Irregular) öngörülemeyen şekilde oluşarak, değerleri etkileyen doğal afet veya faiz indirimleri gibi olaylardır. Düzensiz hareketler değerlerin artma veya azalma yönünde etkili olabilir.

Zaman serileri istatistik yönden analizinden serinin durağan olması istenmektedir. Zaman serilerinde durağanlık verinin sürekli artma veya azalma eğiliminde olmaması -trende sahip olmaması-, mevsimsel etkilerden temizlenmiş olmasıdır. Trend ve mevsimsellik özelliklere sahip serilerde değerler arası farklar düzensiz artıp azalacağı için analizde yanlış sonuçlara sebep olmaktadır.

Durağan serinin ortalaması sabit bir değer civarındadır ve varyansı çok değişken değildir. Yani herhangi andaki değerlerin ortalamaları birbirlerine hemen hemen eşittir. Durağan olmayan seriyi durağan hale getirmek için genellikle serideki değerlerin geçmişteki değerleriyle farkı alınır. Ayrıştırma (Decomposition): Seriyi trend ve mevsimsellik etkilerine ayırma işlemidir.



**Şekil 1.1:** Microsoft Hissesi Kapanış Değerleri



**Şekil 1.2:** Zaman Serisi Bölümleri

### 1.3 Veri Madenciliği Ve Makine Öğrenmesi Farkları

Makine öğrenmesi ile veri madenciliği birbirlerine çok benzemekle beraber işleyiş ve çalışma prensipleri bakımından farklılıkları vardır. Veri madenciliği veriler içerisinde belirli kural ve örüntü çıkarılarak tahmin ve analiz işlemleri yapmaktadır [9]. Makine öğrenmesi veri madenciliği kuralları ile öğrendiği bilgileri de kullanarak yeni yöntem ve anlamlar ile davranışlarını geliştirmektedir. Ayrıca makine öğrenmesi çeşitli kalıp ve algoritmalar da bulmak için bir yöntemdir. Makine öğrenmesi istatistiksel, matematiksel yöntemlerini çeşitli iyileştirme yöntemleri ile birlikte kullanılarak daha ileri seviyeler gelmiştir [10]. Başlıca Resim ve karakter tanımlama olmakla birlikte spam filtreleme, ses tanıma analizler, sağlık sektöründe hastalık tanımlama gibi çok alanda uygulamaları vardır. Makine öğrenmesi genel olarak tek katmanlı bir yapıda iken gelişen CPU ve GPU teknolojisi ile çok katmanlı yapıya sahip olan derin öğrenme ile son yıllarda çok popüler olmaktadır. Derin öğrenimde bulunan çok katmanlı yapı verilerin ve analizlerin doğru yapılması için gereken birçok parametreyi bile kendisi ayarlamaktadır. Makine öğrenmelerinde eğitimlerde girdi ve çıktı verilerek bir amaçta eğitilmesi ile başlayan süreçte, derin öğrenmenin GPU ve bulut teknolojisi sağladığı yüksek işlem gücü ile

sadece verilen bilgiler üzerinden girdileri sınıflandırma gibi uygulamalar yapılmaktadır [11]. Hatta Google'ın Alpha Go uygulaması ile sadece go oyununun kuralları öğretilen bir bilgisayar programı, iki rakip gibi birbirleri ile oynayarak go oyununda uzman olmuşlardır. Alpha Go o kadar başarılı olmuştur ki Go oyununda bilgisayarların insanları yenemeyeceğine dair olan inançları yıkıp geçmiştir [12]. Go oyunu sıra bazlı olmasına rağmen satranç gibi derinlik hesabı ile kolaylıkla hesaplanabilen bir oyun olmayıp, oynayan oyuncunun da karakterinden etkilenen hislerinde önemli olduğu bir oyundur. AlphaGo bu sansasyonel başarıdan sonra Go oyuna olan sevgi ve saygı yüzünden proje çekilmiştir. AlphaGo 'nun başarısından sonra aynı yöntem ile satranç için AlphaZero için çalışma yapılmıştır. Eğitim sonrasında bilinen in iyi satranç motoru olan ve bilinen tüm satranç oyunları ile beslenen Stockfish ile 1000 maçlık bir seri maç yapmış olup toplamda oyun sayısında galip gelip hiç maç kaybetmemiştir [13].

#### **1.4 Veri Madenciliği**

Veri madenciliği veriler içerisinde belirli kural ve örüntü çıkarılarak tahmin ve analiz işlemleri yapmaktadır. Şekil 1.1 de veri madenciliği işleme adımları gösterilmektedir. Veri madenciliğinde adımlar aşağıda yazılmaktadır.

**Veri Seçimi:** Çeşitli farklı kaynaklarından gelen verilerin hangilerinin kullanılacağına karar verme aşamasıdır. Bu bilgiler veri tabanları, siteler, mobil uygulamalar, IOT cihazlarından sağlanabilmektedir. Veri tabanından birden fazla tablonun birleşmesi ile oluşan verileri normalize ederek- diğer tablolardan gelen bilgileri toplama- bir tabloya yazarak ön işleme adımına hazırlanmaktadır.

**Veri Ön İşleme:** Toplanan verilerin, veri madenciliği uygulanmasına geçmeden önce kullanılmak için yapılan hazırlıklardır. Bu hazırlıklar kullanılacak alanları belirleme, boş ve hatalı bilgileri doldurma, gereksiz hatalı verileri temizleme, tekrar edilen veya ağırlığı fazla olan kayıtları düzeltme, verileri normalleştirme, boyut indirgeme gibi veri madenciliği öncesi yapılan işlerdir [14]. Bu adım veri anlamlandırma ve hazırlama konusunda veri madenciliği öncesinde ışık tutmaktadır. Veri madenciliği sürecinin en önemli adımı olup en uzun süren aşamasıdır.

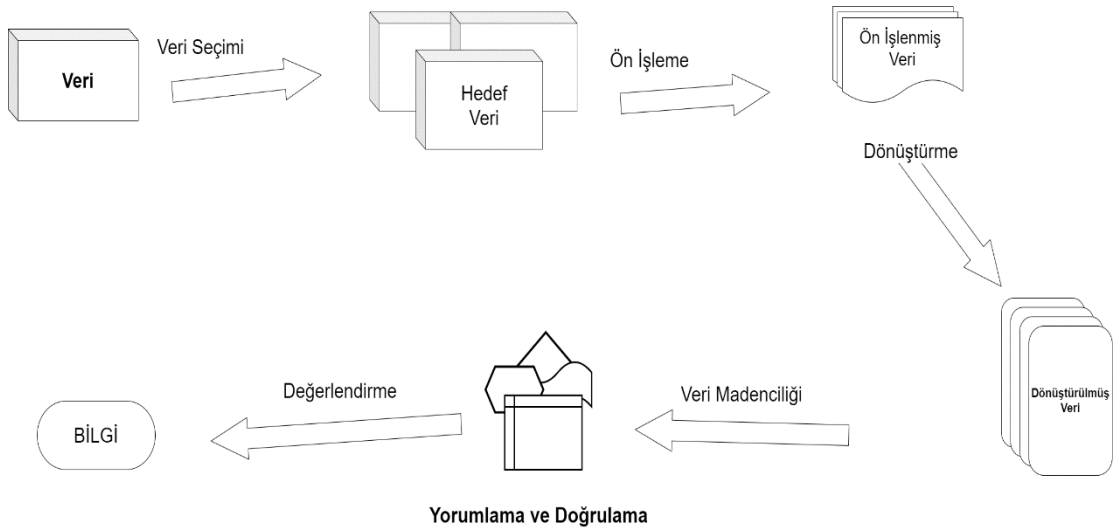


Toplanan verilerdeki eksik bilgi, gürültü (aykırı) bilgiler tespit edilip silinerek veya varsayılan değerler atanarak hatalar düzeltilmektedir. Toplanan veride çok sayıda özellik bulunabilir. Madencilik için önemli özelliklerin tespit edilmesi gerekmektedir. Eğitim sırasında bir grup verinin baskın olarak sonucu aşırı etki etmemesi için kullanılacak veriler seçilir. Bazı durumlarda özellikle karar ağacı ile yapılacak çalışmalar da yaş, ücret gibi sayısal bilgilerin belirli aralıklar ile bir grup bazında temsil edilmesi gerekebilmektedir [15].

**Veri Dönüştürme:** Veri ön işleme sırasında seçilen özelliklerin veri madenciliğinde kullanılmak üzere uygun şekli döndürülüp kaydedilme adımıdır.

**Veri Madenciliği:** Çalışmalar sonucunda elde edilen dönüştürülmüş veriyi yapısı, elde edilmek istenen bilgiye uygun olarak çeşitli algoritmalar ile üzerinde çalışma sürecidir [15]. Eğitim sırasında algoritma çalışması için girdi ve karşılığında çıktı verilmesi durumunda güdümsüz eğitim adı verilmektedir. Eğer sadece girdiler ile çalışma yapılarak bir sonuç elde ediliyor ve eğitim sırasında olması gereken çıktı değerleri verilmiyorsa güdümsüz eğitim adı verilmektedir.

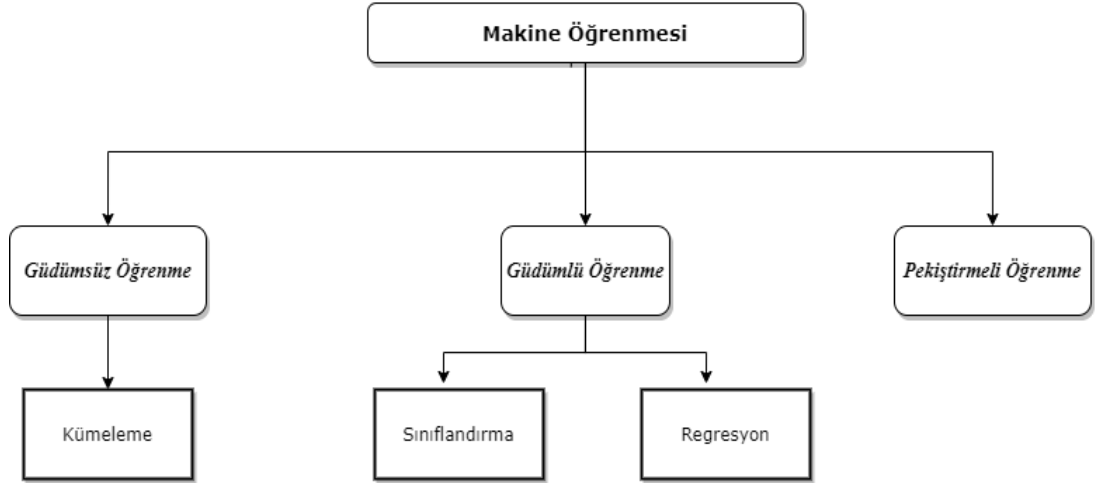
**Veriyi Değerlendirme:** Yapılan çalışmalar sonucunda elde edilen bilgi üzerinden bir tahmin veya örüntüler -bir kural dayanarak devam eden bilgi dizisi- elde edilmeye çalışılmaktadır. Elde edilen kurallara göre test için ayrılan veri ile çalışma ve tahminler yapılır. Bu çalışmaların kullanıcı için anlamlı ve anlaşılabilir olması için çeşitli grafikler ile gösterilmesi de gerekmektedir [16].



**Şekil 1.3:** Veri Madenciliği İşleme Adımları

## 1.5 Makine Öğrenmesi Yöntemleri

Makine öğrenmesi ile birlikte endüstri 4.0 yapısında otonom sistemler, görüntü işleme ile hata ve kalite kontrolü yapan yapay zekâ sistemleri, işleyiş planlama yapan sistemler üretim sistemlerinde de aktif kullanılmaktadır. Gelişen teknoloji ile insansız fabrika oluşumları ile yapay zekanın uygulaması giderek artacaktır [17]. Makine Öğrenmesi veriler üzerinden bilgi çıkarımı yaparak öğrenebilen, model çıkarıp, tahmin gerçekleştirebilen algoritmalar sistemidir [18]. Program talimatlarını takip etmek eğitim verileri üzerinden karar ve tahmin yapan bir model oluşturma sistemidir. Makine öğrenmesindeki öğrenme Şekil 1.2 de gösterildiği gibi Güdümlü Öğrenme, Güdümsüz Öğrenme ve pekiştirmeli Öğrenme yöntemleri uygulanmaktadır.

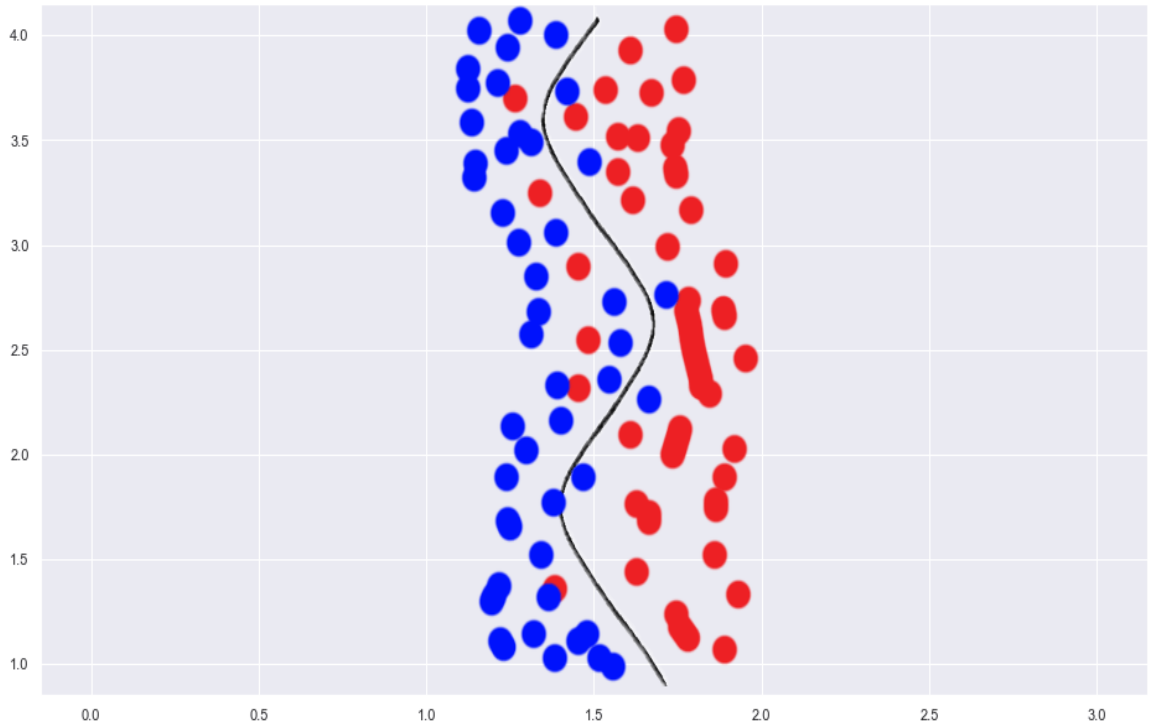


Şekil 1.4: Makine Öğrenme Sınıflandırması

Makine öğrenmesi ile yapılan çalışmalar ise kategori ve sınıf bulma: Sınıflandırma (Classification), verileri inceleyip tahminde bulunma: Regresyon (Regression), verileri aynı özellikli bölgelere ayırma: Kümeleme (Clustering), verilerin özelliklerinin sayısını azaltarak temel duruma getirmek Boyut azalımı (Dimensionality Reduction) işlemleridir. Makine Öğrenmesinde veri ile çalışmak için yapılması gereken adımlar ise ön işleme -veri içerisindeki hataları düzeltme-, özellik çıkarma -veri üzerindeki hangi özellikler kullanılacak-, uygun öğrenme algoritmalarının seçilmesi, model çıkarma -verileri inceleyip eğitim, test verileri oluşturma-, değerlendirme -modelin algoritma başarısını ölçme- adımları vardır [19]. Pekiştirmeli öğrenme, öğrenim öncesi belirlenen bir

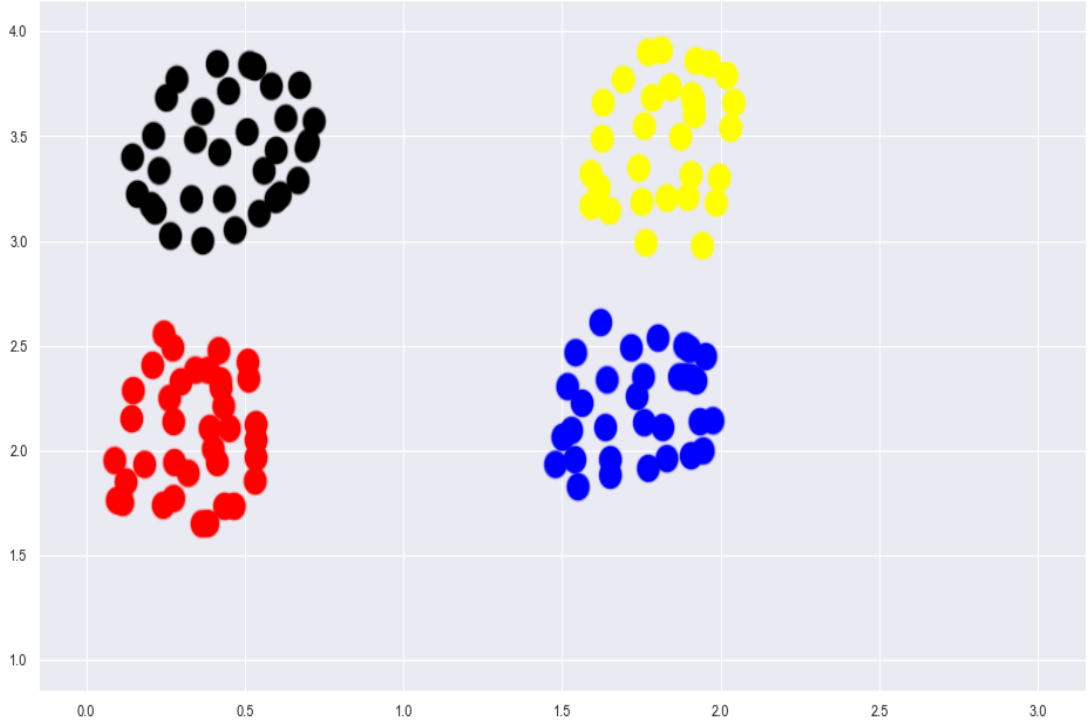
amaca ulaşmak için ne yapması gerektiğini öğrenmeye çalışan ödül ceza temelli bir makine öğrenmesi yöntemlerindedir [20].

**Sınıflandırma:** Toplanan veriler işlenerek çıktılarının birbirinden farklı olarak önceden belirlenmiş kategoriler halinde işlenerek atanmasıdır [21]. Sınıflandırma yöntemine örnek olarak CRM sisteminde müşterinin çalışmayı bırakması veya devam etmesi, kanser tespit çalışmasında ise hücre veya çalışmanın kanserli olup olmadığı örneklerdendir. Şekil 1.3 olduğu gibi veriler arasında bir bölümlenme ile 2 sınıfa ayrılmış olup sınıflar arası bazı geçişlerde bulunmaktadır.



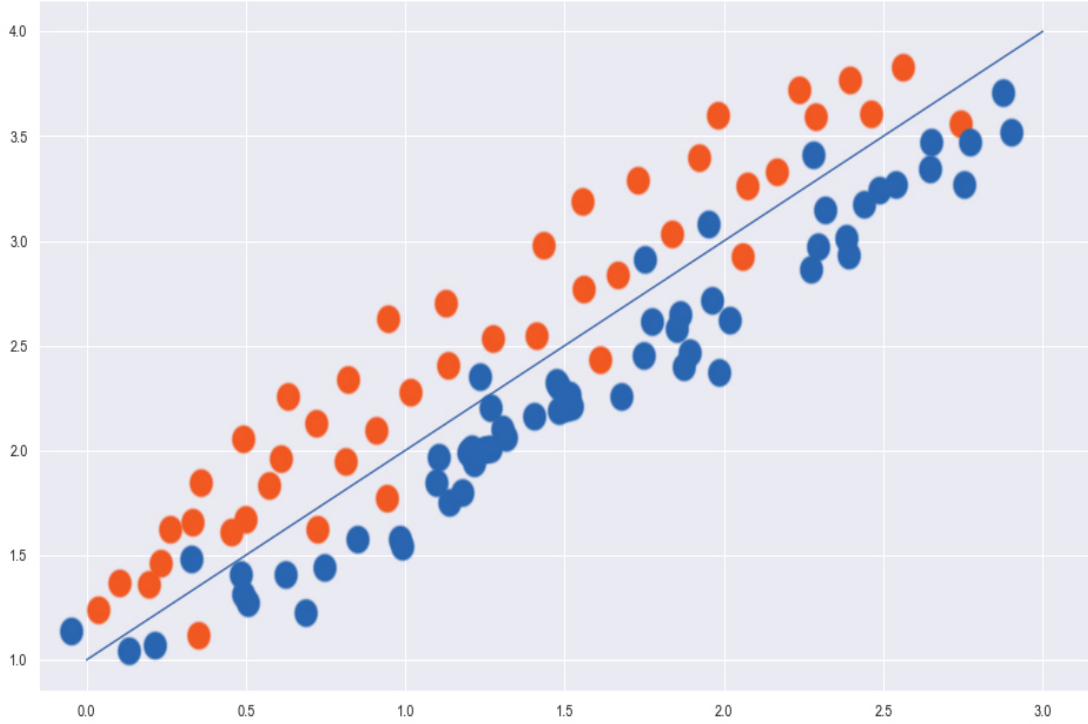
**Şekil 1.5:** Sınıflandırma

**Kümeleme:** Kümeleme analizinde toplanan veriler algoritma tarafından belirlenen özelliklere göre nesnelerin sınıflandırılması için yapılan çalışmadır. Veri sınıflandırması güdümlü bir eğitim yöntemi olup eğitim sırasında önceki durumlar bilgisi verilmektedir [22]. Kümelemede yöntemlerinde güdümsüz eğitim ile yapılan çalışma ile yöntem istenen grup sayısına göre verileri gruplamaktadır. Şekil 1.4 görüldüğü gibi veriler 4 ana kümede gruplanmıştır. Bu gruplamada algoritma verilerin yakınlığına göre hesaplayarak çıkarımda bulunmuştur.



**Şekil 1.6:** Kümeleme

**Regresyon:** Verileri bir veya daha fazla sayıda değişken ile aralarındaki bağlantıyı ve ilişkiyi modellemek için kullanılan bir yöntemdir. Benzerlik doğrusal olarak artan veya azalan bir şekilde modellenmesine Doğrusal Regresyon adı verilmektedir. Benzerlik bir çizgiden daha çok polinom bir şekilde eğriye benzemesi durumunda ise lojistik (Polinomsal) regresyon adı verilmekle birlikte birden fazla değişken ile ifade edilmektedir [23]. Şekil 1.5 de görülen doğrusal regresyon artan bir eğilimde olarak kırmızı ve mavi olarak belirlenen gruptaki değerleri bir çizgi grafiği ile 2 bölüme ayrılmıştır.



**Şekil 1.1:** Regresyon

**Pekiştirmeli Öğrenme:** Pekiştirmeli öğrenme, öğrenim öncesi belirlenen bir amaca ulaşmak için ne yapması gerektiğini öğrenmeye çalışan ödül ceza temelli bir makine öğrenmesi yöntemlerindedir. Bu yöntemde ajan (agent) diye tanımlanan her bir adımda karşılaştığı durumu değerlendirerek bir tepki veren ve sonucunda ödül, ceza sinyalleri ile adımlarını planlayan bir sisteme sahiptir. Deneme yanımlar ile olabilecek durumları rastgele veya ödül değerleri üzerinden olabilecek en yüksek puanı alarak en iyi sonuca varmaya çalışmaktadır [24]. Eğitime başlamadan önce ortam hedef puanları belirlenerek eğitim yapılacak ortamın tanımı yapılmaktadır. İstenen hedefe büyük bir pozitif değer verilir iken kesinlikle ulaşmaması gereken noktalara negatif büyük bir değer verilmektedir. Eğer en kısa yol bulma gibi yapılan her adım bir maliyet olan fakat bu adımdan bazılarının da mutlaka yapılması gereken durumlarda, her bir adım için negatif küçük bir puan verilir. Belirli sayıdaki deneme ve yanılma ile olan eğitimler sırasında her adım için hesaplanan ödül değerleri ile en iyi sonuca erişmeye çalışmaktadır [25].

Pekiştirmeli öğrenmede karar alma aşamasında Markov karar süreçleri adı verilen bir algılama, eylem ve hedef adımları bulunmaktadır. Eğitim öncesinde belirlenmesi gereken hedef ve ödül tablosu yanında sistemin çalışma sırasındaki

keşif ve aldığı sonuçlara veya değerlendirmeye göre hangi adım ve olayın gerçekleşmesini belirlenmesi çok zor ve önemli bir adımdır [26]. Bu sistemde adımları deneyen ajanın topladığı ödül miktarının daha fazla olması için geçmiş deneyimlerinden gelen yüksek değerleri seçerken, önceleri daha az fakat daha ileri adımlarda çok daha yüksek değerler kazanma ihtimali olduğundan ve aynı değerlerle dönerek kısır bir döngüye uğramaması adına rastgele adımlarda denenerek diğer durumlarda karşılanmaya çalışılmaktadır. Pekiştirmeli öğrenmede ile ilgili örnekler satranç gibi oyunlar, robotların çöp toplamada kullanacakları en kısa yol bulma ve çeşitli bilim dallarında karar verme süreçlerinde gibi işler için kullanılmaktadır.

Pekiştirmeli Öğrenme sisteminde ajan ve ortam dışında bulunan unsurlar politika, ödül, Değer/Durum değerler ve çevre modelleridir. Politika ajanın öğrenme sırasında karşılaşacağı duruma göre vereceği tepkiyi temsil edip probleme göre tanımlanması gerekmektedir. Ödül ajan tarafından uzun vadede en yüksek puanı almak için belirlenen politika çerçevesinde yapılan tepkilere karşısında alınan puan değerleridir. Ödül davranışı her adımda farklı değerlere sahip olabilip her adımda farklı tercihlerde bulunabilir. Durum değeri ajanın aktif durumu ve gelecekte olabileceği öngördüğü ödüllerin toplamı olup uzun vadede hangi davranışın doğru olduğuna karar verme durumudur. Ödül anlık davranışlara yönlendiril iken durum değeri toplamda en iyi değeri bulmaya çalışılmaktadır. Çevre modeli ise ajanın hareket edeceği, ödüllerin belirlendiği, çevre koşullarının simülasyonu edilmesidir.

## 1.6 Uygulama Alanları

Veri madenciliğinin kullanıldığı alanlar çok geniş bir kapsama sahiptir. Regresyon, kümeleme ve sınıflandırma yöntemleri ile kullanılan veri madenciliği finans bankacılığa, sağlık sektöründen bilimsel araştırmalara, CRM den ticaret ve e-ticaret gibi en yaygınları olmakla beraber her geçen gün kullanımı giderek popüler hale gelmektedir [27].

**Bankacılık-Finans:** Bankacılık sektöründe veri madenciliği ile tespit edilmek istenen konular şöyledir. Müşterinin kredi kartı ile yaptığı alışverişler doğrultusunda müşteri alışkanlık ve ilgi alanlarını tespit ederek doğru kampanya ve kredi imkanları ile yeni ürünlerin satılmasıdır. Kredi talebinde bulunan

müşterinin ödeyebilme yetkinliği belirleme ve doğru kredi limitinin belirlenmesidir. Borsa, altın, döviz gibi yatırım araçlarının gelecekte olabileceği değerlerin tahmin edilmesi gibi çalışmalar yapılmaktadır. İnternet üzerinden yapılan kredi kartı işlemlerinin gerçek kullanıcı veya dolandırıcılıklarının kullanıp kullanmadığının tespit edilmesidir [28].

**Pazarlama:** Veri madenciliğin en aktif kullanıldığı bir sektörde pazarlama sektörüdür. Müşterilerin ilgilendikleri ürünleri satın alma alışkanlıklarının belirlenmesi ile satış miktarının belirlenmesi. Müşterilerin alışveriş sırasında aldıkları ürün gruplarının belirlenerek uygun ve müşteri ilgisini çeken kampanyalar üretmek için kullanılmaktadır. Satış tahmini ile kritik stok analizleri, temin süreleri gibi konular incelenerek en az stok maliyeti ile sürekli stoka sahip olarak maliyetlerin azaltılmasıdır [29].

**CRM:** Müşterilerimizin hizmet ve ürünlerden memnuniyetini ölçerek kaydedilmesini önler iken, yeni müşteriler kazanmak için müşteri aday profilleri belirlemektir. Müşterilerimizden en iyi seviyede bağlılık ve faydaları sağlamak için yapılan analiz çalışmalarındandır [30].

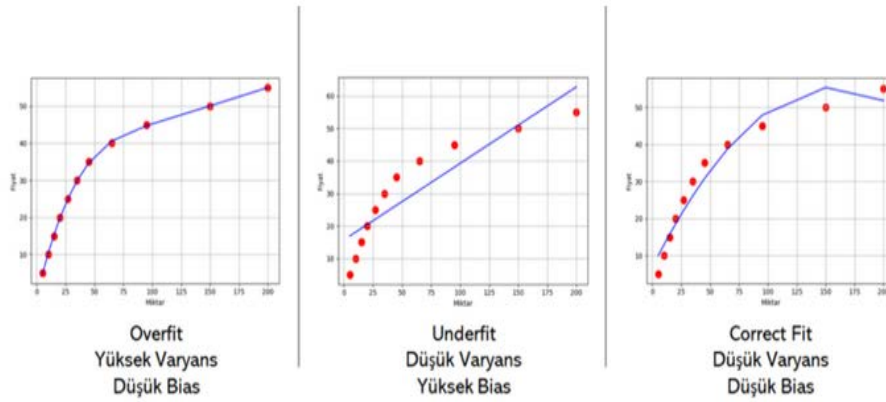
**Bileşim:** Web sitesini ziyaret eden kullanıcıların hangi sayfaları ziyaret edildiği ve site kullanma davranışlarının incelenmesi ile gerekli doğru donanım ve yazılım çalışmalarının yapılmasıdır. Uygulamaların üzerinde çalıştığı sunuculara yapılan saldırıların analiz edilmesi, gerekli durumlarda firewall sistemleri ile birlikte saldırı yapan adres ve bilgisayarların bağlantısının kesilmesi tespitleri, sunucu üzerinden gelen isteklerin saldırı olup olmadığının belirlenmesidir [31].

**Sigortacılık:** Risk ve para yönetimin birlikte kullanıldığı sektör olan sigortacılık da istatistik ve veri madenciliği çok önem kazanmaktadır. Sigorta edilecek olan ev, araç gibi varlıkların uğrayabildikleri zarar ve hasarların olma olasılıkları ve oluşan hasarın maddi değerinin incelenmesidir [32]. Sigorta yaptıran müşterinin yaş, eğitim, yaşadığı bölgede gibi bir özellikleri ile kazaya karışabilme olasılıkları ile doğru sigorta primlerinin tespit edilmesidir. Kötü niyetli müşteriler tarafından yapılabilecek sigorta dolandıracağını önlemek için kazalar üzerinden çalışmalar yapılmaktadır [33].

## 2. YÖNTEM

### 2.1 Polinomsal Regresyon

Polinomsal regresyon lineer regresyonun tek yönlü eğilimi yetersiz kaldığı için eğri fonksiyonu ile ifade edilen lineer regresyonun bir türüdür. Lineer regresyon matematiksel olarak  $y=Q_0+Q_1x$  gibi tek eğilimi iken Polinomsal regresyon  $y=Q_0+Q_1x^2+Q_2x^3+...+Q_nx^n$  gibi eğri fonksiyonu şeklinde olduğu için şekilde görüldüğü gibi daha iyi oturmaktadır [34]. Polinomsal regresyonda seçilen polinom derecesi ideal eğri noktalarını bulunmasını(fitting) sağlar iken derecenin fazla seçilmesi verilerin ezberlenerek (overfitting) testlerde doğru sonuç verir iken, uygulamada sorunlara sebep olabilmektedir [35]. Ortaya çıkan modelden kaynaklanan hataların yanlılık, varyans dengesinin sağlanması ile doğru polinom derecesi bulunabilir. Bias (Yanlılık) verinin eğitim sırasındaki oturma işlemindeki basit ön kabul değerlerinden kaynaklanan hata değeridir. Yüksek yanlılık değeri, doğru eğim modelini bulamadığını (underfitting) ve düşük öğrenme yaptığını gösterir [36]. Variance (Varyans) ise verinin eğitim sırasındaki karmaşık işlemlerden dolayı oluşan hatayı temsil etmektedir. Yüksek Varyans değeri test verilerinin büyük kısmından geçerek ezberlemeye (overfitting) sebep olup bu durumlar Grafik 2.1 de gözükmektedir.



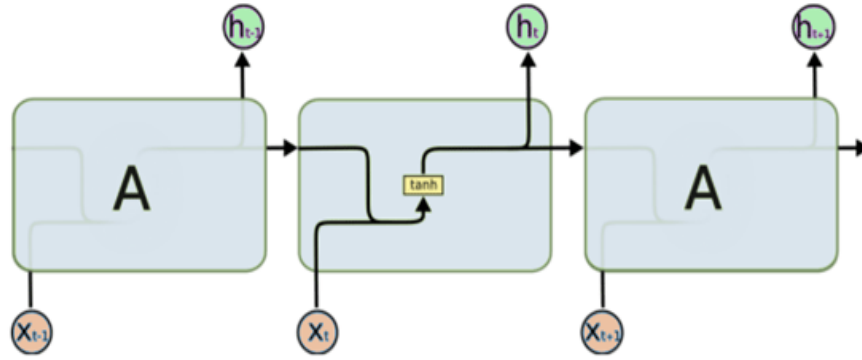
Şekil 2.1: Polinomsal Durumu



## 2.2 LSTM

Birçok popüler yapay zekâ çalışmalarında kullanılan derin öğrenme algoritması, gelişen CPU ve GPU donanımsal yetenekler, neredeyse sınırsız işlem ve saklama yetkinliği sahip bulut teknoloji ise birçok uygulamada karşımıza çıkmaktadır. Verilerden kural çıkarmakta yetkin olan derin öğrenmede daha fazla miktarda gerekli olan veri miktarını işlemede CPU'dan daha güçlü olan GPU'lar kullanımı tercih edilmektedir.

Derin Öğrenme çok miktardaki verileri insan beynindeki nöron yapısını, birden fazla farklı işlevleri olan katmanlar ile sinir ağlarını taklit ederek kurallar çıkarmayı hedeflemektedir [37]. Özellikle algoritmanın en iyi sonuçları alması için gereken parametreleri algoritma tarafından bulunması çok tercih edilme sebeplerindedir.



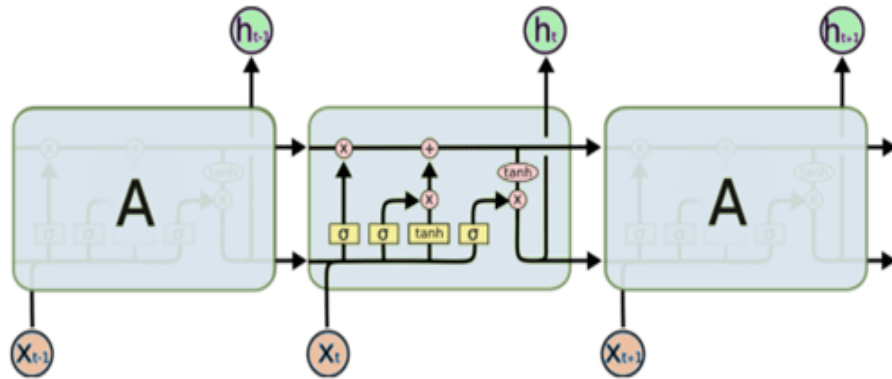
Şekil 2.2: RNN Çalışma Adımları

RNN sıralı bilgilerin giriş yapıldığı tekrarlanan yapıdaki sinir ağlarından oluşan sıralı ve çok katmanlı bir sinir ağı modeline sahip olup bu model Şekil 2.2 de olduğu gibidir. [38]. Bu tip ağlarda bilginin giriş sıralaması çok önemli olup bir önceki bilgiyi de dikkate alarak sanki bir hafızası varmış gibi işlem yapmaktadır. Bu hafıza sayesinde zaman serilerinde olduğu gibi önceki verilerden sonraki değerleri etkilediği gibi, bir cümle tamamlama uygulamasında da bir sonraki gelecek kelimeler sıra ile önceden girilen kelimelerden etkilenmektedir.

Geri bildirimli olarak tekrarlayan kısa süreli bellek yöntemi giriş, gizli ve sonuç katmanlarında farklı aktivasyon fonksiyonları içeren çeşitli ağırlıklara sahip bir

ağ modelidir [39]. Bu aktivasyonlar sadece ilerden gelen verilerden başka çıktı verilerinden de beslenmektedir.

LSTM -Uzun Kısa Vadeli Hafıza Ağları- RNN'lerin daha uzun süreli bağımlı veriler ile öğrenebilen daha özel bir türüdür. Hochreiter & Schmidhuber tarafından 1997 yılında tanıtılarak birçok uygulama ve çalışmadan kullanılmıştır [40]. LSTM özellikle uzun süreli bağımlılıkta olan bazı sorunları çözmek üzere yeniden tasarlanmış bir yapıya sahiptir. LSTM de hücreler davranış ve devralmadaki ezberleme sorununu çözmek için kapı ismi verilen 4 katman bulunmaktadır. Bu katman yapısı Şekil 2.3 de gösterilmiştir. Normal veri bağlantıları üzerinden gelen bilgilerin yanında dışarıdan da bilgi alabilen ve bu bilgileri depolama, hücreye yazma veya silme işlemi yapmaya karar veren kapılara sahip bir yapıdır [41]. Sinir ağlarından olduğu gibi gelen bilgiyi ağırlığına inceleyerek gözden geçirip bu verinin hücrede saklanması veya silinmesi gerekmesini öğrenmektedir. LSTM farklı ihtiyaçlarda ise kapı modelleri değişimi yaparak çıktını adımların farklı değerleri alması sağlanabilmektedir.



Şekil 2.3: LSTM Çalışma Adımları

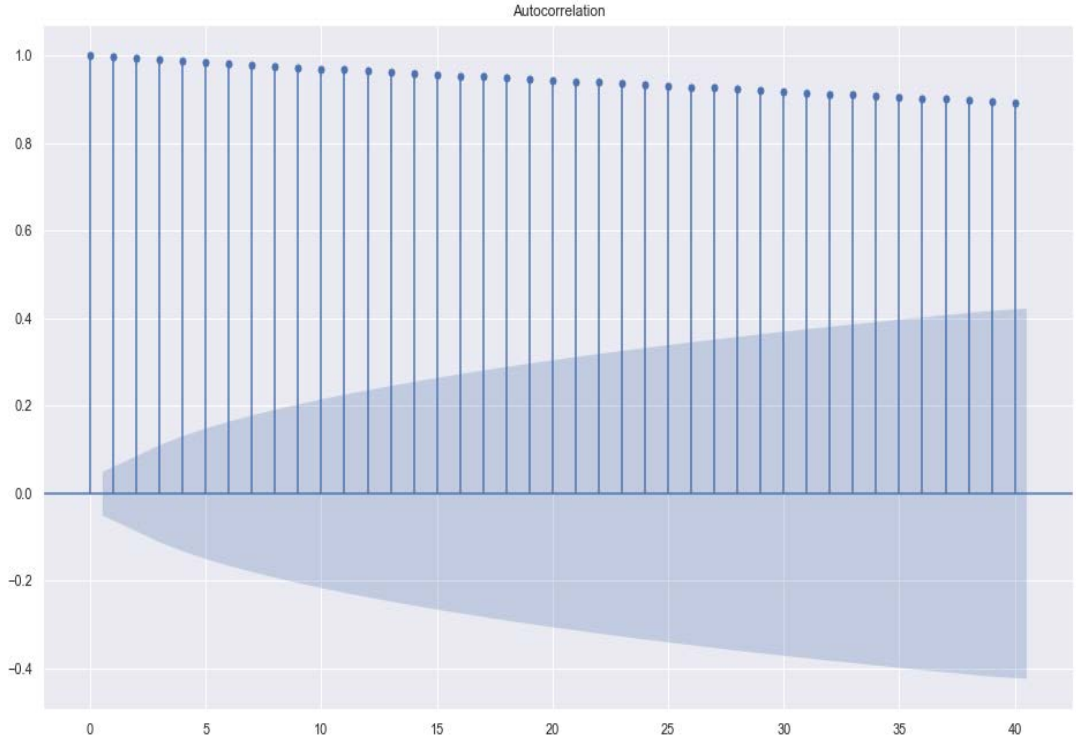
### 2.3 Arima

Birçok popüler yapay zekâ çalışmalarında kullanılan derin öğrenme algoritması, gelişen CPU ve GPU donanımsal yetenekler, neredeyse sınırsız işlem ve saklama yetkinliği sahip bulut teknoloji ise birçok uygulamada karşımıza

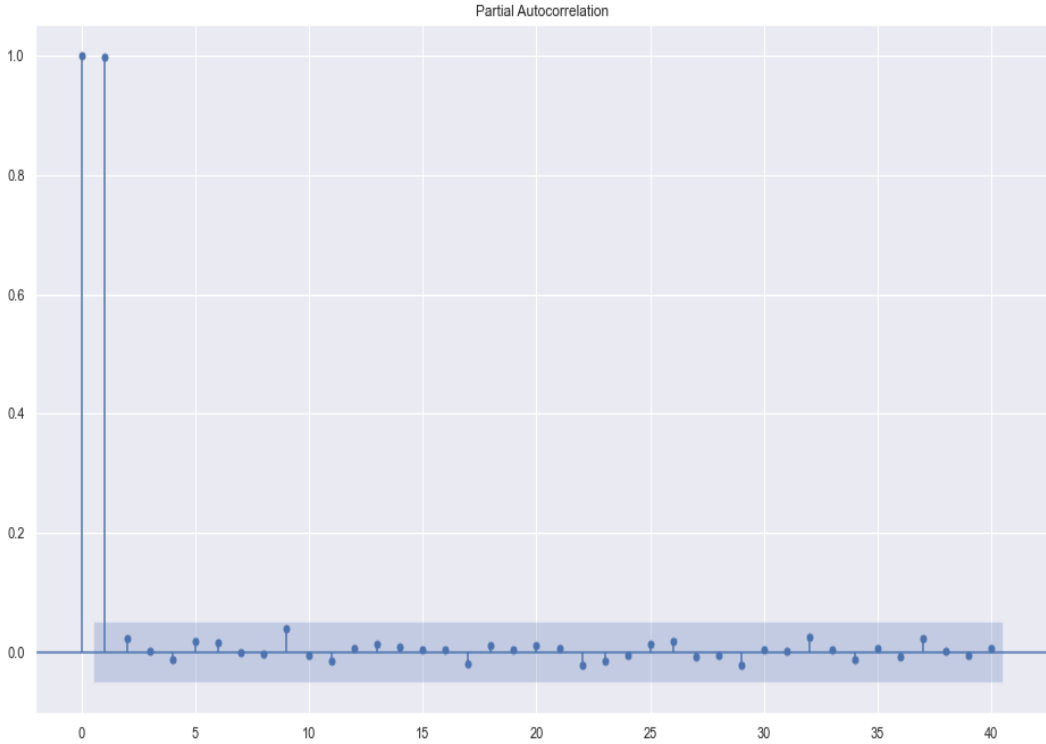
çıkılmaktadır. Verilerden kural çıkarmakta yetkin olan derin öğrenmede daha fazla miktarda gerekli olan veri miktarını işlemede CPU'dan daha güçlü olan GPU'lar kullanımı tercih edilmektedir [42].

1970 yılında George Box ve arkadaşı Gwilym Jenkins ile beraber geliştirip önerilen bir model olan Box Jenkins yöntemi bilim dünyasına kazanılmıştır [43]. Zaman serileri analizlerinde verinin durağan olması gerekliliğinden dolayı ekleme veya çarpma yöntemleri ile durağanlaştırma yöntemleri uygulanmalıdır. Box Jenkins yöntemi veri yapısında göre durağan verilerde ARMA, durağan olmayan durumlarda ARIMA ve mevsimsel uygulamalarda SARMA alt yöntemlerini kullanmaktadır [44]. Arima (Autoregressive Integrated Moving Average Model) zaman serileri analizinde kullanılan serilerin trend ve hareketleri üzerinde parametreler ile düzenlemeler yaparak inceleyen bir yöntemdir [45].

Arima Modeli AR(Autoregression) model oluşumdaki gecikmeli değerleri arasındaki ilişki kurma, I (Integrated) zaman serilerinin incelemesi için gerekli olan durağanlaştırma işlemleri için alınan fark katsayısı, MA (Moving Average) -Hareketli ortalama- verilerin parametreye göre alınan hareketli ortalama ile hatalar arasındaki ilişkinin incelenmesidir. İki veya ikiden daha fazla değişken değerleri arasındaki doğrusal ilişki bakımından yön ve gücü temsil eden bir değere korelasyon adı verilmektedir [46]. Otokorelasyon ise zaman serisinde bulunan değerlerin korelasyon ne kadar olduğunu gösteren bir grafikdir [47]. Grafik 3 Microsoft Hisse Senedi Otokorelasyon Grafiği görünmektedir. Kısmi otokorelasyon ise korelasyonun etkisi temizlenip kaldırıldıktan sonra geriye kalan korelasyon değeri olmakla birlikte Grafik 4 Microsoft Hisse Senedi Kısmi Otokorelasyon Grafiği bulunmaktadır.



**Şekil 2.4:** Microsoft Hisse Senedi Otokorelasyon Grafiği



**Şekil 2.5:** Microsoft Hisse Senedi Kısmi Otokorelasyon Grafiği

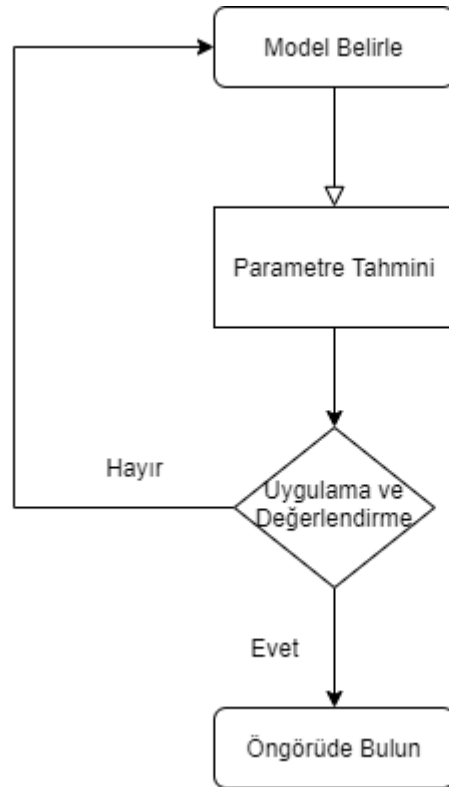
Hareketli ortalama verinin belli dönem parçaları halinde geçmişteki hareket yönünü veren hesaplama yöntemi olmaktadır. Arima modelinde mevsimsellik ve sadece tek değişken ile yapılan modelde ARIMA (p, d, q) tanımı kullanılıp p oto

regresyon alma derecesini, d seriyi durağan hale getirmek için gerekli olan katsayıyı, q ise hareketli ortalama için dikkate alınacak katsayısı temsil etmektedir [43]. Zaman serileri üzerindeki başarılarından dolayı üzerinde birçok gelişmeler olmakla birlikte, birçok framework de verilen test verileri ile en az hata değerlerine sahip parametreleri bulmak için yardımcı metotlar bulunmaktadır. Bu yardımcı metotlar sayesinde araştırmacı tek tek değerleri deneyip değerlendirmek yerine ilgili metot verilen parametre aralıklarından değerleri deneyerek uygun değeri vermektedir.

Bu çalışmada Python da bulunan pmdarima framework üne ait `auto_arima(self.train, start_p=1, start_q=1,max_p=8, max_q=8, m=1,start_P=0, seasonal=False,d=1,`

`D=1,trace=False,error_action='ignore',suppress_warnings=True, stepwise=False)` kullanılmıştır.

ARIMA da model birçok yapay zekâ yönteminde de olduğu gibi, model belirleme, parametre belirleme, yöntem uygulama ve sonuçların değerlendirilmesi ve öngörü yapımı adımlarından oluşmakta olup bu adımlar Şekil 2.4 de görünmektedir.



Şekil 2.6: Arima Çalışma Şekli

## 2.4 Facebook Prophet

Prophet Facebook tarafından geliştirilmiş, kolay kullanım ve performans hedefli zaman serilerinde tahmin yapan bir yazılım kütüphanesidir [48]. Prophet diğer yöntemlerden ayrılan önemli özelliklerin bir tanesi özellikle satış ve talep gibi bazı özel zamanlarda çok farklı değerler alabilen verilerde bu özel zamanları tanımlayarak tahminler yapabilmektedir [49]. Diğer yöntemlerde veriyi temizleme işlemlerinden sonra da yöntemde kullanmak için değişiklik yapmak gerekir iken Prophet de veri yapısında zaman ve değer olması yeterlidir. Bu kolaylık analiz yapan araştırmacının işini kolaylaştırmaktadır. Facebook Prophet altyapı olarak kolay tutulmuş olup doğrudan dış veri tabanları ile bağlantı kurup, özel bir çalışma gerektiren bir modele sahip olmadığı için çok uygundur. Diğer zaman analizi yöntemlerinde ayrıca veriler incelenip uygun model hazırlanmak gerekir iken Facebook Prophet de bu yazı tarih ve veri olmak üzere çok basittir. Facebook Prophet verileri inceleyip her durum için en uygun yöntemi kullanmaktadır. Verileri aylık, yıllık hatta saatlik bilgilere sahip olanlarda saatlik, günlük, özel tatil günlerini de dikkate alarak aykırı değerlere, trendlere dikkat ederek incelemektedir [50]. Prophet verileri analiz etmek için 4 bileşen parçasını kontrol ederek bir model analizi yapar. Modelin doğrusal veya lojistik regresyon eğrisini bulmaya çalışır. Yıllık değişim modeli için Fourier serilerini kullanır iken kukla değişkenler ile haftalık parçaları incelemektedir [50]. Diğer yöntemlerden farklı olarak da özel gün ve hafta tatilleri analize eklenmektedir.

## 2.5 XGBoosting Yöntemi

XGBoost (Extreme Gradient Boosting) Ekstrem Gradyan Arttırmanın olarak adlandırılan Tianqi Chen ve Carlos Guestrin'in 2016 yılında yayınladıkları Gradient Boosting algoritmasının gelişmiş bir versiyonudur [51]. Algoritma paralel çalışma, budama, aşırı öğrenmeyi engelleme gibi işlemler ile iyileştirilmiştir. XGBoost algoritması karar ağacı temelli bir en iyi sonuçları bulmak için kendini iyileştiren bir yapıdadır [52]. XGBoost atası olan Gradient Boosting yöntemi zayıf öğrencilerin adım adım değişimler ile güçlü öğrenciye gelişme yöntemidir. Boosting algoritmaları birbirlerine göre farkları zayıf öğrencilerin nasıl belirlenip bu duruma göre işlem yapmasıdır. Gradient

Boosting ilk adımda karar ağacı için ilk yaprakları oluşturur. Sonra iterasyonlar ile hataları da dikkate alarak yeni ağaçlar oluşturur [53].

Karar ağacı bir sınıflandırma algoritması olmakla beraber istenen sonuçlara varmak için belirli koşullara göre dallandırılması ile oluşan yapısı ağaca benzeyen bir yapıdır. Bir karar ağacı, belirli kurallara göre küçük alt veri kümelerine bölerek gruplandırılan bir yapıdır [54]. Karar ağacının yapısında, içerisinde çok sayıdaki kayıtlar işlenmek için küçük kümelere bölünerek bazı kurallar çerçevesinde bölünen bir yapıda olan bir sistemdir. Büyük miktardaki veriler üzerinden karar verebilmek için daha küçük yapılara dönüştürülmüş halidir.

## 2.6 Hata Ölçme Yöntemi

Yapılan ölçümleri değerlendirmesi çalışmanın doğru ve en uygun değerlere ulaşması için büyük öneme sahiptir [55]. Hata değerlerini ölçmek için tahmin değeri ile gerçek değer arasındaki değişimi ölçmek incelemek gereklidir. Tahmin değeri ile gerçek değer arasındaki değişimi ölçmek için Ortalama Mutlak Hata, Ortalama Kare Hata ve Kök Ortalama Kare Hata ölçme yöntemleri kullanılır.

### 2.6.1 Ortalama Mutlak Hata

MAE (Mean Absolute Error, Ortalama Mutlak Hata) tahmin ve hedef dizileri arasında yön farkında dikkate almaksızın değerlerin farklarının ortalaması bulunarak ölçülmektedir. Tüm dizilerdeki elemanların farklılıkları aynı seviyede olduğundan olayı ortalama diye ifade edilip formülü aşağıda görüldüğü gibidir [56].

$$mae = \frac{\sum_{i=1}^n abs(y_i - \lambda(x_i))}{n}$$

### 2.6.2 Ortalama Kare Hata

MSE (Mean Squared Error, Ortalama Kare Hata) tahmin ve hedef dizileri arasında farkların karelerinin toplanması ile bulunan değerdir [57].

$$mse = \frac{1}{n} \sum_{t=1}^n e_t^2$$

### 2.6.3 Kök Ortalama Kare Hata

RMSE (Root Mean Square Error, Kök Ortalama Kare Hata) tahmin ve hedef dizileri arasında farkların karelerinin ortalamasının karekökü ile bulunan değerdir [56].

$$rmse = \sqrt{\frac{\sum_{t=1}^n e_t^2}{n}}$$



### 3. LİTERATÜR ÇALIŞMASI

Hacioglu Karasu S. ve arkadaşları destek vektör makinesi ve doğrusal regresyon yöntemiyle, 2012 ile 2018 tarihleri arasında olan bitcoin kapanış değerlerini makine öğrenmesi yöntemleri kullanarak tahmin etme konusunda çalışmışlardır [58]. Bu çalışmada farklı zaman bölümlerine göre hareketli ortalama, doğrusal fonksiyon, Polinomal regresyon ile minimum hata değerleri bulunacak şekilde çalışma yapılmıştır. Khan ve arkadaşları 2001 yılında Yapay sinir ağıları yöntemini kullanarak gen ekspresyon imzalarına dayanarak, küçük yuvarlak mavi hücreli tümörleri tanımlayarak kanser hücrelerini tanınması üzerine çalışma yapmıştır. Klinik tanımları ile kanserleri dört farklı kategoriye ayılarak tahmin yapmaktadır [59]. Kenan KARAGÜL 2013 yılında yaptığı çalışmada IMKB-100 endeksi içerisinde bulunan çimento, gıda ve tekstil gibi farklı sektörlerde faaliyet gösteren 42 şirket için finansal durumları bakımından 3 kategoriye ayrılması üzerine Destek Vektör Makineleri ile çalışma yapılmıştır. Şirketlere ait 2006-2011 yıllarına ait 10 farklı finansal veri dikkate alınarak çalışma yapılmıştır [60]. M. Suphi ÖZÇOMAK ve Murat GÜNDÜZ tarafından 2014 yılında yapılan çalışmada Borsa İstanbul bulunan firmaların kapanış fiyatları üzerinden bir çalışma yapılmıştır. Yapılan çalışmada önce veriler fonksiyonel veriye çevirip fonksiyonel kanonik korelasyon analizi ile kapanış ve işlem miktarı arası ilişki incelenmiştir [61]. Hakan PABUÇCU tarafından yapılan çalışmada 2009-2018 yılları arasındaki Borsa İstanbul günlük kapanış değerlerini kullanarak, endekslerinin hareket yönü tahmini için yapay sinir ağlarının, naive bayes sınıflandırma yöntemi ve destek vektör makineleri yöntemleri kullanılmıştır [62]. Melek ACAR BOYACIOĞLU ve Derya ÇÜRÜK yaptıkları çalışmada Borsa İstanbul 100 Endeksi ile döviz kuru arasındaki etkileşimi hakkında çalışma yapmışlardır. Yapılan çalışmada 2006-2014 yılları arasında ticaret ve imalat sektörlerinde faaliyet gösteren 42 firmanın yıllık hisse senedi kazançları ile döviz kazanç oranları incelenmiştir [63]. Birgül Kutlu ve Bertan Badur 2009 da yaptıkları çalışmada, Borsa İstanbul 100 Endeksi tahmin

üzerine Yapay sinir ağlarının ileri beslemeli yöntemini kullanarak 2 Temmuz 2001 ile 13 Temmuz 2006 tarihleri arasında veriler ile çalışma yapmışlardır [64]. Nurdan Değirmenci ve Ali Akay yaptıkları çalışmada ARIMA yöntemi ile borsa, petrol, altın ve döviz fiyatları 01.02.2009-11.25.2016 tarihleri arasındaki veriler kullanılarak arasındaki ilişkiler incelenmiştir [65]. Önder Aydemir 2008 yılında yaptığı çalışmada Elektroensefalografi (EEG) işaretleri üzerinde destek vektör makinesi ve KNN sınıflandırma yöntemini kullanarak sınıflandırma çalışması yapılmıştır [66]. Yasin Kalkavan ve Feridun Özçakır Arima yöntemi ile araç kiralama verileri üzerinden kiralanmış araç sayısı tahmini hakkında çalışma yapmışlardır [67]. Arima yönteminden başka RNN ve LSTM yöntemleri ile de çalışma yapılmıştır. Bu çalışmadaki amaç doğru hedef tahminleri ile firmaların karlılıklarını yükselterek firmanın devamlılığını sürdürmektir [67]. Özlem Alpay 2000 ile 2018 yılları arasındaki USD/TRY paritesi verileri üzerinden LSTM yöntemini ile zaman serileri üzerine çalışma yapmıştır. LSTM yönteminde bulunan bellek yardımı ile zamansal veriler üzerinde iyi bir başarı elde edilmiştir [68]. Onur Sevlî içinde bulunduğumuz günlerin en önemli olay ve durumlarından olan Covid-19 üzerine Facebook Prophet modeli ile incelemiştir [69]. Yapılan çalışmada vaka ve ölüm sayıları ile iyileşen vaka sayılarının artış tahmini üzerine çalışmıştır [69]. Facebook Prophet 'in kolay kullanımı olması ve parametrelerin yöntem tarafından optimize edilmesi çalışmanın hızlı bir şekilde gerçekleşmesini sağlamıştır. Pandemi sürecinde ve sonrasında bu tahminlerin devam edilecektir.

## 4. ARAŞTIRMA DEĞERLENDİRİLMESİ

### 4.1 Araştırma Metodu

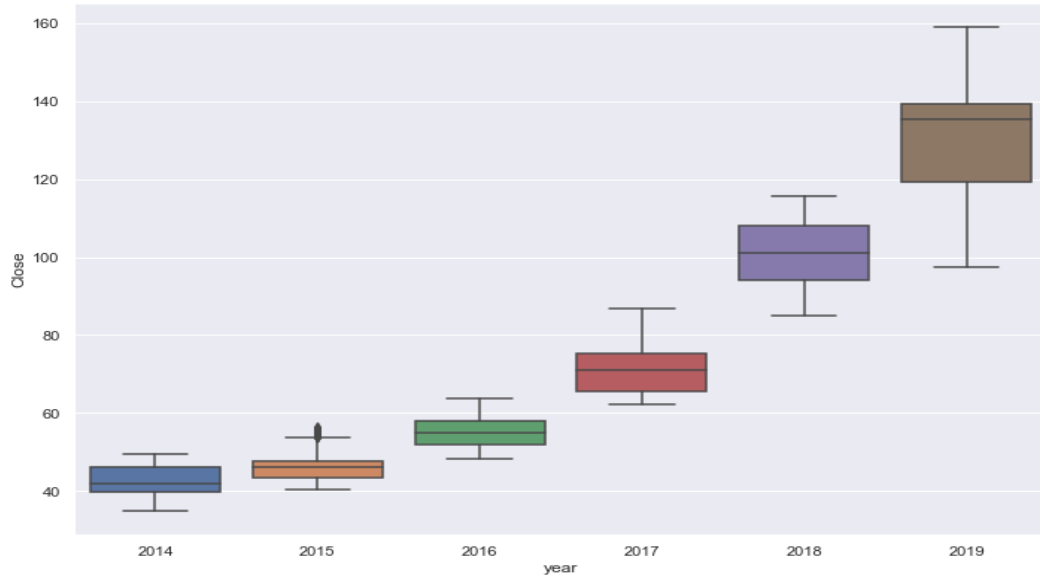
Bu çalışma Yahoo Finans sitesinden alınan 2014 ile 2019 tarihleri arasındaki Microsoft Şirket hisse senedinin günlük kapanış değerleri üzerinden yapılmıştır. Makine öğrenmesi yöntemleri ile ileri tarihteki kapanış değerleri tahmin edilmeye çalışılmıştır. Çalışmada Python üzerinden çeşitli farklı yazılım paketleri kullanılarak, 5 farklı yöntem ile tahmin sonuçları incelenmiştir. Bu çalışmada hisse senedi kapanış fiyatlarının gerçek değerler en yakın sonuçların bulunması amaçlanmıştır. Yatırımcıların kazandıkları varlıklarının değer kazanması için çeşitli yatırım araçlarını tercih ederler. Bu yatırım araçları içinde borsa önemli bir yer tutmaktadır. Fakat borsada hisse senedi fiyatları politik, ekonomik veya sektöre bağlı sebeplerden dolayı çok değişken bir yapıda olabilmektedir. Bu değişkenlik yatırımcıya yüksek miktarda kar veya zarar olarak dönebilir. Bu sebepten dolayı hisse senedinin gelecekteki değerlerini tahmin etmek yatırımcı için çok önem kazanmaktadır. Makine öğrenmesi yöntemleri ile yapılan çalışmalarda hisse senedi kapanış değerleri tahmin modelleri önerilmiştir. Bu çalışmada istatistiksel yöntemler ile yapılan teknik analizlerin yanında, makine öğrenmesi yöntemlerinden zaman serisi analiz yöntemlerine uygun olan 5 farklı algoritma kullanılmıştır. Yöntemlerin performanslarının ölçülmesi için 5, 10 ve 20 günlük tahminler, günlük kapanış değerleri üzerinden herhangi bir grupta veya ortalama alınmadan yapılmıştır. Kullanılan yöntemler ise Polinomal Regresyon, Arima, XGBooster, LSTM (Uzun Kısa Vadeli Hafıza Ağları) ve Facebook Prophet yöntemleri tercih edilmiştir.

Bu çalışmanın yapılmış olduğu dizüstü bilgisayar 16 GB RAM, Intel Core(TM) i7-8550 CPU 1.80GHZ İşlemci donanımına sahiptir. Eğitim için kullanılan veriler, zaman bağımlı olduğu için 20 gün için yapılan tahminde son 20 gün dışındaki veriler eğitim için kullanılmış olup; son 20 gün kapanış değerleri ile

test edilmiştir. Python ile yapılan çalışma sırasında farklı öğrenme algoritmaları ve yapılar kullanılarak minimum hata olanları bulunmaya çalışılmıştır.

#### 4.2 Ölçümlerin Değerlendirilmesi

Bu çalışmada analizi yapılan Microsoft hisse senedi kapanış fiyatları Grafik 4.1 de yıllık bazda kutu grafiğinde görüldüğü gibi yükselen bir trende sahiptir. Bu çalışmada kullanılan Polinomal Regresyon, Arima, XGBooster, LSTM (Uzun Kısa Vadeli Hafıza Ağları) ve Prophet yöntemleri için 5, 10 ve 20 günlük hata sonuçları Tablo 1 Hata Sonuç Tablosu görünmektedir. Sonuçların incelendiğinde grafik-2 de görüldüğü gibi Polinomal en az hata değerini vermiştir.



Şekil 4.1: Yıllık Bazda Kapanış Değişimi

**Çizelge 4.1:** Hata Sonuç Tablosu

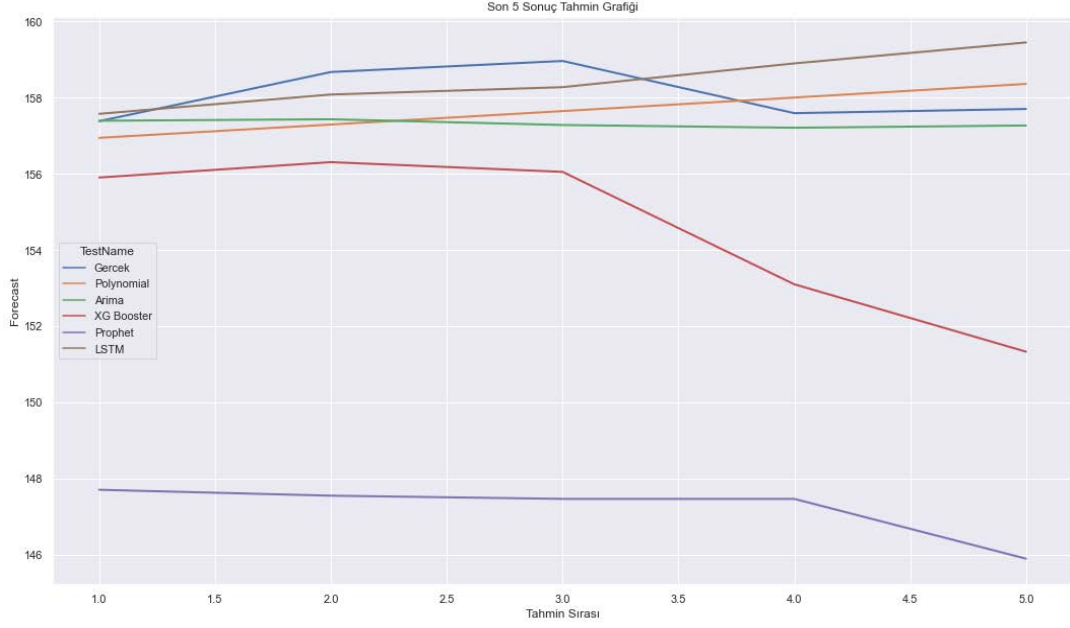
Algorithm	MSE			RMSE			MAE		
	5	10	20	5	10	20	5	10	20
<b>Arima</b>	<b>0,9393</b>	<b>5,4608</b>	<b>19,8731</b>	<b>0,9692</b>	<b>2,3368</b>	<b>4,4579</b>	<b>0,7496</b>	<b>2,0969</b>	<b>3,6710</b>
<b>LSTM</b>	29,8243	10,0034	8,3038	5,4612	3,1628	2,8816	5,3963	3,0374	2,5442
<b>Polinomial</b>	<b>0,8848</b>	<b>0,8238</b>	<b>4,1739</b>	<b>0,9406</b>	<b>0,9076</b>	<b>2,0430</b>	<b>0,8399</b>	<b>0,8114</b>	<b>1,6820</b>
<b>Prophet</b>	118,2301	128,0429	120,4416	10,8734	11,3156	10,9746	10,8429	11,1409	10,3985
<b>XG Booster</b>	<b>15,4017</b>	<b>30,2529</b>	<b>49,4574</b>	<b>3,9245</b>	<b>5,5003</b>	<b>7,0326</b>	<b>3,5235</b>	<b>4,8624</b>	<b>5,7572</b>

MSE: Mean Squared Error

RMSE:Root Mean Square Error

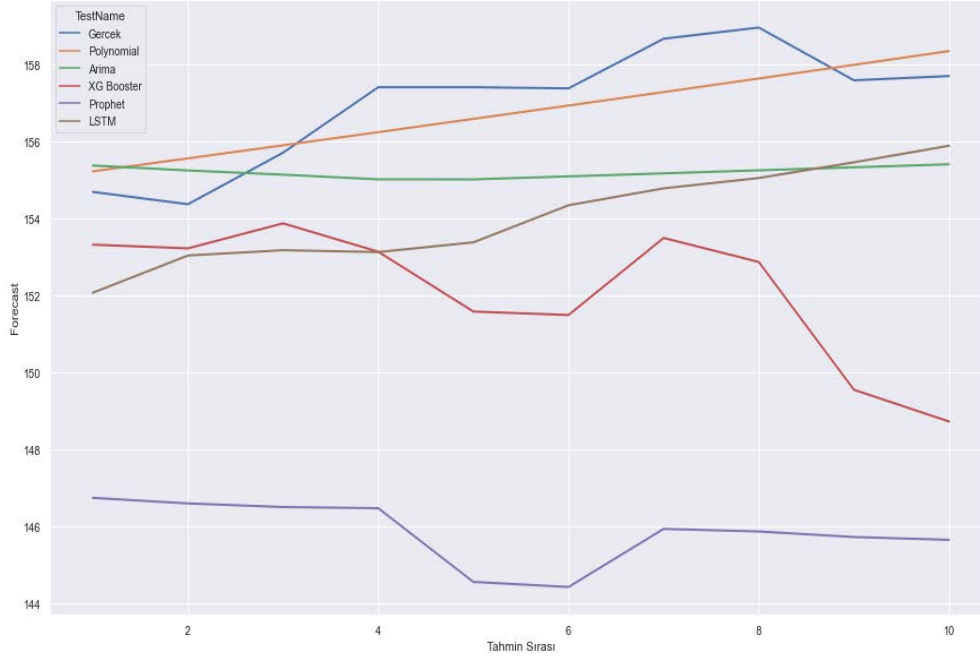
MAE:Mean Absolute Error

Son 5 günlük tahmin değerlerini gösteren Grafik 4-2 grafikte gerçek değere en yakın olan sonuç polinomal regresyon yöntemi olmakla birlikte ARIMA ve LSTM den gerçek değerlere yakın tahminlerde bulunmuştur. İlk 2 günlük tahminde gerçek değerlere yakın olan XGBooster yöntemi özellikle 5 güne doğru daha fazla hata değerleri tahmin etmiştir. Facebook Prophet yöntemi ise neredeyse %10 kadar değerleri düşük tahmin ederek 5 günlük tahmin çalışmasında en kötü değeri tahmin etmiştir.



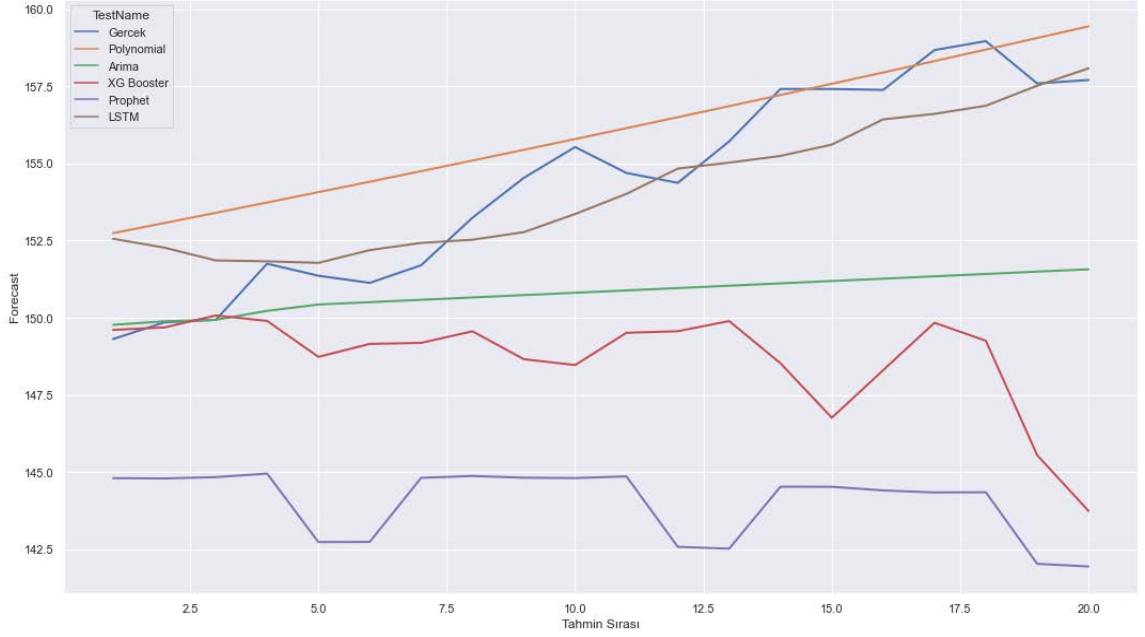
**Şekil 4.2:** Son 5 Gün Tahmin Grafiği

Son 10 günlük değer tahminlerini gösteren Grafik 4-3 deki grafikde yine en iyi Polinominal Regresyon olmakla birlikte LSTM ve ARIMA yöntemleri daha az miktarda hata yapmış olup rahatlıkla kabul edilebilecek bir tahminde bulunmuştur. Gerçek değerlerin günlük hızlı değişimlerine rağmen 10. Gün için yaptıkları tahmin en fazla 3 değerindeki fark ile çok güzel tahminlerde bulunmuştur. XGBooster 5 günlük tahminde olduğu gibi ilk 2 gün için kabul edilebilir miktarda tahminde bulunur iken 10. güne doğru artan gerçek değere rağmen düşüş terendi olan bir tahminde bulunmuştur. Facebook tarafından kolay kullanıma sahip olması için geliştirilen Prohpet yöntemi borsa kapanış tahmini için en kötü tahmini yapmıştır.



**Şekil 4.3:** Son 10 Gün Tahmin Grafiği

Son 20 günlük kapanış değeri tahmininde Grafik 4-4 de görüldüğü gibi gerçek kapanış değeri artan bir trend ile yüksek oynaklığa sahiptir. Bu oynaklık ve 20 adet tahmin istenmesinden dolayı 5 yöntemde farklı tahminde bulunma davranışlarında bulunmuştur. Polinominal Regresyon tahminlerini genel olarak gerçek değerlerden yüksek değerler bulmakla birlikte toplamda yine en iyi tahminlemede bulunmuştur. LSTM yöntemi ilk gün gerçek değerden yüksek tahmin eder iken giderek daha gerçek değerlere yakın tahminlerde bulunarak 20. günde gerçek değere en yakın tahminde bulunmuştur. 5 ve 10 günlük tahminlemede gayet güzel değerlerde tahminlemede bulunan ARIMA 20 günlük tahminde sadece ilk 5 günde yakın değerler tahmin edip hissenin artan trendini tahmin edememiştir. XGBooster yöntemi daha yatay trende gider ilk 3 günde gerçek değerlere yakın tahminde bulunur iken daha sonraki günlerde hisse senedi artış trendine sahip olmasına rağmen düşüş trendi yönünde tahminde bulunmuştur. Facebook Prophet 5 ve 10 günlük tahminlerde olduğu gibi en kötü tahmini yaparak işlemi sonuçlandırmıştır.



**Şekil 4.4:** Son 20 Gün Tahmin Grafiği

Yöntemlerin uzun vadeli performanslarını da incelemek amacı ile son 1 yıllık ait hisse senedi kapanış değer tahmin çalışmaları da eklenmiştir. Tablo 2 de görüldüğü gibi 1 yıllık tahmin için hata tablosunda uzun vadede en iyi sonuçlar LSTM , Prophet ve Polinomial iken Arima ve XG Booster kötü sonuç almıştır. Yapılan tahmin Grafik 4-5 deki tahmin ve gerçek değer detaylı grafiğinde görüldüğü gibi 1 yıl içerisinde 100\$ dan 160 \$ a kadar olan hızlı bir değer artışı olmaktadır. 1 Yıllık eğitim için kullanılan verilerde ise değerler 40 \$ dan 100 \$ a yükselmekte olduğu Grafik 4-6 da görülmektedir. Son 1 yıldaki bu değişim tüm algoritmaların gerçek değerden daha düşük değerleri tahmin edilmesine sebep olmuştur. Tahminin ilk 100 gününde Grafik 4-5 de görüldüğü gibi LSTM , Facebook Prophet ve Polinomial Regresyon gerçeğe daha yakın değerleri tahmin etmiş iken devam eden günlerdeki hızlı artışı yakalayamamıştır. Arima yöntemi ise eğitim verilerinin değişim oranının az olması sebebiyle 100 ve eğitim verilerinin son bölümlerinde azalma trendine sahip olması yüzünden 100 \$ ile 115 \$ aralığında kalmıştır. Karar ağacı temeline sahip olan XG Booster kısa vadeli tahminlerde olduğu gibi uzun vadeli tahminlerde de en kötü sonuçları almıştır.

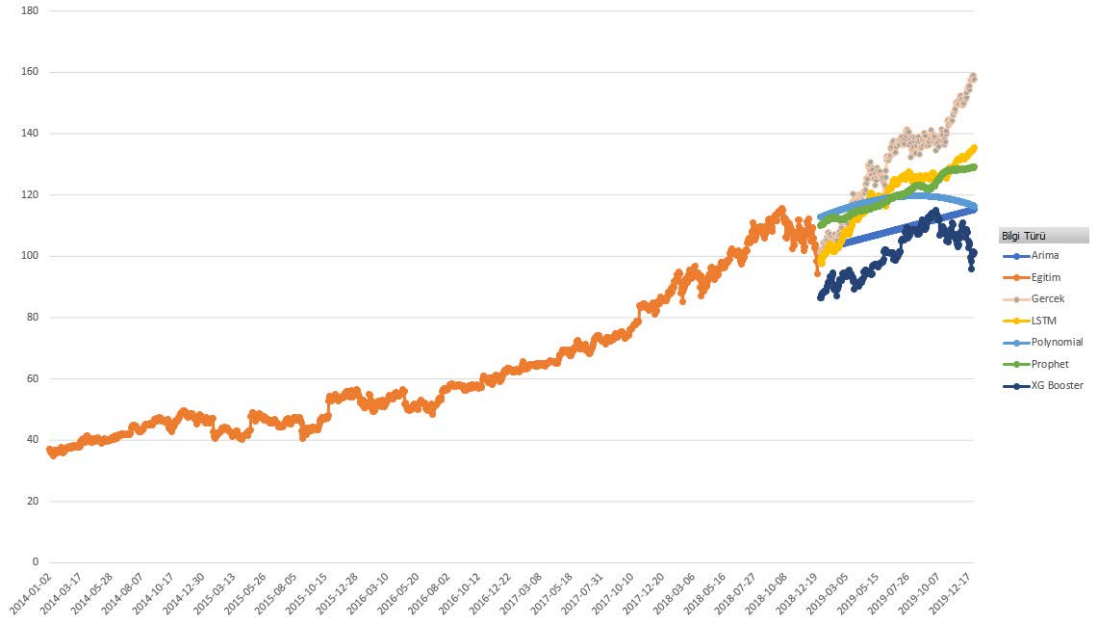


**Çizelge 4.2:** 1 Yıllık Hata Tablosu

Algoritma	MSE	RMSE	MAE
LSTM	133,3537	11,5479	10,2089
Prophet	203,5468	14,2670	12,5386
Polynomial	343,9170	18,5450	15,5097
Arima	595,7021	24,4070	21,7105
XG Booster	958,7950	30,9644	29,3297



**Şekil 4.5:** 1 Yıllık Detaylı Tahmin Grafiği

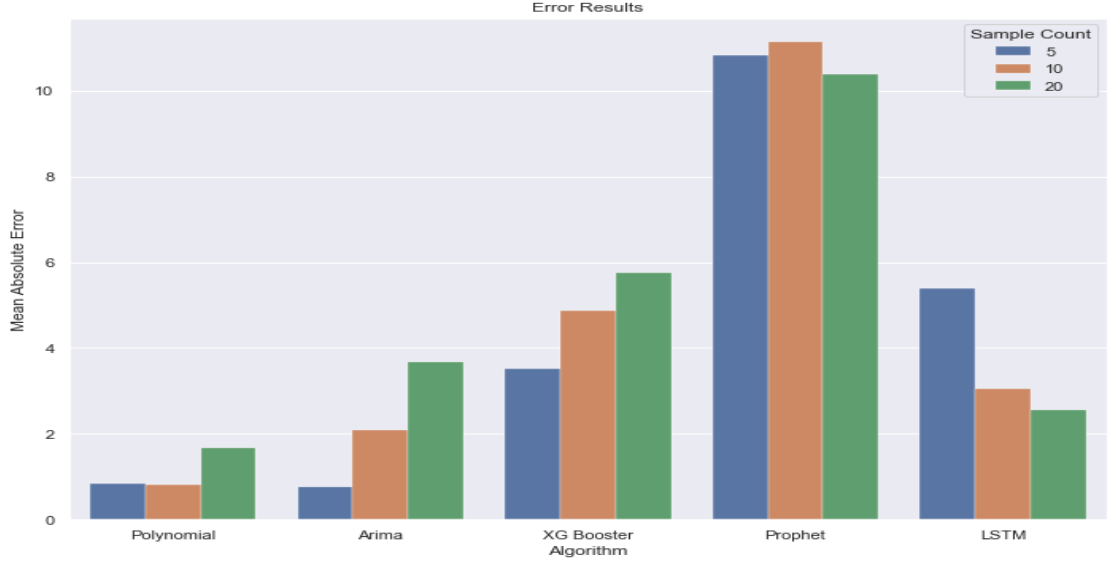


**Şekil 4.6:** 1 Yıllık Tahminde Eğitim -Gerçek-Tahmin Grafiği

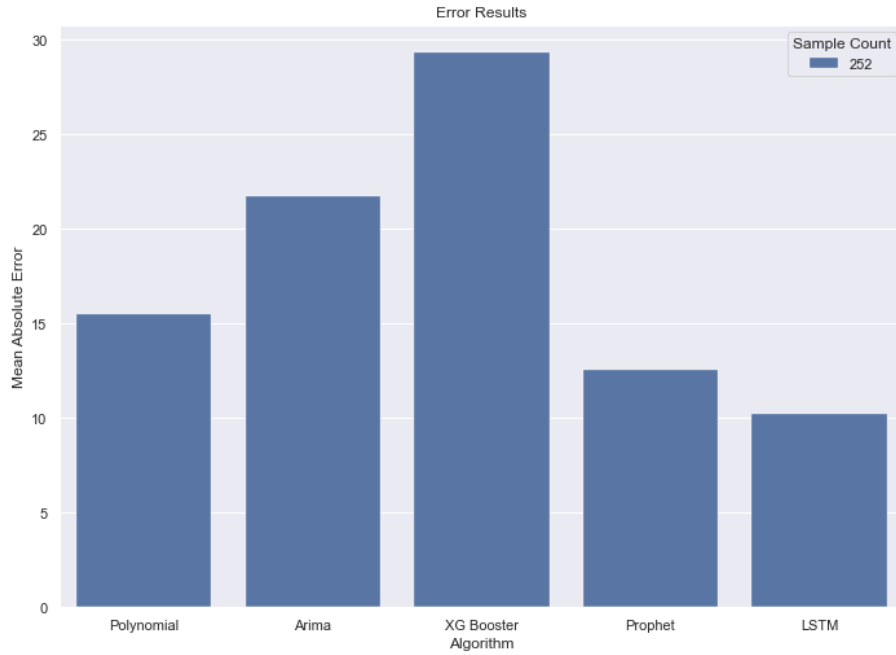
Yapılan çalışmalar sonucunda ortaya çıkan Grafik 4-7 deki grafikte, tahmin yapılan 5 yonteme ait MAE deęerlerin 5,10 ve 20 gnlk hata sonularının karřılařtırılması grnmektedir. Bu alıřma kapsamında ıkan sonular incelendięinde en iyi sonucu Polinominal vermekte iken en kt tahmin ise Facebook Prophet yntemi tarafından yapılmaktadır. ARIMA ve LSTM yntemleri de bařarılı sonular bulmaktadır. XG Booster 2. en kt sonucu veren algoritma olmuřtur. alıřma sırasında eęitim hazırlıęı ařamasında Polinominal, ARIMA ve Facebook Prophet yntemleri iin tarih ve kapanıř deęerleri kullanılarak eęitim yapılmıřtır. Karar Aęacı algoritması zerine alıřan XG Booster ve hafızaya sahip olan derin ęrenme yntemlerinden olan LSTM iin eęitim verisi olarak kapanıř tarihine ait olan tarih; haftanın gn, yılın eyreęi, ay, yıl, yılın gn, ayın gn, yılın hangi haftası bilgilerine dnřm yapılarak, bu bilgiler giriř eęitiminde kullanılmaktadır. LSTM yntemi derin ęrenmede yntem yapısından dolayı eęitim sresi uzun srmektedir. Facebook Prophet kolay kullanımından dolayı ilk incelemede yapmak iin faydalı olup, esas incelemede kabul edilebilir bir tahminde bulunamamıřtır.

Uzun vade iin yapılan son 1 yıla ait deęerlerin tahmininde yntemlerin hata deęerleri Grafik 4-8 de grldęi gibi en iyi sonu LSTM verir iken sonu

Facebook Prophet yöntemi takip etmektedir. Kısa vadeli çalışmada en iyi değerleri veren Polinomial Regresyon 3. sırada bulunmuştur. Kısa vadede iyi sonuçlar veren ARIMA uzun vadede kötü sonuçlar vermiştir. XG Booster ise uzun vadede en kötü değerlere sahip olan yöntemdir.



Şekil 4.7: 5 ,10,20 Günlük Hata Sonuç Grafiği



Şekil 4.8: 1Yıllık Hata Sonuç Grafiği

## 5. SONUÇ

Microsoft Şirketine ait 2014 ile 2019 yılları arasında bulunan günlük hisse senedi kapanış değerleri 5 makine öğrenmesi yöntemleri incelenmiştir. Yapılan incelemelerde Polinomal Regresyon, Arima, XG Booster, Facebook Prophet ve LSTM yöntemleri ile 5 ,10 ve 20 günlük tahminlerde bulunulmuştur. Polinomal Regresyon, Arima, Facebook Prophet yöntemlerinde tarih ve kapanış değerleri üzerinden analiz yapılır iken XG Booster, LSTM yöntemlerinde kapanış tarihlerinin gün, ay, yıl gibi özellikleri üzerinden eğitim yapılmıştır. 5,10,20 günlük tahminlerde en iyi sonuç Polinomal Regresyon yöntemi tarafından yapılır iken, en kötü sonuçlar ise Facebook Prophet yöntemi tarafından yapılmıştır. ARIMA yöntemi tahmin sayısı artıkça, artan hata miktarı olmasına rağmen tüm durumlarda iyi sonuç vermiştir. LSTM yöntemi son 5 ve 10 günün kapanış tahminlerinde hata miktarı yüksek iken 20 günlük tahminde hata miktarı azalarak Polinomal Regresyon dan sonra en iyi sonucu vermiştir. Karar ağacı yapısında olan XG Booster, Facebook Prophet den sonra en kötü tahminde bulunmuştur.

Hisse senedinin artan bir trende sahip olmasından dolayı, Polinomal Regresyon da eğimsel bir artış veya azalış grafiği şekilde olduğu için bu modelde en iyi sonucu vermektedir. Arima ve LSTM bu çalışmada kullanılabilecek diğer yöntemlerdendir. İncelenen dönem aralığında hisse senedinin yüksek fiyat artışlara sahip olması ve tahmin edilmesi istenen artışın en üst kısımlarına denk gelmektedir. Çok sayıda dış parametre tarafından etkilenen borsa üzerine tahminde bulunmak çok zor bir çalışmadır.

Uzun vade için yapılan 1 yıllık tahmin çalışmasında yöntemleri en iyiden en kötüye doğru sıralandığı zaman, LSTM, Facebook Prophet, Polinomal Regresyon, Arima ve XG Booster şeklinde olmaktadır. Özellikle uzun vadeli eğitim çalışmasında tahmin edilmek istenen son 1 yıla ait kapanış değerleri, eğitim verilerine göre çok daha fazla hızlı yükselme trendine sahip olması yüzünden tüm yöntemler gerçek değerlerin altında tahminde bulunulmasına

neden olmuştur. Kısa ve uzun vadeli tahmin çalışmalarında aradaki gün sayı farkının çok olması MAE (ortalama mutlak hata) üzerinden yapılan değerlendirmelerde kendini belli etmiştir. Kısa vade de en kötü hata sonuçlarını veren Prophet 11 değerini alır iken uzun vadede en iyi sonucu veren LSTM de bile bu değer 10 civarlarındadır. Bu hata değerleri, borsada yatırım yapmak için yüksek hata değerler olup sadece hisse senedinin trendi hakkında bilgi verecek seviyededir.

Bu çalışma veri setinde hisse senedine ait açılış, en yüksek, en düşük, kapanış, ortalama kapanış, hacim bilgileri olmasına rağmen sadece kapanış değerleri kullanılmıştır. Gelecekte yapılan çalışmalarda en düşük, en yüksek ve hacim bilgileri de dikkate alınarak daha etkin yöntemler bulunabilir. Ayrıca verileri yükseldi, aynı kaldı, düştü şeklinde gruplayıp daha sonraki günlerdeki hisse senedi artış veya azalma yönündeki davranışlar tahmin edilebilir. Burada incelenen model Microsoft Şirketine ait hisse senedi değerleri iken, başka firmalara ait hisse senedi kapanış fiyatlarında, farklı modeller ile çalışma yapılabilir.

## KAYNAKLAR

- [1] **Yeşildağ, E., & Özen, E.** (2015). Uşak ilindeki hisse senedi yatırımcılarının profili ve yatırım kararlarını etkileyen demografik ve sosyo-ekonomik faktörlerin analizi. *Journal of Accounting, Finance and Auditing Studies*, 1(2), 78-102..
- [2] **Özdi, T., & YILMAZ, C.** (2006). İstanbul Menkul Kıymetler Borsası'nda (İMKB) sektör bazında işlem gören hisse senetlerinin alım-satım kararlarında en yüksek getirili stratejinin belirlenmesi. *Yönetim ve Ekonomi: Celal Bayar Üniversitesi İktisadi ve İdari* .
- [3] **Çetinyokuş, T., & Gökçen, H.** (2002). Borsada göstergelerle teknik analiz için bir karar destek sistemi. *Gazi Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi*, 17(1).
- [4] **Veli, A. K. E. L.** (2015). KIRILGAN BEŞLİ ÜLKELERİNİN HİSSE SENEDİ PİYASALARI ARASINDAKİ EŞBÜTÜNLEŞME ANALİZİ. *Uluslararası Yönetim İktisat ve İşletme Dergisi*, 11(24), 75-96..
- [5] CANBAZ, M. MENKUL KIYMETLER BORSASINDA HİSSE SENEDİ YATIRIMI NASIL YAPILIR?: İSTANBUL MENKUL KIYMETLER BORSASI ÖRNEĞİ. *AKADEMİK ÇALIŞMALAR-2019/2*, 40..
- [6] **Wei, W. W.** (2006). Time series analysis. In *The Oxford Handbook of Quantitative Methods in Psychology: Vol. 2.*
- [7] Yamak, N., & Topbaş, Ö. G. D. F. (2008). STOK YATIRIMLARI VE KONJONKTÜREL DALGALANMALAR. *Ulusal İktisat Kongresi, İzmir.*
- [8] **KIZILGÖL, Ö. A.** (2011). Mevsimsel eşbütünleşme testi: Türkiye'nin makroekonomik verileriyle bir uygulama. *Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 25(2), 13-25..
- [9] <https://qastack.info.tr/stats/5026/what-is-the-difference-between-data-mining-statistics-machine-learning-and-ai>.
- [10] <https://akademi40.org/veri-madenciligi>.
- [11] <https://www.dunyahalleri.com/makine-ogrenmesi-derin-ogrenmenin-populer-kullanim-alanlari/>.
- [12] <https://oyungezer.com.tr/haber/google-in-yapay-zekasi-alphazero-dunyanin-en-iyi-satranc-yapay-zekasinin-canina-okudu/detay>.
- [13] <https://www.chess.com/tr/news/view/alphazeronun-yeni-versiyonu-1-000-oyunluk-macta-stockfishi-ezip-gecti>.
- [14] <https://www.veribilimiokulu.com/blog/buyuk-veri-on-isleme-makale-notlari/>.
- [15] **Silahtaroğlu, G.** (2008). *Veri madenciliği*. Papatya Yayınları, İstanbul..
- [16] **Gülsevin, G., & Türkan, A. H.** (2012). Afyonkarahisar hastanelerinin

etkinliklerinin veri zarflama analizi ile değerlendirilmesi. Afyon Kocatepe Üniversitesi Fen ve Mühendislik Bilimleri Dergisi, 12(2), 1-8..

- [17] <https://www.oracle.com/tr/data-science/machine-learning/what-is-machine-learning/>.
- [18] <https://www.endustri40.com/makine-ogrenimi-nedir/>.
- [19] <https://www.emrealadag.com/makine-ogrenmesi-nedir.html>.
- [20] **ALİYEV, R., & TEKSÖZ, C.** Reinforcement Learning and Optimization Algorithm (Pekiştirmeli Öğrenme ve Optimizasyon Algoritmaları).
- [21] **Coşkun, C., & Baykal, A.** (2011). Veri madenciliğinde sınıflandırma algoritmalarının bir örnek üzerinde karşılaştırılması. Akademik Bilişim, 2011, 1-8..
- [22] **Çakmak, Z.** (1999). Kümeleme Analizinde Geçerlilik Problemi Ve Kümeleme Sonuçlarının Değerlendirmesi. Dumlupınar Üniversitesi Sosyal Bilimler Dergisi, (3)..
- [23] **Savaş, S., Topaloğlu, N., & Yılmaz, M.** (2012). Veri madenciliği ve Türkiye'deki uygulama örnekleri..
- [24] <https://medium.com/deep-learning-turkiye/peki%CC%87%C5%9Fti%CC%87rmeli%CC%87-%C3%B6%C4%9Frenmeye-gi%CC%87ri%CC%87%C5%9Fseri%CC%87si%CC%87-1-8f5c35b6044>, [Çevrimiçi].
- [25] <https://medium.com/@ozcangundes/peki%C5%9Ftirmeli-%C3%B6%C4%9Frenmeye-rl-giri%C5%9F-649a37fa491f>, [Çevrimiçi].
- [26] **Littman, M. L.** (1994). Markov games as a framework for multi-agent reinforcement learning. In Machine learning proceedings 1994 (pp. 157-163). Morgan Kaufmann..
- [27] <https://netuce.com/veri-madenciligi-nedir-hangi-alanlarda-kullanilir>, [Çevrimiçi].
- [28] <https://dralabay.wordpress.com/2014/03/07/veri-madenciligi-ve-isletmelerde-uygulama-alanlari/>, [Çevrimiçi].
- [29] <https://medium.com/@emre.yancizer/veri%CC%87-madenci%CC%87li%CC%87%C4%9Fi%CC%87-ve-t%C3%BCrki%CC%87yede-son-10-yildaki%CC%87-uygulama-%C3%B6rneklere%CC%87-592a8ccadd45>.
- [30] <https://www.gtech.com.tr/veri-madenciligi-nedir-ve-nasil-yapilir/>.
- [31] **ÖZBAY, F. A., & ALATAŞ, B.** Çevrimiçi sosyal medyada sahte haber tespiti. Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi, 11(1), 91-103., [Çevrimiçi].
- [32] **DOĞAN, B., Erol, B., & Buldu, A.** (2014). Sigortacılık sektöründe müşteri ilişkileri yönetimi için birliktelik kuralı kullanılması., [Çevrimiçi].
- [33] **Muslu, D.** (2009). Sigortacılık Sektöründe Risk Analizi: Veri Madenciliği Uygulaması (Doctoral dissertation, Fen Bilimleri Enstitüsü)., [Çevrimiçi].
- [34] <https://veribilimcisi.com/2017/07/18/polinomsal-regresyon-polynomial-regression/>.
- [35] <https://online.stat.psu.edu/stat462/node/158/>, [Çevrimiçi].
- [36] <https://yigitsener.medium.com/polinomsal-polynomial-regresyon-ve-python-uygulamas%C4%B1-f742fb61a158>.

- [37] <https://devhuntery.wordpress.com/2018/07/14/uzun-kisa-sureli-bellek-long-short-term-memory/>.
- [38] <https://globalaihub.com/keras-ile-rnn-ve-lstme-giris/>.
- [39] **KIZRAK, M. A., & BOLAT, B.** (2019). Uçak Motoru Sağlığı için Uzun-Kısa Süreli Bellek Yöntemi ile Öngörücü Bakım. *International Journal of InformaticsTechnologies*, 12(2).
- [40] **SÜZEN, A. A., YILDIZ, Z., & YILMAZ, T.** LSTM tabanlı Derin Sinir Ağı ile Ayak Taban Basınç Verilerinden VKİ Durumlarının Sınıflandırılması. *Bitlis Eren Üniversitesi Fen Bilimleri Dergisi*, 8(4), 1392-1398..
- [41] **Ahmadi, N., Constandinou, T. G., & Bouganis, C. S.** (2019, March). Decoding hand kinematics from local field potentials using long short-term memory (LSTM) network. In *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)* (pp. 415-419). I.
- [42] <https://www.intel.com.tr/content/www/tr/tr/products/docs/processors/cpu-vs-gpu.html>.
- [43] <http://canererden.com/2020/05/06/zaman-serisi-tahminleri-ve-arima-modelleri/>.
- [44] <https://ilkhanaslan.blogspot.com/2016/12/otoregresif-hareketli-ortalama-arima.html>.
- [45] <https://baturnecati.wordpress.com/2017/04/29/timeseries-forecasting-modelleri-arima-ets/>.
- [46] <https://www.hurriyet.com.tr/gundem/korelasyon-nedir-korelasyon-cesitleri-ve-ornekleri-41496009>.
- [47] <https://towardsdatascience.com/time-series-forecasting-arima-models-7f221e9eee06>.
- [48] [https://facebook.github.io/prophet/docs/quick\\_start.html](https://facebook.github.io/prophet/docs/quick_start.html).
- [49] <https://futuraice.com/blog/business-forecasting-with-facebook-prophet>.
- [50] <https://research.fb.com/blog/2017/02/prophet-forecasting-at-scale/>.
- [51] <https://bilisim.io/2020/01/04/xgboost-ile-merhaba-dunya/>.
- [52] <https://xgboost.readthedocs.io/en/latest/tutorials/model.html>.
- [53] <https://teknoloji.org/kaggle-yarismalarinin-en-populer-algoritmasi-xgboost/>.
- [54] veribilimiokulu. [Çevrimiçi]. Available: <https://www.veribilimiokulu.com/blog/xgboost-nasil-calisir/>.
- [55] Gültepe, Y. Makine Öğrenmesi Algoritmaları ile Hava Kirliliği Tahmini Üzerine Karşılaştırmalı Bir Değerlendirme. *Avrupa Bilim ve Teknoloji Dergisi*, (16), 8-15..
- [56] <https://veribilimcisi.com/2017/07/14/mse-rmse-mae-mape-metrikleri-nedir/>.
- [57] **Wang, W., & Xu, Z.** (2004). A heuristic training for support vector regression. *Neurocomputing*, 61, 259-275., [Çevrimiçi].
- [58] **Karasu, S., Altan, A., Saraç, Z., & Hacıoğlu, R.** (2018, May). Prediction of Bitcoin prices with machine learning methods using time series data. In *2018 26th Signal Processing and Communications Applications Conference*



(SIU) (pp. 1-4). IEEE.

- [59] **Khan, J., Wei, J. S., Ringner, M., Saal, L. H., Ladanyi, M., Westermann, F., ... & Meltzer, P. S.** (2001). Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nature medicine*, 7(6), 673-679.
- [60] **Karagül, K.** (2014). İstanbul Menkul Kıymetler Borsası'nda İşlem Gören Firmaların Destek Vektör Makineleri Kullanılarak Sınıflandırılması. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, 20(5), 174-178.
- [61] **ÖZÇOMAK, M. S., & Gündüz, M.** Borsa İstanbul'da İşlem Gören Şirketlerin Kapanış Fiyatları İle İşlem Miktarları Arasındaki İlişkinin Fonksiyonel Kanonik Korelasyon İle Analizi. *Uluslararası İktisadi ve İdari İncelemeler Dergisi*, (13), 233-252.
- [62] **PABUÇCU, H.** BORSA ENDEKSİ HAREKETLERİNİN MAKİNE ÖĞRENME ALGORİTMALARI İLE TAHMİNİ. *Uluslararası İktisadi ve İdari İncelemeler Dergisi*, (23), 179-190.
- [63] **BOYACIOĞLU, M. A., & Çürük, D.** (2016). Döviz kuru değişimlerinin hisse senedi getirisine etkisi: Borsa İstanbul 100 endeksi üzerine bir uygulama. *Muhasebe ve Finansman Dergisi*, (70), 143-156.
- [64] **Kutlu, B., & Badur, B.** (2009). YAPAY SİNİR AĞLARI İLE BORSA ENDEKSİ TAHMİNİ. *Yönetim Dergisi*, 20(63), 45-40.
- [65] **Değirmenci, N., & Akay, A.** (2017). Finansal Verilerin ARIMA ve ARCH Modelleriyle Öngörüsü: Türkiye Örneği. *Eskişehir Osmangazi Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 12(3), 15-36.
- [66] **Aydemir, Ö.** (2008). Beyin bilgisayar arayüzü uygulamalarına yönelik EEG işaretleri için öznelik çıkarma (Doctoral dissertation, Karadeniz Teknik Üniversitesi/Fen Bilimleri Enstitüsü)..
- [67] **Kalkavan, Y., & Özçakır, F.** LSTM DERİN ÖĞRENME YÖNTEMİ İLE ARIMA ZAMAN SERİLERİ YÖNTEMİNİN ARAÇ KİRALAMA SEKTÖR VERİLERİ ARACILIĞIYLA KARŞILAŞTIRILMASI.
- [68] **ALPAY, Ö.** LSTM Mimarisi Kullanarak USD/TRY Fiyat Tahmini. *Avrupa Bilim ve Teknoloji Dergisi*, 452-456..
- [69] **Sevli, O., & GÜLSOY, V. G. B.** (2020). Covid-19 Salgınına Yönelik Zaman Serisi Verileri ile Prophet Model Kullanarak Makine Öğrenmesi Temelli Vaka Tahminlemesi. *Avrupa Bilim ve Teknoloji Dergisi*, (19), 827-835.
- [70] <https://medium.com/@ahmetkuzubasli/neden-ilk-adresiniz-xgboost-75010416a583>.

## ÖZGEÇMİŞ

**Ad-Soyad :**

**Doğum Tarihi ve Yeri:** 15.01.1979 / Afyonkarahisar

**E-posta :** salim\_demiray@hotmail.com,salimdemiray@stu.aydin.edu.tr

### Öğrenim Durumu

**Yüksek Lisans:** : İstanbul Aydın Üniversitesi  
: **Bilgisayar** 2017-2021

**Lisans:** : **Kocaeli Üniversitesi**  
Bilgisayar Öğretmenliği 1997-2001

**Lise:** : **Recep Gençer ATL Bilgisayar Bölümü**  
1993-1997

### Mesleki Deneyim

**Uranium Bilgisayar :** **Yazılım Geliştirme Uzmanı**  
Ağustos 2001-Haziran 2012

**Uyumsoft :** **Kıdemli Yazılım Geliştirme Uzmanı**  
Haziran 2012-