

**GENELLEŐTİRİLMİŐ TOPLAMSAL MODELLER İLE  
BITCOIN İÇİN YÖN ANALİZİ**

**DIRECTIONAL ANALYSIS OF BITCOIN USING  
GENERALIZED ADDITIVE MODELS**

**İLAYDA ARIKAN**

**PROF. DR SERPİL AKTAŐ ALTUNAY**

**Tez DanıŐmanı**

Hacettepe Üniversitesi

Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğinin

İstatistik Anabilim Dalı için Öngördüğü

YÜKSEK LİSANS TEZİ olarak hazırlanmıştır.

2024

## ÖZET

# GENELLEŐTİRİLMİŐ TOPLAMSAL MODELLER İLE BITCOIN İÇİN YÖN ANALİZİ

İlayda ARIKAN

Yüksek Lisans, İstatistik Bölümü

Tez DanıŐmanı: Prof. Dr. Serpil AKTAŐ ALTUNAY

Nisan 2024, 75 sayfa

Küresel ekonomik belirsizlikler ve enflasyon endiŐelerinin artmasıyla birlikte geleneksel banka sistemlerinin merkezi rolü sorgulanmaya başlamıŐtır. Bu durum, kullanıcıların finansal özgürlüğünü kısıtlanmasına neden olmaktadır. Bu bağlamda, Bitcoin ve diđer kripto paraların ortaya çıkmasıyla finansal dünyada büyük bir deęişim yaşandı. Bitcoin'in merkezi olmayan yapısı, bankalara olan bağımlılıęı azaltmış ve kullanıcılara doğrudan ve güvenli bir şekilde finansal işlemler yapma imkânı tanımıŐtır. Microsoft, Tesla, Visa, AXA gibi birçok şirket ve kurumun Bitcoin'i ödeme aracı olarak kabul etmeye başlaması, kullanıcıların günlük alışverişlerinde Bitcoin kullanılabilirliğini artırmıŐtır. Bu durum, Bitcoin'in popülaritesini artırarak finansal dünyada yeni bir dönemin başlatmıŐtır. Popüler olmasına rağmen, Bitcoin gibi kripto paralardaki fiyatlar son derece dalgalıdır ve bu durum fiyat hareketlerini öngörmeyi zorlaŐtırmaktadır. Bu durumda, Bitcoin'in fiyat yönünün tahmini, yatırımcıların stratejik kararlarını belirlemelerinde kritik bir rol oynamaktadır. Yatırımcılar genellikle temel analiz, haber takibi veya teknik analiz gibi yöntemlerden birini kullanarak yatırım yapmayı tercih etmektedirler. Bu bağlamda, tez çalışması, temel analiz, haber takibi ve teknik analiz gibi yöntemleri birleŐtirerek kripto para piyasasında doğru tahminler yapmayı hedeflemektedir.

Bu tez çalışması, kripto para birimi piyasasında Bitcoin fiyat hareketlerinin çok faktörlü etkisini incelemektedir. Finansal piyasalarda fiyat yönünün tahmininde istatistiksel modellerin güvenilirliği temel bir avantaj sağlarken, veri uyumu açısından makine öğrenimi modellerinin tercih edilmesi amaçlanmaktadır. Bu çalışma, Bitcoin ve diğer kripto para piyasalarının fiyat dinamiklerine yönelik güçlü bir model oluşturarak gelecekteki araştırmalara temel oluşturmayı hedeflemektedir. Bu amaçla, çalışmada trend yönünün değiştiği noktalara odaklanılmıştır. Fiyat yönünün hangi noktada değişeceği tespit edilerek, ‘artacak’, ve ‘azalacak’ olmak üzere iki sınıflı tahminleme çalışması yapılmıştır. Analizlerimizde, 17 Ağustos 2017 ile 3 Ocak 2024 tarihleri arasında 'Binance' API'si kullanılarak günlük olarak toplanan Bitcoin verileri üzerinde çalışıldı. Bu verilerle fiyatın kapanışı, günün en yüksek değeri, Teklif Varlık Hacmi, İşlem Sayısı, Bitcoin ticaret hacmi, Teklif Temel Hacmi ve Teklif Hacmi gibi çeşitli bilgiler incelenerek trend tespiti yapıldı. Ayrıca, diğer API kaynaklarından elde edilen verilerle ekonomik-finansal göstergeler de değerlendirildi. Bitcoin piyasasını etkileyen 21 değişken kullanılarak hedef değişkeni olan Bitcoin yönü ile ilgili tahminler elde edilmiştir. Toplanan Bitcoin verileri üzerinde çeşitli istatistiksel ve makine öğrenimi modelleri kullanarak gerçekleştirilmiştir. Zaman serisi modeli olduğu için zaman serisine uygun olacak şekilde eğitim ve test verileri 5 katlı cross validation gerçekleştirildi. Modelleme aşamasında Genelleştirilmiş Toplamsal Modeller (LOESS Lojistik Regresyon Modeli, Eğri Düzleştirmeli Lojistik Regresyon, Lojistik regresyon) ile Makine Öğrenmesi algoritmalarından Karar Ağaçları, Destek Vektör Maineleri ve Gradient Boosting yöntemleri uygulanmıştır. Sonuçlar, kesinlik, duyarlılık, seçicilik, F1 puanı ve doğruluk oranı gibi başarı kriterlerine göre değerlendirilmiştir. En iyi model olarak Eğri Düzleştirmeli Lojistik Regresyon modeli bulunmuştur. En iyi model yardımı ile yakın tarihteki Bitcoin fiyatları üzerinden artış ve azalış tahminleri verilmiştir. Bu bulgular, yatırımcı için finansal teknolojilerin ve yatırım stratejilerinin geliştirilmesinde kullanılabilecek tahminler sunmaktadır.

Kripto para piyasasındaki fiyat dinamiklerini analiz etmek, gelecekteki çabalar için önemli bir adımdır. Kripto para fiyat trendlerinin detaylı incelenmesi, tahmin modellerinin etkinliğini ve piyasa anlayışını önemli derecede iyileştirebilir. Dinamik kripto para piyasaları, finansal analiz ve yatırım stratejilerinde model tahminlerinden faydalanmayı sağlar. Ancak, bu tahminlerin finansal karar alma sürecinde dikkatle

değerlendirilmesi ve piyasa riskleri ile model sınırlılıklarının göz önünde bulundurulması önemli olmaktadır.

**Anahtar Kelimeler:** Bitcoin, Genelleştirilmiş Toplamsal Modeller, Makine Öğrenimi Modelleri, Parametrik Olmayan Lojistik Regresyonlar, Piyasa Davranışları, Yön Analizi, Zaman serisi analizi

# **ABSTRACT**

## **DIRECTIONAL ANALYSIS OF BITCOIN USING GENERALIZED ADDITIVE MODELS**

**İlayda ARIKAN**

**Master's Degree, Department of Statistics**

**Supervisor: Prof. Dr. Serpil AKTAŞ ALTUNAY**

**April 2024, 75 pages**

Global economic uncertainties and rising inflation concerns have challenged the central role of traditional banking systems, limiting users' financial freedom. The emergence of Bitcoin and other cryptocurrencies has revolutionized the financial world. Bitcoin's decentralized structure has reduced reliance on banks and enabled direct and secure financial transactions. Companies like Microsoft, Tesla, Visa, and AXA have begun accepting Bitcoin as a payment method, enhancing its daily use and boosting its popularity. Despite its popularity, the volatility of cryptocurrencies like Bitcoin complicates price predictions. Predicting Bitcoin's price direction is crucial for investors who often rely on fundamental analysis, news tracking, or technical analysis.

This thesis explores the multifaceted impact of Bitcoin price movements in the cryptocurrency market. It leverages statistical models known for their reliability in financial markets and prefers machine learning models for their data compatibility. The study aims to develop a robust model for the dynamics of Bitcoin and other cryptocurrencies, serving as a foundation for future research. It focuses on identifying trend changes and conducting binary predictions on whether prices will rise or fall. The analysis uses Bitcoin data collected daily from the 'Binance' API between August 17,

2017, and January 3, 2024, covering metrics like closing prices, daily highs, and trading volumes. Additionally, economic and financial indicators from other APIs have been integrated.

The study employs time series-specific five-fold cross-validation for training and testing data. It utilizes Generalized Additive Models (including LOESS Logistic Regression and Smoothed Logistic Regression) and machine learning algorithms like Decision Trees, Support Vector Machines, and Gradient Boosting. The models were evaluated based on precision, sensitivity, specificity, F1 score, and accuracy rate, with the Smoothed Logistic Regression model performing best. This model has provided predictions for recent Bitcoin price trends, offering valuable insights for developing financial technologies and investment strategies.

Analyzing the price dynamics in the cryptocurrency market is crucial for enhancing the effectiveness of prediction models and deepening market understanding. The dynamic nature of cryptocurrency markets makes model predictions useful for financial analysis and investment strategies. However, these predictions should be meticulously evaluated in the decision-making process, considering market risks and the limitations of the models.

**Keywords:** Bitcoin, Direction Analysis, Generalized Additive Models, Machine Learning Models, Market Behaviors, Non-Parametric Logistic Regressions, Time Series Analysis

## TEŐEKKÜR

Tez alıőmam boyunca bana her trl destek ve rehberlik sunan, đrencilerine daima pozitif enerji vererek onları gl hissettiren; bilgi birikimiyle alıőmama farklı perspektifler kazandıran ve đrencisi olmaktan srekli bir gurur duyduđum, deđerli akademik danıőmanım Prof. Dr. Serpil AKTAŐ ALTUNAY'a őukranlarımı sunarım. İstatistik blmnde grev yapan tm đretim yelerine ve alıőanlara, akademik yolculuđumda bana verdikleri destekler iin teőekkr ederim. Tez jrimde yer alarak, tezimin daha iyi bir seviyeye ulaőması iin kıymetli grőleri ve yapıcı eleőtirileriyle katkıda bulunan hocalarım Prof. Dr. Meral ETİN ve Do. Dr. Serdar NESLİHANOĐLU' na derin minnettarlıđımı ifade etmek isterim.

Ayrıca, her zaman yanımda olan ve desteđini hibir zaman esirgemeyen sevgili aileme en iten teőekkrlerimi sunarım. Onların sonsuz sevgisi, anlayıőı ve sabrı, bu tezin tamamlanmasında bana g verdi.

Bu srete bana ilham veren ve alıőmalarımı zenginleőtiren tm arkadaşlarıma, meslektaőlarıma ve iő birliđi iinde olduđumuz kurumlara da minnettarlıđımı ifade etmek isterim. Her birinin katkısı, bu alıőmanın daha kapsamlı ve deđerli hale gelmesine yardımcı oldu.

Son olarak, tez srecim boyunca karőılaőtıđım zorluklarda yanımda olan ve bana destek olan herkese teőekkr ederim. Bu tez, sadece bir akademik aba deđer, aynı zamanda birok deđerli insanın desteđi ve yardımıyla őekillenen bir yolculuktur.

# İÇİNDEKİLER

ÖZET .....	i
ABSTRACT.....	iv
İÇİNDEKİLER .....	vii
ŞEKİLLER DİZİNİ .....	ix
ÇİZELGELER DİZİNİ .....	x
SİMGELER VE KISALTMALAR .....	xi
1. GİRİŞ .....	1
2. KRİPTO PARA BİRİMLERİ VE BITCOİN ÜZERİNE KAPSAMLI GİRİŞ ....	3
2.1. Bitcoin ve Kripto Para Birimleri Hakkında Genel Bakış.....	3
2.1.1. Kripto Para ve Bitcoin'in Tarihi Gelişimi.....	3
2.1.2. Kripto Para Piyasalarının Yapısı ve İşleyişi .....	4
2.1.3. Bitcoin'in Finansal Sistemlerdeki Yeri ve Önemi.....	5
3. İSTATİSTİKSEL VE MAKİNE ÖĞRENMESİ MODELLERİ .....	8
2.2. Lojistik Regresyon .....	8
3.2. Genelleştirilmiş Toplamsal Modeller.....	10
3.2.1. Eğri Düzleştirmeli Lojistik Regresyon .....	12
3.2.2. LOESS Lojistik Regresyon .....	15
3.3. Makine Öğrenmesi Modelleri .....	20
3.3.1. Makine Öğrenmesinin Tanımı ve Kategorileri .....	21
3.3.2. Karar Ağaçları (Decision Tree) .....	23
3.3.3. Topluluk Öğrenmesi .....	28
3.3.4. Destek Vektör Makineleri (SVM) .....	35
3.4. Sınıflandırma Performans Ölçütleri .....	38
4. UYGULAMA .....	41
4.1. Veri Kümesi .....	41
4.2. Veri Kaynakları.....	42
4.3. Toplama Süreci .....	45
4.4. Veri Temizleme .....	45
4.5. Veri Dönüştürme.....	46
4.6. Analiz Sonuçları .....	46
4.7. Model Eğitimi ve Test Süreçleri.....	50
4.8. Model Sonuçları.....	51
4.9. İleriye Dönük Sınıflandırma Tahmin Sonuçları.....	61
5. SONUÇ VE TARTIŞMA .....	63
KAYNAKLAR .....	65



EK 1 – Etik Kurul İzin Belgesi .....	73
EK 2 - Tez Çalışması Orjinallik Raporu .....	74
ÖZGEÇMİŞ .....	75

## ŞEKİLLER DİZİNİ

Şekil 3.1. Yaygın kullanılan Makine Öğrenmesi modelleri .....	21
Şekil 3.2. İris veri setinin indirgenmiş bir versiyonu üzerine uygulanan bir sınıflandırma ağacı .....	24
Şekil 3.3. Yaygın bir Topluluk Öğrenmesi yapısı .....	29
Şekil 3.4. Yanlılık ve varyans hatası dengesi .....	30
Şekil 3.5. Topluluk makine öğrenme yaklaşımlarının genel çerçevesi .....	31
Şekil 3.6. Boosting akış diyagramı .....	32
Şekil 3.7. Boosting aşamaları .....	33
Şekil 3.8. Bagging ve Boosting işleyiş mekanizması .....	33
Şekil 3.9. Gradient Boosting aşamaları .....	34
Şekil 3.10. Doğrusal SVM modeli .....	36
Şekil 4.11. En yüksek Bitcoin fiyatının zamana göre değişimi .....	47
Şekil 4.12. En yüksek Bitcoin fiyat dağılımı .....	49
Şekil 4.13. Değişkenler arasındaki ilişkiye ait ısı grafiği .....	50
Şekil 4.14. Modelde kullanılan anlamlı değişkenlerin düzeltme katsayılarına ait güven aralıkları .....	58
Şekil 4.15. Karışıklık matrisi .....	59
Şekil 4.16. Eğri Düzleştirmeli Lojistik Regresyon modeli için ROC eğrisi .....	60
Şekil 4.17. İleri dönük tahmin değerleri için sütun grafiği .....	62

## ÇİZELGELER DİZİNİ

Çizelge 3.1. Karar ağaçlarının avantaj ve dezavantajları .....	24
Çizelge 4.3. Eğitim verisi performans metrikleri ve sonuçları .....	52
Çizelge 4.4. Test verisi performans metrikleri ve sonuçları .....	54
Çizelge 4.5. Modelde kullanılan değişkenlerin düzeltme katsayıları ve anlamlılık düzeyleri.....	57
Çizelge 4.6. İleri dönük tahmin değerleri.....	61

# SİMGELER VE KISALTMALAR

## Kısaltmalar

GLM	Genelleştirilmiş Doğrusal Model
GAM	Genelleştirilmiş Toplamsal Modeller
LOESS	Yerel Ağırlıklı Saçılma Grafiği Düzleştirme
ROC	Alıcı İşlem Karakteristikleri
AUC	Eğri Altındaki Alan
API	Uygulama Programlama Arabirimi

# 1. GİRİŞ

Veri biliminin gelişimi ve büyük veri analitiğindeki ilerlemeler, finansal piyasalarda paradigma değişikliğine yol açtı. Özellikle kripto para birimlerinin ortaya çıkışı ve Bitcoin'in artan popülaritesi, geleneksel finansal piyasaların işleyişine yeni bir boyut katmıştır. Bitcoin piyasasının doğasında var olan yüksek volatilité, yatırımcıları ve analistleri yeni tahmin yöntemleri bulmaya zorunlu kıldı [1].

Kripto para araştırması hızla büyüyen bir alan haline geldi ve mevcut literatür özellikle fiyatın yönünü tahmin etmeye odaklanmaktadır. Ancak sınıflandırmaya dayalı çalışmaların sayısı, regresyona dayalı çalışmalara dayalı çalışmaların sayısı oldukça sınırlıdır ve genellikle bir sonraki dönemin tahmin etmekle sınırlıdır. Güncel araştırmalarda evrişimli sinir ağları (CNN) ve yinelenen sinir ağları (RNN) gibi derin öğrenme modelleri ve pekiştirmeli öğrenme yöntemleri yaygın olarak kullanılmaktadır [2].

Düzeltilme (smoothing) bugün istatistikte saygın bir yeri olduğunda şüphe yoktur. Veri bilimine önemli katkı sağlayan kişiler tarafından birçok makale ve birkaç kitap yayımlandı [3,4,5,6,7]. Gerçek hayat veri setlerinin zengin ve karmaşık yapısı, parametrik modellerle tam olarak modellenemediği için parametrik olmayan modeller giderek daha popüler hale gelmiştir.

Bu tezin amacı, genelleştirilmiş lojistik modellerin, Bitcoin fiyat hareketlerinin yönünü tahmin etmedeki uygulamasını ve etkinliğini incelemektir. Önceki çalışmalardan farklı olarak bu model, mikro ve makro düzeydeki finansal göstergeler ile sosyo-ekonomik verileri birleştirerek Bitcoin fiyat hareketlerinin daha geniş bir bağlamda değerlendirilmesine olanak tanımaktadır. Bu yaklaşımın başarısı, tarihsel veri setleri ve çeşitli finansal performans göstergeleri kullanılarak ayrıntılı olarak değerlendirilecek ve Bitcoin piyasasının yeni ve kapsamlı bir yön analizi sunulacaktır. Bu yaklaşım, kripto para birimi fiyat dinamiklerini daha iyi anlamamıza ve finansal piyasa karar alma sürecine katkıda bulunmamıza yardımcı olacak genel bir çerçeve sağlayacaktır.

Bu çalışmada, Genelleştirilmiş Toplamsal Modeller (LOESS Lojistik Regresyon Modeli, Eğri Düzleştirmeli Lojistik Regresyon, Lojistik Regresyon) ile birlikte Karar Ağaçları,

Destek Vektör Makineleri ve Gradient Boosting gibi Makine Öğrenmesi algoritmaları kullanılmaktadır. Genelleştirilmiş Toplamsal Modeller, veri içindeki yerel varyasyonları efektif bir şekilde yakalayarak modellemenin esnekliğini artırmakta, bu da Bitcoin piyasasının karmaşık yapısını detaylı bir şekilde analiz etme olanağı sunmaktadır. Eğri düzleştirme yöntemi, risk fonksiyonlarının doğrusal olmayan ve düzensiz yapılarını başarılı bir şekilde modellemek için kullanılmıştır. Ayrıca, ikili lojistik regresyon modeli, veri setinin özelliklerini basit ve anlaşılır bir biçimde yansıtacak şekilde tasarlanmıştır, bu sayede model yorumları kolaylaştırılmıştır. LOESS (Locally Estimated Scatterplot Smoothing), modelin doğrusal olmayan ve karmaşık eğilimlerini esnek bir şekilde işlemesine imkân tanırken, veri setinin lokal özelliklerine uyum sağlamakta ve kullanıcı tarafından ayarlanabilen parametreler aracılığıyla görsel olarak anlaşılır sonuçlar üretmektedir. Karar Ağaçları, modelin karar verme süreçlerini görselleştirerek karmaşık veri yapılarını açıklar ve Destek Vektör Makineleri yüksek boyutlu uzaylarda bile etkili sınıflandırma sağlar. Gradient Boosting ise zayıf öğrencileri güçlü tahmin modellerine dönüştürerek genel model performansını artırmaktadır.

Bu çalışma kapsamında, bir Bitcoin veri seti üzerinde bu modellerin nasıl uygulandığı incelenecek, elde edilen istatistiksel açıklamalar ve tahminlerin doğruluğu değerlendirilecektir. Bu analizler, Bitcoin piyasasının daha iyi anlaşılmasına katkıda bulunacak ve ileri düzey finansal analizler için metodolojik bir çerçeve sunacaktır.

## 2. KRİPTO PARA BİRİMLERİ VE BITCOİN ÜZERİNE KAPSAMLI GİRİŞ

### 2.1. Bitcoin ve Kripto Para Birimleri Hakkında Genel Bakış

#### 2.1.1. Kripto Para ve Bitcoin'in Tarihi Gelişimi

Kripto para birimlerinin kökeni, 2008 yılı sonunda "Satoshi Nakamoto" takma adını kullanan bir kişi veya grubun "Bitcoin: Eşler Arası Elektronik Nakit Sistemi" adlı bir makaleye dayanmaktadır. Bu makale, Bitcoin'e ihtiyaç duymadan elektronik işlemlerin gerçekleşmesini sağlayacak bir sistem tasarladı ve bu vizyonun hayata geçirilmesini sağlamıştır. Resmi olarak 2009 yılında, ilk bloğun madenciliği ile başladı [8].

Bitcoin, geleneksel finansal sisteme alternatif olarak ortaya çıkmış bir kripto para birimi kavramıdır. Yeni para birimi, işlemlerin blockchain adı verilen merkezi olmayan bir kayıt sistemi teknolojisi kullanılarak doğrulanmasına ve kaydedilmesine olanak tanımaktadır. Blockchain teknolojisi şeffaflığı, değişmezliği ve sansüre karşı dayanıklılığı nedeniyle finansal işlemlerin güvenliğini ve güvenilirliğini sağlamayı mümkün kılmaktadır [9].

Bitcoin'in yanı sıra her biri farklı işlev ve hedeflere sahip birçok kripto para birimi geliştirildi. Ancak Bitcoin piyasa değeri, kullanıcı sayısı ve kabul edilebilirlik açısından liderliğini sürdürmektedir. Bitcoin'in benimsenmesi ve yayılması zaman içinde inişli çıkışlı bir süreç izlemiştir. İlk günlerinde yalnızca küçük bir teknoloji meraklısı grup tarafından kullanılan Bitcoin, artık spekülasyon bir yatırım aracı ve alternatif ödeme yöntemi olarak geniş çapta bilinmeye başladı.

Akademik literatürde Bitcoin teknolojik bir yenilik ve ekonomik bir olgu olarak analiz edilmiştir. Özellikle küresel finansal krizin ardından Bitcoin'in ortaya çıkışı, insanların merkezi finansal sisteme olan güvensizliğini ve alternatif finansal sistem arayışını yansıtmaktadır. Bitcoin'in gelişimi, finansal piyasaların işleyişinden para politikasına, ekonomik özgürlükten dijital varlıkların düzenlenmesine kadar pek çok alanda yeni tartışmaların ve araştırmaların önünü açmıştır.

Bu bağlamda Bitcoin ve genel olarak kripto para birimlerinin tarihsel gelişimi, dijital çağda ekonomik inovasyonun ve teknolojik ilerlemenin sembolü olarak görülmektedir. Bu gelişme, sosyal ve ekonomik yapıların yanı sıra geleneksel finansal sistemler üzerinde de derin bir etki yarattı ve bu alanda devam eden akademik araştırmaların temelini oluşturdu.

### **2.1.2. Kripto Para Piyasalarının Yapısı ve İşleyişi**

Geleneksel finansal piyasalardan farklı olarak kripto para piyasası merkezi olmayan bir yapıya sahiptir. Bu piyasalarda işlem gören dijital varlıklar, dünya çapındaki çeşitli kripto para borsaları aracılığıyla alınıp satılmaktadır. Blockchain teknolojisi, kripto para piyasasının temelini oluşturarak işlemlerin şeffaf, güvenli ve manipüle edilmesi zor olmasını sağlar [8].

Kripto para piyasaları arz ve talep mekanizmaları tarafından kontrol edilmektedir. Bu piyasalarda kripto para birim fiyatlarının sürekli değişmesinin nedeni, yatırımcıların algıları, piyasa duyarlılığı, teknolojik gelişmeler ve makroekonomik faktörler gibi birçok değişkene bağlı olduğudur [10].

Madencilik, bir blockchain ağındaki işlemlerin doğrulanması ve sisteme yeni kripto para biriminin eklenmesi sürecidir. Ağın güvenliğini sağlamak ve merkezi olmayan yapısını korumak için madencilik süreci gereklidir. Coin arzı, bir kripto para biriminin toplam arzı veya maksimum arzı ile belirlenir. Bu arz, piyasa değeri için önemli bir etkiye sahiptir [3].

Kripto para piyasaları, yüksek oynaklığa sahiptir ve bu, yatırımcılar için hem büyük kazanç fırsatları hem de önemli yüksek riskler sunmaktadır [11]. Piyasalar dünyadaki gelişmelerden hızla etkilenebilmekte, ani ve büyük fiyat hareketlerine neden olabilmektedir. Kripto para piyasalarındaki işlem hacmi, piyasa likiditesi ve yatırımcı ilgisinin bir göstergesidir. Bu hacim, farklı zamanlarda önemli ölçüde değişiklik gösterebilmektedir.

Sonuç olarak, kripto para piyasalarının yapısı ve işlevselliği geleneksel piyasalardan önemli ölçüde farklılık göstermektedir. Bu piyasalar, teknolojik gelişmeler ve kullanıcıların benimseme oranlarından doğrudan etkilenir. Kripto para piyasası, finans



pazarının geleceğini şekillendiren dinamik ve yenilikçi bir alan olmaya devam etmektedir. Bu yapısal ve işlevsel özellikler, kripto para piyasalarını analiz ederken dikkate alınması gereken temel faktörlerden biridir.

### **2.1.3. Bitcoin'in Finansal Sistemlerdeki Yeri ve Önemi**

Bitcoin'in finansal sistemler içindeki durumu, özellikle 21. yüzyılın ilk çeyreğinde, para sistemindeki teknolojik ve ekonomik yaklaşımlarının para biriminin yeniden şekillendirilmesinde önemli bir rol oynamaktadır. Merkeziyetsiz bir dijital para birimi olarak Bitcoin, finansal işlemlerin inşası ve yürütülmesine olanak sağlamamaktadır. Bu nedenle, yenilikçi değişiklikler getirerek geleneksel bankacılık ve aracılık sistemlerine alternatif çözüm modelleri sunmaktadır [12].

### **Finansal İşlemlerde Yenilikçi Yaklaşımlar**

Blockchain teknolojisiyle Bitcoin, işlemlerin doğrulaması ve kaydedilmesi için yeni standartlar belirlenmiştir. Bu teknoloji, işlemlerin değişmez bir şekilde kaydedilmesini mümkün kılar. Böylece dünyada şeffaflığın ve güvenin artmasına yardımcı olmaktadır. Bitcoin'in bu özelliği, özellikle siber güvenlik, veri doğruluğu, tutarlılığı, güvenilirliği ve bütünselliği açısından finansal sistemler için büyük bir gelişme olarak görülmektedir [13].

### **Değer Saklama Aracı Olarak Bitcoin**

Bitcoin, sınırlı arzı ve küresel kabul görmesi nedeniyle, modern zamanların "dijital altını" olarak adlandırılmaktadır [14]. Yeni değer aracı olarak kullanılmasının nedeni, özellikle para birimlerinin değer kaybettiği ve ekonomik belirsizliklerin yaşandığı dönemlerde, Bitcoin'e olan talebi artırmaktadır [14, 15]. Bitcoin'in bu yönü, finansal piyasalarda yeni bir varlık sınıfının ortaya çıkmasına yol açmış ve yatırımcıların portföy çeşitlendirme stratejilerinde önemli bir yer tutmasını sağlamıştır.

## **Bitcoin ve Finansal Kapsayıcılık**

Bitcoin ve kripto para birimlerini, bankacılık hizmetlerine erişim eksikliği yaşayan veya sınırlı olan olması nedeniyle bireyler için finansal işlemler yapma fırsatı sunmaktadır. Bu durum, özellikle gelişmekte olan ekonomilerde finansal hizmetlerin entegre edilmesinde ve ekonomik kalkınmanın desteklenmesinde önemli bir rol oynamaktadır.

## **Yasal Düzenleme ve Kurumsal Kabul**

Bitcoin'in finansal sistemindeki yeri, dünya çapında çeşitli düzenleme ve kurumsal kabul biçimleriyle ortaya çıkmaktadır. Bazı ülkeler Bitcoin ve kripto para birimlerinin yasal ödeme aracı olduğunu tanıırken [16], diğerleri bu yeni varlık sınıfı konusunda daha temkinli yaklaşmaktadır. Bitcoin'in gelecekteki rolü, büyük ölçüde bu düzenlemelere ve finansal kurumlara uyma becerisine bağlı olarak şekillenecektir [17].

Bitcoin'in para sistemindeki yeri ve önemi, bu sektörü yeniden şekillendirme potansiyeli vermektedir. Bitcoin yalnızca teknolojik yenilikler yaratmakla kalmıyor, aynı zamanda mevcut ekonomik, sosyal ve politik sistemlere derinlemesine de etkilemektedir. Dolayısıyla Bitcoin ve benzeri kripto para birimlerinin finansal sistemde rol oynamaya devam etmesi akademik araştırmaların ve politika yapıcıların dikkat etmesi gereken bir konu olmaya devam edecektir.

## **Bitcoin'in Olumsuz Yönleri: Riskler ve Sınırlamalar**

Bitcoin ve genel olarak sanal para birimleri, fintech'in (finansal teknolojinin) yenilikçi bir unsuru haline gelmektedir. Ancak bu yenilikçi yapının, kendine özgü riskleri de vardır. Bu risklerden birincisi, işlemin geri döndürülemezliği. Bitcoin için temel bir özelliğidir. Bu nedenle gerçekleştirilen herhangi bir işlemin hatalı olsa bile geri alınamayacağı anlamına gelmektedir. Merkezi olmayan bir yapıya sahip olan Bitcoin kullanıcılarının, yanlışlıkla yanlış veya bilinmeyen bir alıcıya Bitcoin göndermeleri durumunda, bu işlemleri düzeltmemeleri riskini beraberinde getirmektedir [18].

Siber güvenlik riskleri, Bitcoin ile ilişkili önemli bir diğer endişe kaynağı olmaktadır. Bitcoin cüzdanlarının, siber saldırılar ve hackleme eylemlerine karşı kırılgan olması, bu

riskleri öne çıkarır. Kullanıcıların dijital varlıklarına erişimi sağlayan özel anahtarların yetkisiz kişiler tarafından ele geçirilmesi, Bitcoin'lerin kaybolması veya çalınması gibi ciddi güvenlik ihlallerine yol açabilir. Bu durum, kullanıcıların finansal varlıklarının yanı sıra dijital kimliklerinin de tehlikeye atılmasına neden olabilmektedir [19].

Düzenleyiciler belirsizlikleri, Bitcoin ve diğer sanal para birimlerinin yasal çerçevesi henüz tam olarak oluşturulmaması nedeniyle kaynaklanmaktadır. Bu durum hukuki belirsizliğe ve olası hukuki değişikliklere yol açmaktadır [20].

Bitcoin piyasasının yüksek oynaklığı, yatırımcılar için büyük fiyat dalgalanmaları riski oluşturmakta ve bu da yatırımcıların finansal karar almalarını zorlaştırabilmektedir [21]. Likidite riski (nakde çevrilebilirlik riski) de dikkate alınmalıdır; Bitcoin piyasasında bazen yeterli alıcı veya satıcı bulunmayabileceğinden bu durum varlığın likidite dönüşümünü sınırlandırır [22]. Yasal riskler, Bitcoin'in yasa dışı faaliyetlerle ilişkilendirilmesiyle ilişkilidir ve bu da kullanıcıların yasal soruşturma ile karşı karşıya kalma riskini artırmaktadır [23]. Teknik riskler, Bitcoin protokolündeki olası hatalardan veya güvenlik açıklarından kaynaklanır ve bu durum kullanıcıların varlıklarını riske atabilir [24]. Tüketici korumasının olmaması, Bitcoin'in merkezi veya dahili bir otorite tarafından yönetilmediğinden kullanıcıların işlem sorunlarıyla karşılaştıklarında başvurabilecekleri bir yer olmamasından kaynaklanmaktadır [25]. Rekabet ve yenilik riski, piyasaya giren yeni kripto para birimlerinin Bitcoin'in değerini artırması ve olumsuz sonuçlara yol açmasına neden olmaktadır [26].

Çevresel etki açısından, Bitcoin madenciliği, büyük miktarlarda elektrik tüketen, enerji yoğun bir süreçtir [27]. Bu yüksek enerji tüketimi, özellikle fosil yakıtların kullanımı olmak üzere karbon emisyonlarının artmasına neden olmakta, çevresel sürdürülebilirlik endişelerini artırmaktadır ve iklim değişikliğiyle mücadele çabalarıyla çelişebilmektedir [28]. Bu durumda, Bitcoin ve diğer sanal para birimleriyle ilişkili risklerin ve bunların çevresel sürdürülebilirlikle tutarlılığının akademik ve düzenleyici açıdan kapsamlı bir şekilde incelenmesi, bu yeni finansal ekosistemin sağlıklı gelişimini sağlamak için çok önemlidir [29]. Bu şartlar altında, daha verimli madencilik teknolojilerinin geliştirilmesi ve yenilenebilir enerji kullanımının teşvik edilmesi gibi stratejilerle çözülebilir.

### 3. İSTATİSTİKSEL VE MAKİNE ÖĞRENMESİ MODELLERİ

İstatistiksel modeller ve Makine öğrenimi algoritmaları, veri biliminde önemli bir rol oynamaktadır. Veriden bilgi ve öngörüler elde etmek için temel yöntemlerdendir. Bu yöntemler, karmaşık veri kümelerini anlamamıza ve karar almamız için bilimsel bir temel sağlamamıza olanak tanıyan analitiğin temel taşlarından biridir.

Bu yöntemler finansal analizden sağlık bilimlerine, perakende sektöründen eğitime kadar geniş bir alanda kullanılmaktadır. Tüketici davranışını analiz edilmesi, hastalıkların teşhisine, pazar eğilimlerinin izlenmesi ve özellikle Bitcoin gibi kripto para birimlerinin fiyat hareketlerini tahmin edilmesi gibi alanlarda yaygın olarak kullanılmaktadır.

Bu bölümde, tez çalışmasında kullanılan İstatistiksel modeller ve Makine Öğrenmesi algoritmaları tanıtılmıştır. Genelleştirilmiş Toplamsal Modeller (GAM) sınıfı ile parametrik olmayan lojistik regresyon ve finansal veri analizinde kullanılan diğer yöntemler detaylı bir şekilde incelenmiştir.

#### 2.2. Lojistik Regresyon

Klasik Lojistik regresyon, özellikle biyoistatistik ve epidemiyoloji alanlarında yaygın olarak kullanılan bir istatistiksel yöntemdir. Bu yöntem, ilk olarak 19. yüzyılın sonlarında biyolojik bilimlerde büyüme oranlarını modellemek için kullanılmıştır. Ancak, lojistik regresyonun modern kullanımı ve gelişimi, özellikle 20. yüzyılın ortalarında Joseph Berkson tarafından sağlık bilimleri ve biyoistatistik alanında önemli katkılarla genişletilmiştir. Berkson, bu teknikleri tıbbi araştırmalarda kullanmak için 1944 yılında önemli bir makale yayınladı ve lojistik regresyonun temellerini attı [30].

Lojistik regresyon, bağımsız değişkenlerin bir sonucun olasılığı üzerindeki etkisini modelleyen istatistiksel bir yöntemdir. Bu yöntem özellikle ikili bağımlı değişkenler için uygundur ve finansal piyasalarda Bitcoin gibi kripto para birimlerinin fiyat hareketlerini tahmin etmek için kullanılabilir. Bu yöntem, bağımsız değişkenlerin bir sonucun olasılığı üzerindeki etkisini modelleyen istatistiksel bir yöntemdir. Bu yöntem özellikle ikili bağımlı değişkenler için uygundur ve finansal piyasalarda Bitcoin gibi kripto para birimlerinin fiyat hareketlerini tahmin etmek için kullanılabilir.

Klasik İkili Lojistik Regresyon, bir olayın olasılığını (olabilirliğini) belirlemek için logit dönüşümünü kullanır. Bu model, log-odds'unun bağımsız değişkenlerin doğrusal bir

fonksiyonu olarak modellendiği bir regresyon türüdür. Matematiksel olarak, model Eşitlik 3.1’de ifade edilmektedir:

$$\log(p / (1 - p)) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n \quad (3.1)$$

Eşitlikte yer alan  $p$  olayın gerçekleşme olasılığını ifade ederken,  $(1 - p)/p$ : oranı, olayın gerçekleşmeme olasılığı ile gerçekleşme olasılığı arasındaki oranı, yani odds oranını belirtir.  $\beta_0, \beta_1, \dots, \beta_n$  ise modelin katsayılarını temsil eder; bu katsayılar, bağımsız değişkenlerin  $X_1, X_2, \dots, X_n$  model çıktısı üzerindeki etkilerini niceliklendirmektedir.

Katsayılar ( $\beta$  değerleri), her bir bağımsız değişkenin sonuç üzerindeki göreceli etkisini ve önemini temsil etmektedir. Pozitif bir katsayı, değişkenin olayın olasılığını artıracığı, negatif katsayı ise olayın olasılığını azaltacağı anlamına gelir.

Lojistik regresyon, Bitcoin piyasası gibi finansal alanlarda, özellikle fiyatın artış ya da azalış eğilimini tespit etmek için kullanılmaktadır [31]. Model, teknik göstergeler, makroekonomik faktörler ve piyasa eğilimi gibi çeşitli bağımsız değişkenleri içerebilir. Lojistik regresyon finansal bilgileri analiz etmek için temel bir yöntemdir çünkü sonuçlarının yorumlanması nispeten kolaydır ve değişkenlerin etkilerinin doğrudan tahmin edilmesine olanak sağlamaktadır. Ayrıca model, veri setindeki doğrusal ilişkilerin anlaşılması için değerli bilgiler sunmaktadır. Lojistik regresyon finansal modelleme için güçlü bir araç olmasına rağmen, doğrusal olmayan ilişkileri modellemede sınırlı kalabilir ve çoklu bağlantı (multikolinearite) sorunlarına eğilimlidir. Dolayısıyla modelin uygunluğu ve sonuçlarının doğruluğu, seçilen bağımsız değişkenlerin kalitesine ve veri setinin kalitesine bağlıdır. Kapsamlı bir analitik çerçevede, klasik lojistik regresyon modeli, Bitcoin gibi kripto para birimlerinin fiyat hareketlerini anlamak ve tahmin etmek için diğer istatistiksel ve makine öğrenimi modelleriyle birlikte kullanılabilir. Bu bütünleşmiş yaklaşım, modelin kapsamını genişletir ve finansal piyasaların karmaşık dinamiklerini daha iyi yakalamasını sağlamaktadır.

Genelleştirilmiş Doğrusal Modellerin (GLM) şekilde basit olsa da gerçek hayatta etkiler çoğunlukla doğrusal değildir. Bu durumda sıklıkla yetersiz kalır. Önceki bölümlerde, doğrusal modeller elde etmek için temel fonksiyonlar kullanılan tekniklerinden lojistik regresyondan bahsettik. Bu bölümde, doğrusal olmayan regresyon etkilerini tanımlamak

ve karakterize etmek için daha otomatik ve esnek istatistiksel yöntemler açıklamaktır. Bu yöntemlere "genelleştirilmiş toplamsal modeller" denir [32].

### 3.2. Genelleştirilmiş Toplamsal Modeller

GAM, istatistiksel modellemede kullanılan esnek bir yöntemdir. GLM bir genişlemesi olan GAM, bağımlı değişkenin tahminini açıklayıcı değişkenlerin düzleştirici fonksiyonları olarak ifade eder. Bu modelleme yaklaşımı, doğrusal olmayan ilişkileri yakalamada etkilidir ve veri setindeki karmaşık desenleri ve yapıları ortaya çıkarabilir. Regresyon modeli şeklinde bir GLM modeli Eşitlik 3.2'deki gibi tanımlanır.

$$E(Y | X_1, X_2, \dots, X_p) = \alpha + f_1(X_1) + f_2(X_2) + \dots + f_p(X_p) \quad (3.2)$$

şeklindedir. Burada  $X_1, X_2, \dots, X_p$  bağımsız değişkenleri temsil eder ve  $Y$  ise yanıt değişkeni ya da bağımlı değişkendir.  $f_j$ 'ler tanımlanmamış düzgün parametrik olmayan fonksiyonlardır. Eğer her fonksiyonu baz fonksiyonlarının genişlemesi kullanılarak modelleysek, elde edilen model basit "En Küçük Kareler (EKK)" uygun hale getirilebilirdi. Buradaki yaklaşımımız farklıdır: her fonksiyonu bir dağılım düzleştirici (örneğin, bir kübik düzleştirme eğrisi veya çekirdek düzleştiricisi) kullanırız ve tüm  $p$  fonksiyonlar, eş zamanlı olarak, tahmin etmek için bir algoritma sağlamaktadır.

İki sınıflı sınıflandırma için, ikili veriler için önceki bölümünde tartışılan lojistik regresyon modeli Eşitlik 3.3'te yer almaktadır.

$$\log\left(\frac{\mu(X)}{1 - \mu(X)}\right) = \alpha + \beta_1 X_1 + \dots + \beta_p X_p \quad (3.3)$$

Toplamsal lojistik regresyon modeli, her doğrusal terimi daha genel bir fonksiyonel form ile değiştirir. Bu model, Eşitlik 3.4'te sunulmuştur.

$$\log\left(\frac{\mu(X)}{1 - \mu(X)}\right) = \alpha + f_1(X_1) + \dots + f_p(X_p) \quad (3.4)$$

Burada yine her  $f_j$  tanımlanmamış düzgün bir fonksiyondur. Fonksiyonların  $f_j$  için parametrik olmayan formu modeli daha esnek hale getirirken, toplamsallık korunur ve

modeli önceki gibi yorumlamamıza izin verir. Toplamsal lojistik regresyon modeli, GAM bir örneğidir.

Fonksiyonların  $f_j'$ 'nin doğrusal olmayan olması gerekmez. Doğrusal ve diğer parametrik formları doğrusal olmayan terimlerle kolayca birleştirebiliriz, bu da bazı girdilerin niteliksel değişkenler (faktörler) olduğu durumlarda gereklidir. Doğrusal olmayan terimler yalnızca ana etkilere de sınırlı değildir; iki veya daha fazla değişkende doğrusal olmayan bileşenler veya faktör  $X_k$ 'nin her seviyesi için  $X_j$ 'de ayrı eğriler olabilir.

Parametrik olmayan yöntemler, veri setinin iç yapısını tanımlayan katı model formülasyonlarından bağımsızdır. Bu, modelin veri setinin özgün özelliklerini ve desenlerini doğrudan yansıtmasını sağlamaktadır. Bu teknik, değişkenler arasındaki doğrusal olmayan ilişkileri kontrol eder. Bu nedenle, finansal piyasalar gibi karmaşık ve dinamik yapıları incelemek için ideal hale getirmektedir. Modelleme sürecini doğrudan veri setinin kendisine dayanmaktadır, bu da onu daha doğru ve güvenilir sonuçlar elde edilmesini sağlamaktadır.

Parametrik modellerden farklı olarak parametrik olmayan modeller, belirli dağılımlara veya yapısal formlara dayalı kısıtlayıcı varsayımlara daha az dayanır. Bu, modellerin, başta ekonomik veri serileri olmak üzere çeşitli verilerde bulunan olağandışı ve beklenmedik davranış kalıplarına uyum sağlamasını kolaylaştırır. Parametrik olmayan yöntemler, veri setinin karmaşıklığını hesaba katacak şekilde modelin karmaşıklığını ayarlar. Bu, modelin eğitim verilerine gereğinden fazla uymasını ve genelleme yeteneğinin kaybolmasını önler. Özellikle finansal piyasalarda aşırı uyum, yanıltıcı sonuçlara ve hatalı yatırım kararlarına yol açabilmektedir. Parametrik olmayan modeller yeni verilere hızla uyum sağlayabilir ve genişletilebilmektedir. Bu, finansal piyasaların sürekli değişen doğasıyla uyumlu olup, modelin zaman içinde gelişen piyasa koşullarına uyum sağlamasını ve güncel kalmasına olanak tanımaktadır.

GLM, istatistiksel öğrenme alanında modern tekniklerden biridir ve finans matematiği, hesaplamalı biyoloji, tıp, kimya ve çevre koruma gibi birçok tahmin alanında uygulanabilir [33]. Bu teknikler, özellikle finansal matematiğin bir alt alanı olan kripto para piyasası içinde kullanılmaktadır.

İstatistiksel modelleme, gözlemediğimiz karmaşık görünüşleri anlamak ve açıklamak için kullanılan matematiksel ifadeler bütünüdür. Matematik zarafeti ve genelliği değerlendirilirken, istatistiksel modelleme, gerçek dünya verileriyle doğrudan ilişkilidir ve genellikle araştırmacılar tarafından sorulan somut soruları cevap bulmak için geliştirilir. Ancak, her sorunun net ve genel bir çözümü yoktur; bunun yerine, istatistiksel modelleme, çok sayıda özgün problemi çözmek için esnek yaklaşımlar geliştirir. Bir istatistiksel modellemenin temel özelliği, değişkenliğin olasılık dağılımları aracılığıyla temsil edilmesidir. Bu dağılımlar, modeli inşa etmek için kullanılan temel yapı taşlarıdır ve hem rastgelelik hem de sistematik varyasyonu barındırmalıdır. Rastgelelik, verilerde görünen rastgele dağılımların arkasında yatan olasılıksal yapıyı yakalar, sistematigi ise, genellikle belirli bir teori veya kavramsal çerçeveye dayalı olarak örüntünün üretildiğini ifade etmektedir. Modelin, eldeki soruların yanıtlanmasına veya yeni soruların sorulmasına izin verecek bir denge bulma çabasıdır. Modelin karmaşıklığı, ele alınan soruların doğası ve verilerin yapısına bağlı olarak değişebilir [34].

Bu çalışmada denetimli öğrenme yöntemlerinden yararlanılmış olup uygulanan makine öğrenmesi algoritmaları ile istatistiksel modeller aşağıda detaylandırılmıştır.

### **3.2.1. Eğri Düzleştirmeli Lojistik Regresyon**

Eğri düzleştirmeli lojistik regresyon (Spline Smoothing Logistic Regression) ,istatistik ve veri analizi alanlarında yaygın olarak kullanılan bir yöntem olup, birçok araştırmacının katkı sağladığı bir alandır. Ancak kavrama öncülük eden istatistikçiler arasında Grace Wahba ismi öne çıkmaktadır. Wahba, eğri modelleri ve bunların düzeltme yöntemleriyle ilişkileri üzerine önemli çalışmalar yapmıştır ve bu alandaki katkıları geniş çapta kabul görmektedir. Wahba [35], istatistiksel öğrenme ve regresyon analizi bağlamında eğri modellerinin kullanımına ilişkin kapsamlı araştırmalar yapmış ve bu yöntemlerin teorik temellerini geliştirmiştir.

Düzleştirme eğrileri, gözlemlenen veri noktalarına yakın bir eğri veya fonksiyon oluşturmak için kullanılan bir istatistiksel yöntemdir. Eğri düzeltmeli lojistik regresyon, özellikle verilerin karmaşık yapıları nedeniyle doğrusal modelle ifade edilemeyecek ilişkileri içeren durumlarda faydalıdır. Bağımsız değişkenlerin etkisi, belirli düğüm segmentleri (spline knots) kullanılarak parça parça polinom olan bir regresyon yöntemidir [36].



Bir dizi düğüm (knot) kullanarak bir eğriyi tanımlar ve veriye en iyi uyan bu eğriyi bulmak için düzeltme terimi kullanır. Düğüm noktası, farklı aralıklarda desen davranışında değişiklikler meydana geldiği için ortak bir birleşim noktasıdır. Eğri, düğüm noktalarının yardımı ile keskin bir artış veya düşüş gösteren veri desenini aşmada avantaja sahiptir ve sonuçlanan eğri nispeten düzdür. Düzeltme eğrilerinin temel amacı, veri setindeki gürültüyü azaltırken veri yapısını koruyacak şekilde bir denge sağlamaktır. Böylece, modelin tahmin eğrisi daha esnek hale getirilerek veri setindeki yerel varyasyonları daha iyi yakalanabilir.

Hastie, Tibshirani ve Friedman'a göre düzeltme eğrileri, verilerdeki düzgün temel yapıları yakalamak ve modellemek için özel olarak tasarlanmıştır. Düzeltme eğrilerinin kullanmanın avantajlarından biri de modelin esnekliği ve karmaşık veri yapılarını yakalama yeteneğidir. Düzeltme eğriler, verilere doğal bir düzgünlük sağlama ve aşırı uyuma (overfitting) karşı denge sağlama yetenekleriyle bilinir [36].

Bu yöntem, verilerin üzerinden geçen bir eğriyi modellemek için kullanılan katsayıları belirlemek için bir optimizasyon problemini çözer. Düzeltme eğrileri veri kümesindeki her noktayı hesaba katarak bir kayıp fonksiyonunu en aza indirerek uygun bir eğri bulur. Bu işlem sırasında modelin karmaşıklığını kontrol etmek için bir düzeltme parametresi kullanılır. Bu parametre, eğrinin düzgünlüğünü ve verilerle ne kadar yakından eşleştiğini belirler.

Düzenleştirme eğrilerinin kullanımı genellikle düzeltme parametrelerinin seçimini gerektirir ve bu genellikle çapraz doğrulama veya Akaike Bilgi Kriteri (AIC) gibi yöntemlerle yapılır. Bu parametre, eğrinin düzgünlüğü ile verilere uyum arasındaki dengeyi kontrol etmektedir. Böylece modelin aşırı uyumunu veya yetersiz uyumunu (underfitting) önler.

Bu nedenle düzeltme eğrilerinin kullanılması, veri kümelerindeki karmaşık yapıların modellenmesi ve araştırılması için güçlü ve esnek bir araçtır. Hastie, Tibshirani ve Friedman'ın bu konuyla ilgili çalışması, düzeltme eğrilerinin teorik temellerini ve pratik uygulamalarını kapsamlı bir şekilde kapsamakta ve bu yöntemin istatistiksel öğrenme alanındaki önemini vurgulamaktadır.

Bu model genellikle Eşitlik 3.5'te gibi eğri düzeltmeli lojistik regresyon modeli olarak formüle edilmektedir.

$$\text{logit}(P(Y = 1)) = \beta_0 + \sum_{(i=1)}^n \beta_i X_i + s(Z, \lambda) \quad (3.5)$$

Burada  $P(Y = 1)$  bağımlı değişkenin bir değerini alma olasılığını temsil etmektedir. Logit dönüşümü,  $\frac{\log(P(Y=1))}{(1-P(Y=1))}$  formülüyle ifade edilir ve bağımlı değişkenin bir değer alması ihtimalinin log-odds gösterimini ifade etmektedir.  $\beta_0$ , modelin kesme terimi olup, bağımsız değişkenlerin etkisinin olmadığı durumda bağımlı değişkenin log-odds değerini belirtmektedir.  $X_i$ , modeldeki bağımsız değişkenleri ifade etmektedir. Her  $\beta_i$  katsayısı, ilgili  $X_i$  bağımsız değişkeninin bağımlı değişken üzerindeki etkisini nicel olarak tanımlamaktadır.  $s(Z, \lambda)$  ise,  $Z$  değişkenine bağlı olarak uygulanan düzeltme eğrisi ve bu eğri esnekliğini kontrol eden  $\lambda$  parametresini belirtmektedir.

Bu modelde, doğrusal olmayan ilişkileri modellemek amacıyla kullanılan düzeltme eğrisi, veri üzerinde esnek bir eğri sağlayarak modelin karmaşık ilişkileri yakalayabilmesine olanak tanır.  $\lambda$  parametresi, eğrinin düzgünlüğünü kontrol eder ve aşırı uyum olasılığını azaltır. Bu model özellikle, bağımlı değişkenin kategorik olduğu durumlar için, yani lojistik regresyon durumlarında tercih edilmektedir.

Bu model, finansal serilerin oynaklığı gibi doğrusal olmayan kalıpları modellemek için kullanılabilir. Özellikle kripto para birimi fiyat hareketleri gibi bir dizi değişken ve karmaşık yapılar için etkilidir.

Eğri düzeltmesi eklenen her bir lojistik regresyon modeli, doğru bir şekilde ayarlanması gereken ek parametrelere sahiptir. Modelin karmaşıklığı, seçilen düğümlerin sayısına ve konumlarına bağlıdır. Ek olarak, uygun eğrilerin seçilmesi ve modelin gereğinden fazla uyum sağlamasına (overfitting) karşı dikkatli olunması gereken bir süreçtir.

Yapay sinir ağları gibi diğer karmaşık modellerle karşılaştırıldığında, eğri düzeltmeli lojistik regresyon, modelin yorumlanabilirliğini korurken esneklik sağlamaktadır. Ancak modele çok sayıda düğüm eklemek, modelin anlaşılmasını zorlaştırabilir ve veri

kümesinin ötesinde genelleme yeteneğini azaltabilir. Bu modeli kullanırken uygun model karmaşıklığı ile genelleme performansı arasında bir denge kurulmalıdır [37].

### 3.2.2. LOESS Lojistik Regresyon

LOESS (Yerel Ağırlıklı Saçılma Grafiği Düzleştirme) yöntemi, 1979 yılında William S. Cleveland tarafından geliştirilmiştir. Ancak bu yaklaşımın lojistik regresyonla birleştirilmesi daha sonraki bir gelişmedir ve bu özel kombinasyonu ilk kimin geliştirdiği belirsizdir. Cleveland, veri noktalarına yakın bir eğriye uyacak şekilde yerel ağırlıklandırılmış regresyon olarak kullanan bir yöntem olan LOESS'i keşfetti ve tanıttı.

Lojistik regresyonu LOESS ile birleştirmek, veriler arasındaki doğrusal olmayan ilişkileri modellemek için daha karmaşık bir yöntemdir ve bu alanda çeşitli istatistikçiler ve araştırmacılar tarafından çalışmalar yapılmıştır. Bu nedenle, LOESS Lojistik regresyonun kim tarafından bulunduğunu belirlemek yerine, bu metodolojiye katkıda bulunan birçok istatistikçi ve araştırmacının çalışmalarını incelemek daha uygun olur. Bu konuda yapılan çalışmalar, genellikle istatistik ve veri bilimi literatüründe çeşitli araştırmacılar tarafından genişletilmiş ve derinleştirilmiştir.

Jacoby W.G. göre LOESS, bağımlı bir değişken ile bir veya daha fazla bağımsız değişken arasındaki ilişkiyi tanımlamak için kullanılır. Bu sürecin ayırt edici özelliği, verilerin "kendi adına konuşmasına" izin vermesidir. LOESS parametrik olmayan bir yöntem olduğundan, uygun eğriler verilerde mevcut olabilecek herhangi bir yapının kesin olarak önceden belirlenmiş spesifikasyonlarına göre değil ampirik olarak elde edilir. Bu nedenle, LOESS dinamik dağılım grafikleri, geleneksel istatistiksel modelleme prosedürleri tarafından kolayca gözden kaçırılan nispeten karmaşık ilişkileri sıklıkla ortaya çıkarabilir.

LOESS ve diğer parametrik olmayan tahmin stratejileri sosyal bilim araştırmalarında faydalıdır çünkü mevcut temel teoriler genellikle ampirik verilerde bulunması gereken yapı türleri hakkında çok az ayrıntı sağlamaktadır. Başka bir deyişle, hipotez hangi değişkenlerin birbiriyle ilişkili olması gerektiğini ve çoğu zaman bu ilişkinin yönünü de belirtir: Örneğin, "eğitim seviyeleri oy verme katılımıyla pozitif olarak ilişkili olmalıdır." Ancak, bu tür ifadelerin ötesinde, genellikle fonksiyonel formlar hakkında tahminler yoktur. Bu nedenle, teoriye dayalı karşıt yönlendirmelerin yokluğunda, araştırmacılar

basit spesifikasyonlara başvururlar; bu durum Beck ve Jackman'ın "varsayılan doğrusallık" olarak adlandırdığı bir durumdur [38]. Bu, özellikle seçimler, oy verme ve kitlesel siyasi davranış alanlarında doğrusal olmayan ilişkilerin yaygın olduğunu öne süren ayrıntılı teoriler mevcut olduğunda ciddi bir sorun olabilir. Bu nedenle, LOESS gibi parametrik olmayan teknikler, bu tür doğrusal olmayan durumların tespit edilmesi ve formlarının yorumlanması için çok faydalı olmalıdır [39].

### **Saçılım Grafiği Düzleştirme (Scatterplot Smoothing)**

İki boyutlu dağılım grafiği, iki değişkenli verilerin temel grafiksel temsilidir. Aynı zamanda, dağılım grafikleri, çok değişkenli verilerin daha karmaşık grafiksel gösterimleri için "yapı taşlarıdır" [40]. Dağılım grafiğinin en büyük avantajlarından biri, gösterime dahil edilen değişkenler arasındaki ilişkilerin veya işlevsel bağımlılıkların görsel bir değerlendirmesini sağlamasıdır. İki değişken birbiriyle ilişkiliyse (yani fonksiyonel olarak ilişkiliyse), çizilen noktalar çizim alanı boyunca eşit şekilde dağılmayacaktır. Bunun yerine nokta bulutu tanınabilir bir desen veya şekil oluşturacaktır.

Dağılım grafiklerindeki işlevsel bağımlılıkların değerlendirilmesi basit bir süreç olmalıdır. Ancak pratikte bu genellikle oldukça zordur. Sorun, gürültülü veri değerlerinin, seyrek veri noktalarının ve zayıf karşılıklı korelasyonun herhangi bir modelin görsel olarak tanınmasını engelleyebilmesidir. Ayrıca, bir grafikte genel bir model belirlenebilse bile, yalnızca dağılım grafiğinin görsel incelemesi yoluyla modelin kesin doğasını karakterize etmek neredeyse imkansızdır.

Önceki sorunla başa çıkmak için yararlı bir strateji, dağılım grafiğindeki noktaları düzleştirme bir eğri yerleştirmektir. Eğrinin amacı,  $Y$  değişkeninin,  $X$  değişkeninin dağılımı içindeki farklı konumlardaki dağılımının merkezi eğilimini özetlemesidir. İki değişken arasında korelasyon yoksa, düzgün eğri düz bir çizgi olacaktır ( $X$ 'in değeri ne olursa olsun  $Y$  dağılımının merkezi değişmeyecektir). İki değişken birbiriyle ilişkiliyse düzgün eğri yatay olmayan başka bir şekil almalıdır.

Düzgün eğrileri uyum sağlamak için iki genel strateji vardır: parametrik uyum ve parametrik olmayan uyum [41]. Bunlardan ilki, analistin ilişkinin işlevsel biçimini önceden belirlemesini gerektiren parametrik uyum sağlamasıdır. Uyum algoritmaları

tipik olarak, ikisi arasında en iyi uyumu sađlayan denklem katsayıları kümesini tahmin ederek belirli bir form ile ampirik veriler arasındaki uyumu optimize eder. Regresyon analizi bugüne kadar en iyi bilinen parametrik yumuřatma prosedürüdür. Bir dizi veri noktasına düz bir çizgi sığdırmak için en küçük kareler kriterini kullanır. Parametrik uyum, verideki düzleştirici algoritmasının uyum sađladığı fonksiyon tipine uygun olduđunda ilişkileri özetlemenin çok etkili bir yoludur. Ancak en azından analizin başlangıcında "dođru" fonksiyonel biçim bilinmemektedir. Bu nedenle arařtırmacılar, veri içindeki yapıyı bozan düzleştirici eğriler uyum riskiyle karşı karşıdadır.

Parametrik olmayan düzleştirici, önceki sorunu dođrudan çözer. Deđişkenler arasındaki işlevsel ilişkileri önceden belirlemek zorunda kalmadan, veri noktaları arasındaki düzgün eğrileri bulmak için kullanılabilirler. Bunun yerine, uyum algoritması basitçe çizilen noktaların ampirik yoğunluđunu izlemeye çalışır. Ortaya çıkan uyum "çizgisi", veri alanının en yoğun olduđu dağılım grafiđi alanından geçmelidir.

řu anda en popüler parametrik olmayan düzleştirici LOESS [42]. William S. Cleveland'ın belirttiđi gibi, " LOESS bazı son derece arzu edilen istatistiksel özelliklere sahiptir ve hesaplanması ve kullanımı kolaydır" [41]. " LOESS " terimi, yerel ağırlıklı regresyonun kısaltmasıdır. Bu yöntemde bahsedilen teknik, yerel ağırlıklı saçılma grafiđi düzeltme için "LOWESS" tekniđinin genelleştirilmesidir [43]. Kısacası, LOESS Lojistik Regresyonu, veri setindeki her nokta için yerel bir model oluşturur ve bu noktalar için tahminler üretir. Bu yaklaşımla, özellikle heterojen veri setlerinde farklı alt grup veya örüntüler içeren yapıların ifade edilmesini mümkün kılar.

Aynı zamanda, LOESS yönteminin tüm sorunlara evrensel bir çözüm sađlamadıđına dikkat çekmek önemlidir. Bu yöntemin, belirli veri yapıları ve özel analiz gereksinimleri için uygun olduđu durumlar olmasına karşın, her türlü veri seti veya analitik durum için ideal olduđunu iddia etmek bilimsel bir yanılıđ olurdu. LOESS, özellikle düşük boyutlu veri setlerinde ve veri noktaları arasındaki ilişkilerin karmaşık olmadığı durumlarda etkili bir şekilde çalışırken, büyük veri setleri veya daha karmaşık yapılar söz konusu olduđunda sınırlamaları olan bir yöntemdir. Dolayısıyla, bu yöntemin kullanımı, spesifik durumların ve veri özelliklerinin dikkatlice deđerlendirilmesini gerektirir. LOESS gibi parametrik olmayan düzleştiriciler, geleneksel parametrik düzleştirme yöntemlerine (örn. en küçük kareler yöntemi) bir tamamlayıcı olarak görülmelidir, bir yerine geçme olarak

değil. Bu iki genel yaklaşım- parametrik ve parametrik olmayan düzleştirici ile kullanılmalıdır, çünkü her birinin kendi ayırt edici güçlükleri ve zayıflıkları vardır.

Parametrik düzleştirici güçlükleri şunları içerir: Bu yaklaşıma dayanan prosedürler genellikle, belirlenen model yeterince iyi uyduğu varsayılırsa, verilerin altındaki yapının özünü veren, öz ve kolay anlaşılır bir denklem sonucu verir. Uyum yöntemleri iyi bilinir ve anlaşılır. Ve genellikle, örneklem değişkenliğini düzgün uyumun hassasiyetine dahil etmek için kullanılabilir istatistiksel teori mevcuttur.

Parametrik düzleştirici stratejilerinin bu güçlükleri, bazı zayıflıklarla dengelenir. Parametrik uyum prosedürleri çok esnek değildir. Yani, yanlış belirtileri iyi bir şekilde ele alamazlar. Aynı zamanda, global yerine yerel uyum rutinlerine dayanırlar. Bu, son pürüzsüz eğriyi, veri kümesinin sınırlı bir alt kümesi içinde var olabilen aykırı değerler, uyuşmayan gözlemler ve desenlere göre nispeten hassas hale getirir.

Parametrik olmayan düzleştirici gücü, değişkenler arasındaki ilişkinin kesin doğası konusunda çok esnek olmalarıdır. Ayrıca, modellenmiş eğrinin herhangi bir  $X$  değerleri yakınlığında pozisyonu ve yönlenmesi esas olarak o yakınlıktaki veri noktalarına bağlı olduğundan, nispeten yereldirler. Nispeten uzakta olan  $X$  değerlerine sahip gözlemler, o eğri segmenti üzerinde az veya hiç etkiye sahip değildir (elbette, kendi yakınlıklarındaki düzleştirici eğrinin yerleştirilmesini etkilerler). Bu özellik, Parametrik olmayan düzleştiricilerin, dağılım grafiği içindeki uyuşmayan noktalara karşı daha az hassas hale getirir.

Parametrik olmayan düzleştirici başlıca zayıflığı, verileri basit bir denklemle karakterize edememeleridir. Ayrıca, parametrik olmayan düzleştirici algoritmaları için istatistiksel teori, daha geleneksel uyum yöntemlerinden daha az gelişmiştir. Son olarak, Parametrik olmayan uyum yöntemleri için analistin, uyum parametrelerine ilişkin bazı öznel kararlar almasını gerektirir

Teoriler, genellikle doğrusal olmayan ilişkilerin var olabileceğini öngörür ancak çoğu zaman bu ilişkilerin neye benzediği hakkında açık bilgi sağlayamaz. LOESS yaklaşımı, bu belirsiz koşullar altında ilişkilerin nasıl görüldüğünü anlamamıza yardımcı olabilmektedir. LOESS modeli, veri noktalarına yakın bir eğriyi yerel olarak

uyumlayarak çalışır. Matematiksel olarak, LOESS modeli belirli bir  $x$  noktası için tahmini değeri ( $\hat{y}$ ), Eşitlik 3.6'daki gibi formüle edilmektedir.

$$\hat{y}(x) = \sum_{i=1}^n w_i(x) \times y_i \quad (3.6)$$

Bu kısımda  $n$ , veri noktalarının sayısıdır;  $y_i$ ,  $i$ 'nci veri noktasının hedef değeridir;  $w_i(x)$ ,  $x$  noktasında  $i$ 'nci veri noktasının ağırlığıdır ve  $w_i(x)$ , genellikle  $x$  noktasına göre her veri noktasının mesafesine bağlı olarak belirlenen bir ağırlık fonksiyonu ile hesaplanmaktadır. Ağırlık fonksiyonu genellikle Eşitlik 3.7'de gibi bir çekirdek fonksiyonu ile ifade edilir:

$$w_i(x) = \frac{K((x - x_i))}{h} \quad (3.7)$$

$K$  çekirdek fonksiyonudur (örneğin, üçgen, dikdörtgen, Gauss, vb.),  $x_i$   $i$ 'nci veri noktasının bağımsız değişken değeridir,  $h$  bant genişliği veya pürüzsüzlük parametresidir, bu da ağırlık fonksiyonunun ne kadar geniş bir alan üzerinde etkili olacağını belirler.

LOESS modeli, her bir  $x$  noktası için bu ağırlıkları kullanarak, o noktaya en yakın veri noktalarını daha ağır basacak şekilde yerel bir regresyon hesaplar. Bu, veri setindeki genel trendi ve yerel varyasyonları yakalamayı amaçlar. LOESS, modelin karmaşıklığını kontrol etmek için kullanılan bant genişliği (bandwidth) parametresi ile esnek bir şekilde yerel yapıları uyumlayabilir. Bu özellik, LOESS'i doğrusal olmayan veri yapılarını analiz etmek için özellikle yararlı kılar.

LOESS Lojistik Regresyon, Bitcoin değer tahmininde özellikle veri setlerindeki özgü karmaşıklıkları ve doğrusal olmayan trendleri ve desenleri yakalama yeteneği ile ön plana çıkar. Bu yöntem, ekonomik, politik ve sosyal faktörlerden etkilenen Bitcoin fiyatlarının doğrusal olmayan eğilimlerini belirlemek, yüksek volatilitiyi analiz etmek, fiyat grafiklerindeki dönüş noktalarını ve olağandışı hareketleri tespit etmek, makroekonomik faktörlerin etkisini analiz etmek ve piyasa davranışının zaman içindeki değişimlerini anlamak için kullanılabilir. Bu çeşitli uygulamalar, LOESS Lojistik Regresyonu, Bitcoin ve benzeri finansal piyasalardaki değer değişimlerini anlamak ve gelecekteki fiyat hareketlerine ilişkin tahminlerde bulunmak için değerli bir araç haline getirmektedir.

LOESS modelinin ana sınırlamalarından biri hesaplama yoğunluğudur. Büyük veri kümeleri için modelin uygulanması uzun zaman alabilir. Ayrıca, model parametrelerinin seçimi (örneğin, bant genişliği veya yerel polinom derecesi) modelin başarısı üzerinde önemli bir etkiye sahiptir ve aşırı uyum riski vardır. Özellikle veri setindeki gürültü göz önüne alındığında, model seçimi ve ayarlaması dikkatli yapılmalıdır. Yerel regresyon modellerinin veri seti dışındaki noktalara genelleştirilmesi zor olabilir, dolayısıyla LOESS modellerinin kullanımı genellikle veri açısından zengin ortamlarla sınırlıdır [44].

### 3.3. Makine Öğrenmesi Modelleri

Yapay zekâ (AI), bilgisayar bilimlerinin bir alt dalı olan ve bilgisayarların akıllıca davranmasını sağlayarak akademi ve endüstride kendine yer bulan bir araştırma alanıdır [45]. Son on yılda, veriden öğrenme, bilimin her alanında çok önemli hale geldi. Makine öğrenmesi, bilgisayarların belirli bir veri kümesindeki örüntüleri otomatik olarak tanımlamasına, öğrenmesine ve bunları kullanarak gelecekteki verileri tahmin etmesine izin veren bir yapay zekâ dalıdır [46].

Makine öğrenmesi hakkında ilk çalışmayı ortaya koyan Alan Turing, 1950 yılında “Computing Machinery and Intelligence” adlı makalesinde günümüzde Turing testi olarak bilinen kavram ile ifade etmektedir. 1952 yılında, yapay zekanın öncüsü Arthur Samuel bir bilgisayar oyunu yazarak bu bilgisayar oyunundaki tüm hamleleri, stratejileri kendi programına dahil etmeyi başarılı şekilde gerçekleştirmiştir. 1957’de Frank Rosenblatt, insan beyninin düşünce yapısından etkilenecek bu düşünce sürecini simüle eden yapay sinir ağlarını ortaya koymuştur [47].

Makine öğrenmesi, büyük miktarda veri toplayabilen ve analiz edebilen teknolojilerin ortaya çıkmasıyla birlikte son yıllarda oldukça popüler hale gelmiştir [48]. Temelde veri kümelerindeki kalıpları belirlemek ve gelecekteki veri noktaları hakkında tahminlerde bulunmak için kullanılan algoritmik teknikler bütünüdür.

Makine öğrenmesi, teknik analiz için kullanılan göstergeleri otomatik olarak analiz etmek ve öngörülerini yapmak için kullanılabilir. Bu, yatırımcıların daha iyi kararlar vermelerine yardımcı olabilir ve ticaret stratejilerini otomatikleştirebilir. Örneğin, bir makine öğrenmesi algoritması, belirli bir teknik göstergeye dayalı olarak gelecekteki fiyat hareketlerini tahmin edebilir veya belirli bir piyasa koşulunda hangi teknik göstergelerin



en iyi sonuçları verdiğini belirleyebilir. Bu tür analizler, daha büyük ve daha karmaşık veri kümelerinin analiz edilmesine izin verir ve yatırımcıların daha hızlı ve daha doğru kararlar almalarına yardımcı olur. Bununla birlikte, makine öğrenmesi algoritmaları, doğru şekilde eğitilmedikleri takdirde hatalı sonuçlar verebilir veya yanlış kararlar verilmesine yol açabilir. Bu nedenle, makine öğrenmesinin kullanımı dikkatli bir şekilde yapılmalı ve uygun eğitim ve test verileri kullanılmalıdır [49].

### 3.3.1. Makine Öğrenmesinin Tanımı ve Kategorileri

Makine öğrenmesi, algoritmalar aracılığıyla veri setlerinden öğrenen ve bu öğrenilen bilgiyi yeni veri üzerinden tahminlerde bulunma sürecidir. Şekil 3.1’de Makine Öğrenmesi temel olarak sınıflandırma sistematığına göre dört ana kategoriye ayrılmıştır [50].



Şekil 3.1. Yaygın kullanılan Makine Öğrenmesi modelleri (Kaynak: [50])

**Denetimli Öğrenme (Supervised Learning):** Model, etiketli veriler üzerinde eğitilmektedir. Bu model daha sonra yeni, görülmemiş veri üzerinden tahminler yapmak için kullanılmaktadır. Denetimli öğrenme sınıflandırma ve regresyon olarak iki temel problemden bahsedilebilir [49]:

- i. **Sınıflandırma:** Bir veri kümesindeki örnekleri belirli bir sınıfa atama işlemidir. Örneğin, bir resim veri kümesindeki her resmin bir hayvan resmi mi yoksa bitki resmi mi olduğunu belirlemek isteyebiliriz. Sınıflandırma problemleri genellikle etiketli veri kümelerinde çalışır, yani verilerin hangi sınıfa ait olduğu önceden belirlenmiştir. Sınıflandırma algoritmaları, veri kümesindeki özelliklerin bir kombinasyonunu kullanarak, veri örneklerini belirli sınıflara doğru şekilde atama yapmaya çalışır.
- ii. **Regresyon:** Bir veri kümesindeki örneklerin bir çıktı değerini tahmin etme işlemidir. Regresyon problemleri sürekli bir çıktı değişkeni üzerinde çalışır. Örneğin, bir evin fiyatını, o evin özellikleri (odaların sayısı, konumu, vb.) kullanarak tahmin edebiliriz. Regresyon algoritmaları, bağımlı değişkenin bağımsız değişkenlere olan bağımlılığını hesaplayarak, gelecekteki veriler üzerinde tahminler yapabilmemizi sağlar.

**Denetimsiz Öğrenme (Unsupervised Learning):** Etiketlenmemiş veri kullanılır. Bu algoritma veri setindeki kalıpları, grupları veya yapıları kendi başına keşfetmeye çalışmaktadır. Kümeleme yöntemi, denetimsiz makine öğrenmesi analizlerinden biridir. Çok değişkenli analiz yöntemlerinden biri olan kümeleme analizi, veri setinde yer alan gözlemleri benzerleriyle aynı gruba taşıma işlemi gerçekleştirilmektedir. Temel amaç, gözlemlerin sahip olduğu karakteristik özelliklere göre gruplamaktır.

### **Yarı Denetimli Öğrenme (Semi-Supervised Learning)**

Yarı denetimli öğrenme, büyük bir veri kümesinin yalnızca küçük bir kısmının etiketlendiği durumlarda kullanılan bir yöntemdir. Bu yöntem, etiketlenmemiş verileri doğru bir şekilde tahmin etmek ve sınıflandırmak için etiketli verileri kullanmayı amaçlamaktadır. Yarı denetimli öğrenme, deneme yanılma yoluyla hataları azaltır ve geçmiş deneyimlerden öğrenerek performansı artırır. Bu süreç, modelin birden fazla iterasyon yapmasını ve her seferinde elde edilen bilgilerin kendisini geliştirmek için kullanılmasını içerir.

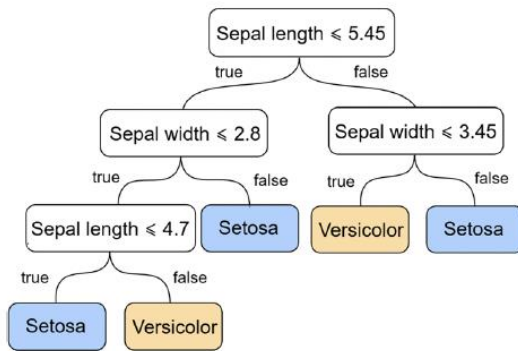
**Pekiştirmeli Öğrenme (Reinforcement Learning):** Bu model, bir hedefe ulaşmak için bir dizi karar almaktadır. Bu kararların sonuçlarına göre ödül veya ceza alarak öğrenmektedir. Bu yöntem yaygın olarak oyun programlarında kullanılmaktadır.

Yukarıda verilen yöntemlerin her biri, farklı veri tipleri ve öğrenme problemleri için uygundur. Makine öğrenmesi uygulamaları, genellikle bu yöntemlerin bir veya daha fazlasını kullanarak en iyi sonuçları elde etmek için bir araya getirilir [47].

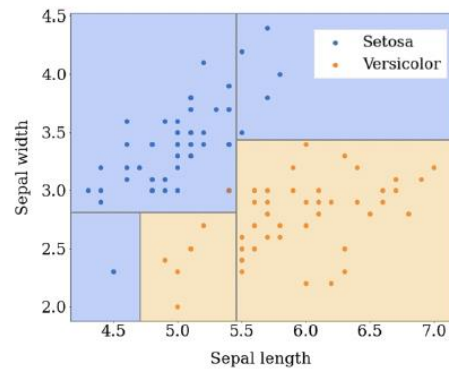
### 3.3.2 Karar Ağaçları (Decision Tree)

Bir veri kümesinin karar yapısını ve olası sonuçlarını gösteren ağaç benzeri modeller oluşturmak için karar ağaçları olarak bilinen denetimli öğrenme algoritmaları kullanılır. Sınıflandırma ve regresyon görevleri bu yaklaşımı kullanır. Karar ağaçları (DT) aynı zamanda sağlamlığı ve yorumlanabilirliğiyle bilinir. İlk önerilmesinden neredeyse 60 yıl sonra, bu konuyla ilgili araştırmalar hala devam ediyor ve son on yılda çok sayıda araştırmacı önemli soruları çözdü [51].

Bu modeller genellikle akış şemasına benzer bir şekilde gösterilir; burada her dahili düğüm mantıksal bir test (bölünme olarak bilinir) ve her yaprak bir tahmindir (bkz. Şekil). Çıkarım sırasında her gözlem kökten başlar ve yapraklardan birinde biter ve kullanıcı için tamamen şeffaf olan bir yol izleyerek biter. Bu yöntemin sağlamlığı, basitliğine dayanmaktadır. İris veri kümesinin kullanıldığı bir karar ağacı örneği, Şekil 3.2' de verilmiştir.



(a) Tree visualization



(b) Partitioning visualization

Şekil 3.2. İris veri setinin indirgenmiş bir versiyonu üzerine uygulanan bir sınıflandırma ağacı (Kaynak: [51])

Ağaçlar, düşük hesaplama maliyeti, negatif değerleri işleme yeteneği ve kullanıma hazır karma veri işleme gibi birçok ek avantaja sahiptir [32]. Karar ağaçları uzun süredir yaygın olarak kullanılmaktadır, ancak son zamanlarda bu modeller her zamankinden daha fazla ilgi görüyor. Makine öğrenimi ve otomatik karar verme sistemlerinin popülerliği arttıkça, insanların bir dereceye kadar anlayabileceği modellerin geliştirildiği yorumlanabilir makine öğrenimine olan ilgi de artıyor. Karar ağaçları yorumlanabilirliğiyle bilindiğinden, birçok araştırmacı bu alanda çalışmaya başladı [52, 53]. Bu alana artan ilgi, olağan yayın sürecine ek olarak daha geniş bir incelemeden fayda sağlayacak zengin bir katkı koleksiyonuna yol açtı. Karar ağaçlarının avantaj ve dezavantajları Çizelge 3.1’de listelenmiştir [54].

Çizelge 3.1. Karar ağaçlarının avantaj ve dezavantajları

<b>Avantajları</b>	<b>Dezavantajları</b>
Anlaşılması ve anlatılması kolaydır.	Karar ağaçları kararsızdır; sonuçlar ağacın hiyerarşik yapısına bağlı olarak farklılık gösterir.
Hem niceliksel hem de kategorik değişkenler için kullanılabilir.	Aşırı öğrenme veya aşırı adaptasyon riski vardır.
Regresyon ve sınıflandırma problemlerine uygundur.	Sürekli değişkenlerin değerlerini tahmin etmede daha az başarılıdır.
Eksik gözlemlere sahip veri kümelerini başa çıkabilme yeteneğine sahiptir.	Ağaç oluşturma ve budama süreçleri karmaşık ve zordur.
	Sınıf sayısı fazla ve öğrenme kümesi sayısı az olduğunda model başarısız olabilir.

## **Karar Ağaçları Algoritmaları**

Tarih boyunca karar ağaçları çok sayıda algoritma kullanmıştır. Bu algoritmalar AID, CHAID, CART, ID3, C4.5 ve QUEST'tir.

### **a) AID Algoritması**

AID (Otomatik Etkileşim Dedektörü) algoritması, Michigan Üniversitesi'nde Morgan ve Sonquist, 1970'lerin başında AID (Otomatik Etkileşim Dedektörü) algoritmasını geliştirdi. Karar ağaçlarına dayalı ilk algoritma ve yazılımdı. Bu algoritma sayesinde, en iyi tahminleri yapmak için bağımlı ve bağımsız değişkenler arasındaki olası tüm ilişkileri inceleyebilirsiniz. Birçok araştırmacı, Einhorn'un bu incelemeyi yaparken anlamlı ve anlamsız ilişkileri ayırt edemeyecek kadar saldırgan olduğunu söyledi [55].

### **b) CHAID Algoritması**

Ki-Kare Otomatik Etkileşim Dedektörü olarak da bilinen CHAID algoritması, 1980'de G.V. Kass tarafından geliştirilen AID algoritmasının geliştirilmiş bir devamıdır. Bu algoritma, CHAID algoritmasına benzer, ancak CHAID birden fazla ağaç üretir. Bu algoritmanın amacı, kategorik değişkenlerden oluşan bir veri setini en iyi şekilde açıklamak için homojen alt gruplara bölmektir. Bu alt gruplar, küçük sayıda öngörülerden oluşur. Daha sonra seçilmiş bu yordayıcılar bağımlı değişkeni tahmin etmek için kullanılır. Bağımlı değişken kategorik olduğunda "Ki-kare" testini kullanın ve bağımlı değişken sürekli olduğunda F testini kullanın. Ki-kare olarak bilinen algoritma, istatistiksel olarak anlamlı oranlar üzerinde çalıştığı için bir dizi çapraz tablo kullanıyor. Hem sürekli hem de kategorik değişkenleri işleme yeteneği ve birden fazla ağaç oluşturma yeteneği nedeniyle şu anda tercih edilen algoritmadır [55, 56, 57].

### **c) CART Algoritması**

Breiman, Friedman, Olshen ve Stone, 1984 yılında CART (Sınıflandırma ve Regresyon Ağaçları) algoritmasını geliştirdi. Bu algoritmada kullanılan sınıflandırma ve regresyon ağaçları, bağımlı değişken kategorik ise sınıflandırma ağacı ve sürekli ise regresyon ağacı olarak adlandırılır. CART algoritması, ikili karar ağaçları oluşturmak için kullanılır.

Başka bir deyişle, kök düğümünden iki alt düğüm oluşturarak izomorfik bir ağaç oluşturmaya çalışın. İkili karar ağaçlarını bölerken, sürekli değişkenler için en küçük kareler sapması, kategorik bağımlı değişkenler için Gini ve Twoing indeksleri kullanılır. En büyük eksikliği, sadece iki farklı duruma göre karar vermesidir. Kâr, maliyet değerleri ve değişken kategorileri arasındaki önceliği belirlemede sağladığı kolaylık nedeniyle algoritma günümüzde de kullanılmaktadır [55,58].

#### **d) ID3 Algoritması**

1986 yılında Quinlan, ID3 algoritmasını keşfetti. Bu algoritma, karar ağacı algoritması olarak kabul edilebilir. ID3 algoritması, bilgi kazancını bölme kriteri olarak kullanır. Tüm numuneler tek bir hedef özellik değerine ait olduğunda veya ideal bilgi kazancı sıfırdan büyük olmadığında büyüme durur. ID3 algoritması, sayısal değişkenlere veya eksik değerlere budama yapmaz [59].

#### **e) C4.5 Algoritması**

Quinlan'ın 1993'te önerdiği C4.5 algoritması, ID3 algoritmasının geliştirilmiş bir versiyonudur. Geliri bölünme kriteri olarak kullanır. Segmentasyon, belirli bir sınırın altına düştüğünde sona erer. Hataya dayalı budama, büyüme aşamasının ardından yapılır. C4.5 algoritması sayısal özellikleri işleyebilecek kapasiteye sahiptir. Ayrıca, algoritma, düzeltilmiş kazanç oranı kriteri aracılığıyla eksik değerleri içeren bir eğitim setinden başlar [59].

#### **f) QUEST Algoritması**

Loh ve Shih, 1997 yılında QUEST (hızlı, tarafsız ve verimli istatistiksel ağaç) algoritmasını geliştirdi. CART algoritması, CHAID ve CART algoritmalarından farklı olarak, ikili karar ağacı oluştururken değişken seçimini ve ayırma noktası seçimini ayrı ayrı ele alır. CART algoritmasının tüm avantajlarına sahip olmasına rağmen, aynı zamanda bir o kadar da yavaştır [56].

Bir karar ağacı, her düğümün veri kümesini daha küçük, daha eşit alt kümelere böldüğü bir dizi karar kuralından oluşur. Karar ağacı, veri özelliklerine göre dallara ayrılır ve kök düğümünden başlar. Yapraklı düğümler, regresyon için sayısal değerleri veya sınıf

etiketlerini gösterir. Karar ağaçları, veri kümelerini en iyi şekilde farklılaştıran özelliklerin ve eşiklerin seçilmesiyle başlar.

### **Karar Ağacı Bölünme Kuralları**

Yukarıdan aşağıya teknikler tipik olarak ağgözlü bir yaklaşımla uygulanır. Bu yöntem, her aşama için belirli kriterlere göre dallara ayrılabilir. Bu hesaplamayı kolaylaştırır da her zaman en iyisi olmayabilir. Dallara ayırırken, test koşuluna hangi değişkenin uygulanacağı, en iyi bölmeyi sağlamak için hangi standartların kullanılacağı ve bölmenin ne zaman sonlandırılacağı gibi konuları düşünmelisiniz. Dallarırken hangi değişkeni seçeceğiniz çok önemlidir. Bilgi kazancı ve Gini katsayısı, bu değişkenin seçilmesinde kullanılan kriterlerdir [60]. Tezde ele alınan karar ağacı modelinde bölünme kuralı olarak entropi kullanılmaktadır; bu nedenle, ilgili bölümde entropinin kullanımı detaylı bir şekilde incelenecektir.

### **Entropi**

Entropi, bilgi ile ilgili bir ölçümdür ve 0-1 arasında değer alan rastgele değişkenlerin belirsizliğini ölçer. Değer 1 yaklaştıkça belirsizlik giderek artıyor. Belirlilikleri azaltmak için veriler hakkında daha fazla bilgi gereklidir. CART, ID3 ve C4.5 algoritmaları entropiyi en iyi ayırt edici özelliklere sahip değişkenleri belirlemek için kullanır [60]. Daha düşük entropiye sahip değişkenler daha fazla bilgi sağlar.

$D$  veri kümesindeki bir gözlemin  $i$  sınıfına ait olma olasılığı  $p$  olsun. Bu nedenle beklenen entropi Eşitlik 3.8'deki gibi hesaplanmaktadır.

$$E(D) = - \sum_{(i=1)}^m p_i \times \log_2(p_i) \quad (3.8)$$

Bu denklemde kullanılan  $E(D)$ , hesaplanan entropi değeri temsil etmektedir.  $p_i$ , her bir durumun olasılığını ifade etmektedir.

Finansal piyasa analizi, tıbbi teşhis, müşteri segmentasyonu, kredi puanlama ve karar ağaçları çok çeşitli alanlarda kullanılmaktadır. Finansal piyasalarda yaygın olarak

kullanılırlar, özellikle Bitcoin ve diğer kripto para birimlerinin fiyat hareketlerini tahmin etmede.

Aşırı uyum eğilimi ve veri değişikliklerine duyarlılık karar ağaçlarının başlıca sorunlarıdır. Ağacın budanması ve ansamble teknikleri bu sorunları giderir. Karar Ağaçları, özellikle doğrusal olmayan ilişkileri bulmada ve modellerin anlaşılmasını kolaylaştırmada etkilidir. Bununla birlikte, bu modellerin Random Forest veya Gradient Boosting gibi topluluk (grup) yöntemlerinin bir araya getirilmesiyle en iyi sonuçlar genellikle elde edilir. Sonuç olarak, bir sonraki bölümde Gradient Boosting topluluk algoritmalarından bahsedilecektir.

### 3.3.3 Topluluk Öğrenmesi

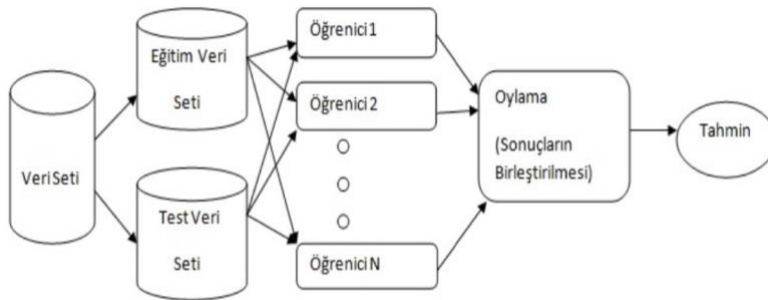
Makine öğrenimine olan ilginin artmasıyla birlikte, çoklu sınıflandırma sistemleri olarak bilinen bütünleşmiş sistemler son zamanlarda popüler hale geldi. Makine öğreniminde bütünleşmiş öğrenme, çeşitli öğrenme algoritmalarının bir araya getirilmesiyle elde edilen sonuçların performansını artırmak anlamına gelir. Problem analizi ve çok yönlü yaklaşım, topluluk öğreniminin pratik uygulamalarını öne çıkarıyor. Bu konuyla ilgili araştırmalar, doktorların bir hastalığa teşhis koyarken ortak teşhis koymas, bir kişinin hastalığı hakkında birden fazla doktordan görüş alması, bir ürünü satın almadan önce kullanıcı yorumlarını okuması ve şirketlerin stratejik karar verme sürecinde yer alması gibi olabilir. Ortak oylama sonuca karar verir. Sonuç olarak, topluluk öğrenmesi metotları, karar verme sürecinde farklı görüşleri değerlendirmeyi öğrenir ve böylece doğru kararlar verebilirler [61,62].

Topluluk metotları, aynı sorunu çözmek için birden fazla öğrenciyi eğitir. Topluluk metotlarının amacı, bir öğrenciyi eğitim verisinden oluşturmak yerine birden fazla öğrenciyi birleştirmektir [63]. Çekirdek öğrenenler, topluluk yapısının bir parçasıdır. Temel öğrenme algoritmaları genellikle karar ağaçları, sinir ağları veya diğer çeşitli öğrenme algoritmaları kullanılarak eğitim verilerine dayanırlar. Homojen bir topluluk oluşturmak için çoğu topluluk yöntemi bir temel öğrenen algoritması kullanır. Aynı şekilde, farklı türlerden heterojen öğrenenler vardır. Yani, birden fazla öğrenme algoritması kullanmak, tek bir öğrenme algoritması kullanılarak oluşturulamayacak



şekilde heterojen öğrenenler oluşturur. Grubun genelleme yeteneği, temel öğrenenlerinkinden daha iyidir.

Zayıf öğrenenler, rastgele tahmin yapmaktan biraz daha iyi olan güçlü öğrenenlere dönüşebilir. Bu nedenle, zayıf öğrenenler de temel öğrenenler olarak bilinir. Araştırmacılar genellikle zayıf öğrenenler üzerinde çalışır ve güçlü algoritmalar geliştirerek zayıftan güçlüye performansı artırmak için çalışır. Sonuç olarak, Adaboost, Bagging ve diğer yaklaşımlar ortaya çıktı ve zayıf öğrenenlerin güçlü öğrenenler olmak için neden ve nasıl geliştiklerine dair teorik bir anlayış geliştirildi [63]. Şekil 3.3'te, topluluk öğrenme yapısını göstermektedir. Bu durumda veri seti iki kategoriye ayrılır: eğitim veri seti ve test veri seti. Bu iki veri seti daha sonra çok sayıda öğrenciye uygulandı ve daha iyi öğrenenler elde edildi. Oylama sonuçlarını en iyi şekilde tahmin etmek amaçlanmıştır.

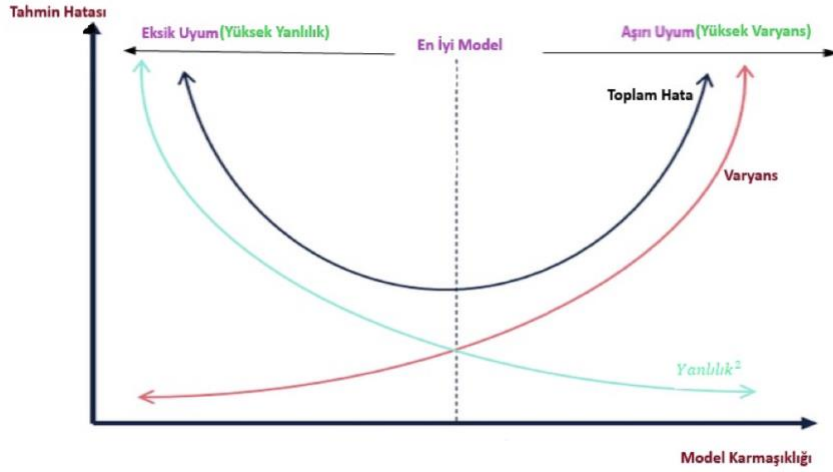


Şekil 3.3. Yaygın bir Topluluk Öğrenmesi yapısı (Kaynak: [54])

Sınıflandırma hatasının iki kontrol edilebilir bileşeni vardır. Yanlılık, sınıflandırıcının doğruluğunu ifade ederken, varyans, sınıflandırıcının farklı eğitim setlerinde nasıl hassas olduğunu gösterir. Bu iki öge arasında ters bir bağlantı vardır.

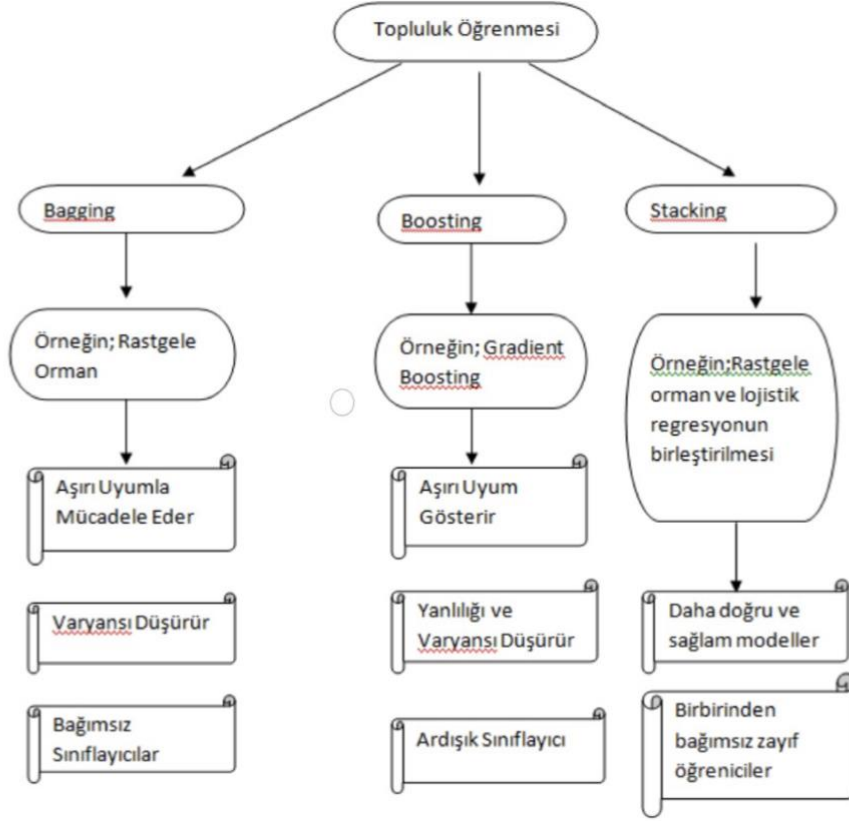
Ortalamanın varyans azaltıcı etkisi nedeniyle topluluk sistemlerinin amacı, nispeten belirlenmiş ya da benzer hatalı birkaç sınıflandırıcı oluşturarak, bu sınıflandırıcıların verilerini bir araya getirip ortalamasını belirleyerek varyansını azaltmaktır [62].

Yanlılık ve varyans hata dengesi arasındaki ilişki Şekil 3.4'te gösterilmektedir. Bu durumda, yanlılığın karesi artı varyansın toplamına eşit olan toplam hata, Ek olarak, varyans ile yanlılığın karesi arasında ters bir ilişki vardır.



Şekil 3.4. Yanlılık ve varyans hatası dengesi (Kaynak: [54])

Topluluk öğrenmesi olarak bilinen bileşik bir model, daha iyi bir sınıflandırıcı oluşturmak için birkaç düşük performanslı sınıflandırıcıyı birleştirir. Torbalama (bagging), güçlendirme (boosting) ve istifleme (stacking) topluluk makine öğrenimi yöntemlerinin temel bileşenleridir. İstifleme yöntemlerini kullanarak doğru tahminlerin olasılığını artırırken, torbalama yöntemlerini kullanarak varyansı ve yanlılığı azaltır [64]. Şekil 3.5'te, topluluk öğrenme yaklaşımını göstermektedir.

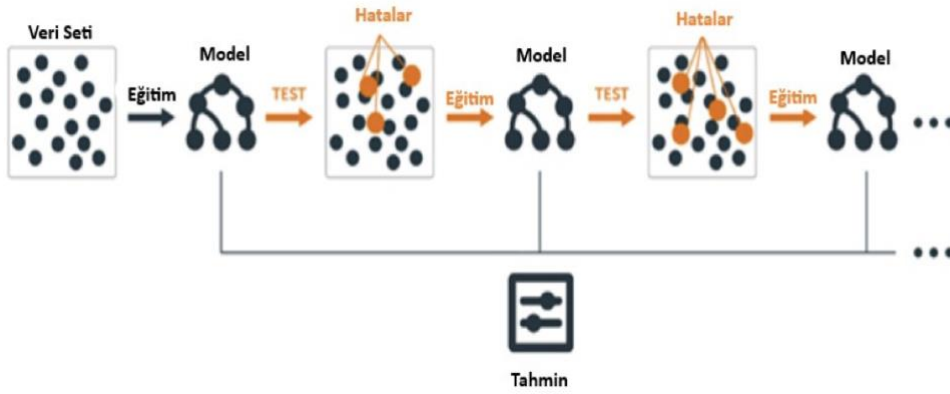


Şekil 3.5. Topluluk makine öğrenme yaklaşımlarının genel çerçevesi (Kaynak:, [54])

### Boosting Algoritması

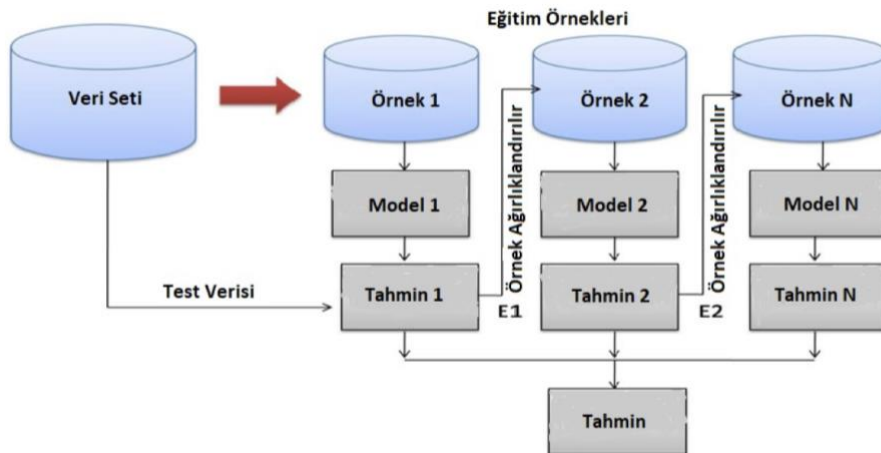
1989 yılında Schapire, güçlendirme yöntemini geliştirdi [65]. "Yeni araştırmalar kapsamında, Freund ile Schapire 1997 yılında [66] ve Schapire 2001 yılında [67] bu yaklaşımı geliştirmiştir. Quinlan'ın yeni geliştirdiği C4.5 ve C5 araç setleri, güçlendirme için çeşitli yaklaşımlara dahil edildi [68]. Torbalamaya benzer şekilde çalışan Güçlendirme, örneğe dahil edilecek gözlemleri seçmek için farklı olasılıklar kullanır [69]. Veri bilimi veya makine öğrenimi yarışmalarında güçlendirme algoritmaları popüler hale geldi. Güçlendirme algoritması, düşük hassasiyetli birden fazla modeli birleştirerek yüksek hassasiyetli bir model oluşturur. Pazarlama, kredi, sigorta ve satış gibi çok çeşitli sektörlerde kullanılmaktadır. Veri bilimi yarışmalarını kazanmak için AdaBoost, Gradient Boost ve XGBoost gibi güçlendirme algoritmaları sıklıkla kullanılır [64]. Güçlendirmenin amacı, hedefi tahmin etme yeteneğine sahip gözlemleri seçme şansını artırmaktır. Bir gözlemin bir hedefi ne kadar iyi tahmin edebileceğini değerlendirmek,

gözlemin tahmin edilen sonuçlarının gerçek sonuçlarla karşılaştırılmasına bağlıdır [69]. Güçlendirme algoritmalarının amacı, sonuçları elde etmek için sınıflandırıcıları sırayla deneyerek, daha sonraki her modelin önceki hatalarını düzeltmektir. Şekil 3.6'da, Boosting algoritmasının adımlarını gösterilmektedir.



Şekil 3.6. Boosting akış diyagramı (Kaynak: [54])

Güçlendirme algoritması ilk olarak tek bir model oluşturur. Ardından, uygun cevap vermeyen alanlara ağırlıklar ve örnekler ekleyerek ikinci bir model oluşturur. Birinci modelin başarısız olduğu örneklere odaklanan ikinci model, daha sonra üçüncü bir model oluşturur ve bu şekilde süreç devam eder. Bilinmeyen durum çıktısı, ağırlıklı ortalama ile gösterilir. Performansı her modelin ağırlığını belirler. Modelin kalitesi ağırlığı artırır [70]. Güçlendirme algoritmasının aşamaları Şekil 3.7'de gösterilmektedir.



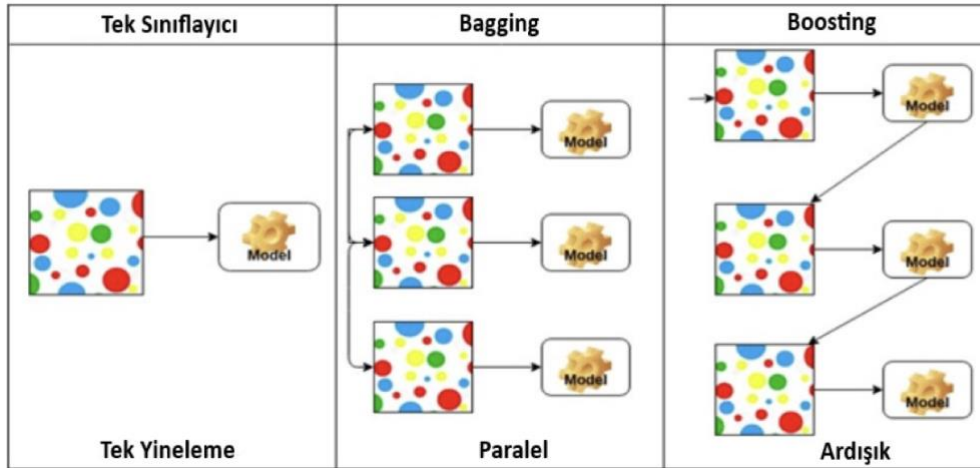
Şekil 3.7. Boosting aşamaları (Kaynak: [54])

Boosting algoritmasında aşağıdaki adımlar takip edilir.

- i. Öncelikle eğitim veri setinden rastgele örnekler oluşturulur.
- ii. Bu örnekte, bir sınıflandırıcı (örn. Model 1) tüm eğitim seti üzerinde eğitilir ve test edilir.
- iii. Her örnek tahmininin hatasını hesaplayın. Bir örneğin yanlış sınıflandırılması durumunda o örneğin ağırlığı artırılarak başka bir örnek oluşturulur.
- iv. Yüksek doğruluğa ulaşana kadar bu işlemleri tekrarlayın.

Güçlendirme algoritmaları AdaBoost, Gradient boosting modeli, XGBoost, Light GBM, CatBoost yöntemlerinden oluşmaktadır.

Şekil 3.8’de torbalama ve güçlendirme olarak bilinen topluluk öğrenme yöntemlerinin işlevlerini göstermektedir. Torbalama modeli oluşturma eğitim aşamasında paralel olarak yürütülür. Her model kendi başına oluşturulurken, yeni öğrenenler Güçlendirme aşamasında ardışık olarak oluşturulur. Gradient Boosting (Gradyan Güçlendirme) modelleri ve rastgele orman torbalama topluluk öğrenimi örnekleridir.



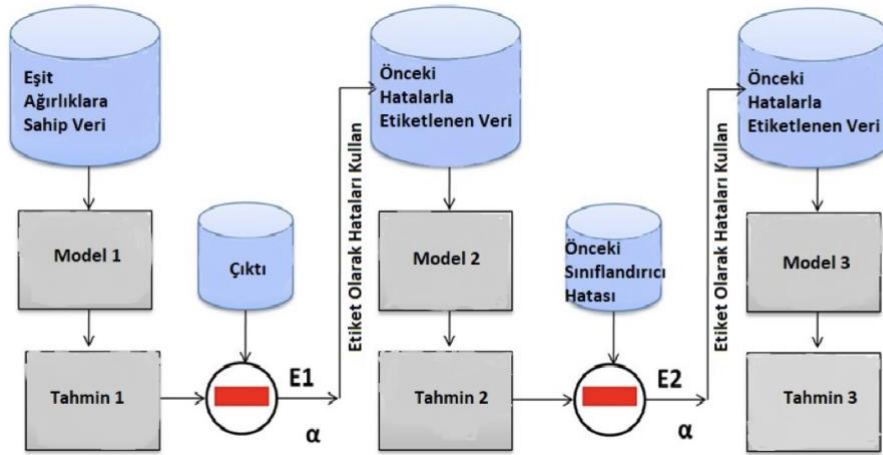
Şekil 3.8. Bagging ve Boosting işleyiş mekanizması (Kaynak: [54])

Güçlendirme, Torbalamaya göre hata oranı daha düşük olan bir sınıflandırma algoritmasıdır [71].

## Gradient Boosting Algoritması

Friedman, 2001 yılında Gradyan Güçlendirme olarak bilinen güçlü bir makine öğrenme yaklaşımı önerdi. Güçlendirme teknikleri, gradyan güçlendirmeye dayanmaktadır. Amacı, güçlülleştirme yöntemlerine dayandığından, bir dizi temel modeli sırayla oluşturmak ve bunları karmaşık bir model haline getirmek. Karar ağaçlarında da gradyan güçlendirme yaygın olarak kullanıldığından, buna gradyan güçlendirme ağaçları denir [72]. Gradyan güçlendirme aşaması Şekil 3.9'da gösterilmektedir. Bu aşamalar;

- i. İlk olarak, ilk veri setine dayalı bir karar ağacı oluşturun ve tahmin ile çıktı arasındaki hatayı bulun.
- ii. Bu hata, veri seti örneği için yeni çıkış değeri olarak kullanılır.
- iii. Veri kümesindeki hataları (örneğin etiketleri) kullanarak yeni bir karar ağacı oluşturun ve ağacı, önceki ağaç tarafından oluşturulan hataları yeniden üretecek şekilde eğitin.
- iv. Önceki çıktı ile mevcut tahmin arasındaki hata istenilen seviyeye ulaşana kadar sürece ağaç eklemeye devam edilir.



Şekil 3.9. Gradient Boosting aşamaları (Kaynak: [54])

Gradient Boosting, model optimizasyonunu zorlaştırabilecek birçok parametreye sahiptir. Bu nedenle, model seçerken ve hiperparametreleri ayarlarken çok dikkatli olun. Bu yöntemin başarısı büyük ölçüde doğru öğrenme oranının seçilmesine ve verilere göre hiperparametrelerin ayarlanmasına bağlıdır.

### 3.3.4. Destek Vektör Makineleri (SVM)

SVM, denetimli bir makine öğrenme algoritması çeşidi olup hem sınıflandırma hem de regresyon (SVR) amacıyla kullanılabilen oldukça basit ve etkili bir algoritmadır [73].

SVM, özellikle sınıflandırma amaçlı kullanılan bir makine öğrenmesi algoritmasıdır. Temel prensibi, sınıflar arasındaki sınırları en iyi şekilde belirleyen bir hiperdüzlem oluşturarak verileri bu düzleme yansıtmaktır. Bu uzayda, örneklerin sınıflandırılması için bir hiperdüzlem bulmak, aslında bir optimizasyon problemidir.

SVM, sınıfların doğru şekilde ayrılabilmesi için en uygun hiperdüzlemi oluşturmak için veri noktalarının bir kısmını kullanır. Bu veri noktaları destek vektörleri olarak adlandırılır ve hiperdüzlemi belirlemek için kullanılan en önemli verilerdir [74]. SVM, yeni veri örneklerinin (diğer) daha güvenilir bir şekilde sınıflandırılabilmesi için en büyük sınıf ayırma marjına sahip katmanı arar.

SVM sınıflandırıcı denklemini Eşitlik 3.9'deki gibidir. Verilen bir etiketli eğitim veri seti için, her  $(x_i, y_i)$ ,  $i$ 'nci eğitim örneğinin sınıf etiketi olan  $x_i$  ve öznitelik vektörü  $x_i$  ile temsil edilir.  $y_i$ , sınıf etiketi (-1 veya +1) olacak şekilde, optimal bir hiperdüzlem şu şekilde tanımlanabilir:

$$w^T x + b = 0 \quad (3.9)$$

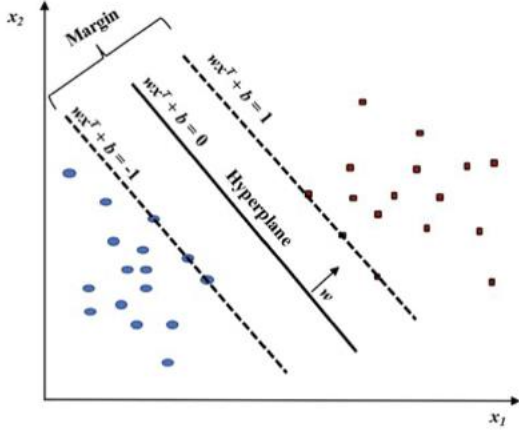
Burada  $w$  ağırlık vektörü,  $x$  girdi öznitelik vektörü ve  $b$  yanlılıktır (bias).  $w$  ve  $b$ , eğitim setinin tüm elemanları için Eşitlik 3.10 ve Eşitlik 3.11'deki eşitsizlikleri sağlamalıdır:

$$w^T x + b \geq +1 \text{ eğer } y = +1 \text{ ise} \quad (3.10)$$

$$w^T x + b \leq -1 \text{ eğer } y = -1 \text{ ise} \quad (3.11)$$

Bir SVM modelini eğitmenin amacı,  $w$  ve  $b$ 'yi bulmak böylece hiperdüzlem verileri ayırır ve marjı  $\frac{1}{\|w\|^2}$  maksimize eder.  $y_i (w^T x_i + b = 1)$  olan vektörler için  $x_i$ , destek vektörü olarak adlandırılacaktır. Kısacası, SVM'de karar sınırı, farklı sınıflardan en yakın veri noktalarına olan mesafeyi maksimize edilerek oluşturulur [75]. SVM algoritması orijinal olarak 1963 yılında Vapnik tarafından doğrusal bir sınıflandırıcı oluşturmak için

önerilmiştir. Doğrusal SVM, doğrusal olarak ayrılabilen veri kümeleri için kullanılır. Şekil 3.10'de anlaşılacağı üzere düz bir çizgi iki grubu etkili bir şekilde ayırıp sınıflandırma yapabiliyor. İki sınıf (kırmızıya karşı mavi) sınıflandırıldı.



Şekil 3.10. Doğrusal SVM modeli (Kaynak: [74])

SVM için alternatif bir kullanım, bize daha yüksek boyutlu, doğrusal olmayan modelleri modellememize olanak tanıyan çekirdek yöntemidir. Doğrusal olmayan bir problemde, bir çekirdek fonksiyonu ham verilere ek boyutlar eklemek ve bu şekilde daha yüksek boyutlu sonuç uzayında doğrusal bir problem haline getirmek için kullanılabilir. Özetle, bir çekirdek fonksiyonu, bazı hesaplamaların aksi takdirde yüksek boyutlu uzayda gerçekleştirilmesi gerekeceği durumlarda hesaplama sürecini hızlı yapılmasına yardımcı olmaktadır. İlgili Eşitlik 3.12'de gibi tanımlanmıştır.

$$K(x, y) = f(x) \times f(y) \quad (3.12)$$

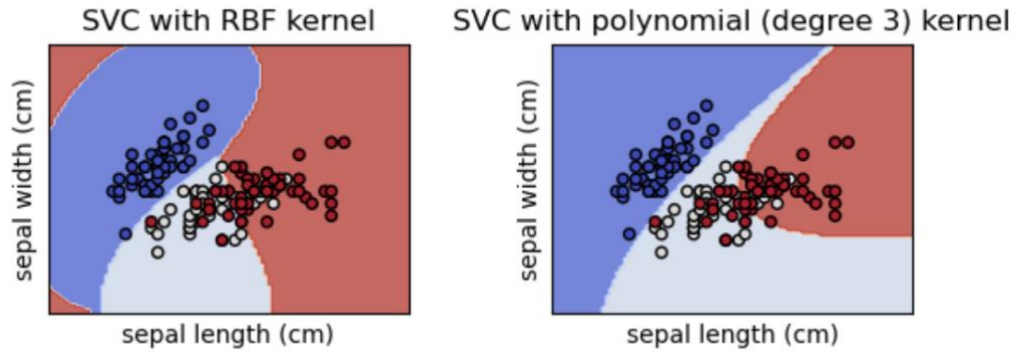
Burada K çekirdek fonksiyonu,  $x$  ve  $y$   $n$  boyutlu girdilerdir.  $f$  fonksiyonu, girdiyi  $n$  boyutlu uzaydan  $m$  boyutlu uzaya eşler.  $\times$ , çarpımını temsil eder. Bu yaklaşım, özellikle yüksek boyutlu veri analizinde ve makine öğreniminde geniş bir kullanım alanına sahiptir. Çekirdek fonksiyonları, orijinal girdi uzayındaki iki veri noktası arasındaki skaler çarpımı, yüksek boyutlu bir özellik uzayında hesaplamak için kullanılır. Bu işlem, çekirdek hilesi (kernel trick) olarak bilinir ve çoğu zaman yüksek boyutlu uzayda doğrudan hesap yapmaktan çok daha verimlidir.

Radial Basis Function (RBF) çekirdeği, yaygın kullanılan çekirdekler arasında özel bir yere sahiptir. Örneğin, RBF çekirdeği Eşitlik 3.13'te tanımlanır:



$$(K_{RBF}(x, y) = e^{(-\gamma \times \|x - y\|^2)}) \quad (3.13)$$

$\Gamma$ , (gama) bir ölçekleme parametresidir.  $\|x - y\|^2$  ifadesi,  $x$  ve  $y$  vektörleri arasındaki Öklid mesafesinin karesini belirtir. Karşılık gelen öznitelik vektörü sonsuz boyutludur. Çekirdek çeşitleri 'doğrusal', 'polinomal', 'radyal tabanlı fonksiyon', 'sigmoid' ve 'precomputed' olmak üzere beş çeşittir, ancak en çok kullanılanlar 'polinomal' ile 'radyal tabanlı fonksiyon'dur (Şekil 3.11).



Şekil 3.11. Iris veri seti üzerinde farklı SVM çekirdeklerinin performans analizi (Kaynak: [https://scikit-learn.org/stable/auto\\_examples/svm/plot\\_iris\\_svc.html](https://scikit-learn.org/stable/auto_examples/svm/plot_iris_svc.html))

Polinomal çekirdek, polinomdan gelmektedir ve eğitim verisinin polinomları üzerindeki benzerliğini temsil ederek, doğrusal olmayan modellerin öğrenilmesini sağlar.

Radyal tabanlı (RBF) çekirdek, Gauss dağılımına benzerliği nedeniyle en yaygın kullanılan çekirdeklerden biridir. Veri kümesindeki iki noktanın konumlarına göre birbirlerine benzerliklerini hesaplar.

Bir SVM modelinin performansını etkileyebilecek diğer faktörler de mevcuttur. Ancak, belirli bir desen tanıma problemi için en iyi çekirdek fonksiyonu seçmek için hiçbir yolu yoktur. En iyi çekirdeği seçmenin tek yolu denemeler yapmaktır. Basit bir 'standart' SVM ile başlayabilir ve ardından çeşitli 'standart' çekirdek fonksiyonları ile deney yapabiliriz. Problemin doğasına bağlı olarak, bir çekirdek diğerinden daha iyi olabilir. İstatistiksel olarak çapraz doğrulama kullanarak en optimal çekirdek fonksiyonu seçilir [74].

SVM, birçok farklı alan ve probleme uygulanabilir. Finansal piyasalarda, özellikle fiyat hareketlerinin yönünü tahmin etmek için kullanılmaktadır. SVM'nin avantajı, yüksek boyutlu verileri işleyebilmesi ve aşırı öğrenmeye direnebilmesidir. Bununla birlikte, doğru hiperparametre ayarları ve çeşitli çekirdek kombinasyonlarının test edilmesi gereklidir, modelin eğitimi büyük veri setleri için zaman alıcı olabilir. Son olarak, iç işleyişleri anlaşılması zor olabilir çünkü temel modeller karmaşık matematiksel sistemlere dayanmakta ve sonuçlar yorumlaması zordur [74].

SVM, hiperparametre seçimine duyarlıdır ve uygun çekirdek (kernel) seçimi model performansı açısından önemlidir. Ayrıca SVM, özellik alanındaki mesafelere duyarlı olduğundan veri ön işleme ve ölçeklendirme gerektirir. Akademik araştırmalarda SVM sıklıkla karşılaştırmalı analiz için temel bir model olarak kullanılır.

### 3.4 Sınıflandırma Performans Ölçütleri

Sınıflandırma performansını ölçmenin yaygın yollarından biri hata matrisi veya karmaşıklık matrisi kullanmaktır. Bu matris, modelin tahminlerinin gerçek etiketlerle karşılaştırılmasını sağlar ve şu terimleri içerir [46].

Gerçek Pozitifler (TP): Modelin pozitif olarak doğru tahmin ettiği durum sayısı.

Gerçek Negatifler (TN): Modelin negatif olarak doğru tahmin ettiği durum sayısı.

Yanlış Pozitifler (FP): Modelin pozitif olarak yanlış tahmin ettiği durum sayısı (I. tip hata).

Yanlış Negatifler (FN): Modelin negatif olarak yanlış tahmin ettiği durum sayısı (II. tip hata).

Bu temel terimler daha karmaşık ölçülerin hesaplanmasının temelini oluşturur [75]:

Doğruluk oranı (Accuracy), modelin hem pozitif hem de negatif sınıfları doğru tahmin etme oranıdır ve doğruluk oranının hesaplanması denklemi Eşitlik 3.14'te verilmiştir.

$$\text{Doğruluk oranı} = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (3.14)$$

Esas köşegenin toplama oranı olarak da ifade edilebilir.

Kesinlik (Precision), pozitif olarak tahmin edilen durumların gerçekten pozitif olduğu oranıdır. Kesinlik oranının hesaplanması Eşitlik 3.15 verilmiştir.

$$Kesinlik = \frac{(TP)}{(TP + FP)} \quad (3.15)$$

Duyarlılık (Recall), gerçek pozitiflerin doğru olarak ne kadarının tahmin edildiğidir. Seçicilik olarak da isimlendirilmektedir. Duyarlılık oranının hesaplanması denklemi Eşitlik 3.16'de verilmiştir.

$$Duyarlılık = \frac{(TP)}{(TP + FN)} \quad (3.16)$$

F1 Skoru (F1 Score), kesinlik ve duyarlılığı dengelemek için F1 puanı performans ölçüsünden yararlanır. Kesinlik ve duyarlılığın harmonik ortalamasıdır ve matematiksel olarak formülü Eşitlik 3.17'de ifade edilmiştir.

$$F_1 \text{ Skoru} = \frac{(2 \times Kesinlik \times Duyarlılık)}{(Kesinlik + Duyarlılık)} \quad (3.17)$$

Seçicilik (Specificity), gerçek negatiflerin doğru olarak ne kadarının tahmin edildiğidir. Seçiciliğin matematiksel ifadesi Eşitlik 3.18'de verilmiştir.

$$Seçicilik = \frac{(TN)}{(TN + FP)} \quad (3.18)$$

### **ROC Eğrisi (Receiver Operating Characteristic)**

ROC Eğrisi, modelin tüm olası eşik değerlerindeki performansını gösterir. Bu, modelin sınıflandırma yeteneğini, yanlış pozitiflerin artması pahasına doğru pozitiflerin nasıl arttığını göstererek değerlendirir. Eğri altında kalan alan (AUC Area Under the Curve), model performansının genel bir ölçüsünü verir. AUC değeri 1'e ne kadar yakınsa modelin performansı o kadar iyidir [46].

ROC Eğrisi, yatay ekseninde yanlış pozitif oranı (FPR) ve dikey ekseninde gerçek pozitif oranı (TPR) olacak şekilde çizilir [75].

Gerçek pozitif oranı (TPR), modelin pozitif olarak doğru tahmin ettiği oranıdır ve Eşitlik 3.19 ile hesaplanır.

$$TPR = \frac{(TP)}{(TP + FN)} \quad (3.19)$$

Yanlış pozitif oranı (FPR), modelin yanlışlıkla pozitif olarak tahmin ettiği oranıdır ve Eşitlik 3.20 ile hesaplanır.

$$FPR = \frac{(FP)}{(FP + TN)} \quad (3.20)$$

Yüksek TPR değerleri ve düşük FPR değerleri modelin iyi performans gösterdiğini göstermektedir. Bu performans ölçümleri, özellikle tıbbi teşhis, spam filtreleme ve finansal dolandırıcılık tespiti gibi alanlarda kullanılması önemlidir. Modelin yanlış olması, yüksek maliyetlere veya risklere yol açabileceği durumlarda özellikle değerlidir. Örneğin, bir kredi kartı sahtekarlığı tespit modelinde, yanlış pozitifler müşteri memnuniyetsizliğine yol açabilirken, yanlış pozitifler finansal kayıplara yol açabilir. Farklı modelleri veya aynı modelin farklı yapılandırılmasını karşılaştırmak için kullanılır.

Performansın kapsamlı bir değerlendirmesi için birden fazla ölçümün aynı anda dikkate alınması tavsiye edilir. ROC eğrileri ve AUC, genel model performansını değerlendirmek için faydalıdır. Ancak belirli iş ihtiyaçlarına, maliyetlere ve sonuçlar üzerindeki etkiye göre değişikliğine bağlı olarak, performansını doğru bir şekilde değerlendirmek için diğer metrikler ile kullanılmalıdır. Örneğin, ROC eğrileri özellikle veri seti dengesizse yanıltıcı olabilir. Bu durumda Kesinlik- Duyarlılık eğrisi gibi diğer metrikler daha bilgilendirici olabilir.

## 4 UYGULAMA

### 4.1. Veri Kümesi

Bu çalışmada, Bitcoin'in fiyat hareketlerini analiz etmek ve tahmin etmek için 17 Ağustos 2017 ile 3 Ocak 2024 yılları arasında elde edilen günlük fiyat verileri kullanılmıştır. Veri kümesi, Bitcoin piyasası ve çeşitli ekonomik göstergelerle ilgili detaylı bilgiler içermektedir. Bitcoin ile ilgili sütunlar, geçmişteki en yüksek fiyat, işlem hacmi, kapanış fiyatı, alıntı varlık hacmi, işlem sayısı, TB baz hacmi ve TB alıntı hacmi gibi önemli finansal göstergeleri sunmaktadır. Ek olarak, 10 yıllık Hazine Getirisi Endeksi, Fed Fonları Hedef Oranı, Ticaret Ağırlıklı dolar Endeksi, Gecelik Federal Fon Oranı, Gayri Safi Yurtiçi Hasıla Büyüme Oranı, İşsizlik Oranı ve Tüketici Fiyat Endeksi gibi makroekonomik göstergeler tezimde ele alınacaktır. Ayrıca, benzin ve altın fiyatlarının en yüksek seviyeleri ve işlem hacimleri, döviz kurları ve piyasa endekslerinin en yüksek seviyeleri ve işlem hacimleri analizlerde yer alacak, bu veriler aracılığıyla piyasalar arası dinamikler ve ekonomik göstergelerin etkileşimleri incelenecektir. Bu veriler, kripto para piyasasındaki büyük borsalardan biri olan Binance'in web sitesindeki bir API (Application Programming Interface) aracılığıyla toplanmıştır.

Tez, Bitcoin yön tahminlemesi, reel zamanlı veri erişiminin kripto para piyasalarındaki hızlı değişikliklere adaptasyon için oynadığı kritik rol üzerinde durulmaktadır. Dünya'nın en büyük Bitcoin borsası Binance, 14 Temmuz 2017 tarihinde kuruldu. Binance, yüksek işlem hacmi ve geniş kullanıcı tabanıyla kripto para piyasasının önemli bir merkezidir. Binance API'sinin sağladığı zengin tarihsel veri setlerinin, uzun vadeli trend analizleri yapma ve geçmiş verilere dayalı modeller geliştirme potansiyelini ele alacaktır. Bu yöntem, belirli zaman aralıklarında Bitcoin ile ilgili verileri sorgulayarak, bu verilerin sistemli bir şekilde toplanmasını ve kaydedilmesini sağlamaktadır. Bu tez, kripto para piyasasının daha iyi anlaşılmasına ve yatırım kararlarının etkin bir şekilde alınmasına imkân tanıyacak bilgiler sunmayı hedeflemektedir.

## 4.2. Veri Kaynakları

**Bitcoin Fiyat Verileri:** Bitcoin'in günlük açılış, kapanış, yüksek ve düşük fiyatlar gibi temel Bitcoin piyasası verilerinden oluşmaktadır. Bu veriler, zaman içindeki Bitcoin fiyat dinamiklerini anlamak için temel olarak kullanılır.

**İşlem hacmi (Volume):** Bitcoin ve diğer kripto paralarda "volume", belirli bir zaman diliminde bir piyasada işlem gören birimlerin sayısını ifade eder. Yani, bir varlığın ne kadar sık el değiştirdiğinin ölçüsüdür. İşlem hacmi, alım-satım işlemleri arasında bir alıcı ve bir satıcı olduğunda belirlenen bir fiyatta anlaşmaya varıldığında kaydedilir. Hacim, hisse senetleri, tahviller, fiat para birimleri veya kripto paralar gibi herhangi bir ticaret varlığında ifade edilebilir. Örneğin, biri 5 Bitcoin'i 20 USD'den sattığında, o işlemin hacmi ya 100 USD ya da 5 Bitcoin olarak ifade edilebilir, bu, hacmin hangi varlık biriminde ifade edildiğine bağlıdır. Aynı zamanda, yüksek bir hacim genellikle güçlü bir trendin varlığını ve yönünü gösterirken, düşük hacim daha zayıf bir eğilimi gösterebilir [76].

**En yüksek işlem fiyatı (High):** Belirli bir zaman dilimi içinde kaydedilen en yüksek işlem fiyatı. Bu, piyasanın o dönemde ulaştığı maksimum değeri gösterir.

**Kapanış fiyatı (Close):** Zaman diliminin sonunda gerçekleşen son işlem fiyatı. Bu, belirlenen zaman diliminin sonunda piyasanın kapandığı fiyatı belirtir.

**Alıntı Varlık Hacmi (Quote Asset Volume):** Alıntı yapılan varlık (genellikle fiyatlandırma yapılan para birimi) cinsinden hesaplanan işlem hacmi. Bu, işlemlerin değer bazında ne kadar büyük olduğunu gösterir.

**Toplam İşlem Sayısı (Number of Trades):** Belirli bir zaman diliminde gerçekleştirilen toplam işlem sayısı. Bu değer, piyasanın likiditesi ve aktivitesi hakkında bilgi verir.

**Alıcı Baz Hacmi (TB Base Volume):** "Taker Buy Base Volume" olarak da bilinir. Bu, piyasa yapıcı olmayan alıcılar tarafından alınan baz varlık miktarını gösterir. Yani, fiyatı belirleyen alıcıların aldığı miktarı ifade eder. Bu, piyasa emirlerini (market orders) kullanan alıcıların piyasada ne kadar etkili olduklarını gösterir ve genellikle piyasa hareketlerinin gücünü analiz etmek için kullanılır.

**Alıcı Alıntı Hacmi (TB Quote Volume):** "Taker Buy Quote Volume" olarak da bilinir. Bu, piyasa yapıcı olmayan alıcılar tarafından alıntı yapılan para birimi cinsinden ödenen toplam miktarı gösterir. Bu, alımların para birimi değerini yansıtır. Örneğin, bir BTC/USD döviz çiftinde, alıcıların USD cinsinden ne kadar harcama yaptığını gösterir. Bu değer, piyasa emirlerini kullanarak alış işlemi yapan yatırımcıların piyasaya ne kadar para akıttıklarını gösterir ve piyasa hareketlerinin mali büyüklüğünü anlamada kullanılır.

**Ekonomik Göstergeler:** FED faiz oranları, büyüme oranları, işsizlik oranları ve enflasyon oranları gibi makroekonomik veriler ile daha kapsamlı bir analiz sağlaması amacıyla modele eklenmiştir. Özellikle FED faiz oranları gibi faktörler, yatırımcıların risk algısını ve dolayısıyla daha riskli varlıklar olan kripto paralara olan ilgisini doğrudan etkileyebilir. Bu göstergeler, makroekonominin Bitcoin fiyat hareketleri üzerindeki olası etkisini değerlendirmek için yol gösterici niteliğe sahiptir.

**Enerji ve Değerli Maden Fiyatları:** Petrol ve altın fiyatları gibi küresel ekonomik koşulların ve risk iştahının göstergesi olarak modele dahil edilmektedir. Bu tür varlıklar, ekonomik belirsizlik zamanlarında sığınak olarak görüldüğü için, piyasalar üzerinde doğrudan etkili olabilir. Örneğin, Öte yandan, altın fiyatlarındaki artış, yatırımcıların daha güvenli varlıklara yönelme eğiliminde olduklarını ve genel piyasa riskinden kaçınma isteklerini ortaya koyar.

Ayrıca Çizelge 4.2'de, analizlerde kullanılan değişkenlerin İngilizce ve Türkçe adlarına yer verilmiştir.

Çizelge 4.2. Değişkenlerin İngilizce ve Türkçe adları

<b>İngilizce Değişken Adları</b>	<b>Türkçe Değişken Adları</b>
hist_df_High	Geçmiş Yüksek Fiyat
hist_df_Volume	Geçmiş Hacim
hist_df_Close	Geçmiş Kapanış Fiyatı
hist_df_Quote Asset Volume	Geçmiş Alıntı Varlık Hacmi
hist_df_Number of Trades	Geçmiş İşlem Sayısı
hist_df_TB Base Volume	Geçmiş TB Baz Hacmi
hist_df_TB Quote Volume	Geçmiş TB Alıntı Hacmi
data_T10YIE	10 Yıllık Enflasyon Beklentisi Oranı
data_FEDFUNDS	Federal Fonlar Faiz Oranı
data_DTWEXB	Dolar Ticaret Ağırlıklı Döviz Endeksi
data_DFF	Günlük Federal Fon Oranı
gdp_growth_data	GSYİH Büyüme Oranı
unemployment_data	İşsizlik Oranı
inflation_data_FPCPITOTLZGUSA	Enflasyon Oranı
gas_prices_data_High	Benzin Yüksek Fiyat
gas_prices_data_Volume	Benzin Hacmi
exchange_rate_data_DEXUSEU	Dolar/Euro Döviz Kuru
endeks_verisi_High	Endeks Yüksek Değeri
endeks_verisi_Volume	Endeks Hacmi
gold_price_data_High	Altın Yüksek Fiyatı
gold_price_data_Volume	Altın Hacmi
Target	Hedef (0: "Azalacak", 1: "Artacak")

Bu tür değişkenler, Bitcoin'in fiyat hareketlerini tahmin etmede sıkça kullanılır ve çeşitli makine öğrenimi teknikleri ile bu tahminlerin gerçekleştirilmesi mümkündür. Fiyat tahmin modellerinin doğruluğunu artırmak için geçmiş verilerden öğrenen algoritmalar kullanılabilir [77].



### 4.3. Toplama Süreci

Veri toplama süreci, Python programlama dili kullanılarak Binance'nin API'si aracılığıyla Bitcoin verilerinin toplanmıştır. API kullanılarak veri çekilmesinin birçok faydası olmuştur ve bunlar şu şekilde özetlenebilir: API kullanımı ile, verilerin manuel olarak toplanmasına kıyasla çok daha hızlı ve otomatik bir şekilde toplanmasını sağlar. Bu, veri toplama sürecini basitleştirir ve insan hatasını azaltır. Böylelikle toplanan verilerin güvenliği konusunda yüksek standartlara sahiptir. Verilerin Binance'dan alınmasının sebebi, dünyanın en büyük kripto para borsalarından biri olduğu için, sunulan verilerin hem güncel hem de doğru olduğunu garanti ettiğindedir. Bu, piyasa analizi ve karar verme süreçleri için kritik öneme sahiptir. Dahası API'ler, fiyatlar, hacimler, işlem geçmişi ve çok daha fazlası gibi çeşitli veri türlerine erişim sağlar. Bu çeşitlilik, kapsamlı ve çok yönlü analizler yapmayı mümkün kılar. En sonunda da reel zamanlı analizi mümkün kılar. Daha sonra yapılacak olan Bitcoin yönünün tahminlemesi için, anlık veriler sayesinde, fırsatları değerlendirilebilir.

### 4.4. Veri Temizleme

Veri işleme ve temizleme süreciniz, bir dizi kritik adımdan oluşmaktadır. Elde edilen ham veriler, çeşitli aşamalarından geçer ve bu da onu Python programlama dilinde yazılmış script dosyaları kullanılarak analize uygun hale getirmiştir. Öncelikle, Binance API'si üzerinden kripto para birimi verileri toplanmış, bu verilerde fiyat, hacim ve diğer işlem detayları gibi bilgiler bulunmaktadır, Elde edilen verileri işlerken, her bir veri çerçevesinden gelen sütunlara ilgili veri çerçevesinin adı eklenerek veriler birleştirilmiştir. Her biri tarih sütununu baz alınarak birleştirildi. Bu adım, farklı kaynaklardan gelen veriler arasındaki ilişkilerin daha iyi anlaşılmasına olanak tanır. Verilerin farklı periyotlarda güncellendiği ve bazılarının aylık, bazılarının yıllık ve bazılarının da haftalık olarak kaydedildiği için eksik veriler mevcuttur. Örneğin, işsizlik oranı aylık olarak güncellenirken, Gayrisafi Yurtiçi Hasıla (GDP) büyüme oranı çeyreklik periyotlarda güncellenir. Bu nedenle, eksik veriler, ilgili periyotlardaki önceki değerlerle 'fillna' metodu kullanılarak doldurulmuştur. Böylece veri setindeki boşluklar doldurularak analizlerin doğruluğu artırılmıştır. Bu işlemten sonra kalan eksik değerler silinip modellere hazır hale getirilmiştir. Toplanan tarihsel veriler datetime formatına çevrilmiş ve numerik sütunlar sayısal tipe dönüştürülerek veriler analiz için uygun hale

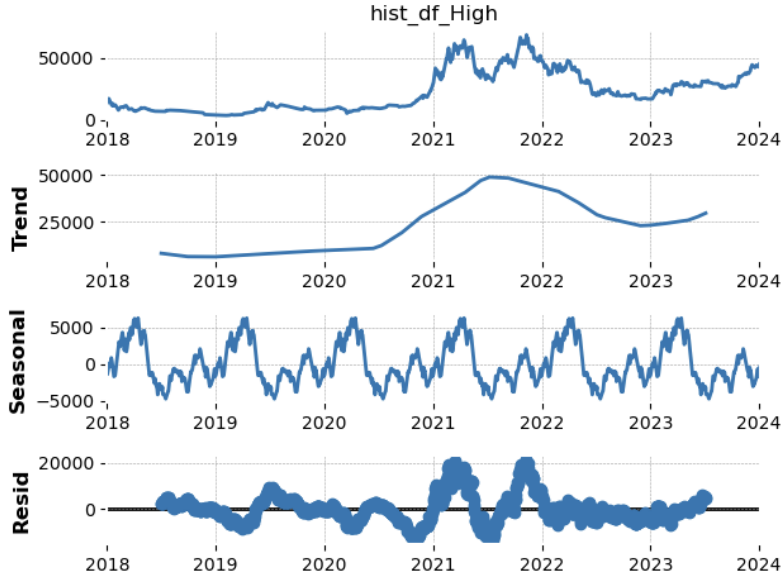
getirilmiştir. Son olarak, seçilen sütunları içeren bir DataFrame oluşturularak veriler modellenme ve daha ileri analizler için yapılandırılmıştır. Bu adımlar, veri bilimi projelerinde veri toplama, temizleme, entegrasyon ve hazırlık işlemlerini kapsayarak projenin analiz aşamasına geçişi için verileri hazır hale getirilmiştir. Z skoru kullanarak veri setindeki anomalileri tespit etmek için bir işlem gerçekleştirdik. İlk adımda, scipy kütüphanesinden stats modülünü ve numpy kütüphanelerini kullandık. Daha sonra, veri çerçevesindeki 'En yüksek işlem fiyatı' sütunundaki değerlerin Z-skorlarını hesapladık ve 'z\_score' adında yeni bir sütun oluşturduk. Ardından, Z-skoru 3'ten büyük olan değerleri içeren bir 'anomalies' veri çerçevesi oluşturdu ve herhangi bir anomali tespit edilemedi

#### **4.5. Veri Dönüştürme**

Veri dönüştürme süreci, özellik mühendisliği olarak bilinir. Bu süreç, verilerin analiz için daha uygun bir formata dönüştürülmesini içerir. Örneğin, tarih ve saat bilgileri zaman serisi analizi için zaman damgalarına dönüştürülebilir veya kategorik veriler modellenme sürecinde kullanılmak üzere sayısal forma dönüştürülebilir. Örneğin, veriler sürekli nicel değişken iken kullanacağım model haline getirmek için bağımlı değişkenimi şu şekilde hazırlanmıştır. Öncelikle kapanış fiyatlarının bir önceki güne göre değişimini hesaplamak için diff() fonksiyonu kullanılıp ve sonuç 'Change' sütununa kaydedilir. Ardından, analiz yönüne göre olduğu için 'Target' sütunu yani bağımlı değişkenimiz oluşturulur. Bu sütunda, artış durumu için 1 ve azalış veya değişmeme durumu için 0 olarak kodlama yapılır. Son olarak, gereksiz olan Change sütunu düşürülür. Bu işlem, fiyatların trendini veya yönelimini belirlemek için kullanılabilir için yapıldı. Yani, fiyatların bir önceki güne göre arttığı durumları veya azaldığı durumları tespit etmek için hazırlandı. Bu tür bir özellik mühendisliği, modelime fiyat hareketlerini tahmin etme yeteneğini kattı.

#### **4.6. Analiz Sonuçları**

Şekil 4.11'de, zaman serisi verilerini analiz etmek için kullanılan çeşitli bileşenler gösterilmektedir. Dört alt grafikten oluşan bu görsel, bir finansal veya ekonomik veri serisinin çeşitli özelliklerini temsil etmektedir.



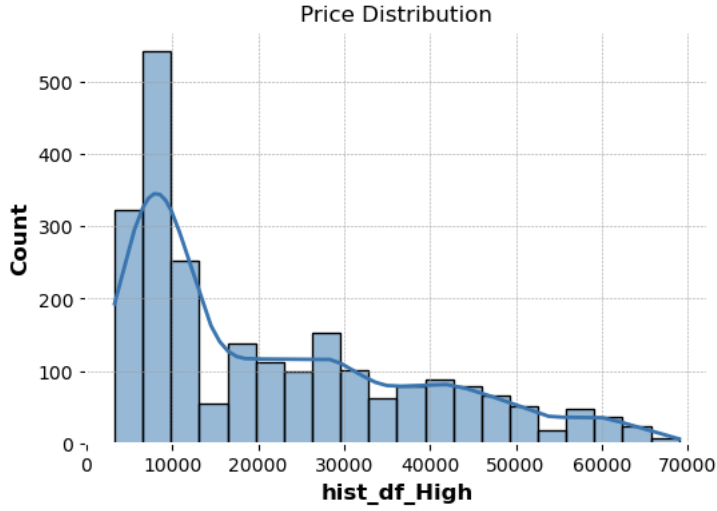
Şekil 4.11. En yüksek Bitcoin fiyatının zamana göre değişimi

1. **Zaman Serisi Grafiği:** Bu grafik zaman içinde Bitcoin fiyatının değişimini göstermektedir. 2018'den 2024'e kadar olan süreçte genel bir artış eğilimi göstermektedir. Bu süre zarfında, özellikle 2021 yılında belirgin bir artış göze çarpmaktadır ve bu artış 2022'ye kadar devam etmiştir. 2022 yılından itibaren ise veriler daha stabil bir seyir izlemeye başlamıştır. Pandemi sürecinin başladığı 2020 ve 2021 yılları arasında finansal piyasalarda ve çeşitli sektörlerde ciddi volatilité gözlemlenmiştir, bu da grafikteki dalgalanmaları açıklayabilir.
2. **Eğim (Trend) Grafiği:** Bu, verilerin detrended (eğimden arındırılmış) hali olup, genel eğilimi gösterir. 2020 ve 2022 yılları arasında bir zirve noktasına ulaştıktan sonra, trendde bir düşüş başlamış. Bu, veri setinin uzun vadeli hareketini daha net bir şekilde görmemizi sağlamaktadır. Grafikte 2020 yılı itibariyle gözlemlenen büyük dalgalanmalar dikkat çekici. Bu, belki de COVID-19 pandemisi gibi olağan dışı ve tahmin edilemeyen olayların etkisini yansıtıyor olabilir. Ayrıca 2020 yıl sonrası artan sıklıkta görülen doğal afetler (yangınlar, sel baskınları) da ekonomik aktiviteler üzerinde etkili olmuş ve global piyasalardaki dalgalanmaları tetiklemiş olabilir.

3. **Mevsimsellik Grafiđi:** Mevsimsel dalgalanmaları gösteren bu grafik, yıl içindeki belirli zamanlarda verilerin nasıl deđiřtiđini gösterir. Dalgaların, yılın belirli dönemlerinde artış veya azalış gösterdiđi, mevsimsel etkileri veya düzenli dalgalanmaları temsil ediyor olabilir.
4. **Artık Grafiđi:** Artık bileřeni gösterir ve model tarafından açıklanamayan rastgele varyasyonları içerir. Bu grafikteki dalgalanmalar, veri setindeki diđer modeller veya tahminler tarafından hesaba katılmayan oynaklıkları yansıtır. Genel olarak, kalıntıların düşük ve istikrarlı olduđu dönemler modelin verileri iyi tahmin ettiđini gösterirken, dalgalanmalar modelin bazı dinamikleri yakalayamadıđı veya verideki olađan dıřı deđiřiklikleri gösteriyor olabilir. 2019, 2021 ve 2022 yıllarındaki artışlar oldukça dikkat çekici. 2019 yařanan ABD/Çin Ticaret Savařı sonucu bu dönemde bařlayan ticaret gerilimleri, global ticaret akıřlarını ve piyasaları etkiledi. Özellikle 2021'de, ip krizinin otomotiv ve teknoloji sektörlerindeki üretim kesintileri bu grafikte belirgin dalgalanmalar yarattı. 2022 yılında Rusya-Ukrayna Savařı (2022 bařlangıcında yařanan bu çatıřma, global enerji piyasaları ve tahıl ihracatında önemli kesintilere neden oldu.

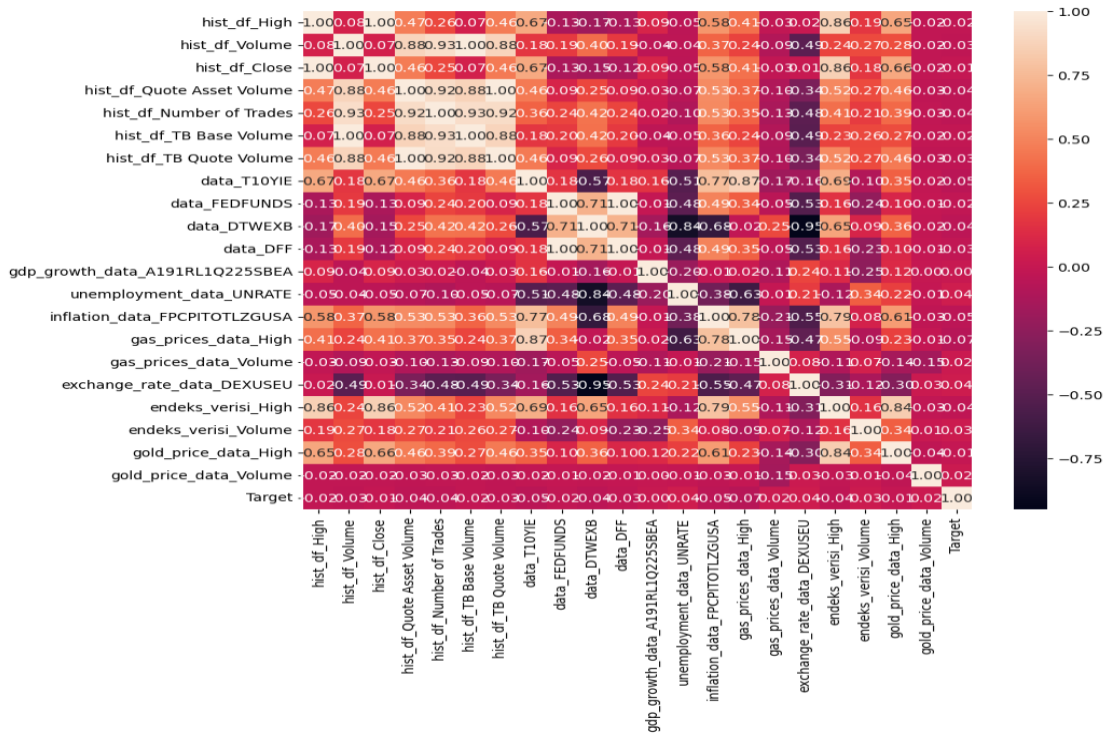
Genel olarak, bu grafikler bir veri serisinin zaman içindeki hareketlerini, trendlerini, mevsimsel etkilerini ve rastgele dalgalanmalarını ayrıntılı bir řekilde analiz etmek için kullanılabilecek yararlı araçlar sunar. Bu tür bir analiz, özellikle ekonomik veriler, finansal piyasalar veya diđer süreklilik gösteren veri serilerinde karar verme süreçlerinde önemli bir role sahiptir.

řekil 4.12'de fiyatın dađılımını gösterilmektedir. Bu grafikte, fiyatların hangi aralıklarda ne sıklıkta olduđunu görmek mümkün.



Şekil 4.12. En yüksek Bitcoin fiyat dağılımı

Özellikler arasındaki ilişkileri anlamak için bir korelasyon matrisi oluşturulmuştur. Bu matris, özellikler arasındaki ilişkileri ve bu ilişkilerin gücünü görsel olarak gösterir. Ayrıca, Şekil 4.13’te belirli bir hedef değişkenle (örneğin, “Target”) özellikler arasındaki korelasyon hesaplanmış ve istatistiksel olarak anlamlı olan özellikleri belirlenmiştir. Örneğin, bağımsız değişkenimiz olan Target kategorik bir değişken olduğu için, diğer değişkenlerimizde sürekli nicel değişkenler olduğu için iki değişken arasındaki ilişkiye Çift serili korelasyon ile hesaplanmıştır [78].



Şekil 4.13. Değişkenler arasındaki ilişkiye ait ısı grafiği

Sonuç olarak, Federal Fon Oranı ve Federal Fon Etkin Oranı değişkenleri, %5 anlamlılık düzeyinde bağımlı değişken 'Hedef' ile anlamlı bir ilişki göstermiştir.

#### 4.7. Model Eğitimi ve Test Süreçleri

Eğitim verisi, modelin öğrenme sürecinde kullanılan veri setidir ve modelin parametrelerinin ayarlanmasında temel bir rol oynamaktadır. Bu veri seti, modelin gerçek dünya verileri üzerinde nasıl performans göstereceğini öğrenmesini sağlamaktadır. Eğitim süreci sırasında, model yapılandırılır ve optimize edilirken, genelleştirme yeteneği ve aşırı uyum gibi potansiyel sorunlar açısından temel bir değerlendirme mekanizması işlevi görür.

Bu kapsamda, test verisi modelin eğitim sürecinde görmediği verilerle performansını ölçmek için ayrılarak, modelin gerçek dünya verileri üzerindeki performansının daha doğru bir şekilde tahmin edilmesini sağlar. Test verisi kullanımı, modelin genelleştirme yeteneğinin değerlendirilmesinde kritik bir öneme sahiptir.

Analizlerde zaman serisi verilerinin doğası göz önünde bulundurularak "Time Series Split" yöntemi kullanılmıştır. Bu yöntem, verileri zaman sırasına göre ayrılmasını sağlar, modelin zamanla değişen eğilimleri ve mevsimsellikleri öğrenmesine olanak tanımaktadır. Geleneksel çapraz doğrulama yöntemleri, zaman serisi verileri için uygun olmayabilir; verilerin rastgele karıştırılmasına dayanır ve bu durum zaman serilerinin yapısını bozabilmektedir. Time Series Split, verileri kronolojik sırayla eğitim ve test kümelerine ayırmaktadır. Analizde, Time Series Split yöntemi kullanılarak veri seti ile beş farklı bölüme ayrılmıştır. Bu yöntem, beş aşamalı bir çapraz doğrulama sürecine (5 katlamalı çapraz doğrulama) imkân tanır. Her bir bölme işlemi, veri setini bir öncekinden farklı bir şekilde eğitim ve test seti olarak iki kısma ayırır. Çapraz doğrulamanın her adımında, ilk bölme, veri setinin ilk %20'lik dilimini test seti olarak kullanmaktadır. Geri kalan %80'lik kısım ise eğitim seti olarak ayrılır. İkinci bölme işlemi, veri setinin ilk %40'ını eğitim seti olarak kullanırken, sonraki %20'lik bölümü test seti olarak ayırır. Bu süreç, her bölme için eğitim setinin kapsamını artırarak ve test setini sonraki bölüm olarak belirleyerek devam etmektedir.

#### **4.8 Model Sonuçları**

Analiz kapsamının temel amacı, Bitcoin fiyat eğilimlerini ve piyasa dinamiklerini tahmin etmektir. Bu nedenle çeşitli istatistiksel yöntemler ve makine öğrenme teknikleri kullanılmıştır. Klasik lojistik regresyon, eğri düzleştirmeli lojistik regresyon ve LOESS lojistik regresyon, karar ağaçları (CART algoritması), destek vektör makineleri (SVM) ve gradient boosting, gibi modelleri kullanarak analizler gerçekleştirilmiştir.

Bu analizlerin eğitim verisine ilişkin performans metrikleri ve sonuçları, Çizelge 4.3'te detaylı bir şekilde sunulmuştur.

Çizelge 4.3. Eğitim verisi performans metrikleri ve sonuçları

Model Adı	Doğruluk Oranı	Seçicilik	Duyarlılık	F1 Skoru
Lojistik Regresyon	0.683	0.680	0.731	0.7045
Eğri Düzleştirmeli Lojistik Regresyon	0.8805	0.8756	0.8962	0.8858
LOESS Lojistik Regresyon	0.4829	0.5000	0.5189	0.5093
Karar Ağaçları	1.0	1.0	1.0	1.0
SVM	0.517	0.517	1.0	0.6817
Gradient Boosting	0.961	0.967	0.958	0.9621

Çizelge 4.3'te yer alan modellerin eğitim veri seti üzerindeki performansı farklılıklar göstermektedir.

Klasik Lojistik Regresyon modeli, dengeli bir performans sergileyerek doğruluk oranında %68.3, seçicilikte %68.0 ve F1 skorunda %70.45 gibi oranlar elde etmiştir. Duyarlılık bakımından %73.1 ile diğer metriklere göre nispeten yüksek bir sonuç vermiştir.

Eğri Düzleştirmeli Lojistik Regresyon modeli, incelenen tüm modeller arasında en yüksek performansı göstermiştir. Bu model, %88.05 doğruluk oranı, %87.56 seçicilik ve %89.62 duyarlılık ile dikkat çekicidir. Ayrıca F1 skoru da %88.58 ile oldukça yüksektir.

Diğer yandan, LOESS Lojistik Regresyon modeli bu performans ölçütleri arasında en düşük sonuçları almıştır; %48.29 doğruluk oranı, %50.0 seçicilik ve %51.89 duyarlılık ile sınırlı kalmıştır. Bu modelin F1 skoru da %50.93 olarak hesaplanmıştır.

Karar Ağaçları modeli mükemmel bir performans sergilemiş ve bütün metriklerde %100 oranında sonuç elde etmiştir. Ancak bu durum, modelin eğitim verisine aşırı uyum sağladığını ve gerçek dünya verilerinde beklenen performansı gösteremeyeceğini düşündürülebilir.



Destek Vektör Makineleri (SVM) modeli düşük doğruluk (%51.7) ile sınırlı bir başarı gösterirken, duyarlılıkta %100 ile tam başarı sağlamıştır. Fakat bu, modelin öngörülerinin dengesizliğine işaret edebilir. SVM'in F1 skoru %68.17 olarak hesaplanmıştır.

Son olarak, Gradient Boosting modeli, %96.1 doğruluk oranı ve %96.21 F1 skoru ile kuvvetli bir performans sergilemiştir. Seçiciliği %96.7 ve duyarlılığı %95.8 olarak belirlenmiştir.

Bu sonuçlar, modellerin eğitim verisi üzerindeki öğrenme kapasiteleri ve genelleştirme yeteneklerinin değerlendirilmesinde önemli farklılıklar göstermektedir. Özellikle, Karar Ağaçları modelinin mükemmel sonuçları aşırı uyuma işaret edebilir, bu da modelin yeni verilere karşı kırılgan olabileceği anlamına gelir. Eğri Düzleştirmeli Lojistik Regresyon ve Gradient Boosting modellerinin sağlam performansları, bu tekniklerin karmaşık veri yapıları ile başa çıkma konusundaki potansiyellerini ortaya koymaktadır. Bununla birlikte, bu modellerin test verisi üzerindeki performanslarına da bakılmalıdır, çünkü bu, genelleştirme hataları ve aşırı uyum hakkında daha fazla bilgi sağlayacaktır.

Bu analizlerin test verisine ilişkin performans metrikleri ve sonuçları, Çizelge 4.4'te detaylı bir şekilde sunulmuştur.

Çizelge 4.4. Test verisi performans metrikleri ve sonuçları

<b>Model Adı</b>	<b>Doğruluk Oranı</b>	<b>Seçicilik</b>	<b>Duyarlılık</b>	<b>F1 Skoru</b>
Lojistik Regresyon	0.7927	0.6562	0.7778	0.7119
Eğri Düzleştirmeli Lojistik Regresyon	0.8171	0.7727	0.6296	0.6939
LOESS Lojistik Regresyon	0.3293	0.3793	1.0000	0.4954
Karar Ağaçları	0.5244	0.3500	0.5185	0.4179
SVM	0.3293	0.3293	1.0000	0.4954
Gradient Boosting	0.6951	0.5556	0.3704	0.444

Test veri seti üzerindeki performans metriklerini içeren Çizelge 4.4, modellerin gerçek dünya verileri üzerindeki genelleştirme kapasitelerini ortaya koymaktadır. Klasik Lojistik Regresyon modeli, %79.27 doğruluk oranı ile iyi bir genelleştirme yaparken, seçiciliği %65.62 ve F1 skoru %71.19 olarak ölçülmüş, bu da modelin dengeli bir performans sergilediğini göstermektedir. Duyarlılık ise %77.78 ile oldukça yüksektir.

Eğri Düzleştirmeli Lojistik Regresyon modeli, test setinde %81.71 doğruluk oranı ile en yüksek doğruluğu göstermiş, ancak duyarlılıkta %62.96 ile eğitim setine göre düşüş gözlemlenmiştir. Bu durum, modelin belirli sınıfları öngörmeye zorlandığını işaret edebilir. Seçiciliği %77.27 ve F1 skoru %69.39 ile yine sağlam bir performans sergilemiştir.

LOESS Lojistik Regresyon ve SVM modelleri, %32.93 doğruluk oranı ile düşük performans sergilemişlerdir. Her iki model de duyarlılıkta %100 elde etmişlerdir, bu da tüm pozitif sınıfları doğru olarak belirlediklerini, ancak bu yüksek duyarlılığın düşük seçicilik ile dengelendiğini gösterir. Bu, modellerin aslında birçok yanlış pozitif öngöründe bulunduğu ve genel doğruluğun düşük olduğuna işaret eder. F1 skorları ise sırasıyla %49.54 olarak hesaplanmıştır.

Karar Ağaçları modeli, test setinde %52.44 doğruluk oranı ile orta seviye bir performans göstermiş, seçicilik ve duyarlılıkta sırasıyla %35.00 ve %51.85 ile daha düşük değerler elde etmiştir. F1 skoru %41.79 olarak belirlenmiştir.

Gradient Boosting modeli, %69.51 doğruluk oranı ile iyi bir genelleştirme yeteneği sergilemiş ancak duyarlılıkta %37.04 ile sınırlı kalmıştır. Seçicilik %55.56 ve F1 skoru %44.44 ile diğer modellere kıyasla daha düşük bir performans göstermiştir.

Genel olarak, Eğri Düzleştirmeli Lojistik Regresyon modeli test seti üzerinde en yüksek doğruluk oranını sağlamıştır ve Klasik Lojistik Regresyon modeli ile karşılaştırıldığında, genel performans açısından daha dengeli bir sonuç ortaya koymuştur. LOESS Lojistik Regresyon ve SVM modellerinin yüksek duyarlılıkları, belki de yanlış pozitifleri çok sayıda tahmin ederek tüm pozitif durumları yakaladıklarını göstermektedir, bu da pratik uygulamalarda sınırlı kullanımlarını işaret edebilir. Gradient Boosting ve Karar Ağaçları modelleri, diğer modellere göre daha düşük genelleştirme performansı sergilemişlerdir. Bu sonuçlar, her bir modelin gerçek dünya verilerindeki uygulanabilirliğini değerlendirmede önemli bir rol oynar.

Eğitim veri seti üzerindeki performansa kıyasla test veri setinde düşen performans, bir modelin aşırı uyum sergilediğini gösterir. Aşırı uyum, modelin eğitim verilerini 'ezberlemesi' ve yeni verilere iyi genelleme yapamaması durumudur. Yüklene tablolarından yola çıkarak, her modelin eğitim ve test performansları arasındaki farka bakarak hangi modellerin aşırı uyum gösterip göstermediğini belirleyebiliriz.

Karar Ağaçları modeli eğitim setinde %100 doğruluk, seçicilik ve F1 skoru ile mükemmel bir performans sergilemişken, test setinde bu oranlar sırasıyla %52.44, %35.00 ve %41.79'a düşmüştür. Bu büyük düşüş, Karar Ağaçları modelinin aşırı uyum sergilediğini göstermektedir.

Diğer modellerin eğitim seti ve test seti performansları karşılaştırıldığında, örneğin Gradient Boosting modeli eğitim setinde %96.1 doğruluk oranına sahipken, test setinde %69.51'e düşmüştür. Bu da bir miktar aşırı uyumun varlığına işaret edebilir, ancak Karar Ağaçları modelindeki kadar belirgin değildir.

Eđri Düzleřtirmeli Lojistik Regresyon modelinin eđitim setindeki dođruluk oranı %88.05 iken, test setinde %81.71 olarak ölçülmüş, bu da modelin bazı ölçüde aşırı uyum sergileyebileceđini, ancak yine de diđer modellere göre daha iyi genelleřtirme yaptığını göstermektedir.

Klasik Lojistik Regresyon modeli eđitim ve test seti performansları arasındaki fark en az olan modeldir ve bu durum, daha sađlam bir genelleřtirme kapasitesine sahip olduđunu gösterebilir. LOESS Lojistik Regresyon ve SVM modelleri için, her iki setteki performanslar benzer olduđundan, aşırı uyumdan ziyade, genel olarak düşük performans sergiledikleri sonucuna varabiliriz. Bu modellerin yüksek duyarlılık oranları, sınıf dengesizliđinden veya modelin yapılandırılmasından kaynaklanan diđer sorunlardan ötürü olabilir.

Bu çalışmada, Bitcoin para biriminin fiyat hareketleri, klasik istatistiksel yöntemlerin yanı sıra çeřitli makine öğrenmesi modelleri kullanılarak incelenmiştir. Elde edilen bulgular, analiz edilen modellerin Bitcoin pazarındaki fiyat hareketlerini tahmin etme konusunda çeřitli başarı düzeylerine sahip olduđunu ortaya koymaktadır. Eđitim ve test veri setlerinde gerçekleştirilen analizler sonucunda, özellikle Eđri Düzleřtirmeli Lojistik Regresyon modelinin, diđerlerine kıyasla hem eđitim hem de test veri setlerinde daha dengeli ve tutarlı bir performans sergilediđi gözlemlenmiştir. Ancak, Karar Ađaçları modelinde görülen performans kaybı, aşırı uyumun somut bir örneđi olarak karşımıza çıkmakta ve modelin genelleřtirme kapasitesinin sınırlı olduđunu göstermektedir.

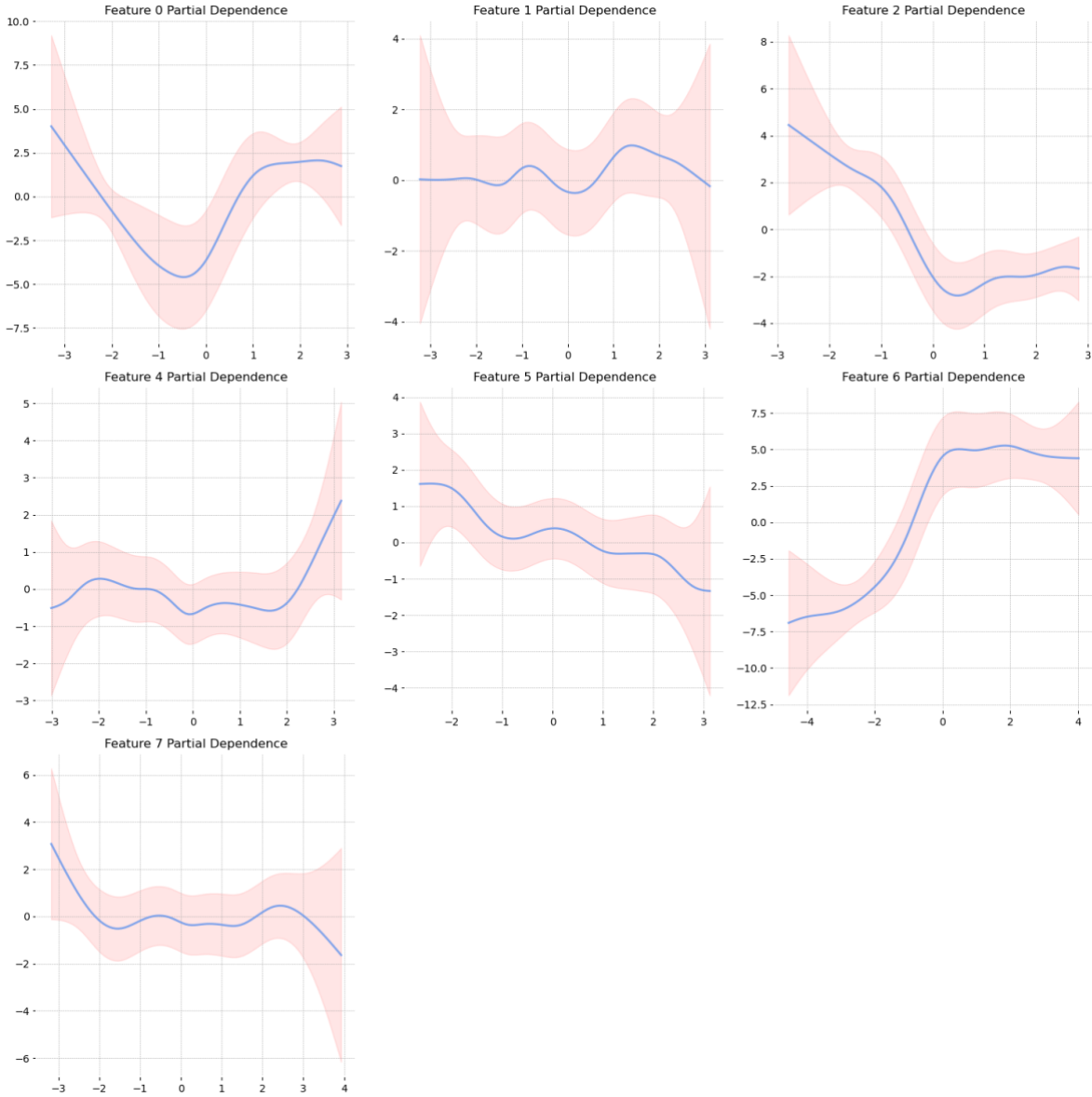
Diđer paketlerde olduđu gibi, bu pakette “summary ()” fonksiyonu kullanılarak katsayı deđerlerine ulaşamıyoruz. Bunun yerine farklı bir yöntem denendiđinde, 421 satır içerisinden katsayı deđerleri elde edilmektedir. Çizelge 4.5’te modelde kullanılan deđişkenlerin düzeltme katsayılarının ve anlamlılık düzeyleri verilmiştir.

Çizelge 4.5. Modelde kullanılan değişkenlerin düzeltme katsayıları ve anlamlılık düzeyleri

<b>Değişkenler</b>	<b>p değeri</b>	<b>Anlamlılık Düzeyi</b>
Geçmiş Yüksek Fiyat	3.12e-07	***
Geçmiş Hacim	5.17e-04	***
Geçmiş Kapanış Fiyatı	8.78e-09	***
Geçmiş Alıntı Varlık Hacmi	5.46e-01	
Geçmiş İşlem Sayısı	1.26e-02	*
Geçmiş TB Baz Hacmi	1.58e-03	**
Geçmiş TB Alıntı Hacmi	8.74e-03	**
10 Yıllık Enflasyon Beklentisi Oranı	2.14e-02	*
Federal Fonlar Faiz Oranı	1.06e-01	
Dolar Ticaret Ağırlıklı Döviz Endeksi	3.10e-01	
Günlük Federal Fon Oranı	8.91e-01	
GSYİH Büyüme Oranı	6.76e-01	
İşsizlik Oranı	8.09e-01	
Enflasyon Oranı	9.12e-01	
Benzin Yüksek Fiyat	9.32e-01	
Benzin Hacmi	3.89e-01	
Dolar/Euro Döviz Kuru	8.70e-01	
Endeks Yüksek Değeri	2.11e-01	
Endeks Hacmi	9.66e-01	
Altın Yüksek Fiyatı	3.10e-01	
Altın Hacmi	9.73e-01	
Sabit Terim	6.23e-02	.

(Bu tablo, modelinizin her özelliği için  $p$  –değerleri ve istatistiksel anlamlılık düzeylerini göstermektedir. İstatistiksel anlamlılık kodları şunlardır: \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ , .  $p < 0.1$ .)

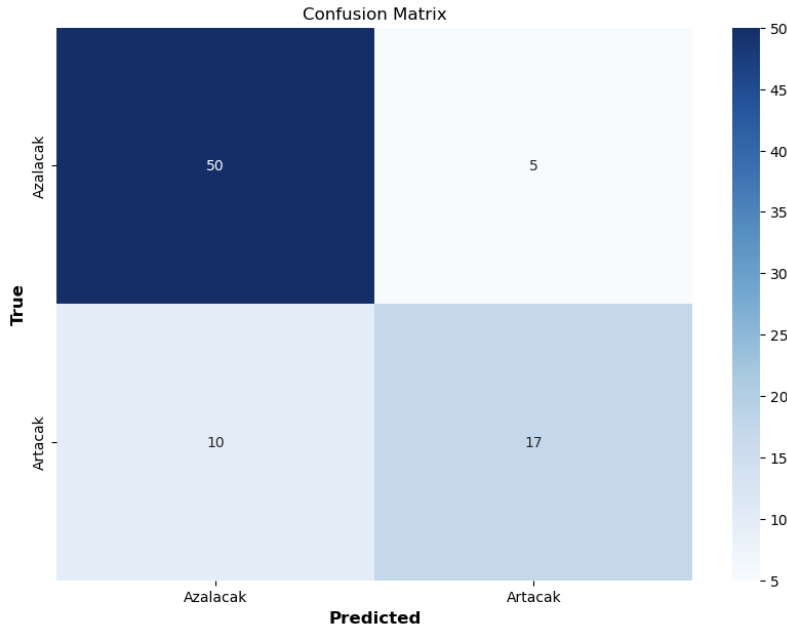
Ayrıca, her bir anlamlı değişken için düzleştirilmiş eğrilerin güven aralıkları Şekil 4.14'te sunulmuştur. Toplam değişken sayısı kadar grafik olmasına karşın, yalnızca anlamlı olanlar verilmiştir. Bu görseller, düzeltme fonksiyonlarının modelin yapısal esnekliğini ve doğrusal olmayan ilişkileri ne ölçüde yakalayabildiğini göstermektedir.



Şekil 4.14. Modelde kullanılan anlamlı değişkenlerin düzeltme katsayılarına ait güven aralıkları

Şekil 4.15 de sunulan karışıklık matrisi (Confusion Matrix), modelin "Azalacak" sınıfını "Artacak" sınıfına kıyasla daha etkin bir şekilde tahmin ettiğini göstermektedir. Ancak, modelin "Artacak" sınıfı için yüksek sayıda yanlış negatif tahminler yapması, bu sınıfın

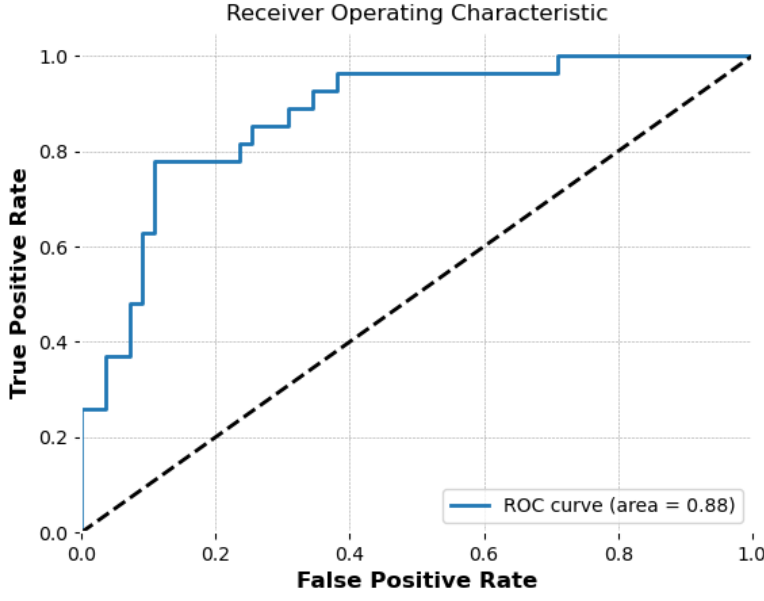
duyarlılık oranının düşük olduğuna işaret etmektedir. Bu, modelin "Artacak" durumları doğru bir şekilde belirleme kapasitesinin zayıf olduğunu ve bu alanda iyileştirmeler yapılması gerektiğini ortaya koymaktadır. Özellikle, "Artacak" sınıfını yanlış sınıflandırma riskini minimize edecek şekilde model ayarlarının gözden geçirilmesi önem taşımaktadır.



Şekil 4.15. Karışıklık matrisi

Sonuç olarak, bu karışıklık matrisi, modelin genel performansını ve yanlış tahminlerin niteliğini değerlendirmede önemli bir araçtır. İlgili düzeltmeler ve parametre ayarlamaları yaparak, modelin her iki sınıfı da dengeli bir şekilde tahmin etmesini sağlamak mümkündür. Bu ayarlamalar, özellikle duyarlılık düşük olan sınıflar için modelin tahmin yeteneğini artırma potansiyeline sahiptir.

Şekil 4.16'te sunulan grafik, sınıflandırma modelinin ROC eğrisini göstermektedir. Bu ROC eğrisi, çeşitli eşik değerlerinde modelin doğru pozitif oranı ile yanlış pozitif oranı arasındaki ilişkiyi görsel olarak ifade etmektedir. ROC eğrisi ve AUC değerinin değerlendirilmesi, modelin performansını ölçmede kritik öneme sahiptir. Bu analiz, modelin sınıflandırma yeteneğinin genel etkinliğini ve doğruluk seviyesini belirlemekte yardımcı olmaktadır.



Şekil 4.16. Eğri Düzleştirmeli Lojistik Regresyon modeli için ROC eğrisi

Eğri, 0.88 olarak belirtilen bir AUC değerine sahiptir. Bu değer, modelin pozitif bir örneği rastgele bir negatif örnekten daha yüksek bir olasılıkla doğru bir şekilde sınıflandırma yeteneğini göstermektedir. AUC değeri, 1.0'a yaklaştıkça modelin mükemmel performans sergilediği, 0.5 değeri ise modelin tahminlerinin rastgele olduğunu ifade etmektedir. Bu durumda, 0.88 değeri modelin yüksek performans gösterdiğini işaret etmektedir.

ROC eğrisi, sol üst köşeye doğru önemli bir eğimle ilerler, bu da modelin düşük yanlış pozitif oranlarında bile yüksek doğru pozitif oranlara ulaşabildiğini göstermektedir. Bu şekil, modelin özellikle negatif sınıfları pozitif olarak yanlış sınıflandırma olasılığının düşük olduğunu belirtmektedir.

İdeal bir ROC eğrisi, sol üst köşeden başlar ve y eksenini boyunca hızla yükselir. Görseldeki eğri, bu ideal duruma yakın bir performans sergilemekte olup, özellikle düşük yanlış pozitif oranlarında doğru pozitif oranının hızla yükselmesi modelin duyarlılığını ortaya koymaktadır.

Bu ROC eğrisi, sınıflandırma modelinin pozitif ve negatif sınıfları etkili bir şekilde ayırt edebildiğini göstermektedir. Ancak, modelin daha kapsamlı bir değerlendirmesi ve



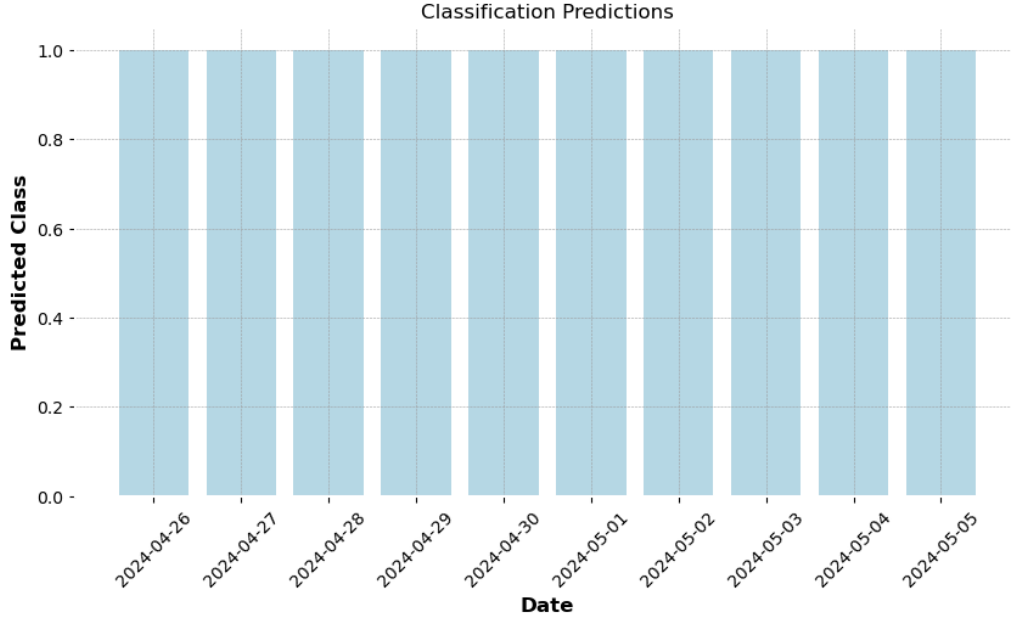
potansiyel iyileştirmeler için, farklı eşik değerlerinin etkilerinin analizi, model parametrelerinin ayarlanması ve eğitim veri setindeki özellik seçiminin optimize edilmesi önerilir.

#### 4.9 İleriye Dönük Sınıflandırma Tahmin Sonuçları

En iyi performans gösteren model olan Eğri Düzleştirmeli Lojistik Regresyonun değerlendirilmesi tamamlanmıştır. Bunu takiben, 26 Nisan 2024 ile 5 Mayıs 2024 tarihleri arasında kapsayan on günlük dönem için ileriye dönük sınıflandırma tahminleri yapılmıştır. Bu tahminlerin sonuçları, Çizelge 4.6 ve Şekil 4.17’te ayrıntılı olarak gösterilmiştir.

Çizelge 4.6. İleri dönük tahmin değerleri

Sıra No	Tarih	Tahmin Edilen Sınıf
0	2024-04-26	1
1	2024-04-27	1
2	2024-04-28	1
3	2024-04-29	1
4	2024-04-30	1
5	2024-05-01	1
6	2024-05-02	1
7	2024-05-03	1
8	2024-05-04	1
9	2024-05-05	1



Şekil 4.17. İleri dönük tahmin değerleri için sütun grafiği

Her bir tarih için tahmin edilen sınıfın sürekli olarak "1" olarak belirlendiği gözlemlenmektedir. Bu durum, modelin öngörülen tüm günler boyunca sabit bir çıktı verdiğini göstermektedir. Eğer modelin devamlı olarak aynı sınıfı tahmin etmesi beklenmiyorsa, bu durum modelin genelleştirme kapasitesinin yetersiz olduğunu veya veri setindeki dengesizliklerden etkilendiğini işaret edebilir. Bu tespitler, modelin daha fazla incelenmesi ve potansiyel iyileştirmelerin yapılması gerektiğini önermektedir.

## 5. SONUÇ VE TARTIŞMA

Bu tez çalışması, 17 Ağustos 2017 ile 3 Ocak 2024 tarihleri arasında toplanan Bitcoin verileri üzerinde çeşitli istatistiksel ve makine öğrenimi modelleri kullanılarak yapılan analizleri kapsamaktadır. Analizlerde Lojistik Regresyon, LOESS, Eğri Düzleştirmeli Lojistik Regresyon, SVM, Karar Ağaçları ve Gradient Boosting gibi modeller kullanılmıştır. Elde edilen sonuçlara göre, sınıflandırma metriklerine göre en iyi performansı gösteren model Eğri Düzleştirmeli Lojistik Regresyon modeli olarak belirlenmiştir.

Analizlerimiz, kripto para piyasalarını etkileyen faktörler ve piyasa davranışlarını derinlemesine anlamak amacıyla gerçekleştirilmiştir. Ayrıca, bu analizler kripto para piyasasının yüksek volatilitésinin kullanılan modellerin tahmin başarısını önemli ölçüde nasıl etkilediğini göstermiştir. Bu nedenle, farklı kripto paraların değerlerini içerecek şekilde veri setinin nitelik ve nicelik açısından zenginliğini genişletmeyi, model performansını optimize etmeyi, yeni modeller geliştirmeyi önerilmektedir. Kripto paraların değerinin anlaşılmasında etkisinin büyük olduğu Blockchain teknolojisinde incelenmesi, bu sayede modelin genelleme yeteneğinin ve tahmin doğruluğunun artırılması hedeflenmektedir. Sonuç olarak, bu gelişmeler tahmin modellerinin etkinliğini artırarak, piyasa davranışlarını daha derinlemesine anlamamıza ve uygun stratejiler geliştirmemize yardımcı olacaktır.

Kripto para piyasalarının dinamik doğası, finansal analizler ve yatırım stratejileri için modellerin sunduğu tahminlerden yararlanmayı mümkün kılarken, bu tahminlerin finansal karar verme süreçlerinde dikkatli bir şekilde değerlendirilmesi gerektiğini hatırlatmaktadır. Piyasa riskleri ve diğer ekonomik faktörlerin yanında modelin sınırlılıklarının da göz önünde bulundurulması önem taşımaktadır. Bu çalışma, kripto para piyasasındaki fiyat dinamiklerini çözümleme çabalarında ileriye dönük bir adım olarak görülebilir ve bu alandaki gelecekteki çalışmalar için sağlam bir temel oluşturmayı hedeflemektedir.

Son olarak, kripto para piyasalarını etkileyen faktörlerin ve piyasa davranışlarının kapsamlı bir profilini çıkarmak için yapılan bu derinlemesine analizler, sadece akademik arařtırmalar için deęil, aynı zamanda finansal sektördeki pratik uygulamalar için de faydalı olacaktır. Bu çalışmada elde edilen bulgular, kripto paraların ve ilgili teknolojilerin anlaşılmasında ve kullanılmasında yeni yollar sunmayı amaçlamaktadır. Bu hedef doğrultusunda, piyasa katılımcılarına ve arařtırmacılara rehberlik edecek sağlam ve bütüncül analizler sunmak, bu tezin odak noktasını oluşturmaktadır.

## KAYNAKLAR

- [1] J. Liang, L. Li, W. Chen, D. Zeng, "Towards an Understanding of Cryptocurrency: A Comparative Analysis of Cryptocurrency, Foreign Exchange, and Stock," The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China, **2019**. Erişim adresi: <https://ieeexplore.ieee.org/abstract/document/8823373>
- [2] Q. Wang, "Cryptocurrency Asset Classification and Pricing Using CNN and RNN," s41060-023-00485-7 Erişim adresi: <https://link.springer.com/article/10.1007/s41060-023-00485-7>
- [3] K Nakayama, GH Silverman, "Serial and parallel processing of visual feature conjunctions," Nature, 320 (**1986**) 264.
- [4] RL Eubank, "Spline smoothing and nonparametric regression," Marcel Dekker, New York, **1988**.
- [5] T Hastie, R Tibshirani, "Exploring the nature of covariate effects in the proportional hazards model," Biometrics, 46 (**1990**) 1005-1016.
- [6] W Härdle, "Applied nonparametric regression," Cambridge University Press, **1990**.
- [7] G Wahba, "Spline models for observational data," SIAM, 90 (**1990**) 59.
- [8] Nakamoto, S. (**2008**). Bitcoin: A Peer-to-Peer Electronic Cash System. [Çevrimiçi] Erişim adresi: <https://bitcoin.org/bitcoin.pdf>
- [9] M. Di Pierro, "What Is the Blockchain?" Computing in Science & Engineering, vol. 19, no. 5, pp. 92-95, **2017**. Erişim adresi: <https://ieeexplore.ieee.org/abstract/document/8024092/>
- [10] Yurttagüler, İpek. "Kripto Para Birimi Piyasalarında GPH Yöntemi ile Uzun Hafıza Analizi: Bitcoin Örneği." Ekonomi Politika ve Finans Araştırmaları Dergisi 9.1 (**2024**): 126. Erişim adresi: <https://dergipark.org.tr/en/download/article-file/3540365>
- [11] Kyriazis, Nikolaos A. "A survey on efficiency and profitable trading opportunities in cryptocurrency markets." Journal of Risk and Financial Management 12.2 (**2019**): 67. Erişim adresi: <https://www.mdpi.com/1911-8074/12/2/67>

- [12] Yurdabak, Merve Kadriye. "Nft: Dijital Sanatta Yeni Bir Perspektif Ve Getirdiği Fırsatlar Üzerine Bir Derleme." Nişantaşı Üniversitesi Sosyal Bilimler Dergisi 10.1 (2022): 143-153. Erişim adresi: <https://dergipark.org.tr/en/pub/nisantasisbd/article/1107687>
- [13] Özyürek, Hamide. "Blockchain teknolojisinin mevcut ve muhtemel kullanım alanları." Anadolu Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi 22.4 (2021): 31-50. Erişim adresi: <https://dergipark.org.tr/en/pub/anadoluibfd/issue/66491/988748>
- [14] J. Taskinsoy, Bitcoin: A New Digital Gold Standard in the 21st Century?, Available at SSRN 3941857, (2021). Erişim adresi: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3941857](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3941857)
- [15] Umar, Muhammad, et al. "Bitcoin: A safe haven asset and a winner amid political and economic uncertainties in the US?." Technological Forecasting and Social Change 167 (2021): 120680. Erişim adresi: <https://www.sciencedirect.com/science/article/pii/S0040162521001128>
- [16] F. Alvarez, D. Argente, D. Van Patten, Science, 10.1126/science.add2844 (2023). Erişim adresi: <https://www.science.org/doi/abs/10.1126/science.add2844>
- [17] A. Sotiropoulou, D. Guégan, Capital Markets Law Journal, 12(4) (2017) 12. Erişim adresi: <https://watermark.silverchair.com/kmx037.pdf>
- [18] D. G. Baur, K. Hong, A. D. Lee, "Bitcoin: Medium of exchange or speculative assets?" Journal of International Financial Markets, Institutions and Money, 54 (2018) 177–189. Erişim adresi: <https://www.sciencedirect.com/science/article/pii/S1042443117300720>
- [19] C. Baek, M. Elbeck, "Bitcoins as an investment or speculative vehicle? A first look." Applied Economics Letters, 22(1) (2015) 30–34. Erişim adresi: <https://www.tandfonline.com/doi/abs/10.1080/13504851.2014.916379>
- [20] M. H. Ali, M. A. Uddin, M. A. F. Chowdhury, M. Masih, "Cross-country evidence of Islamic portfolio diversification: Are there opportunities in Saudi Arabia?" Managerial Finance, 45(1) (2019) 36–53. Erişim adresi: [https://www.emerald.com/insight/content/doi/10.1108/MF-03-2018-0126/full/html?fbclid=IwAR1x3LZBLhAWGRGzKpWfMy1RzRfqvjA6kTBwObPvSuP2TNaiSe\\_yBiusWG8&journalCode=mf&utm\\_source=TrendMD&utm\\_medium=cpc&utm\\_campaign=Managerial\\_Finance\\_TrendMD\\_1&WT.mc\\_id=Emerald\\_TrendMD\\_1](https://www.emerald.com/insight/content/doi/10.1108/MF-03-2018-0126/full/html?fbclid=IwAR1x3LZBLhAWGRGzKpWfMy1RzRfqvjA6kTBwObPvSuP2TNaiSe_yBiusWG8&journalCode=mf&utm_source=TrendMD&utm_medium=cpc&utm_campaign=Managerial_Finance_TrendMD_1&WT.mc_id=Emerald_TrendMD_1)
- [21] B. M. Blau, "Price dynamics and speculative trading in bitcoin." Research in International Business and Finance, 41 (2017) 493–499. Erişim adresi: <https://www.sciencedirect.com/science/article/pii/S0275531917303057>

- [22] E. Bouri, M. Das, R. Gupta, D. Roubaud, "Spillovers between Bitcoin and other assets during bear and bull markets." Erişim adresi: <https://www.tandfonline.com/doi/full/10.1080/00036846.2018.1488075>
- [23] Y. S. Abubakar, A. F. Ogunbado, M. A. Saidi, "Bitcoin and its legality from Shariah point of view." *SEISENSE Journal of Management*, 1(4) (2018) 13–23. Erişim adresi: <https://www.journal.seisense.com/jom/article/view/32>
- [24] J. Bouoiyour, R. Selmi, "What does Bitcoin look like?" *Annals of Economics & Finance*, 16(2) (2015) 449–492. Erişim adresi: [https://www.researchgate.net/profile/Refk-Selmi/publication/283676718\\_What\\_Does\\_Bitcoin\\_Look\\_Like/links/56432c8c08aef646e6c68bcc/What-Does-Bitcoin-Look-Like.pdf](https://www.researchgate.net/profile/Refk-Selmi/publication/283676718_What_Does_Bitcoin_Look_Like/links/56432c8c08aef646e6c68bcc/What-Does-Bitcoin-Look-Like.pdf)
- [25] H. A. Aalborg, P. Molnár, J. E. de Vries, "What can explain the price, volatility and trading volume of Bitcoin?" *Finance Research Letters*, 29 (2018) 255–265. Erişim adresi: <https://www.sciencedirect.com/science/article/pii/S1544612318302058>
- [26] D. Ashraf, N. Mohammad, "Matching perception with the reality—Performance of Islamic equity investments." *Pacific-Basin Finance Journal*, 28 (2014) 175–189. Erişim adresi : <https://www.sciencedirect.com/science/article/pii/S0927538X13000930>
- [27] C. Aloui, B. Hkiri, "Co-movements of GCC emerging stock markets: New evidence from wavelet coherence analysis." *Economic Modelling*, 36 (2014) 421–431. Erişim adresi: <https://www.sciencedirect.com/science/article/pii/S0264999313004033>
- [28] E. Bouri, N. Jalkh, P. Molnár, D. Roubaud, "Bitcoin for energy commodities before and after the December 2013 crash: Diversifier, hedge or safe haven?" *Applied Economics*, 49(50) (2017) 5063–5073. Erişim adresi: <https://www.tandfonline.com/doi/full/10.1080/00036846.2017.1299102>
- [29] E. Bouri, P. Molnár, G. Azzi, D. Roubaud, L. I. Hagfors, "On the hedge and safe haven properties of Bitcoin: Is it really more than a diversifier?" *Finance Research Letters*, 20 (2017) 192–198. Erişim adresi: <https://www.sciencedirect.com/science/article/pii/S1544612316301817>
- [30] Şahin, Orhan. "Lojistik Regresyon ile Ayvalık'ı Turizm Amacıyla Tercih Etmede Önemli Risk Faktörlerinin Belirlenmesi." *Elektronik Sosyal Bilimler Dergisi* 16.61 (2017): 647-660. Erişim adresi: <https://dergipark.org.tr/en/download/article-file/292091>
- [31] Chen, Zheshi, Li, Chunhong, & Sun, Wenjun. "Bitcoin price prediction using machine learning: An approach to sample dimension engineering." *Journal of Computational and Applied Mathematics* 365 (2020): 112395. Erişim adresi: <https://www.sciencedirect.com/science/article/pii/S037704271930398X/pdf>

- [32] T. Hastie, R. Tibshirani, J. Friedman, in *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Vol. 2, Springer, New York, **2009**.
- [33] Levy, N., Garti, N., & Margdassi, S. "Colloids and Surfaces A: Physicochemical and Engineering Aspects." *Colloids and Surfaces A: Physicochemical and Engineering Aspects* 97 (1995): 91. Erişim adresi: <https://www.tandfonline.com/doi/abs/10.1080/02331930701618740>
- [34] A. C. Davison, in *Statistical Models*, Cambridge University Press, **2003**.
- [35] G., Wahba, *Spline Models for Observational Data*. Society for industrial and applied mathematics, **1990**.
- [36] Adha, M.R., Nurrohmah, S., & Abdullah, S. "Multinomial logistic regression and spline regression for credit risk modelling." *Journal of Physics: Conference Series*, 1108.1 (2018). Erişim adresi : <https://iopscience.iop.org/article/10.1088/1742-6596/1108/1/012019/meta>
- [37] Taylan, P., Weber, G.-W., & Beck, A. "New approaches to regression by generalized additive models and continuous optimization for modern applications in finance, science and technology." *Optimization*, 56.5-6 (2007): 675-698. Erişim adresi : <https://www.tandfonline.com/doi/abs/10.1080/02331930701618740>
- [38] N. Beck ve S. Jackman, "Getting the mean right is a good thing: generalized additive models." *Society for Political Methodology Working Papers*, **1997**.
- [39] Jacoby, William G. "Loess: a nonparametric, graphical tool for depicting relationships between variables." *Electoral Studies* 19.4 (2000): 577-613. Erişim adresi: <https://www.sciencedirect.com/science/article/pii/S0261379499000281>
- [40] Jacoby, W.G. *Statistical Graphics for Visualizing Multivariate Data*, vol. 120, Sage, **1998**.
- [41] W. S. Cleveland, *Visualizing data*, Hobart Press, **1993**.
- [42] R. A., Becker, W. S., Cleveland, G., Weil, in *Dynamic Graphics for Statistics*. "The use of brushing and rotation for data analysis." Elsevier, Amsterdam, Chapter 247-275, **1988**.
- [43] W.S., Cleveland, *Journal of the American Statistical Association*, 74 (1979) 829-836. Erişim adresi : [https://www.sciencedirect.com/science/article/pii/S0261379499000281?ref=pdf\\_download&fr=RR-2&rr=8814fcd8af9d2c46](https://www.sciencedirect.com/science/article/pii/S0261379499000281?ref=pdf_download&fr=RR-2&rr=8814fcd8af9d2c46)



- [44] LeCun Y., Bengio, Y., & Hinton, G. "Deep learning." *Nature*, 521.7553 (2015): 436-444. Erişim adresi: <https://www.nature.com/articles/nature14539>
- [45] N. J. Nilsson, *Principles of Artificial Intelligence*, Morgan Kauffmann, San Francisco, 1980. Erişim adresi: <https://stacks.stanford.edu/file/druid:zd294jv9941/zd294jv9941.pdf>
- [46] S. Ürgenç, B. Aşıkgil, "Makine Öğrenmesi Yöntemleri ile Bitcoin Trend Dönüşlerinin Tahmin Edilmesi", Yüksek Lisans Tezi, Mimar Sinan Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul, 2023.
- [47] M. F. Keskenler, E. F. Keskenler, "Geçmişten Günümüze Yapay Sinir Ağları Ve Tarihçesi," *Takvim-İ Vekayi*, 5(2) (2017) 8-18. Erişim adresi: <https://dergipark.org.tr/pub/takvim/issue/33375/346279>
- [48] E. Alpaydın, in *Introduction to Machine Learning*, MIT Press, 2020.
- [49] I. Goodfellow, Y. Bengio, A. Courville, in *Deep Learning*, The MIT Press, 1st ed., 2018.
- [50] Z. Şengül, "Makine öğrenmesi algoritmalarını kullanarak bitcoin fiyat tahmini," Yüksek Lisans Tezi, Trakya Üniversitesi Sosyal Bilimler Enstitüsü, 2022.
- [51] Costa V.G., and Pedreira C.E. "Recent advances in decision trees: An updated survey." *Artificial Intelligence Review* 56.5 (2023): 4765-4800. Erişim adresi: <https://link.springer.com/article/10.1007/s10462-022-10275-5>
- [52] Rudin, C. "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead." *Nat Mach Intell* 1.5 (2019): 206-215. Erişim adresi: <https://www.nature.com/articles/s42256-019-0048-x>
- [53] Roscher, R., Bohn, B., Duarte, M.F., et al. "Explainable machine learning for scientific insights and discoveries." *IEEE Access* 8 (2020): 42200-42216. Erişim adresi: <https://ieeexplore.ieee.org/abstract/document/9007737/>
- [54] Yangın, Gülçin. "XGBoost ve karar ağacı tabanlı algoritmaların diyabet veri setleri üzerine uygulaması." MS thesis, Mimar Sinan Güzel Sanatlar Üniversitesi, Fen Bilimleri Enstitüsü, 2019. Erişim adresi: <https://acikerisim.msgsu.edu.tr/xmlui/handle/20.500.14124/1152>
- [55] Akpınar, H. "Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği." *İ.Ü.İşletme Fakültesi Dergisi*, 29 (2000): 1-22.

[56] Pehlivan, G. "Chaid Analizi ve Bir Uygulama." Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul, **2006**. Erişim adresi: <https://platform.almanhal.com/Details/Thesis/2000248191>

[57] Kıran, Z.B. "Lojistik Regresyon ve Cart Analizi Teknikleriyle Sosyal Güvenlik Kurumu İlaç Provizyon Sistemi Verileri Üzerinde Bir Uygulama." Yüksek Lisans Tezi, Gazi Üniversitesi Fen Bilimleri Enstitüsü, Ankara, **2010**.

[58] Köktürk, F. "K-En Yakın Komşuluk, Yapay Sinir Ağları ve Karar Ağaçları Yöntemlerinin Sınıflandırma Başarılarının Karşılaştırılması." Doktora Tezi, Bülent Ecevit Üniversitesi, Sağlık Bilimleri Enstitüsü, Biyoistatistik Anabilim Dalı, Zonguldak, **2012**. Erişim Adresi: <https://acikbilim.yok.gov.tr/handle/20.500.12812/598126>

[59] Maimon, O., & Rokach, L. "Data Mining and Knowledge Discovery Handbook." Springer Science Business Media, New York, USA, **2010**. Erişim adresi: <https://link.springer.com/content/pdf/10.1007/978-0-387-09823-4.pdf>

[60] Akman, M. "Veri Madenciliğine Genel Bakış ve Random Forests Yönteminin İncelenmesi: Sağlık Alanında Bir Uygulama." Yüksek Lisans Tezi, Ankara Üniversitesi Sağlık Bilimleri Enstitüsü, Biyoistatistik Anabilim Dalı, Ankara, **2010**. Erişim adresi: <https://acikbilim.yok.gov.tr/handle/20.500.12812/75659>

[61] Akpınar, H. "Data Veri Madenciliği Veri Analizi." İstanbul/Fatih, Türkiye: Papatya Yayıncılık Eğitim A.Ş., **2014**.

[62] Zhang, C., & Yunqian, M. "Ensemble Machine Learning Methods and Applications." New York, USA: Springer Science Business Media, **2012**. Erişim adresi: <https://link.springer.com/content/pdf/10.1007/978-1-4419-9326-7.pdf>

[63] Zhou, Z.-H. "Ensemble Methods Foundations and Algorithms." U.S.: Chapman & Hall / CRC Press, **2012**.

[64] Navlani, A. "Datacamp." (Erişim tarihi: 27 Mayıs **2019**) <https://www.datacamp.com/community/tutorials/adaboost-classifier-python>.

[65] R. E. Schapire, in Proceedings of the International Conference on Machine Learning (ICML), Nashville, 8-12 July **1997**, Morgan Kaufmann, San Francisco, 1997, p. 97. Erişim adresi: <http://rob.schapire.net/papers/Schapire97.pdf>

[66] Y. Freund ve R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," Journal of Computer and System Sciences, 55(1) (**1997**) 119-139. Erişim adresi: <https://www.sciencedirect.com/science/article/pii/S002200009791504X>

- [67] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of Statistics*, 29(5) (2001) 1189-1232. Eriřim adresi: [https://www.jstor.org/stable/2699986?casa\\_token=Va2vChCZ8f0AAAAA%3AS24B5sFtG5kclZK8BgxMZhivAmGbSVb\\_nQFluGA6rw848yQ-Zr23Sx0klCXYbGMRbREbFm-jm94GfEAxVlXzRy0eBT-kBcjMexj\\_QZvWO1haX1Tq6vMnaQ](https://www.jstor.org/stable/2699986?casa_token=Va2vChCZ8f0AAAAA%3AS24B5sFtG5kclZK8BgxMZhivAmGbSVb_nQFluGA6rw848yQ-Zr23Sx0klCXYbGMRbREbFm-jm94GfEAxVlXzRy0eBT-kBcjMexj_QZvWO1haX1Tq6vMnaQ)
- [68] J. R. Quinlan, in *C4.5: Programs for Machine Learning*, Elsevier, Amsterdam, Chapter 1, 2014. Eriřim adresi: <https://books.google.com/books?hl=tr&lr=&id=b3ujBQAAQBAJ&oi=fnd&pg=PP1&q=Quinlan+c4.5&ots=sS2rUQDtl6&sig=bMBBcc--HJXoXkgEJUtcqjddLyg>
- [69] De Ville, B. "Decision Trees for Business Intelligence and Data Mining: Using SAS Enterprise Miner." North Carolina, United States of America: SAS Institute Inc., 2006.
- [70] Jankowski, N., Duch, W., & Grańczewski, K. "Meta-Learning in Computational Intelligence." Heidelberg, Berlin: Springer-Verlag Berlin Heidelberg, 2011. Eriřim adresi: <https://link.springer.com/content/pdf/10.1007/978-3-642-20980-2.pdf>
- [71] Garrido, A. P. "QuantDare." (Eriřim tarihi: 28 Mayıs 2019). Eriřim adresi: <https://quantdare.com/what-is-the-difference-between-bagging-and-boosting/>.
- [72] Marksfield, A.E. "Machine Learning in Sociodemographic Segmentation of a Telco Company Customers." Bachelor Thesis, Czech Technical University, Department of Theoretical Computer Science, Prague, 2018. Eriřim adresi: <https://dspace.cvut.cz/handle/10467/76842>
- [73] Kaplan, H.İ. "Destek Vektor Makineleri (Support Vector Machine-SVM)" (Eriřim tarihi: 14 Nisan 2024) Eriřim adresi: <https://medium.com/machine-learning-turkiye/destek-vektor-makineleri-support-vector-machine-svm-ddaa0d003aa3>
- [74] Huang, S., Cai, N., Pacheco, P.P., Narrandes, S., Wang, Y., & Xu, W. "Applications of Support Vector Machine (SVM) Learning in Cancer Genomics." *Cancer Genomics & Proteomics* 15.1 (2018): 41-51. Eriřim adresi: <https://cgp.iijournals.org/content/cgp/15/1/41.full.pdf>
- [75] Grandini, M., Bagli, E., & Visani, G. "Metrics for Multi-Class Classification: An Overview." DOI: 10.48550/arXiv.2008.05756, 2020. Eriřim adresi: <https://arxiv.org/pdf/2008.05756.pdf>
- [76] Anonim, (Eriřim tarihi: 13 Nisan 2024) Eriřim adresi: <https://academy.binance.com/>
- [77] Dimitriadou, A., and Gregoriou, A., "Entropy," 25.5 (2023) 777. Eriřim adresi: <https://www.mdpi.com/1099-4300/25/5/777>

[78] Prof. Dr. Ezel Tavşancıl, TEMEL İSTATİSTİK Korelasyon II (Erişim tarihi: 14 Nisan 2024) Erişim adresi: [https://acikders.ankara.edu.tr/pluginfile.php/169658/mod\\_resource/content/0/8\\_KORELASYON%20II.pdf](https://acikders.ankara.edu.tr/pluginfile.php/169658/mod_resource/content/0/8_KORELASYON%20II.pdf)