



**BİTCOİN FİYAT HAREKETLERİNİN MAKİNE
ÖĞRENME ALGORİTMALARI İLE TAHMİNİ**

Hulusi Mehmet TANRIKULU

**Yüksek Lisans Tezi
İşletme Ana Bilim Dalı
Doç. Dr. Hakan PABUÇCU
2021
(Her Hakkı Saklıdır)**

T.C.
BAYBURT ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
İŞLETME ANABİLİM DALI

**BİTCOİN FİYAT HAREKETLERİNİN MAKİNE ÖĞRENME ALGORİTMALARI
İLE TAHMİNİ**

(Forecasting Bitcoin Price Movements with Machine Learning Algorithms)

YÜKSEK LİSANS TEZİ

Hulusi Mehmet TANRIKULU

Danışman: Doç. Dr. Hakan PABUÇCU

Bayburt
Eylül, 2021

ETİK VE BİLDİRİM SAYFASI

Tez çalışması olarak sunduğum “*Bitcoin Fiyat Hareketlerinin Makine Öğrenme Algoritmaları İle Tahmini*” başlıklı çalışmanın tarafımdan bilimsel etik ilkelerine uyularak yazıldığını ve yararlandığım eserleri kaynakçada gösterdiğimi beyan ederim.

09/09/2021

Hulusi Mehmet TANRIKULU

TEŞEKKÜR

Çalışmalarında katkılarını hiçbir zaman esirgemeyen, bilgi ve deneyimleri ile bana yardımcı olan danışman hocam sayın Doç. Dr. Hakan PABUÇCU'ya, çalışmalarım esnasında desteklerini esirgemeyen merhum Daire Başkanım Mehmet Şerif TOPCU'ya ve tez savunma sınavıma katılan değerli jüri üyelerine teşekkürlerimi sunarım.

Yüksek lisans süresince ve tezimin her aşamasında yanımda olan, beni güçlendiren, varlığı ile hayatımı güzelleştiren, başarılarımın yegâne ortağı eşim Hülya TANRIKULU'ya ve yaşam pınarımız, biricik kızım Defne TANRIKULU'ya hayatımda oldukları ve destekleri için teşekkür ederim.

ÖZ
YÜKSEK LİSANS TEZİ
BİTCOİN FİYAT HAREKETLERİNİN MAKİNE ÖĞRENME ALGORİTMALARI
İLE TAHMİNİ

Hulusi Mehmet TANRIKULU

Eylül 2021, 117 sayfa

Kripto para terimi her ne kadar finans dünyasına geçtiğimiz on yıl içerisinde girmiş olsa da başlangıç döneminde var olan güven kaygıları yenmiş ve bir ticaret ve yatırım aracı olarak piyasalarda yerini almıştır. Kripto para birimlerinin internet kaynaklı bir değer olması, arzının sınırlı olması, bir merkeze bağlı olmaması ve 7/24 anında finansal işlemlerin yapılabilmesi kripto paraları cazip bir yatırım aracı haline getirmektedir. Günümüzde bilinen binin üzerinde kripto para birimi vardır. Bu kripto para birimlerinin ilki ve en yaygın kullanılanı olan Bitcoin'dir.

Son günlerde fiyatlarındaki hareketlilikten dolayı medyada kendinden sıkça söz ettiren Bitcoin, sadece yatırım olmaktan çıkmış günlük ihtiyaçlarında karşılanabildiği genel bir değer haline gelmiştir. Bu çalışmada Blok Zincir, Kripto Para kavramlarına değinilerek, şimdiye kadar Bitcoin özelinde fiyat hareketlerinin tahminlerinde kullanılan makine öğrenme algoritmaları, tahmin için kullanılan analiz göstergeleri(indikatörler) ve geçmiş akademik çalışmalarda izlenen yöntemler ve verimlilik bulguları üzerine incelemeler yapılmıştır.

Yapılan incelemeler sonunda; makine öğrenme yöntemlerinin çok veri ve az veri ile performansların kıyaslanabilmesi adına geçtiğimiz 7 yıllık ve 1 yıllık Bitcoin hareketleri üzerine analizler yapılmıştır. Teknik indikatörlerle beraber günlük Tweet sayıları ve Google Arama Trendleri verisinin dahil edildiği veri setleri 7 farklı makine öğrenme yöntemi ile fiyat yönünün tespiti üzerine performans analizi yapılmıştır. Son olarak eğitilen makine öğrenme yöntemlerinin tahmin performansları yeni bir veri seti üzerinde ölçülmüştür. Bu performans analizleri ile, henüz uzun vadeli verisi oluşmamış yeni kripto paraların fiyat yönü tahmininde kullanılabilecek makine öğrenme yöntemlerinin belirlenmesinde yol gösterici olması hedeflenmiştir.

Anahtar Kelimeler: Kripto para, sanal para, blok zinciri, bitcoin, makine öğrenmesi.

ABSTRACT
MASTER THESIS
FORECASTING BITCOIN PRICE MOVEMENTS WITH MACHINE LEARNING
ALGORITHMS

Hulusi Mehmet TANRIKULU

September 2021, 117 pages

Although the term cryptocurrency has entered the financial world in the past decade, it has overcome the trust concerns that existed in the initial period and took its place in the markets as a trading and investment tool. The fact that cryptocurrencies are an internet-based value, has limited supply, ability to transactions 24/7 and not controlled by any authority makes it an attractive investment tool. There are over a thousand known cryptocurrencies today. The first and most widely used of these cryptocurrencies is Bitcoin.

Bitcoin, which has been frequently mentioned in the media due to the volatility in its prices in recent days now become more than an investment tool that can be used for daily needs. In this study, by referring to the concepts of Blockchain and Cryptocurrency, will be examined machine learning algorithms used in Bitcoin-specific price movements forecasts, indicators used for forecasting analysis and methods followed in past academic studies and efficiency findings.

After the examinations, in order to compare the performance of machine learning methods with big and small data, analyzes have been made on Bitcoin movements for the past 7 years and 1 year. Performance analysis was conducted on the determination of the price direction with 7 different machine learning methods on the datasets that includes technical indicators, daily Tweet numbers and Google Trends data. Finally, the prediction performances of the trained machine learning methods were measured on a new data set. With these performance analyzes, it is aimed to guide the determination of machine learning methods that can be used in predicting the price direction of new cryptocurrencies that have not yet formed long-term data.

Keywords: Cryptocurrency, virtual currency, blockchain, bitcoin, machine learning.

İÇİNDEKİLER

ETİK VE BİLDİRİM SAYFASI.....	II
TEŞEKKÜR.....	III
ÖZ	IV
ABSTRACT	V
İÇİNDEKİLER.....	VI
TABLOLAR DİZİNİ.....	IX
ŞEKİLLER DİZİNİ.....	XI
KISALTMALAR VE SİMGELER DİZİNİ.....	XIII

BİRİNCİ BÖLÜM

Giriş.....	1
Araştırmanın Konusu, Problemi ve Varsayımları	1
Araştırmanın Sınırlılıkları	2
Araştırmanın Amacı	2

İKİNCİ BÖLÜM

Kavramlar	3
Para.....	3
Para Türleri.....	4
İtibari para:	4
Elektronik para:	5
Dijital para:	5
Sanal para:	5
Kripto Para ve Bitcoin.....	6
Blokszincir Teknolojisi.....	8
Blokszincir yapısı:.....	8
Madencilik (Mining):	9
Bitcoin hesap cüzdanı ve bitcoin transferi.....	10
Dünyada ve ülkemizde bitcoin kullanımı.	11
Temel ve Teknik Analiz.....	12

Makine Öğrenme.....	16
Denetimsiz öğrenme (Unsupervised learning).	17
Denetimli öğrenme (Supervised learning).....	18
Sınıflandırma Algoritmaları	18
NB - Naive Bayes:	18
KNN – K-En Yakın Komşu:	18
SVM - Destek Vektör Makineleri:	19
Regresyon Algoritmaları	19
ANN - Yapay Sinir Ağları.....	19
LSTM – Kısa Uzun Süreli Bellek.....	19
RF - Rastgele Orman	19
XGBoost - eXtreme Gradient Boosting.....	20
Makine Öğrenmede Performans Ölçümü ve Değerlendirme	20
Hata Metrikleri	23
Literatür Çalışması	24

ÜÇÜNCÜ BÖLÜM

Bitcoin Fiyat Hareketlerinin Makine Öğrenme Algoritmaları ile Tahmin Edilmesi	31
Veriler.....	31
Yöntem.....	37
Analizler.....	38
Doğrulama yöntemi analizi:	38
Eğitim-Test ayrıştırması (Train-Test split).	38
Katmanlı çapraz doğrulama.	39
Rastgele Orman (RF) Yöntemi İle Analiz.....	41
Kesikli verilerin analizi:	41
Sürekli verilerin analizi:	46
K-En Yakın Komşu (KNN) Yöntemi İle Analiz.....	51
Kesikli verilerin analizi:	51
Sürekli verilerin analizi:	55
XGBoost Yöntemi İle Analiz.....	59
Kesikli verilerin analizi:	59
Sürekli verilerin analizi:	64
Destek Vektör Makinası (SVM) Yöntemi İle Analiz.....	69
Kesikli verilerin analizi:	69
Sürekli verilerin analizi.	73

Naive Bayes (NB) Yöntemi İle Analiz	77
Kesikli verilerin analizi:	77
Sürekli verilerin analizi	81
Yapay Sinir Ağları (ANN) Yöntemi İle Analiz	85
En iyi optimizier için doğruluk:.....	86
En iyi epoch için doğruluk:	87
Uzun-Kısa Vadeli Bellek (LSTM) Yöntemi İle Analiz	91
En iyi optimizier için doğruluk:.....	92
En iyi epoch için doğruluk:	92

DÖRDÜNCÜ BÖLÜM

Bulgular ve Yorumlama.....	98
Kesikli Verilerin Analizi.....	98
Kesikli Tüm Verilerin Genel Analizi	98
Kesikli tüm veri için tahmin testi.....	99
Kesikli 1 Yıllık Verilerin Genel Analizi	101
Kesikli 1 yıllık veri için tahmin testi.....	102
Sürekli Verilerin Analizi.....	104
Sürekli Tüm Verilerin Genel Analizi	104
Sürekli tüm veri için tahmin testi.....	105
Sürekli 1 Yıllık Verilerin Genel Analizi	106
Sürekli 1 yıllık veri için tahmin testi.....	107

BEŞİNCİ BÖLÜM

Sonuç ve Öneriler	110
KAYNAKÇA.....	113
ÖZGEÇMİŞ	117

TABLÖLAR DİZİNİ

Tablo 1. Kesikli Veri Setlerinin Oluşturma Yöntemleri.....	34
Tablo 2. Girdilerin Tanımsal İstatistikleri	36
Tablo 3. Eğitim-Test Ayırıştırması Doğruluk Değişimi.....	38
Tablo 4. Çapraz Doğrulama Katman Doğrulukları.....	40
Tablo 5. Kesikli Tüm Veriler RF Yöntemi Optimizasyon Ayarları.....	41
Tablo 6. Kesikli Tüm Veri RF Tahmininde İndikatörlerin Önem Derecesi.....	43
Tablo 7. Kesikli 1 Yıllık Veriler RF Yöntemi Optimizasyon Ayarları	44
Tablo 8. Kesikli 1 Yıllık Veriler RF Yöntemi İndikatörlerin Önem Derecesi.....	45
Tablo 9. Sürekli Tüm Veri RF Yöntemi Optimizasyon Ayarları	46
Tablo 10. Sürekli Tüm Veri RF Yöntemi İndikatörlerin Önem Derecesi	48
Tablo 11. Sürekli 1 Yıllık Veriler İçin RF Yöntemi Optimizasyon Ayarları	49
Tablo 12. Sürekli 1 Yıllık Veri RF Yöntemi İndikatörlerin Önem Derecesi	50
Tablo 13. Kesikli Verilerde KNN En İyi Komşu Oranı	52
Tablo 14. Kesikli Tüm Veri KNN Yöntemi İndikatörlerin Önem Derecesi	53
Tablo 15. Kesikli 1 Yıllık Veri KNN Yöntemi İndikatörlerin Önem Derecesi	54
Tablo 16. Sürekli Tüm Veri KNN Yöntemi İndikatörlerin Önem Derecesi	56
Tablo 17. Sürekli 1 Yıllık Veri KNN Yöntemi İndikatörlerin Önem Derecesi	58
Tablo 18. Kesikli Tüm Veri XGBoost Optimizasyon Ayarları.....	59
Tablo 19. Kesikli Tüm Veri XGBoost Yöntemi İndikatörlerin Önem Derecesi	61
Tablo 20. Kesikli 1 Yıllık Veriler İçin XGBoost Optimizasyon Ayarları.....	62
Tablo 21. Kesikli 1 Yıllık Veri XGBoost Yöntemi İndikatörlerin Önem Derecesi.....	63
Tablo 22. Sürekli Tüm Veri İçin XGBoost Optimizasyon Ayarları	64
Tablo 23. Sürekli Tüm Veri XGBoost Yöntemi İndikatörlerin Önem Derecesi.....	66
Tablo 24. Sürekli 1 Yıllık Veriler İçin XGBoost Optimizasyon Ayarları.....	67
Tablo 25. Sürekli 1 Yıllık Veri XGBoost Yöntemi İndikatörlerin Önem Derecesi	68
Tablo 26. Kesikli Tüm Veriler İçin SVM Optimizasyon Ayarları.....	70
Tablo 27. Kesikli 1 Yıllık Veriler İçin SVM Optimizasyon Ayarları	71
Tablo 28. Sürekli Tüm Veriler İçin SVM Optimizasyon Ayarları.....	73
Tablo 29. Sürekli 1 Yıllık Veriler İçin SVM Optimizasyon Ayarları	74
Tablo 30. Sürekli 1 Yıllık Veri SVM Yöntemi İndikatörlerin Önem Derecesi	76

Tablo 31. <i>Kesikli Tüm Veriler için en iyi NB Yöntemi</i>	77
Tablo 32. <i>Kesikli Tüm Veriler NB Yöntemi İndikatörlerin Önem Derecesi</i>	78
Tablo 33. <i>Kesikli 1 Yıllık Veriler için en iyi NB Yöntemi</i>	79
Tablo 34. <i>Kesikli 1 Yıllık Veriler NB Yöntemi İndikatörlerin Önem Derecesi</i>	80
Tablo 35. <i>Sürekli Tüm Veriler için en iyi NB Yöntemi</i>	81
Tablo 36. <i>Sürekli Tüm Veriler NB Yöntemi İndikatörlerin Önem Derecesi</i>	82
Tablo 37. <i>Sürekli 1 Yıllık Veriler için en iyi NB Yöntemi</i>	83
Tablo 38. <i>Sürekli 1 Yıllık Veriler NB Yöntemi İndikatörlerin Önem Derecesi</i>	84
Tablo 39. <i>ANN Yöntemi En İyi Optimizer Ayarları</i>	86
Tablo 40. <i>ANN Yöntemi En İyi Epoch Ayarları</i>	87
Tablo 41. <i>LSTM Yöntemi En İyi Optimizer Ayarları</i>	92
Tablo 42. <i>LSTM Yöntemi En İyi Epoch Ayarları</i>	93
Tablo 43. <i>Kesikli Tüm Veriler İçin Makine Öğrenme Yöntemleri Kıyaslaması</i>	98
Tablo 44. <i>Kesikli Tüm Veri Seti ile Eğitilmiş Yöntemlerin Tahmin Sonuçları</i>	99
Tablo 45. <i>Kesikli 1 Yıllık Veriler İçin Makine Öğrenme Yöntemleri Kıyaslaması</i>	101
Tablo 46. <i>Kesikli 1 Yıllık Veri Seti ile Eğitilmiş Yöntemlerin Tahmin Sonuçları</i>	102
Tablo 47. <i>Kesikli Veriler İçin T Testi Sonuçları</i>	103
Tablo 48. <i>Sürekli Tüm Veriler İçin Makine Öğrenme Yöntemleri Kıyaslaması</i>	104
Tablo 49. <i>Sürekli Tüm Veri Seti ile Eğitilmiş Yöntemlerin Tahmin Sonuçları</i>	105
Tablo 50. <i>Sürekli 1 Yıllık Veriler İçin Makine Öğrenme Yöntemleri Kıyaslaması</i>	106
Tablo 51. <i>Sürekli 1 Yıllık Veri Setiyle Eğitilmiş Yöntemlerin Tahmin Sonuçları</i>	107
Tablo 52. <i>Sürekli Veriler İçin T Testi Sonuçları</i>	109
Tablo 53. <i>Makine Öğrenme Yöntemlerinin Genel Sonuçları</i>	109

ŞEKİLLER DİZİNİ

Şekil 1. Dünya geneli nakit dışı para transferi	4
Şekil 2. 2015’den günümüze Bitcoin fiyat hareketleri ve hacmi	7
Şekil 3. Blokzincir çalışma yapısı	8
Şekil 4. Bitcoin madenciliği ile Hash başına günlük kâr oranları	9
Şekil 5. Bitcoin soğuk cüzdanı	10
Şekil 6. Dünya geneli yıllara göre blokzincir cüzdan sayısı	10
Şekil 7. Ülkelere göre Bitcoin’in yasallık durumu	11
Şekil 8. Türkiye'nin ilk Bitcoin ATM'si	12
Şekil 9. Makine öğrenme algoritma türleri	17
Şekil 10. Kümeleme işlemine K-Ortalamaları modeli uygulaması	18
Şekil 11. Denetimli öğrenme ve denetimsiz öğrenme farkı	20
Şekil 12. Karışıklık matrisi	20
Şekil 13. 2014-2021 yılları arası BTC fiyat değişimi	31
Şekil 14. Haziran 2020 - Haziran 2021 arası BTC fiyat değişimi	32
Şekil 15. Normalize edilmiş Kapanış, Tweet ve Google Trend verileri (Son 1 Yıllık)	33
Şekil 16. Sürekli veri seti için tahmin mekanizması	35
Şekil 17. Kesikli veri seti için tahmin mekanizması	36
Şekil 18. Kesikli tüm veriler RF yöntemi optimizasyon sonuçları	42
Şekil 19. Kesikli tüm veri seti RF tahmininde indikatörlerin önem oranı	43
Şekil 20. Kesikli 1 yıllık veriler RF yöntemi optimizasyon sonuçları	44
Şekil 21. Kesikli 1 yıllık veriler RF yöntemi indikatörlerin önem oranı	46
Şekil 22. Sürekli tüm veri RF yöntemi optimizasyon sonuçları	47
Şekil 23. Sürekli tüm veri RF yöntemi indikatörlerin önem oranları	48
Şekil 24. Sürekli 1 yıllık veriler için RF yöntemi optimizasyon sonuçları	49
Şekil 25. Sürekli 1 yıllık veri RF yöntemi indikatörlerin önem oranı	51
Şekil 26. Kesikli tüm veri KNN yöntemi indikatörlerin önem oranları	53
Şekil 27. Kesikli 1 yıllık veri KNN yöntemi indikatörlerin önem oranı	55
Şekil 28. Sürekli verilerde KNN en iyi komşu oranı	55
Şekil 29. Sürekli tüm veri KNN yöntemi indikatörlerin önem oranı	57
Şekil 30. Sürekli 1 yıllık veri KNN yöntemi indikatörlerin önem oranı	58
Şekil 31. Kesikli tüm veri XGBoost optimizasyon ayarları	60
Şekil 32. Kesikli tüm veri XGBoost yöntemi indikatörlerin önem oranları	61
Şekil 33. Kesikli 1 yıllık veriler için XGBoost optimizasyon sonuçları	62

Şekil 34. Kesikli 1 yıllık veri XGBoost yöntemi indikatörlerin önem oranları	64
Şekil 35. Sürekli tüm veri için XGBoost optimizasyon sonuçları	65
Şekil 36. Sürekli tüm veri XGBoost yöntemi indikatörlerin önem oranları	66
Şekil 37. Sürekli tüm veriler için XGBoost optimizasyon ayarları	67
Şekil 38. Sürekli 1 yıllık veri XGBoost yöntemi indikatörlerin önem oranları	69
Şekil 39. Kesikli tüm veriler İçin SVM Optimizasyon Sonuçları.....	70
Şekil 40. Kesikli 1 yıllık veriler için SVM optimizasyon sonuçları	72
Şekil 41. Sürekli tüm veriler için SVM optimizasyon sonuçları	73
Şekil 42. Sürekli 1 yıllık veriler İçin SVM optimizasyon sonuçları.....	75
Şekil 43. Sürekli 1 yıllık veri SVM yöntemi indikatörlerin önem oranları	76
Şekil 44. Kesikli tüm veriler NB yöntemi indikatörlerin önem oranları	79
Şekil 45. Kesikli 1 yıllık veriler NB yöntemi indikatörlerin önem oranları	81
Şekil 46. Sürekli tüm veriler NB yöntemi indikatörlerin önem oranları	83
Şekil 47. Sürekli 1 yıllık veriler NB yöntemi indikatörlerin önem oranları	85
Şekil 48. Kesikli tüm veri için kayıp ve doğruluk oranları (500 Ölçek).....	87
Şekil 49. Kesikli tüm veri için kayıp ve doğruluk oranları (50 Ölçek).....	87
Şekil 50. Kesikli 1 yıllık veri için kayıp ve doğruluk oranları (500 Ölçek)	88
Şekil 51. Kesikli 1 yıllık veri için kayıp ve doğruluk oranları (200 Ölçek)	89
Şekil 52. Sürekli tüm veri için kayıp ve doğruluk oranları (500 Ölçek).....	89
Şekil 53. Sürekli tüm veri için kayıp ve doğruluk oranları (200 Ölçek).....	90
Şekil 54. Sürekli 1 yıllık veri için kayıp ve doğruluk oranları (500 Ölçek)	90
Şekil 55. Kesikli tüm veri için kayıp ve doğruluk oranları (500 Ölçek).....	93
Şekil 56. Kesikli tüm veri için kayıp ve doğruluk oranları (50 Ölçek).....	93
Şekil 57. Kesikli 1 yıllık veri için kayıp ve doğruluk oranları (500 Ölçek)	94
Şekil 58. Kesikli 1 yıllık veri için kayıp ve doğruluk oranları (200 Ölçek)	95
Şekil 59. Sürekli tüm veri için kayıp ve doğruluk oranları (500 Ölçek).....	96
Şekil 60. Sürekli tüm veri için kayıp ve doğruluk oranları (50 Ölçek).....	96
Şekil 61. Sürekli 1 yıllık veri için kayıp ve doğruluk oranları (500 ölçek)	97
Şekil 62. Kesikli tüm verilerin algoritmalarında doğruluğa göre ayrışması	99
Şekil 63. Kesikli tüm veri seti ile eğitim ve tahmin süreleri (saniye)	100
Şekil 64. Kesikli 1 yıllık verilerin algoritmalarında doğruluğa göre ayrışması.....	101
Şekil 65. Kesikli 1 yıllık veri seti ile eğitim ve tahmin süreleri (saniye).....	102
Şekil 66. Sürekli tüm verilerin algoritmalarında doğruluğa göre ayrışması.....	104
Şekil 67. Sürekli tüm veri seti ile eğitim ve tahmin süreleri (saniye)	106
Şekil 68. Sürekli 1 yıllık verilerin algoritmalarında doğruluğa göre ayrışması.....	107
Şekil 69. Sürekli 1 yıllık veri ile eğitim ve tahmin süreleri (saniye)	108

KISALTMALAR VE SİMGELER DİZİNİ

ANN	: Yapay Sinir Ağları
BOLL	: Bollinger Bantları
BTC	: Bitcoin
CCI	: Emtia Kanal Endeksi
DA	: Diskriminant (Ayrırma) Analizi
FN	: Yanlış Negatif
FP	: Yanlış Pozitif
KNN	: K en yakın komşu
LSTM	: Kısa-Uzun Süreli Bellek
MA	: Hareketli Ortalama
MACD	: Hareketli Ortalama Yakınsaması/İraksaması
MAE	: Ortalama Mutlak Hata
MAPE	: Ortalama Mutlak Yüzde Hatası
MSE	: Ortalama Kare Hatası
MOM	: Momentum
NB	: Naive Bayes
RAE	: Bağlı Mutlak Hata
RF	: Rastgele Orman
RMSE	: Kök Ortalama Kare Hatası
RSI	: Relatif Güç Endeksi
SAR	: Stop And Reversal
SVM	: Destek Vektör Makineleri
TP	: Gerçek Pozitif
TN	: Gerçek Negatif
USD	: Amerikan Doları

BİRİNCİ BÖLÜM

Giriş

Geçtiğimiz otuz yıl içinde internetin hayatımıza girmesi ile günlük yaşantımız hızlı bir dijital dönüşüme girmiş, fiziki mekanların ve sınırların ötesinde yeni sanal gerçeklikler yaşantımızın parçası haline gelmiştir. Bu dönüşümün parçası olarak mesai kavramından bağımsız dünyanın her yerinde gün boyu birbiriyle etkileşim içinde olan iş modelleri gelişmeye başlamıştır. Teknolojinin ve internetin gelişimine paralel olarak, para kavramı da büyük ölçüde dijitalleşerek banka hesaplarına bağlı rakamlar haline gelmişlerdir. 2009 yılında tüm bu gelişmelerin yanına bankacılık sistemlerinden bağımsız, herhangi bir devlet veya merkeze bağlı olmayan, fiziki üretilmemiş ve dijital blokların belirli bir sıraya göre dizilmesi ile sınırlı sayıda elde edilebilen Bitcoin (BTC) adında tamamen dijital bir para birimi yeni bir değişim aracı olarak hayatımıza girmiştir.

Dijital dünyanın gelişmesine paralel veri analiz araç ve yöntemleri de aynı şekilde gelişmektedir. Birçok akademik çalışma ve yatırım uzmanı tarafından istatistik biliminin para piyasaları için kullanıldığı temel ve teknik analiz yöntemleri, Makine öğrenme algoritmaları ile birleştirilerek daha iyi tahmin araçları geliştirilmeye çalışılmaktadır. Son yıllarda Bitcoin'in yatırım aracı olarak sıkça işlem görmesi, bu alanda yapılan çalışmaları da arttırmıştır.

Araştırmanın Konusu, Problemi ve Varsayımları

Bu çalışmada Blok Zincir, Kripto Para ve Bitcoin kavramları, finans piyasasındaki yeri, kripto para borsaları, şimdiye kadar Bitcoin özelinde fiyat hareketlerinin tahminlerinde kullanılan makine öğrenme algoritmaları, tahmin için kullanılan analiz göstergeleri (indikatörler) ve geçmiş akademik çalışmalarda izlenen yöntemler incelenmiştir. Bitcoin fiyat tahminine yönelik akademik çalışmaların birçoğunda eğitim verisi olarak açılış, kapanış, en düşük ve en yüksek fiyat verilerinden faydalandığı, bazı çalışmalarda ise teknik indikatörlerin eğitim verisi olarak kullanıldığı gözlenmiştir. Bunların dışında günlük tweet sayıları, google arama trendleri veya dolar, altın, petrol, emtia ve diğer kripto paraların fiyatları ile Bitcoin fiyat korelasyonu üzerine tahmin çalışmalarına rastlanmıştır.

Bu çalışmalar içinde teknik indikatörle yapılan tahminlerin daha iyi sonuçlar verdiği gözlenmiştir. Bu sebeple çalışmamızın veri setinde diğer çalışmalarda sıkça kullanılan teknik indikatörlere yer verilmiştir. Ayrıca diğer çalışmalarda kesikli yapıda kullanımına rastlanmamış örtüşme indikatörlerinden Üstel Hareketli Ortalamalar (EMA), Bollinger Bantları (Boll) ve momentum indikatörlerinden Ortalama Gerçek Aralık (ATR) indikatörleri veri setlerine dahil edilerek tahmin performansına etkisinin ölçümü gerçekleştirilmiştir. Günlük Bitcoin'e ilişkin atılan Tweet Sayıları ve Google Arama Trendleri yine kesikli ve sürekli yapıdaki veri setlerine dahil edilerek tahminlere olan etkisi ölçülmüştür. Veri aralığı olarak Bitcoin'in borsa hacminin arttığı 2014 yılından günümüze kadar olan verilerle 1. veri oluşturulmuştur. Bu veri setinden son 1 yıllık veriler ayrılarak 2. Veri seti oluşturulmuştur. Veri setleri hem kesikli hem de sürekli yapıda oluşturularak, toplamda 4 veri seti ile 7 farklı makine öğrenme algoritması optimize edilerek kıyaslamaları yapılmıştır.

Araştırmanın Sınırlılıkları

Araştırma veri büyüklükleri düşünülerek genel olarak makine öğrenme yöntemleriyle sınırlı kalmaya özen gösterilmiş, derin öğrenme yöntemleri tercih edilmemiştir. Yapay sinir ağları ve tekrarlayan sinir ağları makine öğrenme yöntemlerine eklenmiş ve diğer akademik çalışmalarda sık değinilmeyen eğitim süreleri ve tahmin süreleri üzerine kıyaslanarak yorumlanmıştır. Araştırmada Bitcoin kripto para biriminin günlük dolar fiyatı baz alınmıştır.

Araştırmanın Amacı

Araştırmanın amacı Bitcoin fiyat hareketlerinin 2020 yılının ikinci yarısından 2021 yılının ilk yarısına kadar volatilité açısından 2014 yılından günümüze geçmiş 7 yıl ile gösterdiği benzerlikten dolayı; Bitcoin'e ait teknik indikatörler, Tweet sayıları ve Google Trend verileri ile oluşturulmuş 7 yıllık uzun vadeli ve 1 yıllık kısa vadeli 2 ayrı veri setinin farklı makine öğrenme yöntemlerinin eğitilmesinde kullanılarak performans farklarının ölçülmesidir. Elde edilen sonuçlara göre her gün yenileri eklenen ancak henüz uzun vadeli verisi oluşmamış kripto paraların fiyat yönü tahmininde kullanılabilecek en iyi makine öğrenme yönteminin tespit edilebilmesini amaçlamaktadır.

İKİNCİ BÖLÜM

Kavramlar

Bu bölümde çalışmada yer alan kavramların detaylı açıklamaları ve Bitcoin'in hayatımızdaki yeri üzerine açıklamalarda bulunulacaktır.

Para

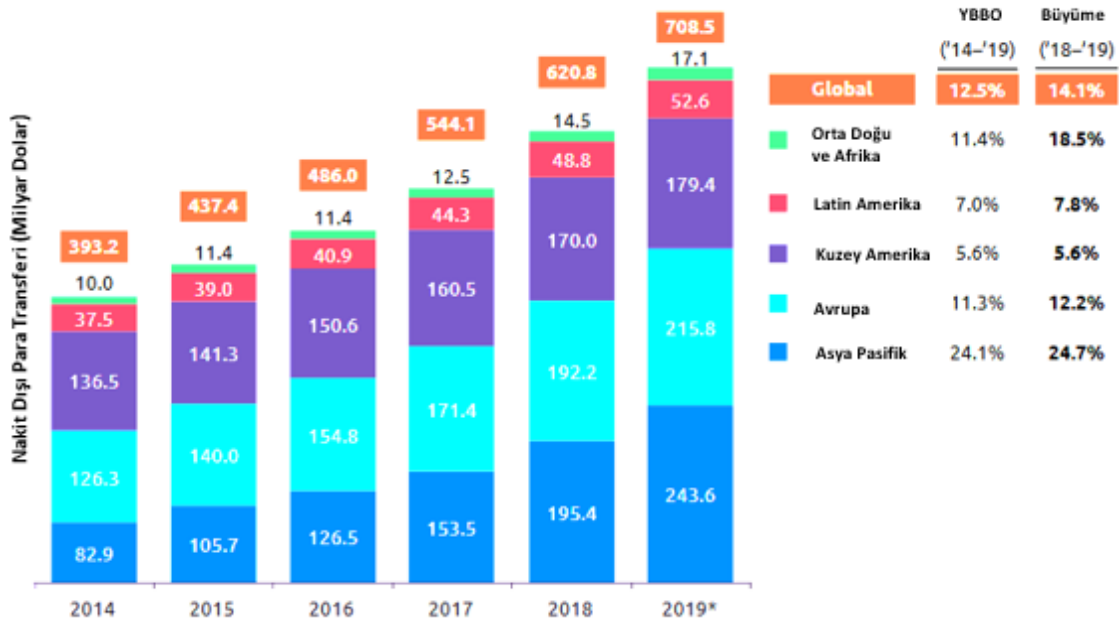
İnsanlık tarihinde takas (trampa) ile başlayan ticaret hayatı, göçebe hayattan yerleşik hayata geçtikten sonra yerini değerli madenlere bırakmıştır. Takasta kullanılan gıda ürünlerinin taşıma, bölme ve saklama gibi zorunlulukları yerine doğada nadir bulunan altın, gümüş veya bakır madenlerin ticarete kullanılmaya başlanmasıyla insanların genelini kapsayan ortak değer kavramı ortaya çıkmıştır (Keş, & Turgut, 2015, s. 33).

Merkezi otoriterin kurulması ile değerli madenlerden üretilen madeni paralar değişim aracı olarak kullanılmaya başlamıştır. 17. yüzyılın sonlarında otoritelere bağlı merkez bankaları kurularak bir senet türü olarak banknot ya da kâğıt paraların basılması ile değerli madeni paralar yerini kâğıt paralara bırakmıştır. Değerli madenden üretilmiş paraların karşılığı madenin değerine göre ölçülebilirken, kâğıt olan paranın değeri onu üreten otoritenin gücü ve güvenilirliğine göre değişkenlik göstermektedir.

Bir nesneye para adını verebilmek için halk tarafından ödeme aracı olarak benimsenmesi gerekmektedir. Daha önceleri para olarak kullanılan bir nesne veya değer, zamanla bu özelliğini yitirebilmektedir. Bazen de bir toplumun ödeme aracı olarak gördüğü bir para, başka bir toplum tarafından değişim aracı olarak görülmeyebilir. Bu sebeple para kavramı, zaman ve mekâna bağlı olarak değişim gösterebilen bir kavram olarak karşımıza çıkmaktadır (Şıklar, 2004, s. 3-4). Günümüzde devletler tarafından basılan paraların değeri, piyasada var olan diğer değerlere (Döviz, Değerli Maden, Emtia) oranlanarak bulunmaktadır.

İnternetin hayatımıza girmesiyle dijitalleşen para, kredi kartlarının da yaygınlaşmasıyla fiziksel olmaktan çıkmış, banka hesaplarında rakamlara dönüşmüştür. Bankalararası Kart Merkezi'nin 2020 yılı raporuna göre, Ülkemizde pandemi sürecinin de etkisiyle 2020 yılında

kartlı ödemeler %17 artarak 1,15 trilyon TL olmuştur. Bu dijital dönüşümün göstergesi olarak dünya geneli nakit dışı para transfer artışı Şekil 1’de görülmektedir.



Şekil 1. Dünya geneli nakit dışı para transferi. (World Payments Report, 2020)

İnternet sayesinde artık Bitcoin ve diğer kripto paralar da değişim aracı ve para olarak değerlendirilebilmektedir. Bir değişim aracının para olarak kabul edilebilmesi için belli özelliklere sahip olması gerekmektedir. Bu özellikler Taşınabilirlik, Dayanıklılık, Bölünebilirlik, Kabul Edilebilirlik ve Taklit Edilememesi olarak beş maddede belirlenmiştir (Öztürk, & Koç, 2006, s. 210).

Para Türleri

Değişim aracı olarak parayı günümüzde itibari para, elektronik para, dijital para ve sanal para olarak dört farklı türde gruplayabiliriz.

İtibari para: 20. yüzyılın ikinci yarısına kadar Devletler değerli maden rezervlerine karşılık göstererek para basmakta iken, günümüzde Merkez Bankaları, gücünü resmi otoriteden yani bağlı olduğu devletin itibarı üzerinden değerlendirilen paralar basmaktadır. İtibari Para, otoriteye bağlı merkez bankalarınca basılan; altın veya gümüş gibi değerli madenden üretilmemiş ve karşılığı bunlar olmayan, ancak altında imzası olan otoriteye bağlı olarak taklit edilmeyeceği güvence altına alınarak mal ve hizmet alımı için piyasaya sürülmüş olan banknot

ve kâğıt paralardır. (Şıklar, 2004) Kâğıt paranın taşınma ve çalınma gibi zorlukları olması sebebiyle itibari paranın büyük kısmı günümüzde banka mevduatlarında saklanmaktadır. (Afşar, Özdemir, & Arslan, 2012)

Elektronik para: Elektronik para üzerinde manyetik yüzey, RFID veya çip bulunan fiziki bir objenin içine değer olarak yüklenmiş değişim aracıdır. Elektronik Para, banka kartı veya kredi kartından farklı olarak banka sistemi içindeki paraları değil; kullanılmak üzere elektroniğe dönüştürülmüş paraları temsil eder. Otobüs kartları, eğlence merkezlerinde kullanılmak üzere bakiye yüklenen kartlar, belirli bir cihazla bakiye alınabilen yemek kartları içinde bulunan değerler elektronik paraya örnek olarak verilebilir. (Karpuz, 2012, s. 145) Yine benzer şekilde bazı kuruluşlar internet sayfalarında kendi kontrollerinde olan bakiye sistemi ile elektronik para üretmekte ve alışveriş için bu paranın kullanılmasına olanak sağlamaktadır.

Dijital para: Günümüz paralarının büyük kısmını oluşturan dijital para, mal ve kıymet ve paraların elektronik veri tabanlarında saklanan ve elektronik olarak transferi yapılabilen, dilendiği zaman fiyatı karşılığı değerler arası dönüştürülebilen para türüdür. Günümüz banka hesaplarında yer alan paraların tamamı dijital para olarak saklanmaktadır. Dijital paralar Merkez Bankaları gibi devlet otoritelerinin kontrolü ve denetimi altındadır. Tüzel veya gerçek kişilerin adlarına tanımlı Banka hesaplarında kayıt altında tutulan dijital paralar, günümüzde internet teknolojisi sayesinde günün büyük bölümünde havale, EFT veya swift gibi farklı yöntemler ile hesaplar, bankalar ve ülkeler arası transfer edilebilmektedir. Ülkemizde 2019 yılında yayınlanan 11. Kalkınma planında yer alan “Blokzincir Tabanlı Dijital Merkez Bankası Parası” kararı sonrası çalışmalar devam etmiştir ve Merkez Bankacınca 2021 Eylül ayının ortasından itibaren pilot testlerin başlayacağı bildirilmiştir. 5 Eylül 2021 tarihli Resmi Gazete’de yayınlanan 2022-2024 Orta Vadeli Programda da dijital liraya dönük testlerin genişleyeceği bildirilmiştir.

Sanal para: Herhangi bir resmi otorite veya devlet tarafından üretilmediği halde değişim aracı olarak değer almış ve kabul görmüş dijital ortamda üretilen para türüne sanal para adı verilmektedir. Diğer para türlerinden en büyük farkı paranın bağlı olduğu bir otoritenin olmaması ve banka gibi araçlar olmadan transferlerin yapılmasına imkân tanınmasıdır. Kripto paralar sanal paralar arasında en bilinen örneklerdendir.

Kripto Para ve Bitcoin

Gerçek paraya alternatif dijital bir takas aracı olarak ortaya çıkmış kripto para, şifrelenmiş olarak kümelenen veri blokundan oluşmaktadır. Kripto Para (Sanal Para) herhangi devlete ya da kurala bağlı olmayan, internet dışında kullanımı olmayan, piyasaya sürülmesi ve denetimi parayı ortaya çıkaranlar tarafından yapılan ve sadece bu kripto para birimini kullanıcıları arasında internet üzerinden takas veya alışveriş yapmak üzere kullanılabilen sayısal para türüdür. Kripto paralar arz-talep içinde fiyatlandırıldığından dolayı dijital emtia olarak adlandırılabilir.

Açık kaynak kodlu olarak Kripto Paraların kütüphaneleri kullanıcılara açıktır ve Blok zincir yapısıyla çalışırlar. Blok zincir tüm kullanıcılarında aynı kopyası olan büyük bir veri tabanı olarak tanımlanabilir. Yapılan işlemler geri alınmaz şekilde Blok olarak veri tabanına kaydedilir ve zincirin sonuna eklenir. Bilgiler veri tabanının sahibi de dahil hiçbir kullanıcı tarafından değiştirilemez ve silinemez.

2008 yılında küresel kriz sonrası, Satoshi Nakamoto adını kullanan biri veya bir grup tarafından “Kullanıcıdan kullanıcıya elektronik nakit sistemi” adlı makale ile adını duyuran Bitcoin, ilk olarak 2009 yılında üretilmeye başlanmıştır. İnternet üzerinden 7/24 işlem yapılabilen, arzı sınırlı sayıda, hızlı transfer edilebilen ve takibi zor olan Bitcoin gün geçtikçe daha çok kişi tarafından tercih edilmeye başlanmıştır. Talep ve spekülasyonlar arttıkça 2010 yılından bu yana değeri katlanarak artan Bitcoin, birçok yatırımcıyı da kendine çekmektedir. (Halaburda, & Sarvary, 2016, s. 97-98)

Bitcoin üst sınır olarak 21 Milyon adetle sınırlandırılmıştır. Yapısı sayısal veri dizisinden oluştuğundan dolayı, her bir diziyi çözmek için bilgisayar işlemcilerinin gücüyle ayıklanması gerekmektedir. İlk çıktığı 2008 yılında ayıklanan blok başına 50 Bitcoin kazanılan yapıda her 4 yılda bir kazanılan oran yarıya indirilmektedir. Günümüzde 6,25’a düşmüş bu oranda Ağustos 2021 itibarıyla 18,79 milyon adet Bitcoin kazılmıştır. Bu sayı toplam Bitcoin’in %89,42’sini oluşturmaz.



Şekil 2. 2015’den günümüze Bitcoin fiyat hareketleri ve hacmi

İlk 5 yılında 1000 USD seviyelerine kadar değer kazanıp, sonraları tekrar değer kaybeden Bitcoin; 2017 yılında Japonya’nın “Sanal Para Birimi Kanunu” ile Bitcoin ve diğer kripto para birimlerinin ödeme aracı olarak kabul etmesi, Amerika ve Kanada gibi devletlerde Bitcoin ATM’lerinin yer almaya başlaması ve büyük yatırımcıların Bitcoin talebinin artması ile 2017 yılı sonrası popüler bir yatırım aracı haline gelmiştir. Şekil 2’de yer alan grafikte de görüleceği üzere 2017 yılından sonra artan talep ve hacimle birlikte fiyatı giderek artan Bitcoin küresel pandemi salgını sonrası hem fiyat hem de hacim olarak 7 kat artmıştır.

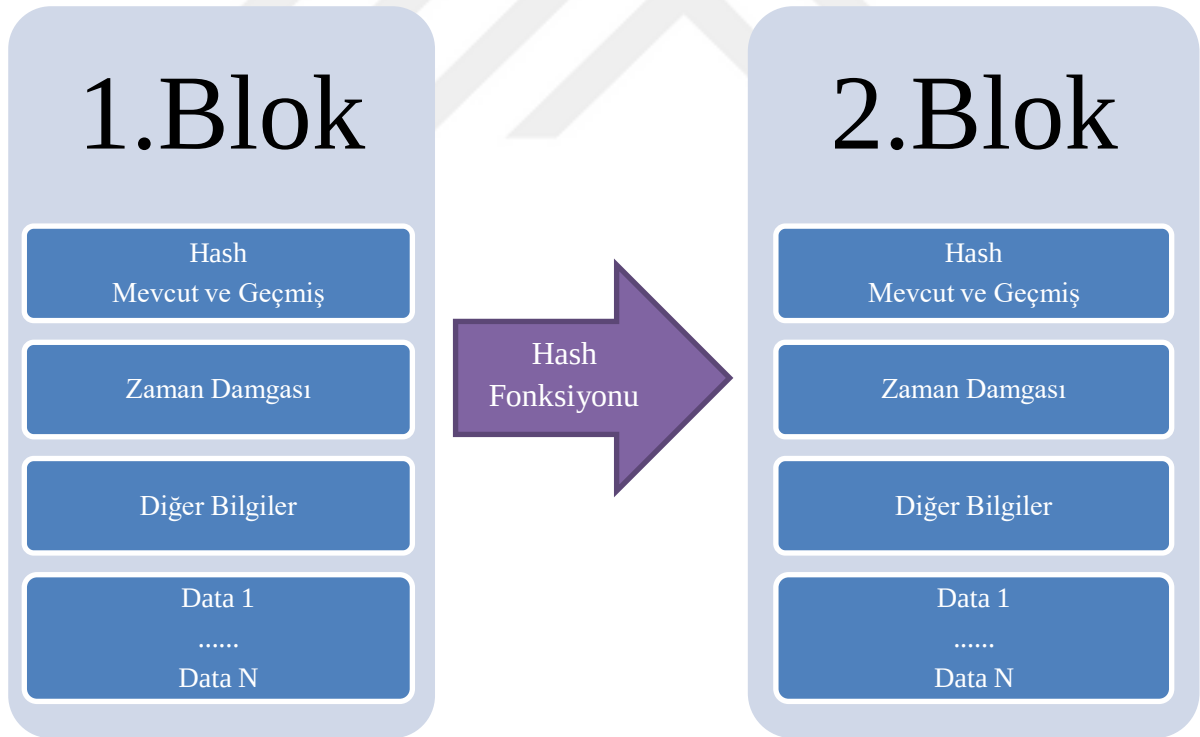
İlk kripto para olan Bitcoin’in ardından, 2011 yılında Litecoin ile başlayan kripto para süreci günümüzde 6000’den fazla çeşitle kripto para ve token borsalarında işlem görmektedir. Tokenların kendine ait bağımsız bir blok zinciri bulunmamaktadır. Yani kendi başına bir kripto varlık olmayıp, kripto paradan üretilen hisse senedi, dijital ürün veya menkul kıymeti ifade edilebilir. Örneğin piyasaya yeni girecek bir kripto para, sermaye toplamak için önden token satışı yapılabilir veya bir firma başka bir kripto para teknolojisi altında ürettiği tokenlara isim vererek bunların satışını yapılabilir. Sene başında 1 trilyon dolardan az seviyede olan kripto para market hacmi Mayıs 2021 itibarıyla toplam 2,48 trilyon dolar seviyesine gelerek ticaretin her alanında karşımıza çıkmaya başlamıştır. Kripto para piyasası içinde halen Bitcoin en büyük hacmi oluşturmaktadır. Bitcoin’den sonra en sık kullanılan kripto paralar 2015 yılında Vitalik Buterin tarafından kodlanan Ethereum(ETH) ve 2017 yılında Charles Hoskinson tarafından kodlanan Cardano(ADA) kripto para birimidir.

Blokzincir Teknolojisi

Blokzincir teknolojisi kriptografi güvenli hale getirilerek birbirine zincir gibi bağlanmış veriler dizisidir. Kriptografiye dayanan ilk blokzincir çalışması Stuart Haber ve W. Scott Stornetta tarafından 1991 yılında açıklanmıştır, ancak Satoshi Nakamoto'nun 2008 yılında ortaya koyduğu Bitcoin projesinde kavramsallaştırılmıştır. Güvenirliğini kanıtlamış blokzincir teknolojisi, kripto para dışında banka kayıtlarının tutulması, telif dosyaları, satışlarla ilgili kayıtların tutulması, kablosuz ağların paylaşım kayıtları, veri depolama, sigorta sektörü gibi birçok alanda kullanılmaya başlanmıştır. Bu çalışmada blokzincir Bitcoin sistemi üzerinden anlatılacaktır.

Blokzincir yapısı:

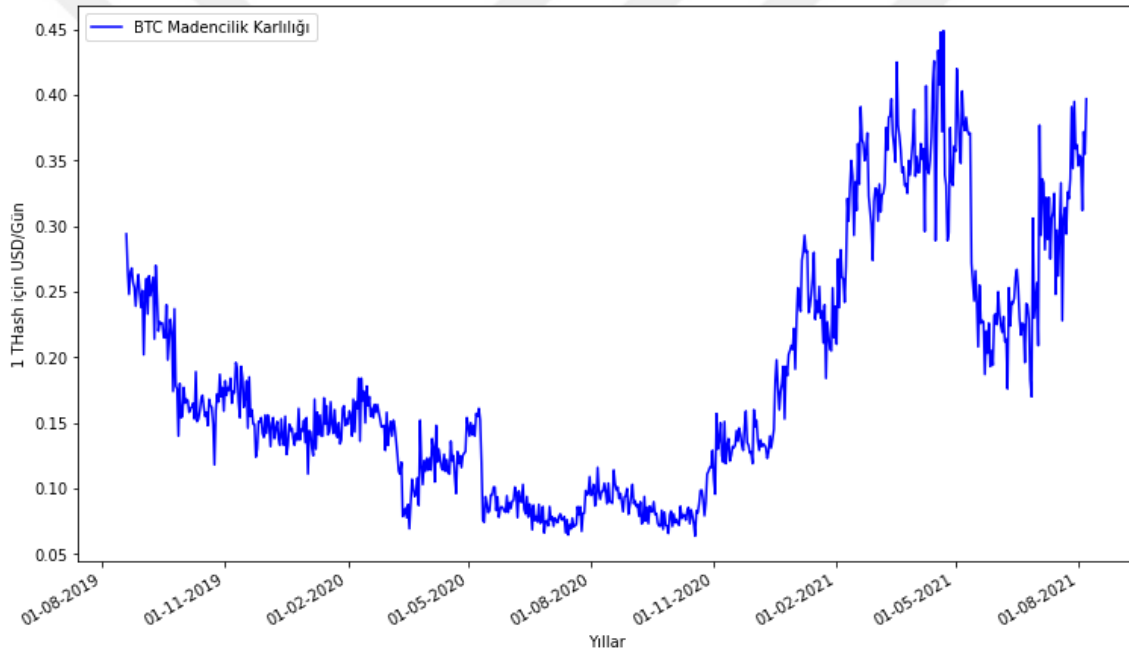
Bitcoin blokzincir yapısında, her bir blokta yer alan veriler verilerin doğruluğuna şahitlik edenler (madenciler) tarafından bloklara zincir şeklinde veritabanına eklenerek tamamlanır. Bu veritabanı sistemin içindeki tüm kullanıcılar tarafından aynı veriye sahip olacak şekilde birbirini doğrulama mantığı ile çalışır. (Nakamoto, 2008)



Şekil 3. Blokzincir çalışma yapısı (Nakamoto, 2008)

Madencilik (Mining):

Madencilik blokzincir içindeki verilerin doğruluğuna şahitlik eden kullanıcıların, bu noter benzeri verdikleri hizmet karşılığı sistem tarafından ödüllendirilmesi işlemidir. Kullanıcılar arasında her bitcoin transferi gerçekleştiğinde bunun diğer kullanıcılar tarafından doğrulanması gerekmektedir. Bu doğrulama işlemi sonucu yeni blok oluşturmak için yine şifreli algoritmanın oluşturulması gerekmektedir. Bu da şahitlik eden kullanıcıların bilgisayarları tarafından yüksek efor ve enerji tüketimi sonucunda elde edilmektedir. (Vranken, 2017) Madencilerin bu işlemi sonucunda matematiksel formülden geçen eşsiz yapı oluşmaktadır. Oluşan blok yapının özeti olan yapıya Hash adı verilmektedir. Bitcoin Hash algoritması SHA-256 kullanmaktadır.



Şekil 4. Bitcoin madenciliği ile Hash başına günlük kâr oranları

Bitcoin madenciliğinde şahitlere iki tip ücret ödenmektedir. Birincisi madencilik sonucu oluşan yeni Bitcoin'den gelen pay, diğeri ise “transaction fee” yani işlem ücretidir. İlk çıktığı 2008 yılında ayıklanan blok başına 50 Bitcoin kazanılan yapıda her 4 yılda bir kazanılan oran yarıya indirilmektedir. Günümüzde bu oran 6,25 seviyesine düşmüştür. Ancak işlem ücretleri, işlem hacmi arttıkça arttığından dolayı sistem madenciler için belli oranda kar sağlamaya devam etmektedir.

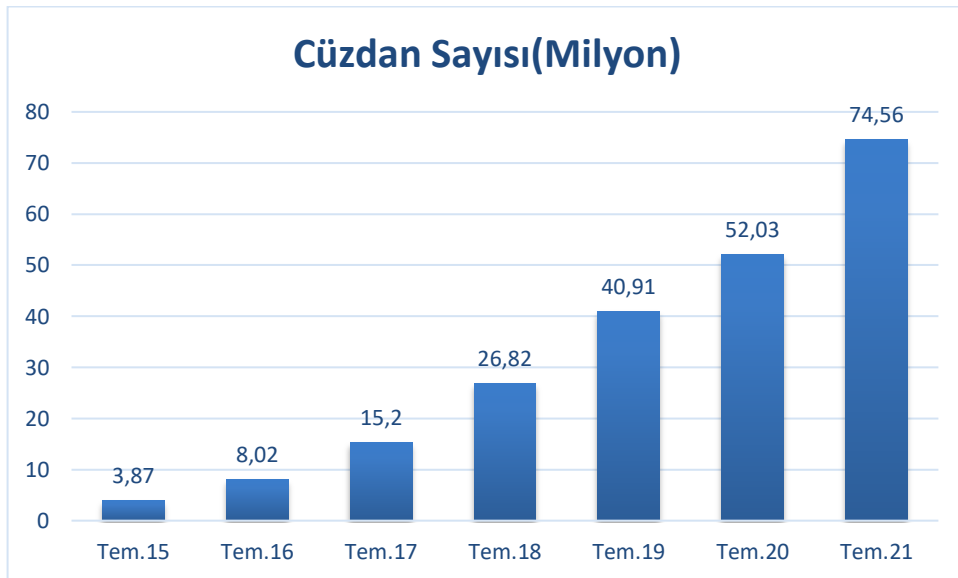
Bitcoin hesap cüzdanı ve bitcoin transferi.

Kişilerin mevcut Bitcoinlerini saklayabildiği sıcak ve soğuk olmak üzere iki tip cüzdan bulunmaktadır. İnternete bağlı yani çevrimiçi şekilde erişilebilen cüzdanlara sıcak cüzdan, çevrimdışı ortamda usb bellek gibi fiziki olarak depolama yapan cüzdanlara ise soğuk cüzdan adı verilmektedir. Sıcak cüzdanlar sunucularda saklandığından dolayı güvenlik açısından soğuk cüzdanlara oranlara daha risklidir, ancak depolama alanı daha büyük olduğundan dolayı daha fazla Blokzincir içinde barındırabilirler.



Şekil 5. Bitcoin Soğuk Cüzdanı (coinsspent.com, 2021)

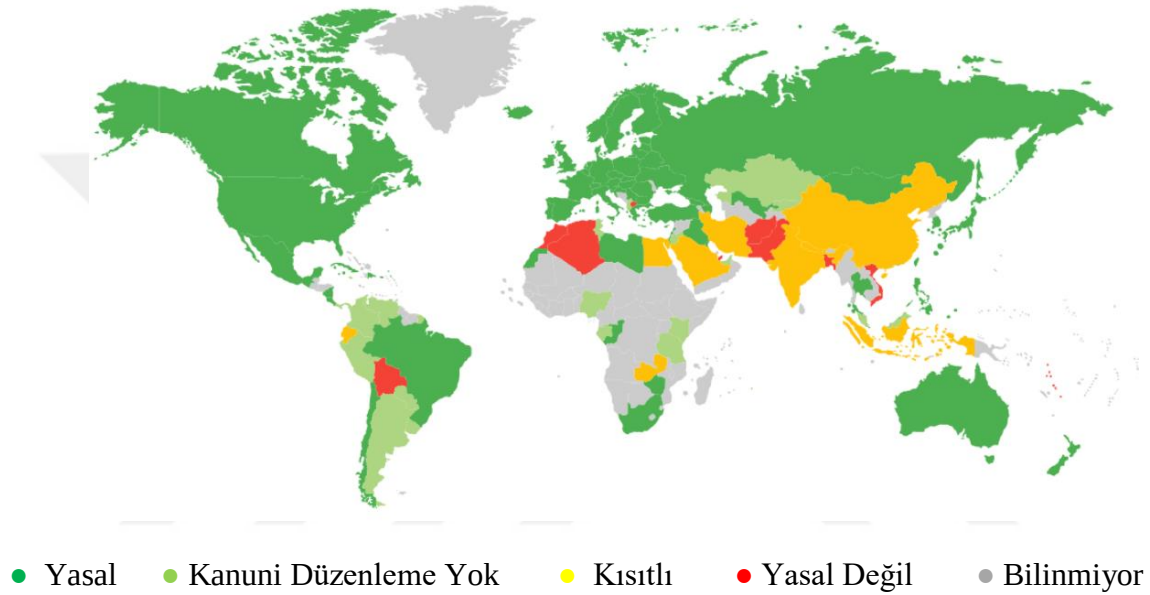
Hesaplar arası Bitcoin transferi için taraflar internete bağlanılarak transferi kendilerine ait dijital imza ile imzalamaları gerekmektedir. Temmuz 2021 itibarıyla 74,56 milyon aktif blokzincir cüzdanı bulunmaktadır.



Şekil 6. Dünya Geneli Yıllara Göre Blokzincir Cüzdan Sayısı (Blockchain.com, 2021)

Dünyada ve ülkemizde bitcoin kullanımı.

Dünyada Bitcoin soğuk ve sıcak cüzdanların arasında transfer dışında büyük ölçüde kripto para borsalarda işlem görmektedir. Bitcoin'in herhangi bir otorite tarafından kontrol edilmiyor olması para akışının izlenmesi ve vergilendirmeye ilişkin Devletlerin kaygısını arttırsa da, insanların kripto paraya yönelik talepleri ticaret hayatında kabul görmesine sebep olmaktadır. Bu sebeple bazı devletler kriptoya yönelik yasaklar getirirken, bazı devletler belirli ölçülerde kullanımına, alım ve satımına müsaade etmektedir.



Şekil 7. Ülkelere göre Bitcoin'in Yasallık Durumu (CoinDance, 2021)

Orta Amerika ülkelerinden El Salvador Haziran 2021'de Bitcoin'i resmi para birimi olarak tanıyan ilk ülke olmuştur. Hindistan'da buna benzer çalışmalar yapılmaktadır. Ülkemizde de 16 Nisan 2021 tarihli Resmi Gazete'de "Ödemelerde Kripto Varlıkların Kullanılmamasına Dair Yönetmelik" yayınlanmış; 30 Nisan 2021 tarihinden geçerli Kripto paraların mal ve hizmet alımında kullanımı yasaklanmıştır ancak kripto para borsalarına yatırımın önünde engel bulunmamaktadır. İlki 2013 yılında Kanada'da açılan Bitcoin ATM'lerinden Dünya üzerinde 71 ayrı ülkede 16 binin üzerinde bulunmaktadır. Borsaya göre komisyon miktarı daha yüksek olan bu ATM'lerden Ülkemizde Ağustos 2021 itibarıyla 9'u İstanbul'da olmak üzere, Ankara, Antalya ve İzmir illerimizde 1'er tane olmak üzere toplamda 12 adet bulunmaktadır. Bu ATM'lerden belirli bir limitle sınırlı olmak üzere cüzdanınızın QR kodunu taratarak Bitcoin alınabilmekte ve satılabilmektedir.



Şekil 8 . (Türkiye'nin İlk Bitcoin ATM'si Hizmete Girdi, 2018)

Temel ve Teknik Analiz

Finansal piyasaların tümünde olduğu gibi Bitcoin piyasasında da fiyatlarının tahmini için temel ve teknik analiz yöntemlerinden faydalanılmaktadır. Bu verileri doğru okumak, piyasanın hareket yönünün ve hacminin anlaşılması açısından önemlidir. Makine öğrenme tahminlerinde de bu teknik analiz göstergelerinin yönüne göre tahmin yürütüldüğü çalışmalar yapılmıştır.

Temel analiz ekonomik, sosyal veya politik etkenlerle finansal değerlerin fiyat tahminidir. Bitcoin açısından teknolojik analizlerde temel analizin parçası haline gelmiştir. Günlük atılan tweet sayıları, trend aramalar, Hashrate oranı, maliyet hesapları vb. veriler temel analiz yöntemleri olarak makine öğrenmesi ile birleştirilebilmektedir. Veri setleri hazırlanırken de bu analizden faydalanılabilir. Örneğin bir devletin Bitcoin'i tanıdığı tarih veya dünyaca ünlü bir firmanın Bitcoin ile ticarete başladığını duyurduğu tarihte gerçekleşen fiyat hareketleri inceleme için baz alınabilir. Bu sebeple Kripto para borsasında spekülasyon iddialarından dolayı çalışmamızda Günlük Tweet Sayıları ve Google Trend Arama verileri teknik indikatör verileri setine dahil edilerek analizler yapılmıştır.

Teknik analizde, finansal değerlerin fiyat hareketi tahmininde, alım satım hacmi veya geçmiş piyasa hareketlerine göre tahmin oluşturulur. Teknik analizler sonucunda temel analize göre daha bilimsel veriler oluşturulur. (Sarı, 1998, s. 1-4) En sık kullanılan teknik analiz indikatörleri şunlardır:

SMA – Basit Hareketli Ortalama - n zamanlık fiyatın n günlük kapanış ortalaması alınarak hesaplanan bir göstergedir. Verideki dalgalanmaları çözmek için kullanılırlar.

$$SMA = \frac{A_1 + A_2 + \dots + A_n}{n}$$

A_n = n zamanında kapanış fiyatı
 n = toplam zaman periyodu

WMA – Ağırlıklı Hareketli Ortalama - n zamanlık fiyatın belirlenen ağırlıklara göre ortalamasının alınarak hesaplanan göstergedir.

$$WMA = \frac{\sum_{t=1}^n W_t * V_t}{\sum_{t=1}^n W_t}$$

V_t = t zamanında Gerçek Değer
 W_t = t zamanında Ağırlıklandırma Değeri
 n = toplam zaman periyodu

EMA – Üstel Hareketli Ortalama - n zamanlık fiyatın ortalamasının alınarak, ağırlıklandırma işleminin yakın dönem fiyat hareketlerine göre yapıldığı göstergedir. Ağırlıklandırma gecikme oranını azaltmak için kullanılmaktadır.

$$EMA = A_t * k + EMA_{t-1} * (1 - k)$$

A = Kapanış fiyatı
 t = Hesabın yapıldığı gün
 $k = 2 \div (\text{toplam zaman periyodu} + 1)$

ATR - Ortalama Gerçek Aralık – Fiyatta oluşan her hareketin en yüksek ve en düşük farkının ortalaması alınarak hesaplanır. Piyasanın volalitesini ölçmek için kullanılır.

Gerçek Aralık - TR = maks(en yüksek, kapanış_t) - min(en düşük, kapanış_t)

t = önceki gün

$$ATR = \frac{1}{n} \sum_{i=1}^n TR_i$$

TR_i = Belirlenmiş gerçek aralık

n = toplam zaman periyodu

Boll - Bollinger Bantları – çok sık kullanılan teknik analiz yöntemlerinden olan Bollinger Bantları, hareketli ortalamanın altına ve üstüne yerleştirilerek oynaklığı(volatilite) gösteren banttır. Volalite standart sapmaya bağlı bir değişkendir. Bollinger Bantlarında 3 adet veri oluşmaktadır. Genel olarak orta bant 20 günlük hareketli ortalamadan, yukarı ve aşağı bantlar ise 20 günlük hareketli ortalamalının iki standart sapma değeri kadar aşağı ve yukarı taşınması ile elde edilir.

$$\delta = \sqrt{\frac{\sum_{j=i}^n (X_j - \bar{X})^2}{n}}$$

$$\text{Üst Bant} = \bar{X} + 2 \delta$$

$$\text{Orta Bant} = \bar{X}$$

$$\bar{X} = \frac{\sum_{j=i}^n X_j}{n}$$

$$\text{Alt Bant} = \bar{X} - 2 \delta$$

ADOSC - Birikim Dağılım Göstergesi(A/D) – Fiyat değişimi ile hacim arasındaki ilişkiyi inceleyen momentum göstergesidir. Fiyat hareketin sürdürülebilirliği olarak yorumlanır.

$$A/D = \frac{H_t - C_{t-1}}{H_t - L_t}$$

H_t = t zamanındaki en yüksek fiyat

C_t = t zamanındaki kapanış fiyatı

L_t = t zamanındaki en düşük fiyat

MACD – Hareketli Ortalama Yakınsaması/İraksaması – Bir momentum indikatörü olup, iki farklı fiyat hareketli ortalamasının arasındaki ilişkiyi gösterir. Genelde 12 günlük üstel hareketli ortalama dan 26 günlük üstel hareketli ortalama çıkarılarak elde edilir ve trend takibi için kullanılırlar.

$$MACD = EMA_{12} - EMA_{26}$$

RSI – Relatif Güç Endeksi – seçilen zaman biriminde kapanış fiyatının, bir önceki kapanış değerleriyle karşılaştırılması sonucu hesaplanır. Kısa vadeli ve orta vadeli trendin yönünü öngörebilmek için kullanılan bir indikatördür. Genel olarak 14 günlük periyodlar halinde alınır ve 30'un altına düştüğünde alış, 70'in üstüne çıktığında satış sinyali üretir.

Eğer $A_n - A_{n-1}$ pozitif ise Pozitif Kapanış (PK), negatif ise Negatif Kapanış (NK) olarak tanımlanır. Bu işlem belirlenen periyodun tamamı için uygulanır.

$$RS = \frac{Ortalama(PK)}{Ortalama(NK)} \quad RSI = 100 - \frac{100}{1 + RS}$$

A_n = n zamanında kapanış fiyatı

LW - Larry Williams – Aşırı alış ve aşırı satış bölgelerinin tespitinde kullanılır. Gösterge eğimi -20 seviyesinin üstünde ise aşırı alış, -80 seviyesinde ise aşırı satış anlamına gelir. Genel olarak 14 günlük zaman periyodu için hesaplanır.

$$LW = \frac{H_n - C_t}{H_n - L_n} * 100$$

H_n = n zaman aralığındaki en yüksek fiyat

C_t = t zamanındaki kapanış fiyatı

L_n = n zaman aralığındaki en düşük fiyat

Mom – Momentum - Belirli bir zaman aralığındaki paritelerde gerçekleşen değişimi yüzdesel olarak gösteren bir teknik analiz indikatörüdür.

$$Mom = \frac{C_t}{C_{t-n}} * 100$$

C_t = t zamanındaki kapanış fiyatı
 C_{t-n} = n zaman önceki kapanış fiyatı

Stoch – Stokastik %K %D - indikatörü belirli bir periyod içerisinde yüksek ve düşük değerlerin mevcut fiyata olan yakınlığını göstermektedir.

$$\%K = \frac{C_t - L_n}{H_n - L_n} * 100 \quad \%D = \frac{H_n}{L_n} * 100$$

C_t = t zamanındaki kapanış fiyatı
 H_n = n zaman aralığındaki en yüksek fiyat
 L_n = n zaman aralığındaki en düşük fiyat

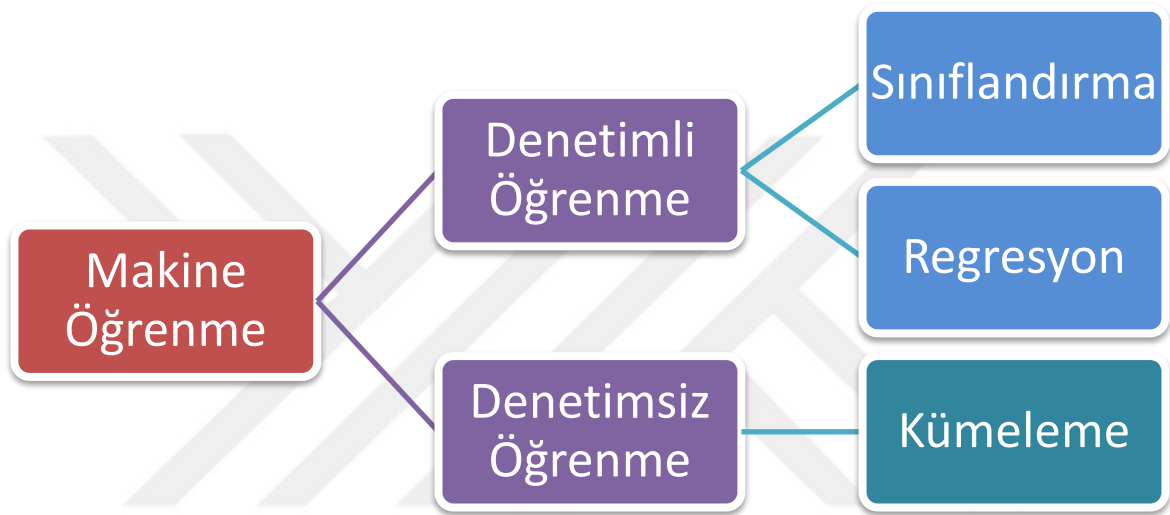
Makine Öğrenme

20. yüzyılın ikinci yarısında düşünebilen makineler üzerine çalışmalar artmaya başlamıştır. Alan Turing'in Hesaplama Makineleri ve Zeka (1950) adlı ünlü makalesi ile hız kazanan serüven, Arthur Samuel'in 1959 yılında makine öğrenme kavramını literatüre katmasıyla devam etmiştir. Bir çeşit yapay zeka olan makine öğrenmesi kısaca mevcut veriler temel alınarak öğrenme ve sonrasında tahmin gerçekleştirebilen algoritmaların yapısı ve çalışma prensiplerini kapsamaktadır. Bu algoritmalar klasik bilgisayar programlamadan farklı olarak, programın talimatlarını takip etmek yerine örnek girdi verilerini modelleyerek, tahmin yapabilir veya karar verebilirler.

Günümüzde makine öğrenme algoritmaları birçok alanda kullanılmaktadır. Örneğin, yüz tanıma sistemleri, e-posta filtrelemeleri, otomatik vitesli araçlar, sıralama problemleri, sözlük uygulamaları, borsa çözümleri, DNA dizilerinin sınıflandırılması, konuşma ve el yazısının tanımlaması, oyunlar, uyarlamalı web sayfaları, hastalıkların teşhisi ve sınıflandırılması gibi birçok alanda kullanılmaktadır. Kısaca her türlü sınıflandırma ve

regresyon probleminin çözümünde makine öğrenmeden faydalanılarak tahmin ve çözümler üretilmektedir.

İstatistik bilimine dayanan makine öğrenme, birçok alanda yerini derin öğrenme modellerine bırakırken; finansal çözümlerde derin öğrenme algoritmaları kadar etkili sonuçlar vermektedir. Özellikle derin öğrenmeye yetecek kadar veri olmayan alanlarda makine öğrenme kullanılması daha iyi sonuçlar vermektedir.

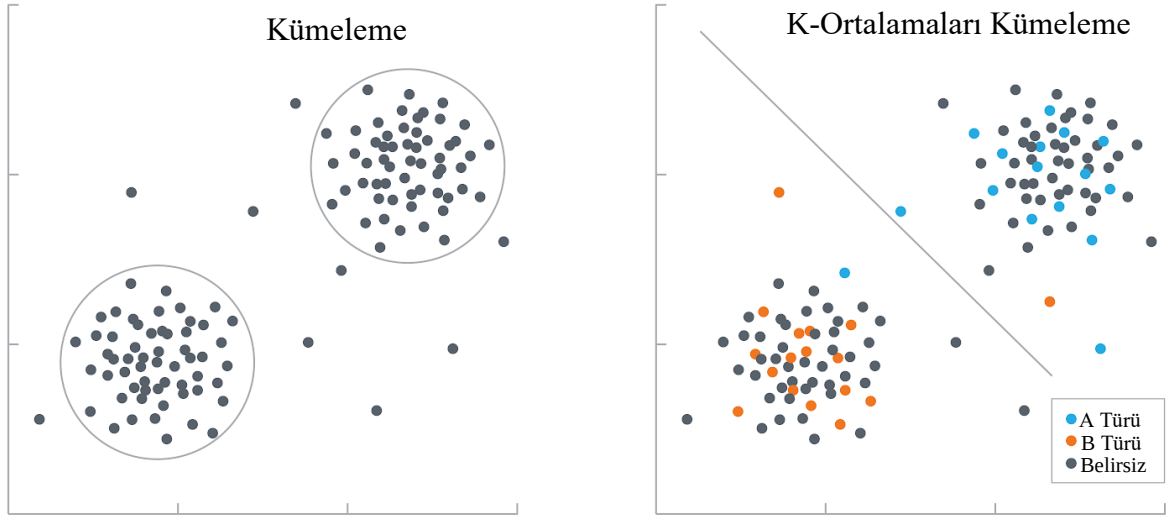


Şekil 9. Makine öğrenme algoritma türleri (Aery, & Ram, 2017, s. 91)

Makine öğrenme algoritmaları Denetimsiz Öğrenme ve Denetimli Öğrenme olarak ikiye ayrılırlar.

Denetimsiz öğrenme (Unsupervised learning).

Denetimsiz öğrenme modelinde büyük veriler üzerinde kümeleme yöntemiyle işlem yapılır. Öğrenme için yalnızca girdilerin kullanıldığı bu yöntemde; verilerin yakınlık, uzaklık, benzerlik gibi ölçütlere göre analiz ederek ortak noktaları bulunması ve bunların kümeleştirme işlemi yapılarak anlamlı sınıflara ayrılmasıdır. En sık kullanılan denetimsiz öğrenme algoritmaları k-ortalamaları ve k medoidler, hiyerarşik kümeleme, gauss karışım modelleri, gizli markov modelleri, kendi kendini düzenleyen haritalar, bulanık c-kümeleme ve çıkarımlı kümeleme modelleri olarak sayılabilir. (Aery, & Ram, 2017, s. 92)



Şekil 10. Kümeleme işlemine K-Ortalamları modeli uygulaması (Jeffares, 2019)

Denetimli öğrenme (Supervised learning).

Denetimli öğrenme modelinde makine kendisine verilen verilerden anlamlı değişkenler oluşturarak sınıflandırma ve regresyon yoluyla öğrenir. Denetimli öğrenmenin denetimsiz öğrenmeden en büyük farkı hem girdi hem de çıktıları kullanarak öğrenme modeli oluştururlar. Denetimli öğrenme modelleri Sınıflandırma ve Regresyon olarak ikiye ayrılırlar. Denetimli öğrenme modeli kullanan makine öğrenme algoritmalarından en sık kullanılanları şunlardır:

Sınıflandırma Algoritmaları

NB - Naive Bayes: algoritmasının amacı, birden fazla sınıf barındıran bir veri setinden bir öğrenim seti oluşmaktır. Oluşturulan bu öğrenim setine göre yeni gelecek verileri doğru bir sınıfa yönlendirmektir. Thomas Bayes'in 1812 yılında koşullu olasılık üzerine geliştirdiği matematik teoreminin makine öğrenme yöntemlerine uyarlanmış halidir. Naive Bayes algoritması yöntem olarak tüm koşullu olasılıkların çarpımıdır. (Pawlak, 2003, s. 488)

KNN – K-En Yakın Komşu: Sınıfları belli olan algortmada sınıflandırılmak istenen yeni verinin daha önceki verilerden k tanesine olan yakınlığına göre sınıflandırılması işlemidir. K en yakın komşu algoritması gürültülü ve basit eğitim veri setlerine karşı başarılı olması sebebiyle en sık kullanılan makine öğrenme algoritmalarından biridir. (Cunningham, & Delany, 2021, s. 3)

SVM - Destek Vektör Makineleri: algoritmasında amaç, tüm eğitim verileri için elde edilen hedeflerden en fazla e sapmasına sahip, aynı zamanda mümkün olduğunca doğrusal $f(x)$ fonksiyonunu bulmaktır. Hatalar e değerinden küçük olduğu sürece yöntem tarafından önemsenmeyecek ancak e değerinden büyük herhangi bir sapma da kabul edilmeyecektir. Destek Vektör Makinaları hem Sınıflandırma hem de Regresyon problemlerinin çözümünde kullanılmaktadır. (Cervantes, Garcia-Lamont, Rodríguez-Mazahua, & Lopez, 2020)

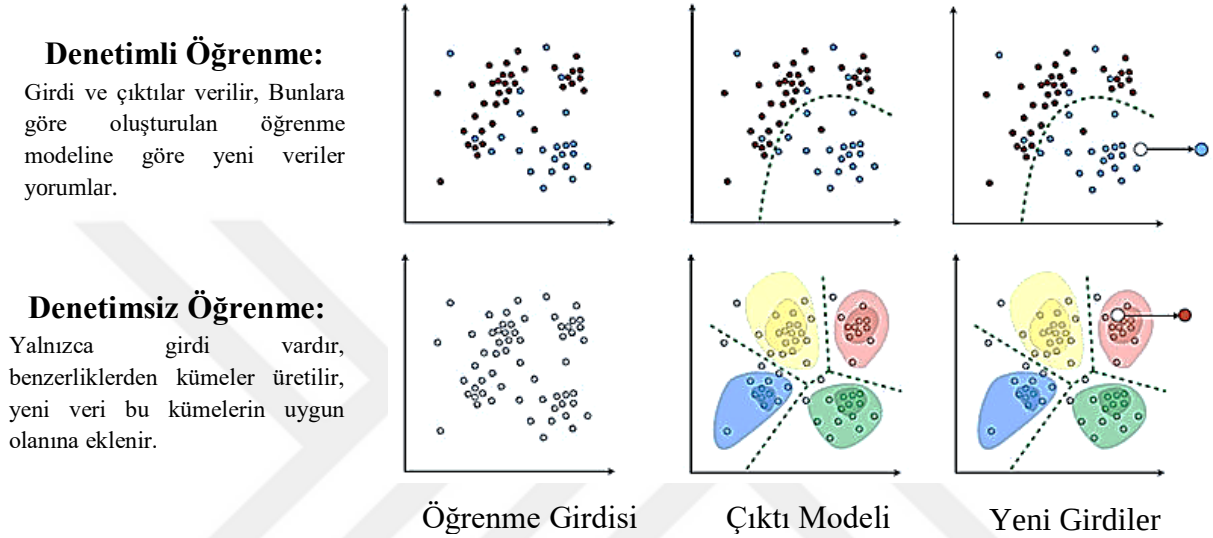
Regresyon Algoritmaları

ANN - Yapay Sinir Ağları, biyolojik sinir sisteminden etkilenecek geliştirilmiştir. Yapay Sinir Ağları, ağırlıklandırılarak birbirine bağlanmış birçok nörondan (işlem birimi) oluşan matematiksel sistemlerdir. Öğrenme, girdi ve çıktı verilerinin işlenerek bağlantı ağırlıklarının belirlenmesi (*weights of the synapses*) ile gerçekleşir. Bu işlem bir yakınsama meydana gelene kadar sistem içinde tekrarlanır. (Öztemel, 2003, s. 29) ANN girdi-çıkı değişkenleri arasındaki tüm doğrusal olmayan ilişkileri modelleyebilir. (Pabuçcu, & Yakıcı Ayan, 2017) ANN kesin olmayan verileri yönetme yeteneğinden dolayı üretim planlama, tıp, imalat ve finans gibi birçok alanda kullanılmasını mümkün kılmaktadır. (Teke, Çolak, Kiraz, & İpek, 2019)

LSTM – Kısa Uzun Süreli Bellek, tekrarlayan sinir ağları (RNN) mimarisine dayanan ve girdileri rastgele aralıklarla hatırlama yöntemiyle öğrenen yapay zeka modelidir. Öğrenme işlemi gerçekleşince değerler saklanır ve değişmezler. Nöronlar arası ileri-geri bağlantılara müsaade eder. Daha çok zaman serilerini işlemek, sınıflandırmak ve tahmin etmek için kullanılırlar. (Hochreiter, & Schmidhuber, 1997, s. 6-8)

RF - Rastgele Orman, algoritması, torbalama (bagging) temelli bir makine öğrenmesi yöntemidir (Breiman, 2001). Rastgele Orman algoritması, birden fazla karar ağacı oluşturarak doğruluğu yüksek ve kararlı modeller oluşturmak için bunları birleştirir. Algoritmada ağacı büyütmek modelin rastgelelik oranını arttırmaktadır. Yöntem olarak düğümleri ayırırken en önemli özellik yerine, rastgele bir özellik alt kümesi içindeki en iyi özelliği aramaktadır. Bu durum geniş bir çeşitlilik yaratarak daha iyi modellerin elde edilmesini sağlamaktadır. (İbrahim, Kashef, & Corrigan, 2021)

XGBoost - eXtreme Gradient Boosting, rastgele orman algoritması gibi çalışan ancak çok daha yüksek performanslı bir algoritmadır. Tahmin gücünün yüksek olması ve aşırı uyma (overfitting) çok yakalanmaması, boş verilerin yönetimi ve tüm bunları çok hızlı yapabilmesi XGBoost'u son yılların önemli makine öğrenme yöntemlerinden biri haline getirmiştir. Karar ağaçları tabanını kullanan algoritmalar içinde verimi en yüksek olanlardan biridir. (Chen, & Guestrin, 2016)



Şekil 11. Denetimli Öğrenme ve Denetimsiz Öğrenme farkı (Langs, ve diğerleri, 2018)

Makine Öğrenmede Performans Ölçümü ve Değerlendirme

Makine öğrenmede tahminlerimizin doğruluğunu ölçmek ve farklı algoritmaların performanslarını kıyaslayabilmek için çeşitli araçlar vardır. Bunlardan ilki, istatistiksel sınıflandırma problemlerinde de kullanılan hata matrisi diğer adıyla karışıklık matrisidir. Karışıklık matrisi bir algoritmanın performansının göstermeye yarayan özel bir tablo düzenidir.

		Gerçek	
		Pozitif	Negatif
Tahmin	Pozitif	Gerçek Pozitif (TP)	Yanlış Pozitif (FP)
	Negatif	Yanlış Negatif (FN)	Gerçek Negatif (TN)

Şekil 12. Karışıklık matrisi (Shivaprasa, 2020)

Gerçek Pozitif (TP): Gerçek değerin ve tahmin ettiğimiz değerin 1 olduğu örneklerdir.

Gerçek Negatif (TN): Gerçek değerin ve tahmin ettiğimiz değerin 0 olduğu örneklerdir.

Yanlış Pozitif (FP): Gerçek değer 0 iken tahmin ettiğimiz değerin 1 olduğu örneklerdir.

Yanlış Negatif(FN): Gerçek değer 1 iken tahmin ettiğimiz değerin 0 olduğu örneklerdir.

$$\text{TOPLAM} = \text{TP} + \text{TN} + \text{FP} + \text{FN}$$

$$\text{GERÇEK POZİTİFLER} = \text{TP} + \text{FN}$$

$$\text{GERÇEK NEGATİFLER} = \text{TN} + \text{FP}$$

Bir algoritmayı en iyi duruma getirmede Karışıklık matrisinin amacı, TP ve TN değerlerini maksimize etmek ve FP ve FN değerlerini en aza indirmektir. (Shivaprasa, 2020)

Doğruluk hesaplanmasında sıkça kullanılan yöntemler şunlardır:

Doğruluk Oranı (Accuracy Rate): Algoritmanın doğru tahmin ettiği alanların toplam veri kümesine oranı ile hesaplanmaktadır. Doğru tahmin sıklık ölçüsüdür.

$$(\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN})$$

Yanlış Sınıflandırma Oranı (Misclassification Rate): Algoritmanın ne sıklıkta yanlış tahmin ettiğini gösterir. Hata Oranı (Error Rate) olarak da bilinir.

$$(\text{FP} + \text{FN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN})$$

Gerçek Pozitif Değerlerin Oranı (True Positive Rate): Algoritmanın gerçek pozitif değeri hangi oranda doğru tahmin ettiğini ölçer. Duyarlılık (Sensitivity), İsabet Oranı (Hit Rate) veya Hatırlama (Recall) olarak da adlandırılır.

$$\text{TP} / (\text{TP} + \text{FN})$$

Gerçek Negatif Değerlerin Oranı (True Negative Rate): Algoritmanın gerçek negatif değeri hangi oranda doğru tahmin ettiğini ölçer. Özgüllük (Specificity) veya Seçicilik (Selectivity) olarak da adlandırılır.

$$\text{TN} / (\text{TN} + \text{FP})$$

Yanlış Pozitif Değerlerin Oranı (False Positive Rate): Gerçek değer 0 iken 1 olarak tahmin edilenlerin oranıdır. Yan Ürün (Fall-out) olarak da adlandırılır.

$$FP / (TN + FP)$$

Yanlış Negatif Değerlerin Oranı (False Negative Rate): Gerçek değer 1 iken 0 olarak tahmin edilenlerin oranıdır. Kayıp oranı (Miss Rate) olarak da bilinir.

$$FN / (TP + FN)$$

Kesinlik (Precision): Pozitif tahminlerin, gerçek pozitiflik oranını verir. Hassasiyet olarak da adlandırılır.

$$TP / (TP + FP)$$

Yaygınlık (Prevalence): Tahmin etme sonucunun ne sıklıkta 1 değeri bulduğunun ölçüsüdür.

$$(TP + FN) / (TP + TN + FP + FN)$$

Boş Hata Oranı (Null Error Rate): Çoğunluk sınıfına ait değerin (0 veya 1) sürekli tahmin edilmesi halinde ne oranda yanlış tahmin edildiğini gösterir. Algoritmaları kıyaslamak için kullanılan oranlardandır. Bazen en iyi sınıflandırmayı yapan algoritmanın hata oranı, boş hata oranından daha yüksek olabilir; buna Doğruluk Paradoksu (Accuracy Paradox) denir.

$$(TN + FP) / (TP + TN + FP + FN)$$

Cohen's Kappa: İki sınıflandırıcının birbiri ile uyumunu değerlendiren bir yöntemdir. Doğruluk ve boş hata oranı arasında büyük bir fark varsa, algoritmanın Kappa puanı yüksek olacaktır. 0,60'ın üzerindeki katsayı puanları iyi uyuma olduğunu göstermektedir.

F Puanı (F Score): Gerçek pozitif değerlerin oranının (recall) ve kesinliğin (precision) harmonik ortalamasıdır. Algoritmanın ne kadar iyi performans gösterdiğinin bir ölçüsüdür ve algoritma kıyaslamasında sıklıkla kullanılır.

$$\frac{2 * Kesinlik * Gerçek Pozitif Değerlerin Oranı}{Kesinlik + Gerçek Pozitif Değerlerin Oranı}$$

ROC Eğrisi (ROC Curve): Algoritmanın tüm olasılıklar üzerindeki performansının özeti olan grafikdir. ROC eğrisinin altında kalan alan (Area Under Curve (AUC)) olarak adlandırılır. AUC bir parametrenin iki sınıf arasında ayırt edilebilirliğini gösteren ölçüdür. AUC 0-1 aralığında değer alır ve değer 1'e ne kadar yakın ise algoritmanın ayırt etme performansı o kadar iyidir.

Hata Metrikleri

Hata metrikleri performans ölçümü için gerçek değerler ile tahmin edilen değerler arasında farkı gösteren göstergelerdir. (VeriBilimcisi, 2017) Makine Öğrenme uygulamalarında performans ölçümü için en sık kullanılan hata metrik yöntemleri şunlardır:

MAE - Ortalama Mutlak Hata(Mean Absolute Error): Ortalama mutlak hata tüm gerçek değerler ile veriye en uygun çizgi arasındaki ortalama yatay ve dikey mesafedir. MAE, iki sürekli değişken arasındaki farkın ölçüsüdür. (Sel, 2020) MAE regresyon ve zaman serisi problemlerinde sıkça kullanılmaktadır. Bunun sebebi değerlerinin kolayca yorumlanabilmesidir. MAE değeri 0'dan ∞ 'a kadar değişebilir. MAE değerinin düşük olduğu tahminler daha iyi performans anlamına gelmektedir.

$$MAE = \frac{1}{n} \sum_{t=1}^n |A_t - F_t|$$

A_t = t zamandaki Gerçek Değer
 F_t = t zamandaki Tahmin Edilen Değer
 n = toplam zaman periyodu

MAPE-Ortalama Mutlak Yüzde Hatası(Mean Absolute Percentage Error): Zaman serisi ve Regresyon modellerinin tahminlerinde doğruluğu ölçmek için sıkça kullanılır. (VeriBilimcisi, 2017) Tahmin değerleri çok düşük olduğunda yüzde hatası %100'ü geçemez, ancak tersi için 0 ile ∞ arası değer alır. Gerçek değerler arasında sıfır içerenler varsa, sıfır ile bölünme hatası oluşacağı için MAPE hesaplanamaz.

$$MAPE = \frac{100}{n} \sum_{t=1}^n \frac{|A_t - F_t|}{|A_t|}$$

A_t = t zamandaki Gerçek Değer
 F_t = t zamandaki Tahmin Edilen Değer
 n = toplam zaman periyodu

MSE - Ortalama Kare Hatası (Mean square error): Bir regresyon eğrisinin bir dizi noktaya ne kadar yakın olduğunu gösteren değerdir. MSE, algoritmanın performansını ölçmede kullanılır ve her zaman pozitif değerlidir. MSE değeri sıfıra yakın olan algoritmaların daha iyi bir performans gösterdiği söylenebilir.

$$MSE = \frac{1}{n} \sum_{t=1}^n (A_t - F_t)^2$$

A_t = t zamandaki Gerçek Değer

F_t = t zamandaki Tahmin Edilen Değer

n = toplam zaman periyodu

RMSE - Kök Ortalama Kare Hatası (Root Mean Square Error): Kısaca ortalama kare hatasının kareköküdür. Tahmin edilen değerler ile gerçek değerler arasında ne kadar fark olduğunu ölçmek için kullanılır. (Sel, 2020) Verilerin regresyon modelinden sapma miktarını göstermek için kullanılır. Verilere en uygun çizginin etrafında oluşan veri yoğunluğunu verir. RMSE değeri 0'dan ∞ 'a kadar değer alabilir. RMSE değeri ne kadar düşükse, algoritma o kadar iyi performans vermiştir. Değerin sıfır olması algoritmanın hiç hata yapmadığı anlamına gelir. RMSE, birçok matematiksel hesaplamada mutlak değer kullanılmasının önüne geçer.

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (A_t - F_t)^2}{n}} \quad RMSE = \sqrt{MSE}$$

RAE- Bağıl Mutlak Hata (Relative Absolute Error): Algoritmanın tahmin performansını ölçmek için kullanılır. Tahmin edilen değerlerin gerçek değerlerden sapma toplamının gerçek değerlerin ortalamadan sapmalarına bölümü ile bulunur.

$$RAE = \frac{\sum_{t=1}^n |F_t - A_t|}{\sum_{t=1}^n |A_t - \bar{A}|}$$

Literatür Çalışması

Bitcoin'in fiyat akışının belirlenebilmesi özellikle yatırımcılar açısından büyük önem arz ettiğinden dolayı, geçtiğimiz yıllarda bu alanda birçok çalışma yapılmıştır. Bu çalışmaların bazılarında makine öğrenme yöntemleri kullanılırken, bazılarında derin öğrenme yöntemleri kullanılmıştır. Bu çalışmalardan makine öğrenme üzerine olanları incelenecektir.

Bitcoin fiyat tahmini üzerine ilk araştırmalardan biri Shah ve Zhang (2014) tarafından Bayes Lineer Regresyonu modeli ile Bitcoin fiyatlarının tahmini üzerine yapılmıştır. 2014 yılının Şubat-Temmuz arası Bitcoin fiyat hareketleri zaman serisi olarak alınmış ve 200 milyondan fazla veri üzerinde k-ortalama (k-mean) kümeleme makine öğrenme yöntemi uygulanmıştır. Bu işlem için güçlü bilgisayarlar kullanılmış, sonraları normalleştirilmiş verilerle Python programlama dili kullanılarak ile aynı tip bir bilgisayarda 1 saniyede çok daha verimli sonuçlar alınabildiği anlatılmıştır. İşlemin çıktısı olarak eğer makine öğrenme sonuçlarına göre yatırım yapılırsa, yatırımı ikiye katlayacağı yönünde olmuştur.

Madan, Saluja ve Zhao (2015) çalışmalarında iki farklı veri seti üzerinden tahmin üzerine makine öğrenme algoritmalarını çalışmışlardır. İlki 5 yıllık günlük kapanış değerleri ile Bitcoin piyasasını etkileyen 26 bağımsız değişken içinden seçtikleri 16 değişkenden oluşan veri seti, ikincisi ise 10 saniye ve 10 dakikalık Bitcoin fiyat hareketlerinden oluşan veri seti olarak belirlenmiştir. Makine öğrenme algoritmalarından Genelleştirilmiş Doğrusal Model (GLM), Destek Vektör Makinesi (SVM) ve Rasgele Orman (RF) yöntemleri ile tahminler gerçekleştirmişlerdir. Sonuçlara Hassaslık, Özgüllük, Kesinlik ve Doğruluk üzerinden değerlendirilmiştir. Çalışmanın genel sonucunda SVM ile kötü tahmin sonuçları alınmışken, GLM ve RF teknikleri ile daha iyi sonuçlar elde etmişlerdir.

McNally (2016) yüksek lisans tez çalışmasında Bitcoin fiyatlandırması ile Borsa fiyatları benzerliğinden faydalananarak Borsa konusundaki makine öğrenme ile tahmin yöntemlerinin Bitcoin'e uyarlamıştır. Veri olarak 2013-2016 yılları arası günlük Bitcoin Açılış, Yüksek, Düşük, Kapanış (OHLC) ve Hash oranı verileri alınırken, teknik indikatör olarak Basit Hareketli Ortalamalar (SMA) ve Arınmış Kapanış Fiyatı (De-noised Closing Price) alınmıştır. Veriler 0-1 aralığına normalize edilmişlerdir. Makine öğrenme yöntemi olarak Yinelenen Sinir Ağları (RNN) ve Uzun Kısa Süreli Bellek (LSTM) ve ARIMA modelleri ile test edilmiştir. Sonuçlar Hassaslık, Kesinlik, Özgüllük, Kesinlik, Doğruluk ve RMSE üzerinden değerlendirilmiştir. Hata oranı olarak ARIMA hariç iki modelde de yüzde 10'un altı, doğruluk olarak ise yüzde 50'nin biraz üzerinde sonuçlar alınmıştır. ARIMA modeli yüzde 50'nin üzerinde hata oranı vermiştir. RNN ve LSTM modellerinin eğitim sürelerinin kıyaslaması da yapılmıştır. Parametrelerin doğru ayarlanmaması durumunda aşırı uyum (overfitting) sorunuyla karşılaşılacağı da belirtilmiştir.

Karasu, Aytac, Sarac ve Hacıoglu (2018) tarafından makine öğrenmesi yöntemlerinden Doğrusal Regresyon (LR) ve Destek Vektör Makinesi (SVM) ile 2012-2018 yılları arasındaki günlük Bitcoin kapanış fiyatlarından oluşan zaman serisi kullanılarak Bitcoin (BTC/USD) kapanış değerleri tahmini yapılmıştır. Tahmin modelinin eğitimi 10-kat çapraz doğrulama yöntemiyle yapılmış ve girişine Hareketli Ortalama (MA) ve Ağırlıklı Hareketli Ortalama (WMA) filtre uzunluğu 2 ile 6 arasında değiştirilerek uygulanarak en düşük hatayı veren model elde edilmesini amaçlamışlardır. Ayrıca, A/D (Accumulation / Distribution) indikatörü tahmin modeline ikinci giriş olarak verilmektedir. Tahmin modelinin çıkışı bir sonraki günün tahmini Bitcoin kapanış fiyatı olarak belirlenmektedir. Elde edilen modelin başarımı, istatistiksel göstergelerden Ortalama Mutlak Hata (MAE), Ortalama Kare Hatası (MSE), Ortalama Hata Kareleri Karekökü (RMSE) ve Pearson korelasyon yardımıyla ölçülmüştür. Elde edilen modellere ait tahmin değerleri birbirleri ile kıyaslanmış ve en iyi sonuca gecikme sayısı 2 olan WMA filtreli P-SVM ile ulaşılmıştır. Sonuç olarak kullanılan yöntemin yüksek başarımlı gösterdiği ve ileride yapılacak çalışmalara yol gösterebileceği belirtilmiştir.

J. Huang, W. Huang ve Ni (2018) yaptıkları çalışmada, 2012-2018 yılları arasındaki 2168 günlük Bitcoin (BTC/USD) açılış, en yüksek, en düşük ve kapanış değerlerini alarak 3 parçaya ayırmışlardır. Bu verilerden teknik analiz araçlarını kullanarak 124 farklı teknik indikatör üretmişlerdir. Birinci parçada ilk 120 günlük veriyi öngörü (predictors) için sisteme tanımlanmış, ikinci kısım olarak sonraki 1539 günlük veriyi eğitim seti (training set) belirlenmiş, üçüncü parça olarak ise geriye kalan 509 günlük veri test seti olarak sisteme tanımlanmıştır. Makine öğrenme yöntemi olarak Sınıflandırma ve Regresyon Ağaçları (CART) kullanılmıştır. Bu çalışmada farklı olarak kapanış değerinin yüzdesel artışı ve getirisi üzerine değerlendirme yapılmış ve yatırımcılar açısından daha verimli bir model çıkarılmaya çalışılmıştır.

Mallqui ve Fernandes (2018) çalışmalarında veri seti olarak Bitcoin açılış, en yüksek, en düşük ve kapanış (OHLC) değerlerini, işlem hacmini (Volume), işlem ücretleri (transaction fees), işlem sayısı, işlem başına ücret ve ortalama Hash Oranı kullanmışlardır. Veri seti olarak 2013-2016 yılları arası değerleri (McNally, 2016)'nin yüksek lisans tezinde kullandığı verilerle kıyaslama adına birinci set belirlemiş, ikinci veri setini ise 2013-2017 yılları arasındaki verileri alarak oluşturmuşlardır. İndikatör olarak hem temel hem de teknik indikatörler kullanılmıştır. Temel İndikatör olarak vadeli ham petrol fiyatları, vadeli altın fiyatları, vadeli S&P500, vadeli NASDAQ ve DAX endeksi kullanılmıştır. Kapanış verileri bir önceki günün kapanış değerlerine göre düşükse 0 yüksekse 1 olarak ayarlanarak hareketler tanımlanmıştır. Tüm verilerin 30 günlük Ağırlıklı Hareket Ortalaması (WMA) hesaplanmıştır. 1. Veri seti %80 train,

%20 test; 2. Veri seti %75 train, %25 test olarak ayrılmıştır. Makine öğrenme yöntemi olarak Yapay Sinir Ağları (ANN) ve Destek Vektör Makinaları (SVM) kullanılmıştır. Çalışmada en iyi performans için Korelasyon Analizi (Corr), Relief Tekniği (ReliefF), Bilgi Kazancı Karar Ağacı (InfoGain), Temel Bileşenler Analizi (PCA) ve Korelasyona Dayalı Özellik Kümesi Seçimi (CFS) yöntemiyle en iyi indikatörler belirlenerek, en iyi sonuç verenler kıyaslanmıştır. Veriler eğitilirken (train) ANN ve SVM ile birlikte Yinelene Sinir Ağları (RNN)'da kullanılarak 3 farklı şekilde veriler işlenmiş ve bunların sonucuna göre Bitcoin hareket yönü belirlenmiştir. Sonuçların karşılaştırılması için MAE, MAPE ve RMSE hata metrikleri kullanılmıştır. Sonuç olarak ANN ve SVM yöntemiyle diğer yöntemlerin birleştirilmesinde sonuçlar çok fazla olumlu yönde etkilemediği görülmüş ve Larry Williams(%R), Göreceli Güç Endeksi (RSI), Stokastik Osilatörler(%K and%D), Hareketli Ortalama Yakınsama/İraksama (MACD) gibi klasik teknik indikatörlerle test etmenin sonucu olumlu yönde değiştirebileceğinden bahsedilmiştir.

Mangle, A. Bhat, Avabrahtha ve N. Bhat (2019) çalışmalarında Bitcoin hareketlerinin tahmini için sınıflandırma modeli olarak Lojistik Regresyon (LR), SVM kullanılmıştır. Bunların dışında Otoregresif Entegre Hareketli Ortalama (ARIMA) ve Tekrarlayan Sinir Ağları (RNN) uygulanıp test edilmiştir. Veri seti olarak 2015-2019 arası Bitcoin kapanış fiyatları kullanılmıştır. Çalışmada Makine Öğrenme Algoritmalarının performansları test edilmiş ve performansını artırmaya yönelik denemeler yapılmıştır. Çalışma sonucunda ARIMA'nın kısa vadeli hesaplamalarda iyi sonuç verdiği ancak, 5-7 günlük tahminlerde düşük kaldığı; RNN'in ise 6 güne kadar iyi sonuç verdiği sonucuna ulaşılmış.

Z. Chen, Li ve Sun (2019) çalışmalarında Bitcoin fiyat yönünü belirleme üzerine çok boyutlu modelleme yapmayı hedeflemişlerdir. Birinci veri seti olarak Bitcoine ait 2017-2019 yılları arası günlük kapanış değeri, İkinci veri seti olarak 5 dakika periyodlu Bitcoin fiyat hareketleri alınmıştır. Veriler %75 öğrenme, %25 test için ayrılmıştır. İndikatör olarak 15 farklı veriden faydalanılmıştır. (Blok büyüklüğü, Hash oranı, Kazılma Oranı, tranfer sayısı, Google Trend aramalar, işlem hacmi, altın fiyatları vd.) Veri setlerine Lojistik Regresyon (LR) ve Doğrusal Diskriminant Analiz (LDA) Rastgele Orman (RF), XGBoost (XGB), İkinci Dereceden Diskriminant Analiz (QDA), Destek Karar Makinesi (SVM) ve Uzun Kısa-Süreli Bellek(LSTM) makine öğrenme algoritmaları uygulanmıştır. Performans analizi için iki veri seti için tüm algoritmaların Doğruluk (Accuracy), Kesinlik (Precision), Hatırlama(Recall) ve F1-Skor oranları kıyaslanmıştır. İlk veri setinde LR %66 Doğruluk ile en iyi sonucu verirken, XGB %48,3'te kalmıştır. 2. Veri seti için LSTM 67.2% ile en iyi sonucu verirken, LR 54.5% ve LDA %51.5 Doğruluk sonucu vermiştir.

Lahmire ve Bekiros (2020) çalışmalarında yedi farklı makine öğrenme algoritmasını Bitcoin verileri üzerinde uygulayarak performanslarını test etmişlerdir. Algoritma olarak Destek Vektör Regresyonu (SVR) ve Poisson Gauss Regresyonu (GRP) modelleriyle birlikte Regresyon ağaçları (RT) ve K-En Yakın Komşu (KNN) kullanılmıştır. Son olarak yapay zeka yöntemlerinden İleri Beslemeli Sinir Ağları (FFNN), Bayes Düzenleme Ağları (BRNN), ve Radyal Tabanlı Fonksiyon Ağları (RBFNN) kullanılmıştır. Veri seti olarak 2016 yılı başından 2018 Mart ayına kadar 5 dakika aralıklı Bitcoin fiyat hareketleri alınmıştır. Verilerin %80'i öğrenme için, %20 test için ayrılmıştır. Ortalama Hata Kareleri Karekökü (RMSE) oranı 1.1139, SVR 0.8674, GRP 1.2231, RT 2.1111, KNN 0.8475, BPNN 0.8443 ve BRNN 1.3165 olarak bulunmuştur.

W. Chen, Xu, Jia ve Gao (2020) çalışmalarında Bitcoin verilerini işlemeden önce Bitcoin'i etkileyen ekonomik (altın, petrol, S&P500, döviz kurları vb) ve teknolojik (Hasrate, ortalama işlem ücretleri, Tweet, Google trendleri vb.) Bitcoin verileri faktörleri sıraladıktan sonra, bu verileri ANN ve RF makine öğrenme algoritmaları ile Bitcoin fiyat hareketlerine göre önemlerini ayarlanmışlardır. Veriler 2011 yılından başlayarak Bitcoin'in önemli ölçüde yükseldiği veya düştüğü periyodlar belirlenerek 4 parça halinde işlenmiştir. Elde edilen verilerden boş veya eksik olanlar temizlenmiştir. Uzun kısa-süreli belleğe (LSTM) önceki Bitcoin kuruna bakmadan sadece bu girdilerle yaptığı tahminleri; önceki Bitcoin kurları ile eğitilen ANFIS, SVR, ARIMA ve LSTM makine öğrenme algoritmaları kıyaslayarak performans ölçümü yapılmıştır. Farklı parametreler kullanılarak algoritmalar en iyi düzeye getirilmeye çalışılmıştır. Son 2 aylık Bitcoin verisi ayrılarak modelin testi için tekrar kullanılmıştır. Çalışmanın sonucuna olarak RMSE, MAPE, ve MAE testlerine göre Teknik ve Ekonomik göstergeler verilerek önceki Bitcoin kurunun tahmin dahil edilmediği LSTM modeli, önceki döviz kuruna göre tahmin tüm diğer yöntemlere göre daha iyi sonuç vermiştir.

Yavuz, Özen, Taş ve Çağlar (2020) çalışmalarında Bitcoin fiyat tahmininin için 3491 günlük kapanış fiyatları yanında İşlem Başına Bitcoin Maliyeti, Bitcoin Tahmini İşlem Hacmi (BTC), Bitcoin Tahmini İşlem Hacmi (USD), Bitcoin Toplam Çıkış Hacmi, Bitcoin Toplam İşlem Sayısı, Dolaşımdaki Toplam Bitcoin Miktarı, madencilik getirisi, Hash oranı vb. toplam 17 değeri alarak veri setini oluşturmuşlardır. Makine öğrenme yöntemi olarak Yapay Sinir Ağları (ANN) kullanmışlardır. 20 gizli katman kullandıkları çalışmada veri setinin %60'ı öğrenme, geri kalanı test ve doğrulama için kullanılmıştır. Çalışma sonucunda Ortalama Kare Hatası (MSE) 0,99977 oranında R değeri vermiş, basit olasılıklı örnekleme modeliyle ağ dışından seçilen 20 veriyi doğru tahmin ederek ağın başarılı olduğu belirlenmiştir.

Mudassir, Bennbaia, Unal ve Hammoudeh (2020) çalışmalarında 2013-2019 yılları arası Bitcoin fiyat verisi ile Yapay Sinir Ağları (ANN), Yığılmış Sinir Ağları (SANN), Destek Vektör Makinası (SVM) ve Uzun Kısa-Süreli Hafıza (LSTM) makine öğrenme yöntemleriyle fiyat hareket yönü tahmini üzerine çalışmışlardır. Veriler önceki günün kapanış değerinin farkına göre 0 ve 1 değerlerine dönüştürülmüştür. İndikatör seçiminde Bitcoin fiyatını etkileyen 700'den fazla gösterge 1, 7, 30 ve 90 günlük tahminler için Rastgele Orman yöntemiyle önem derecesi belirlenerek önemli çıkan göstergeler makine öğrenmesine dahil edilmiştir. Çalışmanın sonucunda kullanılan makine öğrenmeleri yöntemleri Ortalama Mutlak Hata (MAE), Kök Ortalama Kare Hata (RMSE) ve Ortalama Mutlak Oran Hata (MAPE) üzerinden tüm zaman setleri için ayrı ayrı kıyaslanmış; Doğruluk (Accuracy), Eğri Altındaki Alan (AUC) ve F1 Skor testleri yapılarak performansları ölçülmüştür. Günlük tahminlerde SANN iyi sonuç verirken, kısa vadeli tahminlerde SVM, uzun vadeli tahminlerde ise ANN ve LSTM'nin daha iyi doğruluk sonucuna ulaştığı görülmektedir. Toplamda en yüksek yakalanabilen doğruluk oran %64 olmuştur.

Pabuçcu, S. Ogan ve A. Ogan (2020) çalışmalarında veri seti olarak 2008-2019 yılları arası 1935 adet Açılış, Yüksek, Düşük, Kapanış (OHLC) Bitcoin verisi alınmış, veri %75 eğitim, %25 test için kullanılmıştır. Doğrulama verisi olarak 2020 Haziran sonrası 100 günlük veri kullanılmıştır. Teknik indikatör olarak Hareketli Ortalamalar (MA), Ağırlıklandırılmış Hareketli Ortalamalar (WMA), Momentum (Mom), Stokastik (%K) ve (%D), Relatif Güç Endeksi (RSI), Hareketli Ortalama Yakınsaması/İraksaması (MACD), Larry Williams (LW %R) ve Birikim/Dağıtım Endeksi(A/D) olarak alınmış ve ham OHLC verisinden formüller üzerinden hesaplanarak veri seti oluşturulmuştur. Normalize edilen veri seti "Sürekli veri seti" olarak kullanılmış, her bir indikatör için hesaplama yöntemiyle kesikli(binary) yapıya dönüştürülen veri seti "Kesikli veri seti" olarak 2. veri olarak kullanılmıştır. Kesikli veri setinde düşüşler -1 olarak temsil edilmişken, yükselişler +1 olarak alınmıştır. Makine öğrenme yöntemi olarak Destek Vektör Makinası (SVM), Yapay Sinir Ağları (ANN), Naïve Bayes (NB) ve Rastgele Ortman (RF) yöntemleri kullanılarak, kıyaslama için Lojistik Regresyon (LR) kullanılmıştır. İki veri seti için ayrı ayrı Doğruluk (Accuracy), Ortalama Mutlak Hata (MAE), Kök Ortalama Kare Hatası (RMSE) ve Bağlı Mutlak Hata(RAE) performans ölçümleri yapılarak en iyi sonuç veren parametreler bulunmuştur. F-testi ile karışık matrisi sonuçlarına göre modeller kıyaslanmıştır. Bu sonuçlara göre ANN 0.941 ile en iyi başarıyı göstermişken, RF 0.939, SVM 0.938 ve NB 0.902 sonuç vermiştir. Son olarak t-testi ile iki veri setine uygulanan modellerin genel kıyaslama yapıldığı çalışmada, sürekli veri seti açısından RF iyi sonuç verirken, kesikli veri seti için ANN daha iyi sonuç verdiği görülmüştür. RF kullanımının

kolaylığı açısından daha pratik olsa da yeni bir veri setinde farklı sonuçlar elde edebileceği belirtilmiştir.

Ibrahim, Kashef ve Corrigan (2020) çalışmalarında veri seti olarak günümüzden geçmişe iki ayrı Bitcoin veri setinin Açılış, Yüksek, Düşük, Kapanış (OHLC) değerleri kullanmıştır. Veri setlerinden birisi günlük değerlerken, diğeri 5'er dakikalık Bitcoin verisi şeklindedir. Analizde kullanılmak üzere ayrıca Apple, Facebook, Google ve Microsoft hisse değerlerinin 5'er dakikalık değişimleri alınmıştır. 5'er dakikalık değişim verisi olmayan tarihler için program yardımıyla veri oluşturulmuştur. Teknik indikatör olarak Volatilité, MA, Parabolik SAR, MACD, RSI ve Denge İşlem Hacmi (OBV) değerleri hesaplanarak kullanıldığı çalışmada, makine öğrenme algoritması olarak ARIMA, Prophet, Rastgele Orman (RF) ve Çok Katmanlı Algılayıcı (MLP) kullanılarak sonuçlar kıyaslanmıştır. İndikatörler ve algoritma parametreleri en iyi doğruluk sonucuna göre seçilmiştir. Veriler artış-azalışa göre 1 ve 0 olarak tanımlandığı sete ek olarak üst üste kaç kez artış-azalış trendinde olduğu ve artış-azalış yönünde yüzdesel belli bir artıştan bu yana gerçekleşen artış-azalış trendinin hesaplandığı veri seti oluşturulmuştur. Geçmiş 10 ve 20 periyotta artış-azalış fazlalığına göre kısa vadeli değişkenlerde kullanılmıştır. Makine öğrenme yöntemlerinden Doğruluk testi sonuçlarına göre yüzde %54,09 ile MLP en iyi sonucu verirken, ardından Prophet'in en iyi sonucu verdiği görülmüştür. Tüm testler %50 ve üzeri performans vermiştir.

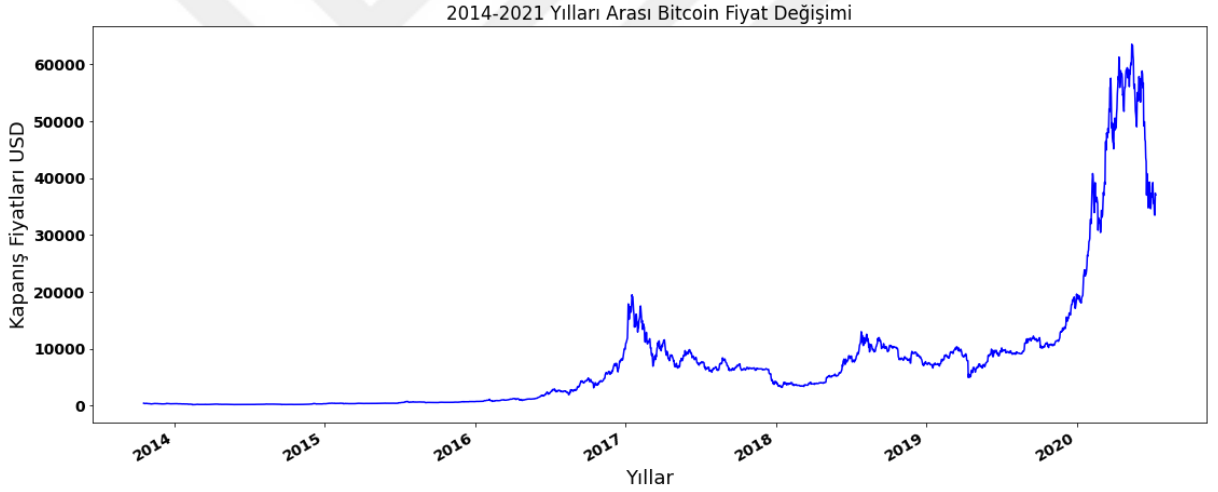
Tekinay (2021) yüksek lisans tez çalışmasında veri setini 2017-2020 yılları arası Bitcoin fiyatları yanında, 8 farklı döviz, dünyaca ünlü 7 borsa, vadeli altın fiyatları ve 10 çeşit altcoin verisinden oluşturmuştur. Tüm veriler üzerinde Engle – Granger Eşbütünleşme Testi ve Granger Nedensellik Testi yapılmıştır. Elde edilen verilere göre 6 çeşit altcoin ile Bitcoin arasında çift yönlü nedensellik bulunmuştur. Her ilişki için toplam 6 adet model oluşturulmuştur. 1075 adet verinin %70'i eğitim seti ve %30'u test seti olarak kullanılmıştır. Yapay sinir ağları modeliyle 10 kez eğitilen modellerden 31, 45 ve 60 günlük tahmin setleri oluşturulmuştur. Tahmin sonuçları MAPE ve RMSE istatistikleri ile kıyaslanmıştır. Çalışmanın sonucuna göre 31 günlükte Decred, 45 günlükte Ethereum ve 60 günlükte BitcoinCash ile en iyi sonuç elde edilmiştir.

ÜÇÜNCÜ BÖLÜM

Bitcoin Fiyat Hareketlerinin Makine Öğrenme Algoritmaları ile Tahmin Edilmesi

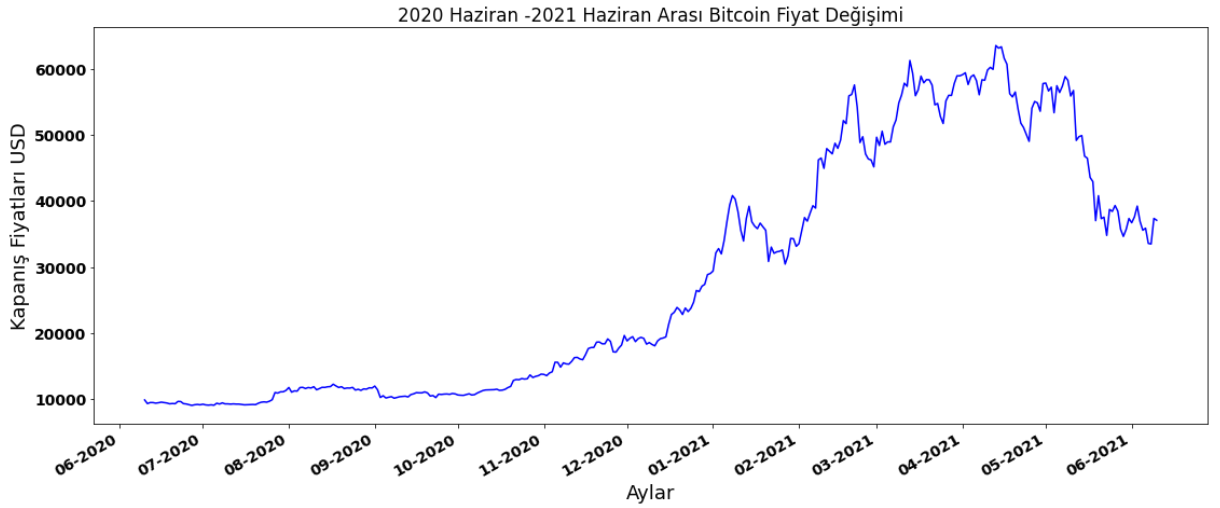
Veriler

Tez çalışmasının kullanılan Bitcoin (BTC) kripto para biriminin USD bazında günlük Açılış (Open), Kapanış (Close), en yüksek değer (High), en düşük değer (Low) ve İşlem Hacmi (Volume) verileri Python programlama dilinin Yahoo Finance kütüphanesi ile 17/09/2014-10/06/2021 tarihleri arası 1. Veri seti, 10/06/2020-10/06/2021 tarihleri arası 2. Veri seti olarak aktararak veri setlerine dönüştürülmüştür. 1. Veri setinin tarih aralığı seçiminde Bitcoin'in yaygınlaşarak Yahoo Finance veritabanına girdiği 2014 yılı ve sonrası tarih seçilmiştir.



Şekil 13. 2014-2021 yılları arası BTC fiyat değişimi

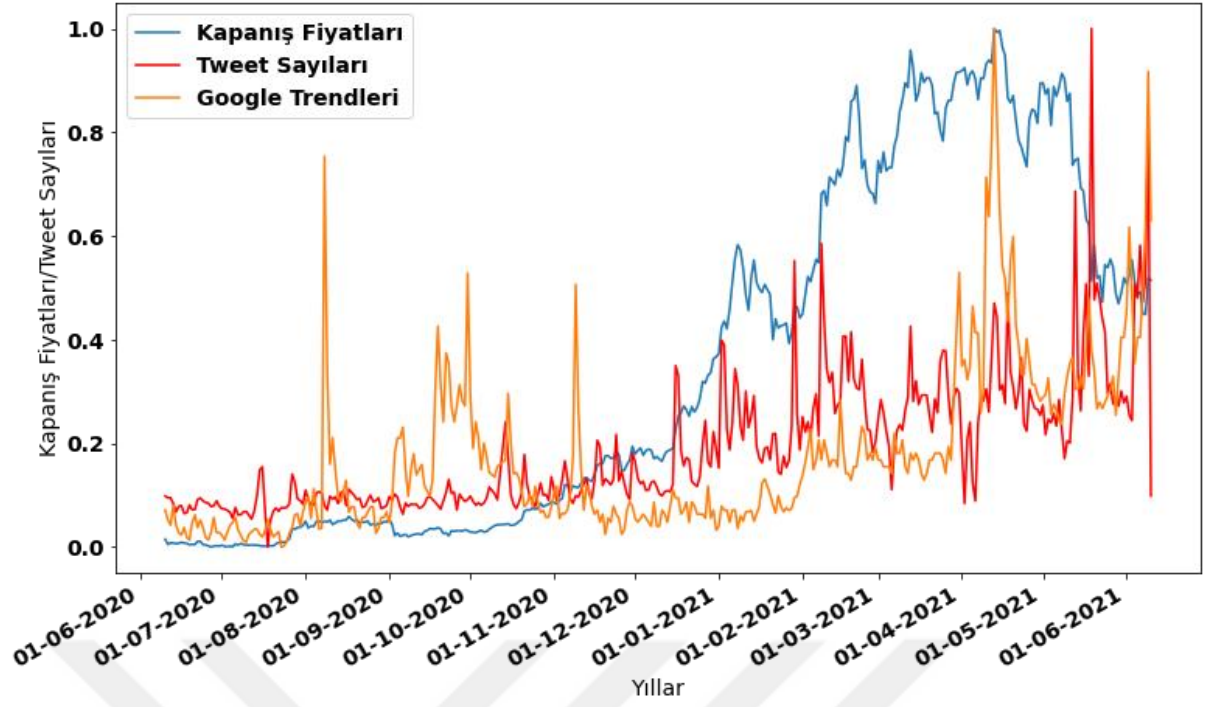
2. Veri seti olarak ise fiyatının 7 aylık süreçte 6 kat kadar arttığı ve son aylarda tekrar fiyatının düştüğü geçtiğimiz 1 yıl aralığı seçilmiştir. Bu seçim ile hızlı yükselme ve düşüş içeren küçük veri seti ile öğrenmenin, büyük veri seti ile öğrenme arasındaki farkların ve hangi yöntemin daha verimli olduğunun gözlemlenmesi hedeflenmiştir. Böylece bu yöntemler sonraki çalışmalarda henüz büyük verileri oluşmamış Altcoinlerin fiyat yönü tahmini için referans olabilecektir.



Şekil 14. Haziran 2020-Haziran 2021 arası BTC fiyat değişimi

Belirlenen tarih aralığındaki veriler Pandas kütüphanesi ile tablolaştırıldıktan sonra, teknik analizlerde sıkça kullanılan Hareketli Ortalamalar (SMA), Ağırlıklı Hareketli Ortalama (WMA), Üstel Hareketli Ortalama (EMA), Ortalama Gerçek Aralık (ATR), Bollinger Bantları (Boll), Momentum (Mom), Stokastik indikatörler (K ve D), Relatif Güç Endeksi (RSI), Hareketli Ortalamanın Yakınlaşması ve Uzaklaşması (MACD), Larry Williams (LW), Birikim Dağılım Göstergesi (A/D) indikatörler, 2. bölümde açıklanmış formüller Technical Analysis Library (TA-Lib) kütüphanesinden de faydalanılarak hesaplamaları yapılmış ve Pandas Kütüphanesi ile veri tablosuna eklenmiştir. Bitcoine ilişkin atılan günlük Tweet sayıları ve Google Arama Trend verileri <https://bitinfocharts.com/> adresinden Python 3.6'da yazılan kodlarla BeautifulSoup kütüphanesi ile veri kazıma yöntemi ile elde edilerek veri setine dönüştürülmüştür. Daha sonra bu veri seti fiyat ve indikatör bilgilerinin yer aldığı veri setine katılarak yeni bir veri setine dönüştürülmüştür.

Elde edilen verilerden 2014-2021 yılları arası Tweet sayılarının önceki güne göre yükseldiği günlerde kapanış fiyatının da yükseldiği günlerin oranı %52,7; Google Trend aramaların ise %50,3 olduğu hesaplanmıştır. 2020 yılı 6. Ayından 2021 yılı 6. ayna kadar Tweet sayılarının önceki güne göre yükseldiği günlerde kapanış fiyatının da yükseldiği günlerin Pearson korelasyon katsayısı 0,689; Google Trend aramaların ise 0,533 olarak hesaplanmıştır. Bu oranın etkisi çalışma sırasında gözlemlenecektir.



Şekil 15. Normalize edilmiş Kapanış, Tweet ve Google Trend verileri (Son 1 Yıllık)

Teknik indikatörler ve kapanış fiyatından oluşan 17 sütun ve 2455 satırlık ilk veri setimiz 0-1 aralığına normalize edilerek sürekli veri seti oluşturulmuştur. Kesikli veri setini ise sürekli veri setini normalize etmeden önce aşağıda açıklanan kurallar doğrultusunda artış trendi için 1, azalış trendi için 0 olarak kodlanan verilerden oluşturulmuştur. Kesikli ve sürekli veri setlerinin son 365 satırları son 1 yıllık 3. ve 4. veri setlerimizi oluşturmaktadır.

Kesikli veri seti oluşturulurken Kapanış(Close) fiyatı bir önceki günden yüksek ise 1, düşük ise 0 değeri verilmiştir. Kısa vadeli tahmin için kullanılan Basit Hareketli Ortalama(SMA), Ağırlıklı Hareketli Ortalama(WMA) ve bunların birleşimiyle oluşmuş Üstel Hareketli Ortalama(EMA) 14 günlük olarak alınmış ve dönüşümlerinde Kapanış değeri bu değerlerden büyükse trendin yukarı yönlü olduğu anlamına geldiği için 1 değeri, küçük ise 0 değeri verilmiştir. Fiyat hareketlerinin ölçümünü veren Ortalama Gerçek Aralık(ATR) indikatörünün dönüştürülmesinde eğer bir önceki gün olan değerinden yüksek ise trendin yönü yukarı olacağından 1, küçükse 0 değeri verilmiştir. Diğer çalışmalarda rastlanmayan indikatörlerden Bollinger Bantları bu çalışmada deneysel olarak dahil edilmiştir. Diğer çalışmalarda yer almadığından dolayı kesikli yapıya dönüştürürken finansal yorumlama yapılarak kapanış değeri alt banda yaklaştıysa yukarı yönlü trend beklentisi olduğundan 1 değeri, yukarı banda yaklaşırsa düşüş trendi olarak 0 değeri atanmıştır. Bu şekilde hesaplanan

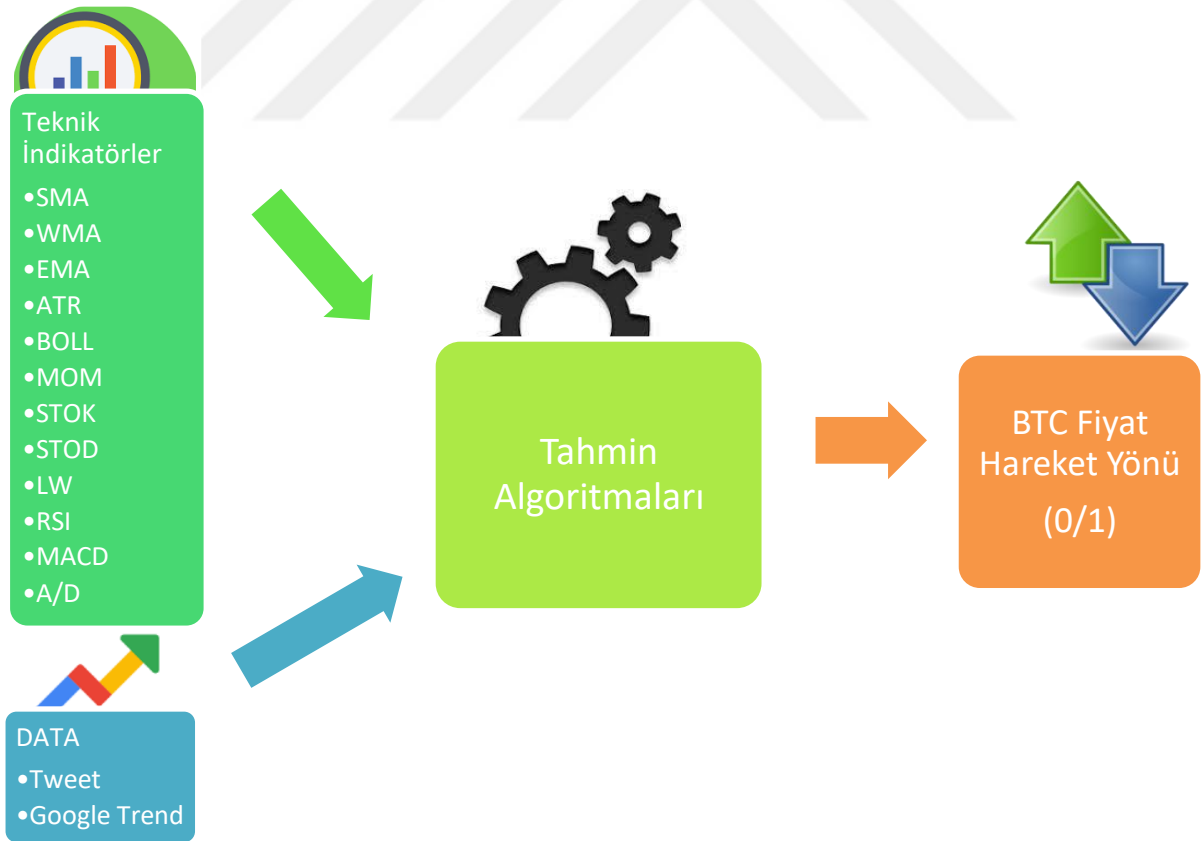
Bollinger Bandlarının Kapanış Değerin hareketine benzerlik oranının %71,3 olduğu hesaplanmıştır. Momentum değeri trendin sürdürülebilirliğini bildirdiğinden dolayı pozitifse 1, negatifse 0 değeri atanmıştır. Stokastik indikatörler olan %K, %D ve Larry Williams(LW) fiyat değişim hızını ve aşırı alım satımını gösteren göstergeler olduğundan dolayı önceki gün olan değerinden fazla ise 1 değeri, düşük ise trend aşağı yönlü olacağı yorumlandığından sıfır değeri atanmıştır. RSI değeri yorumlanırken eğer değer 30'dan küçükse trend yukarı yöne gideceği yorumlandığından 1, 30-70 aralığında bir değerse bir önceki günden büyükse 1, küçükse 0 değeri atanır. Eğer RSI 70'den büyükse trend aşağı yönlü olacağı düşünüldüğünden 0 değerini alır. MACD'nin yorumlanmasında eğer MACD artış azalışının fiyatında artışı azalışı anlamına geldiğinden dolayı, MACD değeri bir önceki günden yüksekse 1 değeri, düşük ise 0 değeri atanmıştır. A/D Osilatörü içinde aynı şekilde atama yapılarak Osilatör değeri önceki günün değerinden yüksek ise 1, düşük ise 0 değeri atanmıştır. Günlük atılan Tweet sayısı ve Google Arama Trendi içinde aynı şekilde bir önceki gün ile kıyaslanarak yüksek ise 1, düşük ise 0 sıfır değeri atanmıştır.

Tablo 1. Kesikli Veri Setlerinin Oluşturma Yöntemleri

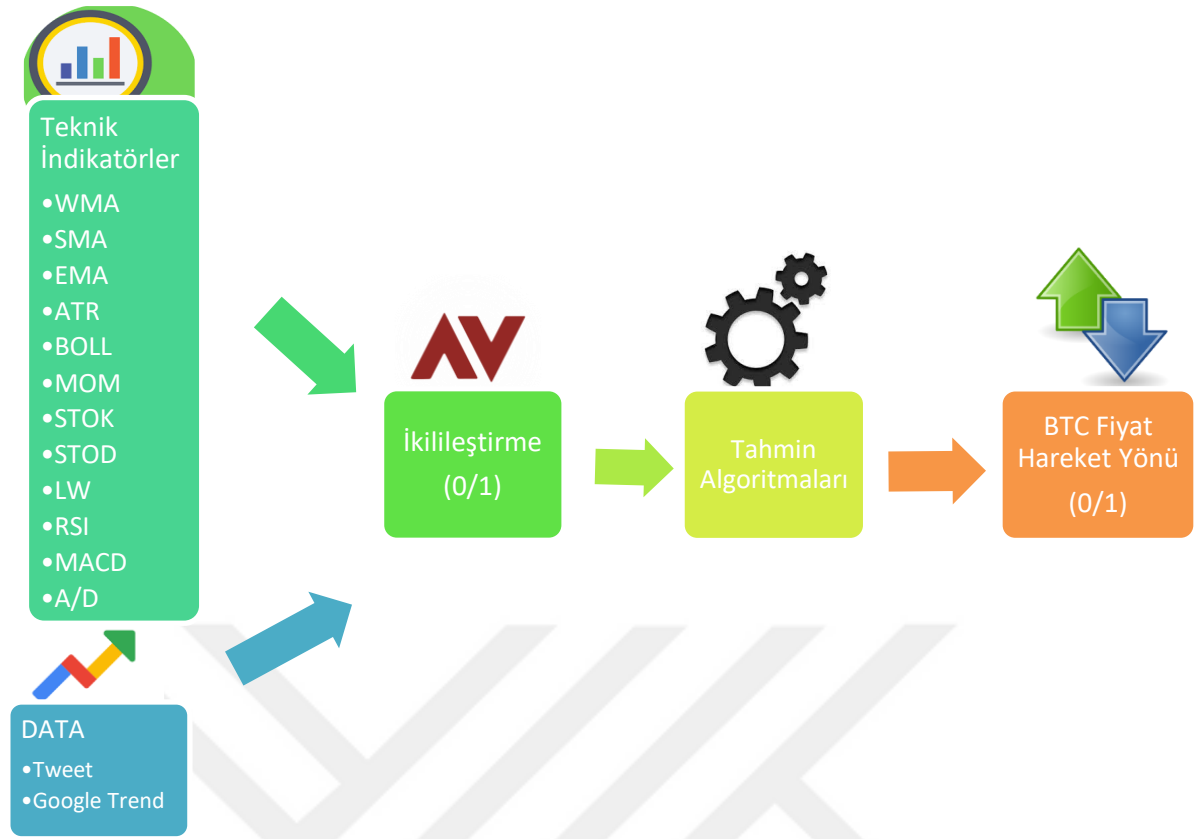
Veri	1 Değeri (Yukarı Yönlü Trend)	0 Değeri (Aşağı Yönlü Trend)
Kapanış Fiyatı (Close)	Bir önceki günden büyük ise	Bir önceki günden küçük ise
Basit Hareketli Ortalama(SMA)	Kapanıştan küçük ise	Kapanıştan büyük ise
Ağırlıklı Hareketli Ortalama(WMA)	Kapanıştan küçük ise	Kapanıştan büyük ise
Üstel Hareketli Ortalama(EMA)	Kapanıştan küçük ise	Kapanıştan büyük ise
Ortalama Gerçek Aralık(ATR)	Bir önceki günden büyük ise	Bir önceki günden küçük ise
Bollinger Bantları	Kapanış değeri alt çizgiye yakınsa	Kapanış değeri üst çizgiye yakınsa
Momentum	Pozitif ise	Negatif ise
Stokastik %K	Bir önceki günden büyük ise	Bir önceki günden küçük ise
Stokastik %D	Bir önceki günden büyük ise	Bir önceki günden küçük ise
Larry Williams(LW)	Bir önceki günden büyük ise	Bir önceki günden küçük ise
RSI	30'dan küçükse veya 30-70 aralığında ise bir önceki günden büyükse	70'den büyükse veya 30-70 aralığında ise bir önceki günden küçükse
MACD	Bir önceki günden büyük ise	Bir önceki günden küçük ise
A/D Osilatörü	Bir önceki günden büyük ise	Bir önceki günden küçük ise
Tweet	Bir önceki günden büyük ise	Bir önceki günden küçük ise
Google Trend	Bir önceki günden büyük ise	Bir önceki günden küçük ise

Bitcoin 2014-2021 yılları arasında 250 kata fiyat hareketliliğine uğradığından dolayı Veriler kesikli yapıya dönüştürülürken yapılacak hesaplamalarda normalize edilmiş verilerle gerçek veriler üzerinden yapılan dönüştürme işleminde ölçekten kaynaklı farklar oluşmaktadır. Bu oluşan farkın tahmin algoritmasının doğruluğunu da etkilediği tespit edilmiş ve bulgular kısmında belirtilmiştir. Bu sebeple çalışmada kesikli yapıya dönüştürme işleminde normalize edilmemiş veriler kullanılmıştır.

Sonuç olarak tez çalışmasında kullanılmak üzere 2014-2021 tarihleri 2455 günlük ve 2020-2021 tarihleri arası 363 günlük Bitcoin (BTC) USD bazında verilerinden teknik indikatörler yardımıyla hazırlanmış sürekli ve kesikli veri setlerinden oluşan 4 ayrı veri setimiz oluşmuştur. Oluşturulan tüm veri setleri makine öğrenme yöntemleri kullanılmak ve tekrar çağırabilmek üzere CSV formatında fiziksel dosyalara aktarılmıştır. Doğrulama için Çapraz Doğrulama (Cross Validation) yöntemi kullanılacağı dolay, ayrıca bir doğrulama veri seti oluşturulmamıştır.



Şekil 16. Sürekli veri seti için tahmin mekanizması



Şekil 17. Kesikli veri seti için tahmin mekanizması

Tablo 2. Girdilerin Tanımsal İstatistikleri

	Minimum	Maksimum	Ortalama	Standart Sapma
SMA14	222.873	60050.151	7706.481	11654.864
WMA14	219.543	60904.757	7738.658	11692.165
EMA14	225.270	60397.746	7704.597	11626.316
ATR	3.556	5003.094	450.985	833.745
MOM	-21229.875	14372.156	148.338	2094.090
StoK%	0	100	56.147	29.923
StoD%	2.714	99.199	56.148	28.172
Boll-Upper	211.193	66836.082	8288.583	12601.241
Boll-Middle	200.878	62382.740	7776.278	11759.605
Boll-Lower	161.955	60159.319	7263.974	10941.897
RSI	9.920	94.302	54.293	14.288
LW	-100	0	-43.853	29.923
ADOSC	-53660.135	119838.185	5269.312	14876.947
Tweets	7300	363566	37187.716	28701.176
GoogleTrends	5.268	616.867	48.825	60.356

Yöntem

Yapılacak tahmin çalışmasında günlük Bitcoin (BTC) fiyatı ve analiz verilerinden oluşan kesikli ve sürekli veri setleri Python ortamına çağırıldıktan sonra makine öğrenme yöntemlerinde kullanılmak üzere SKLearn kütüphanesinin Cross_Validate modülü ile 20 katmanlı olarak Çapraz Doğrulama ile ayrılarak sınanmıştır. Çapraz doğrulama bölümünde detaylı anlatılan bu yöntem ile verileri normal bölme yöntemi ayırmış olsak %75 eğitim, %25 test için ayrılacakken, 10 katmanlı testlerimizde tüm veriler katmanlı olarak doğrulama ile teste dahil edilerek gerçek doğruluk oranı yükseltilmiştir. Çapraz doğrulama katman sayısı olarak 5 katmanlı ve 20 katmanlı testlerde yapılmış ancak 5 katman için doğruluk düşerken, 20 katmanda ise öğrenme süresi çok uzun olmaktadır. Kesikli veri setlerinde eğitim için toplam 14 sütun, sürekli veri setlerinde Bolling Bantlarının yukarı, orta ve aşağı bölümleri olduğundan dolayı toplamda 16 sütun bulunmaktadır.

Tüm veri setlerinde çıktı, günlük kapanış fiyatının hareket yönü anlamına gelen kesikli verisi olarak tanımlanmıştır. Bu verinin 0 ise aşağı yönlü trend, 1 ise yukarı yönlü trend anlamına gelmektedir. Çıktı verisi de Eğitim ve Test setleri ile aynı oranda bölünerek kullanılmıştır.

Makine öğrenme yöntemlerinden K-En Yakın Komşu (KNN), Rastgele Orman (RF), XGBoost, Destek Vektör Makinası (SVM), Naive Bayes (NB), Yapay Sinir Ağları (ANN) ve Kısa-Uzun Süreli Bellek (LSTM) kullanılmıştır. Analizler kısmında her bir makine öğrenme yönteminin kesikli ve sürekli veri setlerinde tahmin performansının doğruluk testi sonuçlarına göre optimizasyonu, Teknik analiz yöntemlerinin makine öğrenmede ki önem derecesi, Ortalama Hata Karesi, Hata Matrisi (Confusion Matrix) sonuçları ve son olarak makine öğrenme yöntemlerinin performans kıyasları yapılacaktır.

Makine öğrenme yöntemlerinde Python Programlama dilinin scikit-learn, xgboost, keras ve tensorflow kütüphaneleri ve ilgili eklentileri ile uygulamalar gerçekleştirilmiştir. Matematiksel hesaplamalar için numpy kütüphanesi, görselleştirme ve grafikler için Matplotlib ve Seaborn kütüphaneleri kullanılmıştır. Hata matrisinin oluşturulması ve performans kıyaslarında yine scikit-learn kütüphanesi içinde yer alan metrics, confusion_matrix eklentilerinden faydalanılmıştır. Testler AMD A12-9720P 2.70 GHz İşlemci, 16 GB DDR4 RAM, 256 GB M.2 SSD, AMD Radeon 535 Ekran Kartı bulunan sistemde gerçekleştirilmiştir.

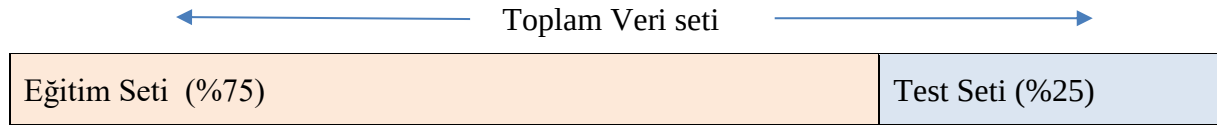
Analizler

Doğrulama yöntemi analizi:

Makine öğrenme yöntemleri test edilirken verilerin belirli bir kısmı eğitim için, belirli bir kısmı ise test için ayrılmaktadır. Bunun için kullanılabilecek yöntemler doğrulama yöntemleri bölümünde açıklanmış olan Eğitim-Test Ayırıştırması (Train-Test Split) ve Çapraz Doğrulama (Cross Validation) yöntemleridir. Veri setimize en uygun yöntemi seçmek için iki yöntemde test edilmiştir.

Eğitim-Test ayırıştırması (Train-Test split).

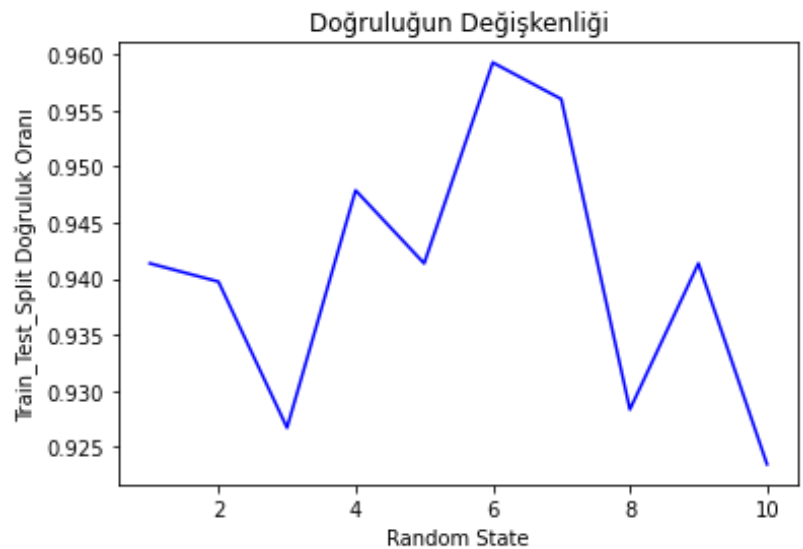
Bu yöntemi verilerimizi standart ayarda %75 eğitim seti, %25 test seti olarak ayırmaktadır. Ancak eğer sabitleme yapılmaz ise bu ayırıştırma veri içinden rastgele yapılmaktadır. Bu durum her ayırıştırmada farklı doğruluk sonucu vermektedir. Bu rastgele durumu (random state) dilersek sabitleyebilmekteyiz.



Aşağıdaki tabloda Kesikli tüm verilerden oluşan 1. Veri setimiz K-en yakın komşu yöntemi ile 1-10 arası farklı sabitlere göre doğruluğun değişkenliği ve ortalaması görülmektedir.

Tablo 3. Eğitim-Test Ayırıştırması Doğruluk Değişimi

Random State	Doğruluk
1	0.941368
2	0.939739
3	0.926710
4	0.947883
5	0.941368
6	0.959283
7	0.956026
8	0.928339
9	0.941368
10	0.923453
Ort	0.940554



Tablo 3’te yer alan verilerde her sabitte doğruluk oranı değiştiği gözlenmiştir. Bu durum yöntem her test edilirken farklı sonuçlarla karşılaşmamız anlamına gelebilmektedir.

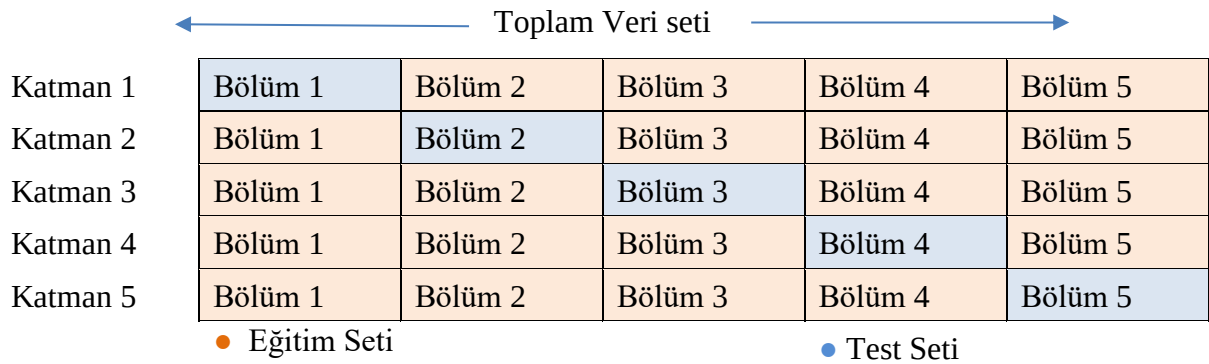
Eğitim-Test Ayrışması Sınıflandırma Raporu:

	Kesinlik	Hatırlama	F Skoru	Destek
0	0.94	0.94	0.94	282
1	0.95	0.95	0.95	332
Doğruluk			0.94	614
Makro Ort.	0.94	0.94	0.94	614
Ağırlıklı Ort.	0.94	0.94	0.94	614

Sınıflandırma raporunda yer alan verilere göre Eğitim-Test Ayrıştırması yönteminde Doğruluk olarak %94’lük doğruluk ve 2455 veriden oluşan setin %25’i yani 614’lük kısmı test için kullanılabilmektedir.

Katmanlı çapraz doğrulama.

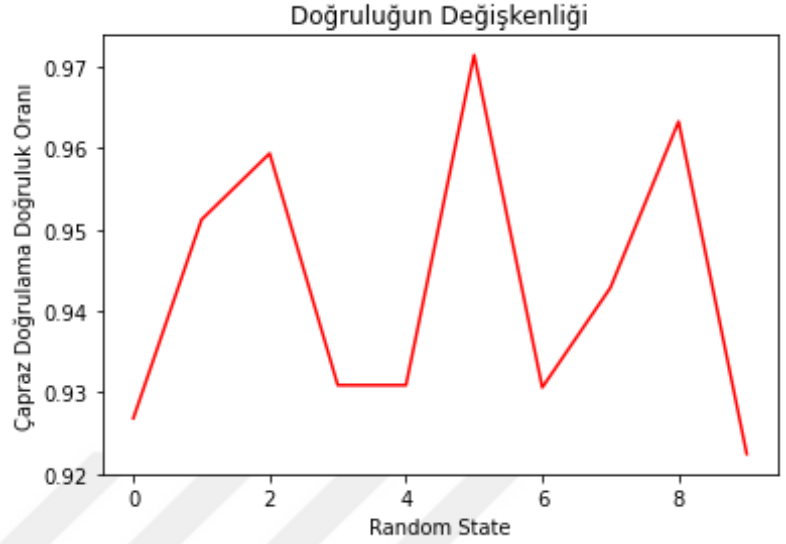
Katmanlı çapraz doğrulama yönteminde veri setimiz belirlediğimiz katman sayısına bölünerek bu katmanların her biri eğitim ve doğrulama için kullanılmakta ve sonuç olarak bu sonuçların ortalaması vermektedir. Bu yöntem özellikle kümeleme problemlerinde iyi sonuç vermektedir.



Tablo 4’te kesikli tüm verilerden oluşan 1. veri setimiz K-en yakın komşu yöntemi ile 10 katman doğruluk ayarının her katmanında oluşan doğruluk ve bize verilen sonucu görmekteyiz.

Tablo 4. Çapraz Doğrulama Katman Doğrulukları

Katman	Doğruluk
1	0.92682927
2	0.95121951
3	0.95934959
4	0.93089431
5	0.93089431
6	0.97142857
7	0.93061224
8	0.94285714
9	0.96326531
10	0.92244898
Sonuç	0.94297992



Katmanlı Çapraz doğrulama yöntemi sonuç olarak tüm katmanlarda oluşan doğruluğun ortalamasını vermektedir.

Çapraz Doğrulama Sınıflandırma Raporu:

	Kesinlik	Hatırlama	F Skoru	Destek
0	0.94	0.94	0.94	1120
1	0.95	0.95	0.95	1335
Doğruluk			0.94	2455
Makro Ort.	0.94	0.94	0.94	2455
Ağırlıklı Ort.	0.94	0.94	0.94	2455

Sınıflandırma raporunda yer alan verilere göre Katmanlı çapraz doğrulama yönteminde Doğruluk olarak %94'lük doğruluk yani Eğitim-Test Ayırıştırması ile aynı sonuçları alırken ve 2455 veriden oluşan setinin tamamı eğitim ve test için kullanıldığı görülmüştür. Bu sebeple çapraz doğrulama yöntemi kullanması uygun olan K-En Yakın Komşu, Rastgele Orman, XGBoost ve Destek Vektör Makinası yöntemlerde çapraz doğrulama, Yapay Sinir Ağları ve Kısa-Uzun Süreli Bellek yöntemlerinde Eğitim-Test Ayırıştırması yöntemi ile testler gerçekleştirilmiştir.

Rastgele Orman (RF) Yöntemi İle Analiz

Rastgele Orman Makine Öğrenme yönteminde en iyi doğruluk sonucuna ulaşabilmek için 4 farklı derinlik ayarında 10-500 arası Ağaç Sayısı 20 katmanlı Çapraz Doğrulama Yöntemi ile Doğruluk değerlerinin ortalaması üzerinden test edilmiştir. Önem dereceleri hesaplanırken Feature Importance modülünden faydalanılmıştır.

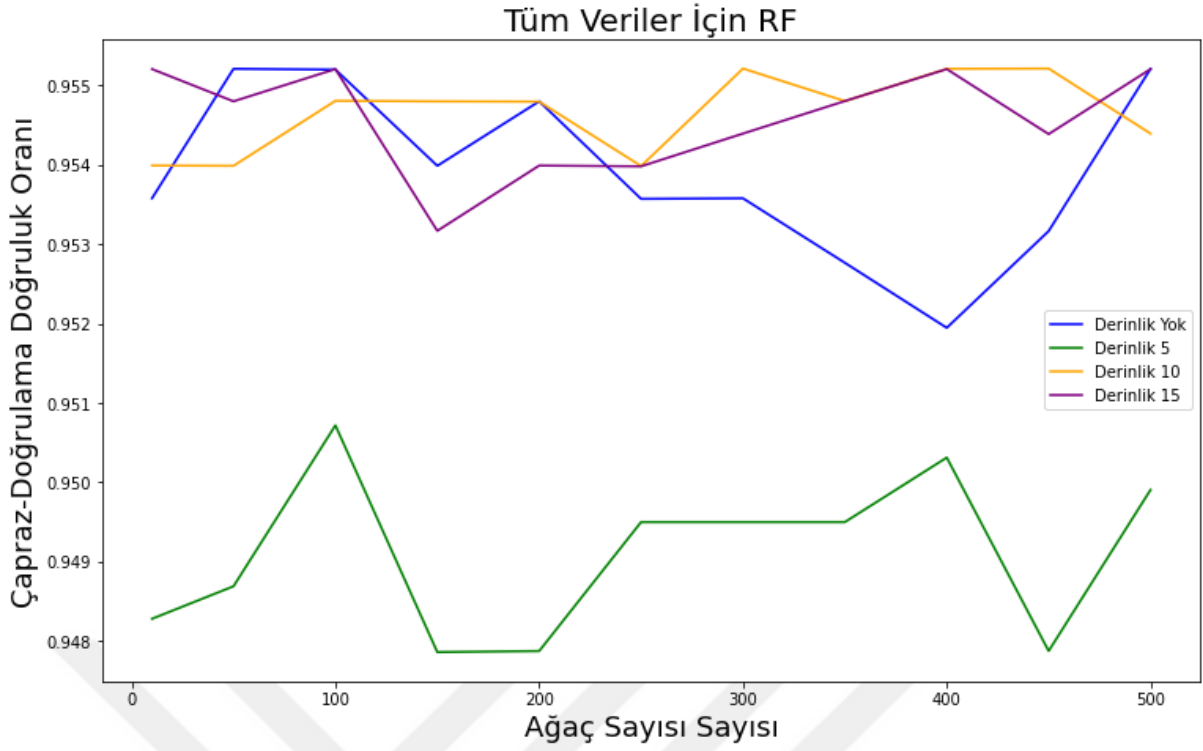
Kesikli verilerin analizi:

Kesikli tüm veri.

Kesikli yapıdaki tüm verilerden oluşan 1. veri seti için 300 ağaç ve 10 derinlikte en iyi sonuca ulaşılmıştır.

Tablo 5. Kesikli Tüm Veriler RF Yöntemi Optimizasyon Ayarları

Ağaç Sayısı	Derinlik Yok	Derinlik 5	Derinlik 10	Derinlik 15
10	0.953579	0.948277	0.953992	0.955208
50	0.955211	0.948687	0.953988	0.954801
100	0.955201	0.950716	0.954808	0.955208
150	0.953988	0.947858	0.954801	0.953169
200	0.954801	0.947871	0.954798	0.953992
250	0.953572	0.949497	0.953985	0.953982
300	0.953579	0.949497	0.955215	0.954395
350	0.952766	0.949497	0.954805	0.954805
400	0.951946	0.950310	0.955211	0.955208
450	0.953165	0.947871	0.955215	0.954388
500	0.955211	0.949903	0.954395	0.955208



Şekil 18. Kesikli Tüm Veriler RF yöntemi Optimizasyon Sonuçları

Ağaç Sayısı = 100, derinlik 10, çapraz doğrulama 20 katman için yapılan tahmin işleminde:

Karışıklık Matrisi:

	Pozitif	Negatif
Pozitif	1062	58
Negatif	54	1281

Sınıflandırma Raporu:

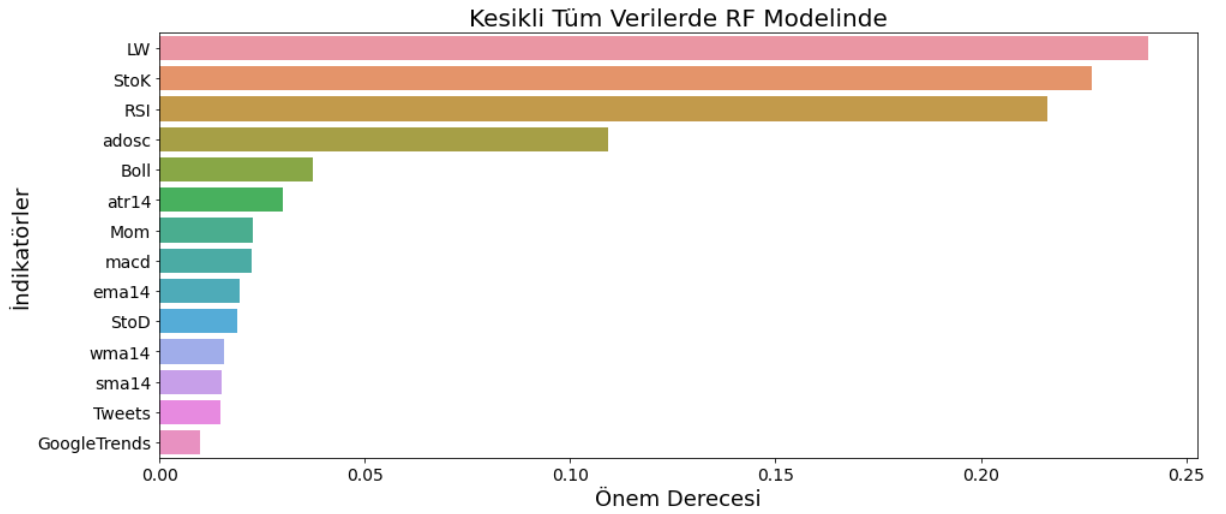
	Kesinlik	Hatırlama	F Skoru	Destek
0	0.95	0.95	0.95	1120
1	0.96	0.96	0.96	1335
Doğruluk			0.95	2455
Makro Ort.	0.95	0.95	0.95	2455
Ağırlıklı Ort.	0.95	0.95	0.95	2455

Kesikli tüm veriler için indikatörlerin önem derecesi.

Kesikli tüm veriler için RF yönteminde Larry Williams, Stokastik %K, ve RSI indikatörlerinin diğer verilere göre tahmin için ön plana çıktığı görülmüştür.

Tablo 6. Kesikli Tüm Veri RF Tahmininde İndikatörlerin Önem Derecesi

İndikatör	Önem Derecesi
LW	0.240790
StoK	0.226936
RSI	0.216358
adosc	0.109182
Boll	0.037206
atr14	0.029977
Mom	0.022760
macd	0.022549
ema14	0.019516
StoD	0.019048
wma14	0.015826
sma14	0.015087
Tweets	0.014955
GoogleTrends	0.009808



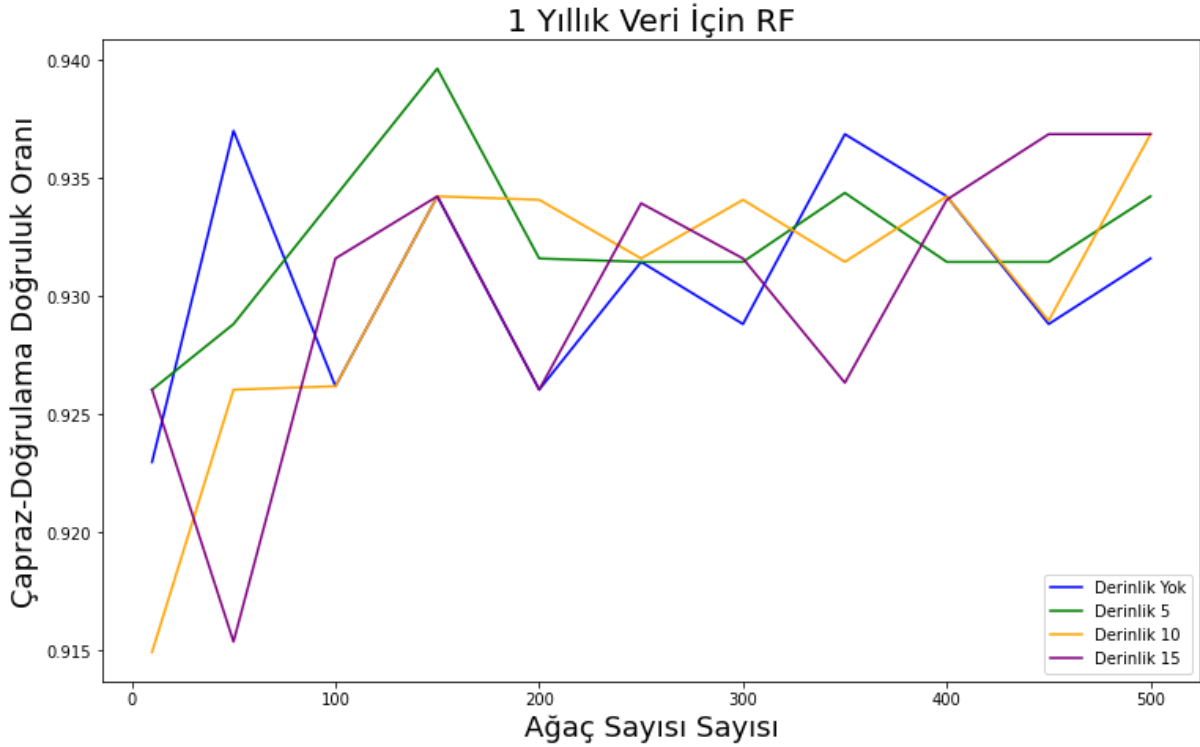
Şekil 19. Kesikli tüm veri seti RF tahmininde indikatörlerin önem oranı

Kesikli 1 yıllık veri.

Kesikli yapıdaki 1 yıllık verilerden oluşan 2. veri seti için 150 ağaç ve 5 derinlikte en iyi sonuca ulaşılmıştır.

Tablo 7. Kesikli 1 Yıllık Veriler RF Yöntemi Optimizasyon Ayarları

Ağaç Sayısı	Derinlik Yok	Derinlik 5	Derinlik 10	Derinlik 15
10	0.922953	0.926023	0.914912	0.926023
50	0.936988	0.928801	0.926023	0.915351
100	0.926170	0.934211	0.926170	0.931579
150	0.934211	0.939620	0.934211	0.934211
200	0.926023	0.931579	0.934064	0.926023
250	0.931433	0.931433	0.931579	0.933918
300	0.928801	0.931433	0.934064	0.931579
350	0.936842	0.934357	0.931433	0.926316
400	0.934211	0.931433	0.934211	0.934064
450	0.928801	0.931433	0.928947	0.936842
500	0.931579	0.934211	0.936842	0.936842



Şekil 20. Kesikli 1 yıllık veriler RF yöntemi optimizasyon sonuçları

Ağaç Sayısı = 150, derinlik 5, çapraz doğrulama 10 katman için:

Karışıklık Matrisi:

	Pozitif	Negatif
Pozitif	156	10
Negatif	14	183

Sınıflandırma Raporu:

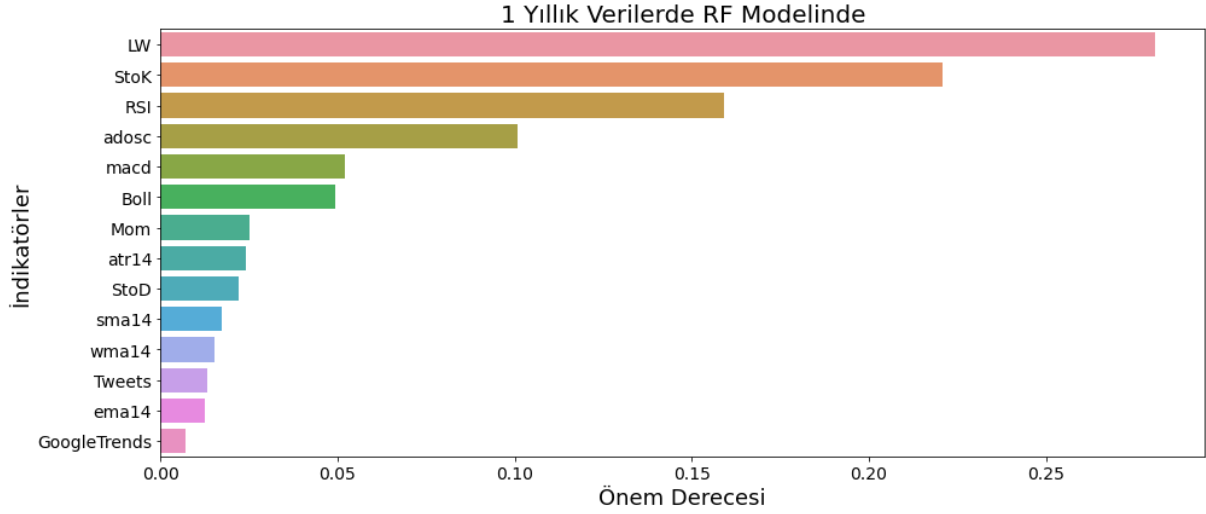
	Kesinlik	Hatırlama	F Skoru	Destek
0	0.92	0.94	0.93	166
1	0.95	0.93	0.94	197
Doğruluk			0.93	363
Makro Ort.	0.93	0.93	0.93	363
Ağırlıklı Ort.	0.93	0.93	0.93	363

Kesikli 1 yıllık veriler için indikatörlerin önem derecesi.

Kesikli 1 yıllık veriler için RF yönteminde, tıpkı kesikli tüm verilerde olduğu gibi Larry Williams, Stokastik %K ve RSI indikatörlerinin diğer verilere göre tahmin için ön plana çıktığı görülmüştür. Bu durum önem açısından bazı indikatörlerin kapanış değerlerine daha fazla yakınlık gösterdiği olarak yorumlanabilir.

Tablo 8. *Kesikli 1 Yıllık Veriler RF Yöntemi İndikatörlerin Önem Derecesi*

İndikatör	Önem Derecesi
LW	0.280857
StoK	0.220896
RSI	0.159213
adosc	0.100936
macd	0.052214
Boll	0.049330
Mom	0.025186
atr14	0.024183
StoD	0.021928
sma14	0.017110
wma14	0.015352
Tweets	0.013128
ema14	0.012583
GoogleTrends	0.007085



Şekil 21. Kesikli 1 yıllık veriler RF yöntemi indikatörlerin önem oranı

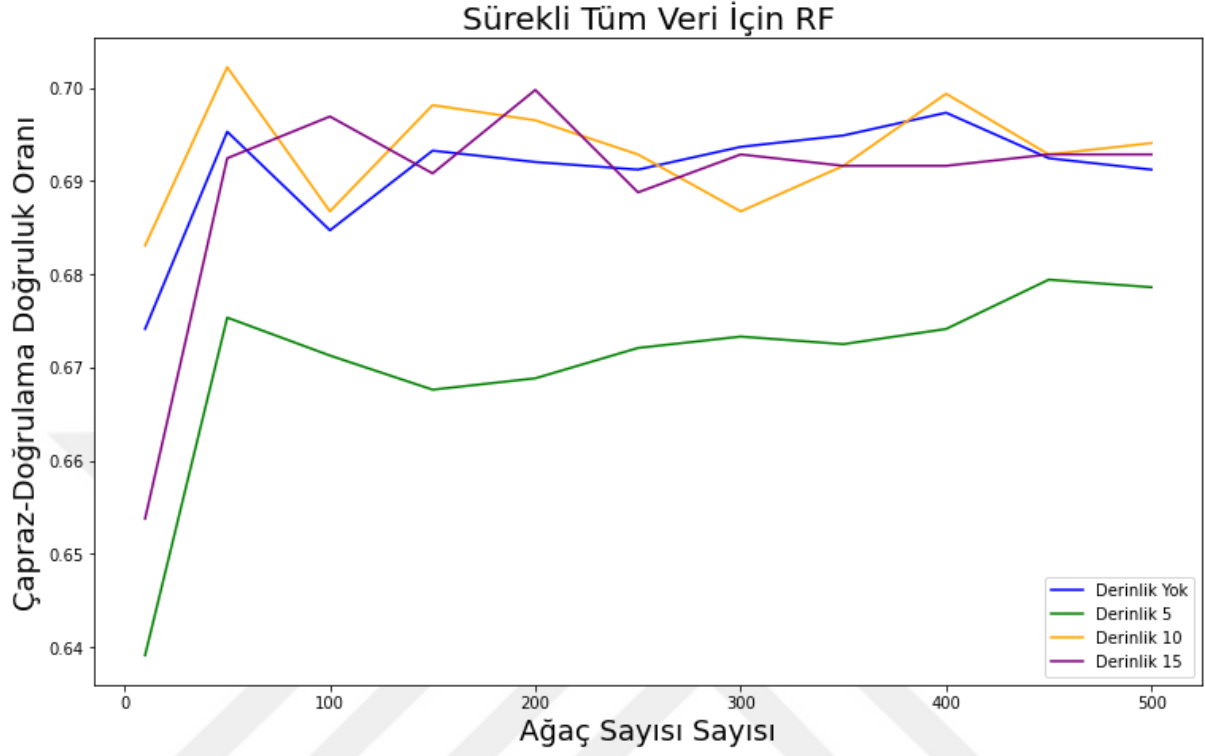
Sürekli verilerin analizi:

Sürekli tüm veri.

Sürekli yapıdaki tüm verilerden oluşan 3. veri seti için 50 ağaç ve 10 derinlikte en iyi sonuca ulaşılmıştır.

Tablo 9. *Sürekli Tüm Veri RF Yöntemi Optimizasyon Ayarları*

Ağaç Sayısı	Derinlik Yok	Derinlik 5	Derinlik 10	Derinlik 15
10	0.674134	0.639104	0.683096	0.653768
50	0.695316	0.675356	0.702240	0.692464
100	0.684725	0.671283	0.686762	0.696945
150	0.693279	0.667617	0.698167	0.690835
200	0.692057	0.668839	0.696538	0.699796
250	0.691242	0.672098	0.692872	0.688798
300	0.693686	0.673320	0.686762	0.692872
350	0.694908	0.672505	0.691650	0.691650
400	0.697352	0.674134	0.699389	0.691650
450	0.692464	0.679430	0.692872	0.692872
500	0.691242	0.678615	0.694094	0.692872



Şekil 22. Sürekli tüm veri RF yöntemi optimizasyon sonuçları

Ağaç Sayısı = 50, derinlik 10, çapraz doğrulama 5 katman için:

Karışıklık Matrisi:

	Pozitif	Negatif
Pozitif	804	316
Negatif	444	891

Sınıflandırma Raporu:

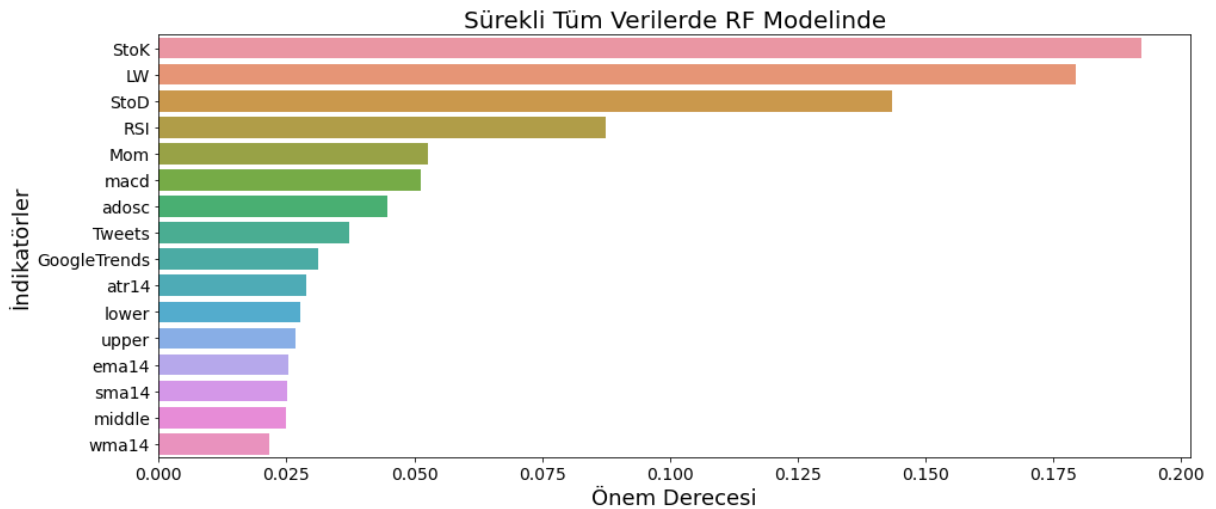
	Kesinlik	Hatırlama	F Skoru	Destek
0	0.64	0.69	0.68	1120
1	0.74	0.65	0.70	1335
Doğruluk			0.69	2455
Makro Ort.	0.69	0.69	0.69	2455
Ağırlıklı Ort.	0.70	0.69	0.69	2455

Sürekli tüm veriler için indikatörlerin önem derecesi.

Sürekli tüm veriler için RF yönteminde Stokastik %K, Larry Williams ve Stokastik %D indikatörlerinin diğer verilere göre tahmin için daha önemli çıktığı görülmüştür. Kesikli verilerde RSI indikatörü daha ön plandayken bu kez 4. Sırada yer aldığı gözlenmiştir. Bu durum RSI'in kesikli yapıya dönüştürülürken uygulanan formülün etkili olduğu göstermektedir.

Tablo 10. *Sürekli Tüm Veri RF Yöntemi İndikatörlerin Önem Derecesi*

İndikatör	Önem Derecesi
StoK	0.192210
LW	0.179475
StoD	0.143429
RSI	0.087425
Mom	0.052615
macd	0.051295
adosc	0.044638
Tweets	0.037346
GoogleTrends	0.031233
atr14	0.028772
lower	0.027650
upper	0.026789
ema14	0.025340
sma14	0.025164



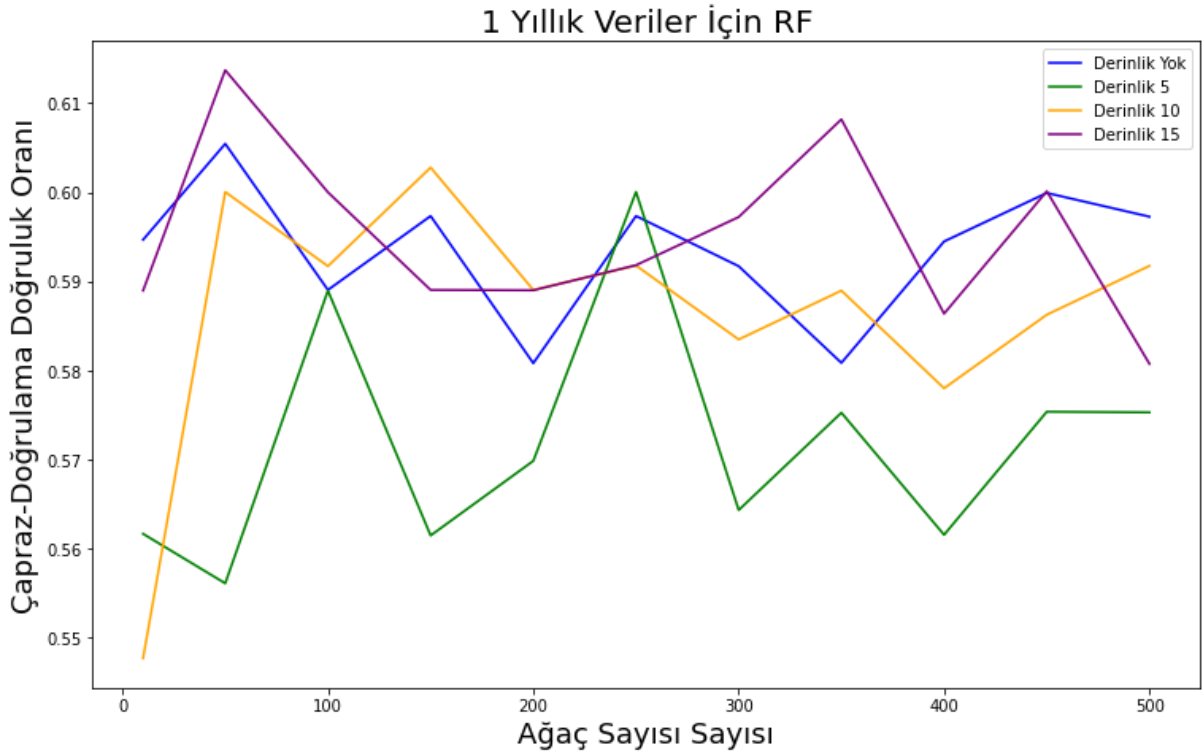
Şekil 23. Sürekli tüm veri RF yöntemi indikatörlerin önem oranları

Sürekli 1 yıllık veri.

Tablo 11. Sürekli 1 Yıllık Veriler İçin RF Yöntemi Optimizasyon Ayarları

Ağaç Sayısı	Derinlik Yok	Derinlik 5	Derinlik 10	Derinlik 15
10	0.594673	0.561682	0.547717	0.588965
50	0.605441	0.556126	0.600038	0.613699
100	0.589041	0.588965	0.591705	0.600000
150	0.597336	0.561492	0.602778	0.589041
200	0.580822	0.569863	0.589041	0.589003
250	0.597336	0.600038	0.591781	0.591819
300	0.591705	0.564346	0.583486	0.597222
350	0.580860	0.575266	0.588965	0.608181
400	0.594482	0.561568	0.578006	0.586377
450	0.599924	0.575381	0.586263	0.600114
500	0.597260	0.575304	0.591743	0.580746

Sürekli yapıdaki 1 yıllık verilerden oluşan 4. veri seti için 50 ağaç ve 15 derinlik ayarı ile en iyi sonuca ulaşılmıştır.



Şekil 24. Sürekli 1 yıllık veriler için RF yöntemi sonuçları

Ağaç Sayısı = 50, derinlik 15, çapraz doğrulama 5 katman için:

Karışıklık Matrisi:

	Pozitif	Negatif
Pozitif	91	75
Negatif	79	118

Sınıflandırma Raporu:

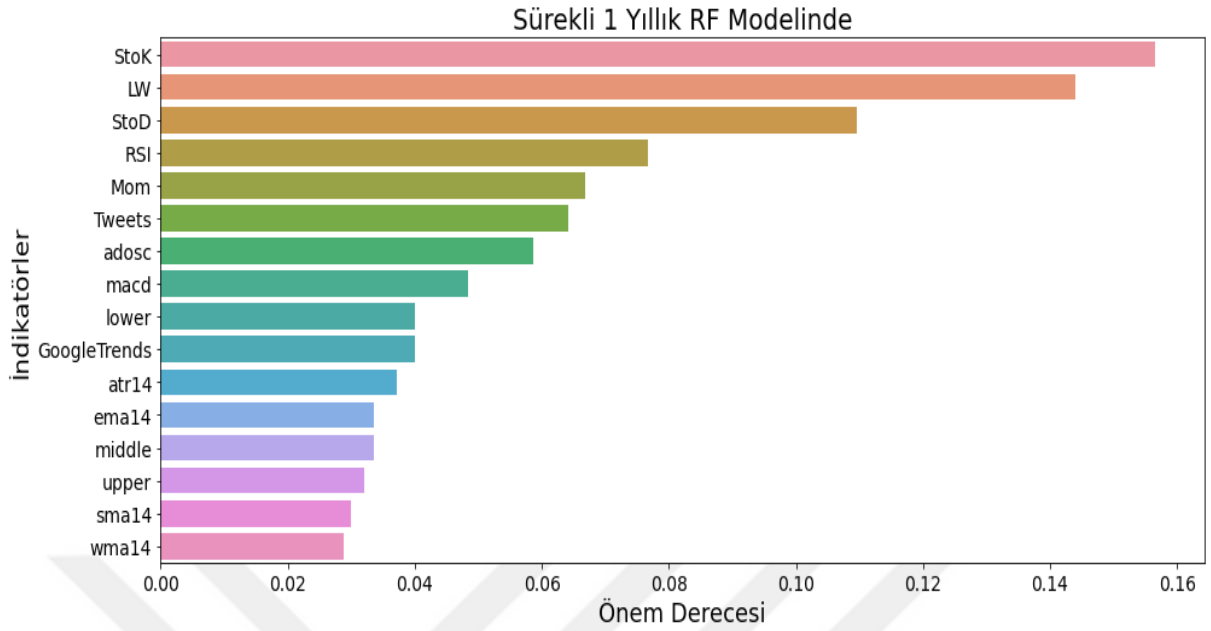
	Kesinlik	Hatırlama	F Skoru	Destek
0	0.54	0.55	0.54	166
1	0.61	0.60	0.61	197
Doğruluk			0.58	363
Makro Ort.	0.57	0.57	0.57	363
Ağırlıklı Ort.	0.58	0.58	0.58	363

Sürekli 1 yıllık veriler için indikatörlerin önem derecesi.

Sürekli 1 yıllık veriler için RF yönteminde Stokastik %K, Larry Williams ve Stokastik %D indikatörlerinin tıpkı sürekli tüm verilerde olduğu gibi diğer verilere göre tahmin için daha önemli çıktığı görülmüştür. Hem kesikli hem de sürekli yapıda verilerin RF yöntemi ile tahmininde tüm veriler ve 1 yıllık verilerde önem derecesi paralellik göstermiştir. Bu durum RF yönteminin çok ve az verinin yorumlanmasında önem açısından ağırlıkların değişmediğini ve çözümü bir regresyon problemi olarak yorumladığını göstermektedir.

Tablo 12. *Sürekli 1 Yıllık Veri RF Yöntemi İndikatörlerin Önem Derecesi*

İndikatör	Önem Derecesi
StoK	0.156518
LW	0.143925
StoD	0.109569
RSI	0.076713
Mom	0.066734
Tweets	0.064069
adosc	0.058654
macd	0.048446
lower	0.040118
GoogleTrends	0.040012
atr14	0.037108
ema14	0.033580
middle	0.033529
upper	0.032115



Şekil 25. Sürekli 1 yıllık veri RF yöntemi indikatörlerin önem oranı

K-En Yakın Komşu (KNN) Yöntemi İle Analiz

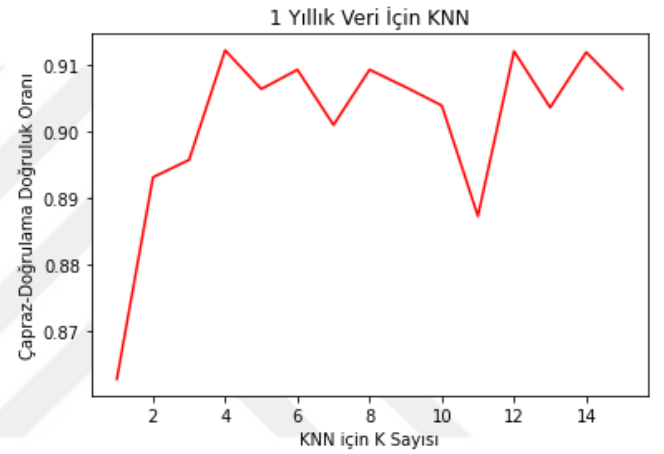
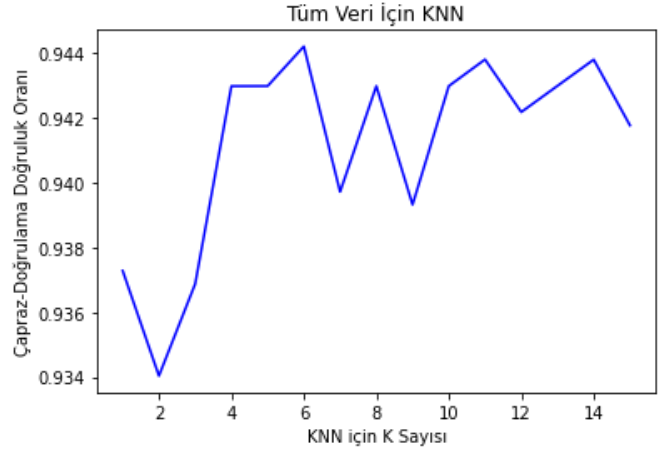
K-En Yakın Komşu(KNN) Makine Öğrenme Yönteminde en iyi sonuç için komşu sayıları kıyaslanmış ve kesikli verilerde ve sürekli verilerde 200 komşuya kadar testler yapılmıştır. 20 katmanlı Çapraz Doğrulama Yöntemi ile Doğruluk değerlerinin ortalaması alınarak test edilmiştir.

Kesikli verilerin analizi:

Yapılan test sonucunda, Tüm Veriler için %94,4 ile en iyi komşu oranı 6 olurken, 1 Yıllık Veri en iyi komşu sayısına %91,2 ile 4 komşu sonucunda ulaşılmıştır. 15 komşudan fazla komşu sayısı denemeleri yapılmış ancak daha yüksek bir sonuç alınamamış. Önem dereceleri hesaplanırken Permutation Importance modülünden faydalanılmıştır.

Tablo 13. Kesikli Verilerde KNN En İyi Komşu Oranı

K Sayısı	Tüm Veri	1 Yıllık Veri
1	0.937278	0.862719
2	0.934036	0.893129
3	0.936872	0.895760
4	0.942973	0.912281
5	0.942976	0.906433
6	0.944199	0.909357
7	0.939714	0.901023
8	0.942976	0.909357
9	0.939318	0.906725
10	0.942976	0.903947
11	0.943796	0.887281
12	0.942173	0.912135
13	0.942983	0.903655
14	0.940247	0.921429
15	0.939703	0.928571



Kesikli tüm veri için:

K = 6 için:

Karışıklık Matrisi:

	Pozitif	Negatif
Pozitif	1068	52
Negatif	85	1250

Sınıflandırma Raporu:

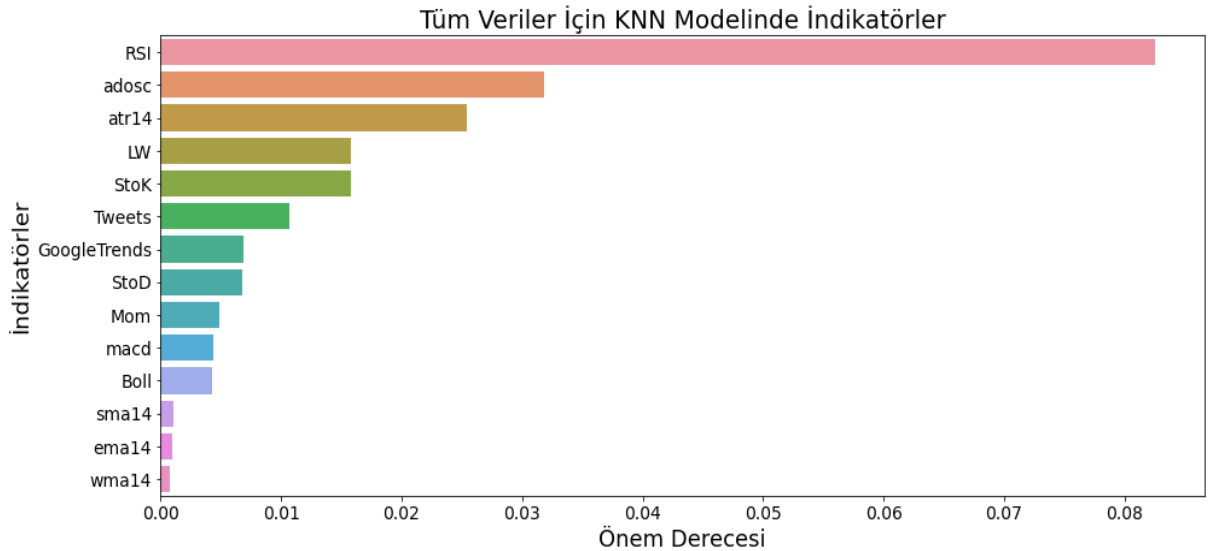
	Kesinlik	Hatırlama	F Skoru	Destek
0	0.93	0.95	0.94	1120
1	0.96	0.94	0.95	1335
Doğruluk			0.94	2455
Makro Ort.	0.94	0.94	0.94	2455
Ağırlıklı Ort.	0.94	0.94	0.94	2455

Kesikli tüm veriler için indikatörlerin önem derecesi.

Kesikli tüm verilerin KNN yöntemi ile tahmininde RSI indikatörünün belirgin olarak tahminde rol oynadığı görülmektedir. Bunun sebebinin RSI indikatörünün kesikli yapıya dönüştürülürken uygulanan formül sonucunun sınıflandırma açısından kapanış değerine gösterdiği benzerlikten kaynaklanabilmektedir.

Tablo 14. Kesikli Tüm Veri KNN Yöntemi İndikatörlerin Önem Derecesi

İndikatör	Önem Derecesi
RSI	0.082525
adosc	0.031853
atr14	0.025418
LW	0.015804
StoK	0.015804
Tweets	0.010672
GoogleTrends	0.006843
StoD	0.006762
Mom	0.004888
macd	0.004399
Boll	0.004236
sma14	0.001059
ema14	0.000978
wma14	0.000815



Şekil 26. Kesikli Tüm Veri KNN Yöntemi İndikatörlerin Önem Oranları

Kesikli 1 yıllık veri için:

K = 4 için:

Karışıklık Matrisi:

	Pozitif	Negatif
Pozitif	152	14
Negatif	18	179

Sınıflandırma Raporu:

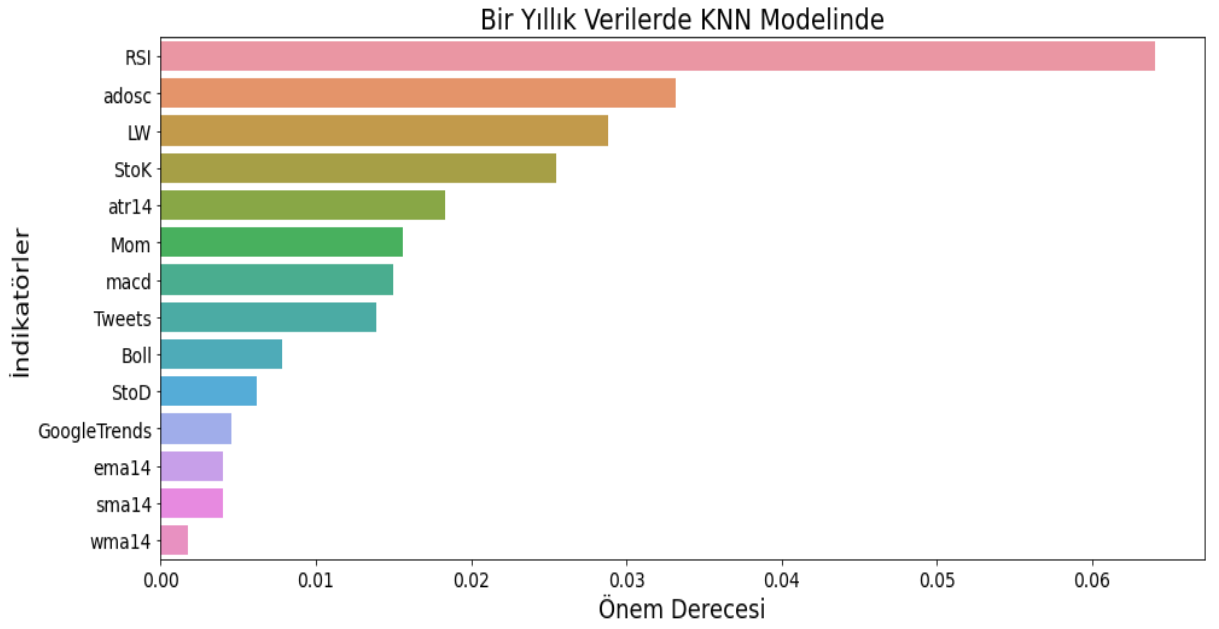
	Kesinlik	Hatırlama	F Skoru	Destek
0	0.89	0.92	0.90	166
1	0.93	0.91	0.92	197
Doğruluk			0.91	363
Makro Ort.	0.91	0.91	0.91	363
Ağırlıklı Ort.	0.91	0.91	0.91	363

Kesikli 1 yıllık veriler için indikatörlerin önem derecesi.

Kesikli 1 yıllık verilerin KNN yöntemi ile tahmininde RSI indikatörü kesikli tüm verilerin tahmininde olduğu gibi tahminde rol oynadığı görülmektedir. LW indikatörü tüm veriye göre bir miktar daha fazla önem kazanmıştır, ancak genel olarak kesikli tüm verilerin tahmin önem derecesi ile paralellik göstermektedir. Verilerin önem derecesi sonuçlarının pozitif yorumlanabilmesi için tüm ağırlıklara 0.004 ekleme yapılmıştır.

Tablo 15. *Kesikli 1 Yıllık Veri KNN Yöntemi İndikatörlerin Önem Derecesi*

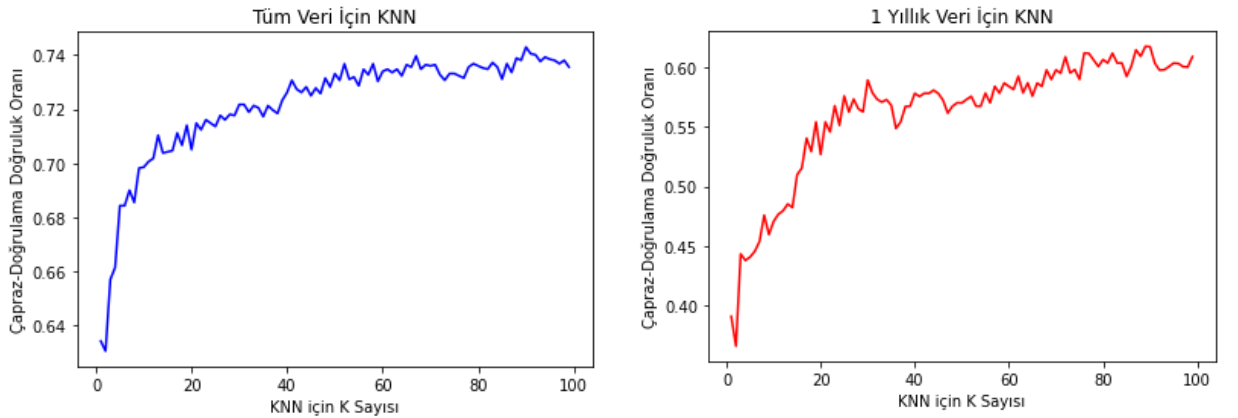
İndikatör	Önem Derecesi
RSI	0.064055
adosc	0.033201
LW	0.028793
StoK	0.025488
atr14	0.018325
Mom	0.015570
macd	0.015019
Tweets	0.013917
Boll	0.007857
StoD	0.006204
GoogleTrends	0.004551
ema14	0.004000
sma14	0.004000
wma14	0.001796



Şekil 27. Kesikli 1 Yıllık Veri KNN Yöntemi İndikatörlerin Önem Oranı

Sürekli verilerin analizi:

Yapılan test sonucunda, Tüm Veri için en iyi komşu sayısı %74,29 doğruluk ile 89; 1 Yıllık Veri için ise en iyi komşu sayısına %61,72 ile 88. Komşu sonucunda ulaşılmıştır. Daha fazla komşu sayısı denemeleri yapılmış ancak daha yüksek bir sonuç elde edilememiştir.



Şekil 28. Sürekli Verilerde KNN en iyi komşu oranı

Sürekli tüm veriler için:

K = 89 için:

Karışıklık Matrisi:

	Pozitif	Negatif
Pozitif	719	401
Negatif	242	1093

Sınıflandırma Raporu:

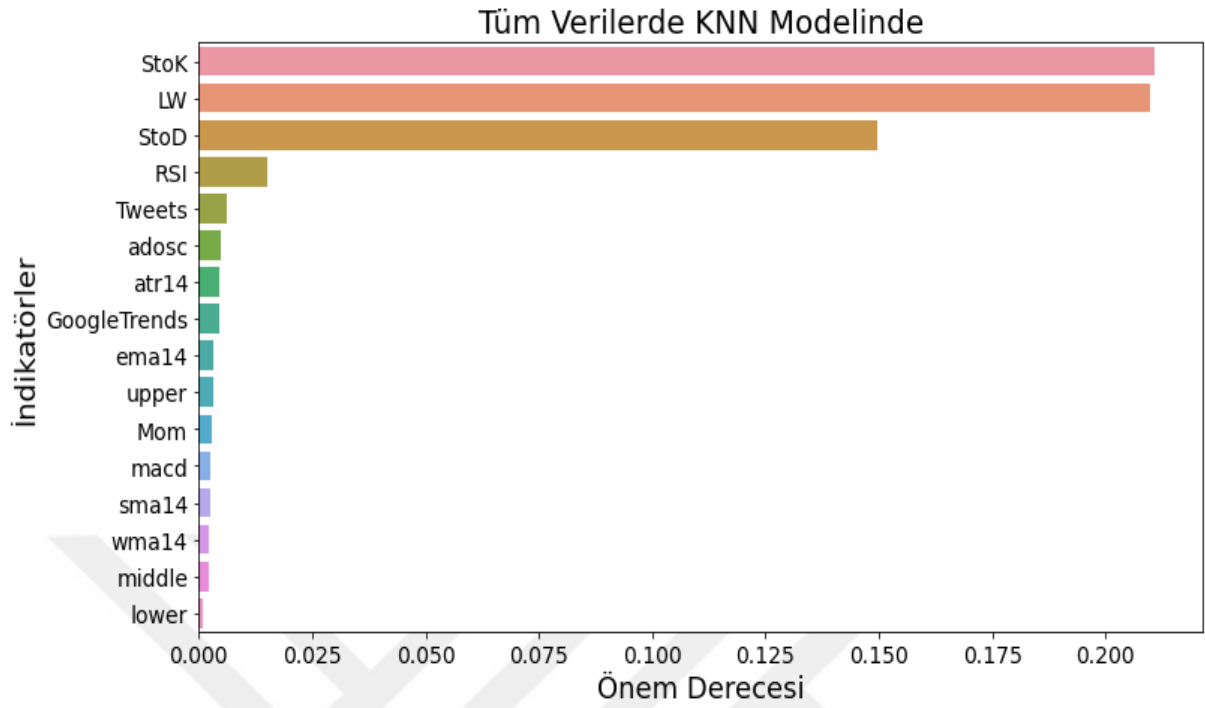
	Kesinlik	Hatırlama	F Skoru	Destek
0	0.75	0.64	0.69	1120
1	0.73	0.82	0.77	1335
Doğruluk			0.74	2455
Makro Ort.	0.74	0.73	0.73	2455
Ağırlıklı Ort.	0.74	0.74	0.74	2455

Sürekli tüm veriler için indikatörlerin önem derecesi.

Sürekli tüm verilerin KNN yöntemi ile tahmininde kesikli yapıya göre tamamen farklı sonuçlar elde edilmiştir. Sürekli tüm yapıdaki verilerin tahmininde Stokastik %K, Larry Williams ve Stokastik %D indikatörleri ön plana çıkmışken, RSI baskınlığını kaybetmiştir. Bunun sebebinin KNN algoritmasının sınıflandırma yönteminin kesikli ve sürekli yapıdaki verilerin sınıflandırmasındaki farktan kaynaklandığı düşünülmektedir.

Tablo 16. *Sürekli Tüm Veri KNN Yöntemi İndikatörlerin Önem Derecesi*

İndikatör	Önem Derecesi
StoK	0.210843
LW	0.209703
StoD	0.149825
RSI	0.015161
Tweets	0.006037
adosc	0.004815
atr14	0.004489
GoogleTrends	0.004407
ema14	0.003267
upper	0.003267
Mom	0.003022
macd	0.002615
sma14	0.002534
wma14	0.002289



Şekil 29. Sürekli Tüm Veri KNN Yöntemi İndikatörlerin Önem Oranı

Sürekli 1 yıllık veriler için:

k = 88 için:

Karışıklık Matrisi:

	Pozitif	Negatif
Pozitif	91	75
Negatif	67	130

Sınıflandırma Raporu:

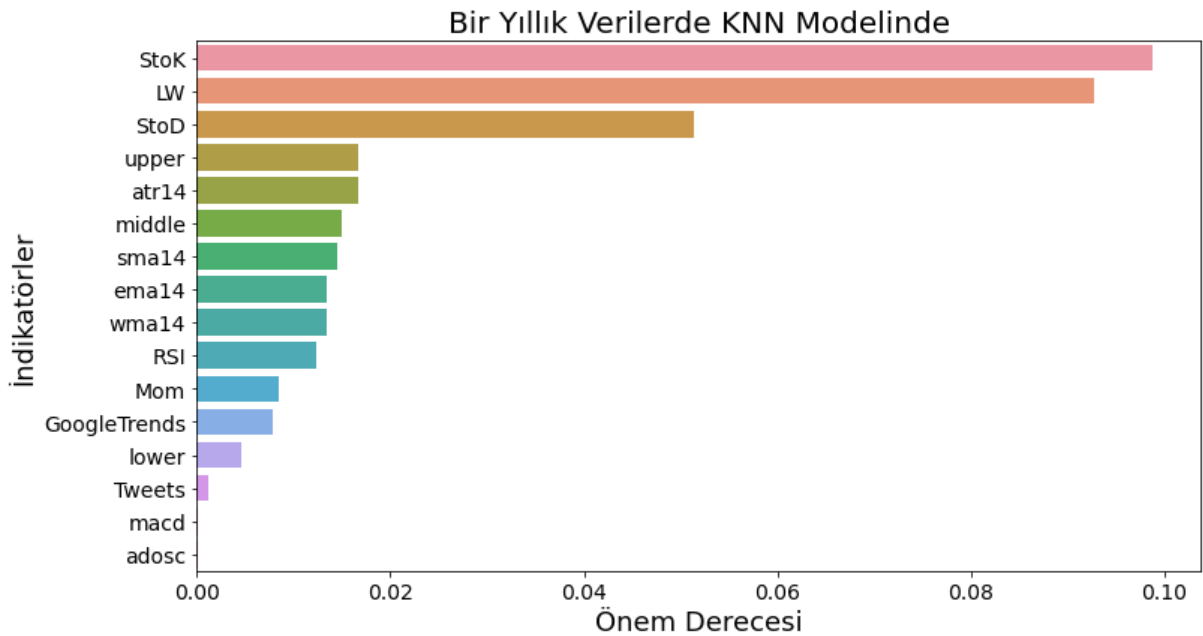
	Kesinlik	Hatırlama	F Skoru	Destek
0	0.58	0.55	0.56	166
1	0.63	0.66	0.65	197
Doğruluk			0.61	363
Makro Ort.	0.61	0.60	0.60	363
Ağırlıklı Ort.	0.61	0.61	0.61	363

Sürekli 1 yıllık veriler için indikatörlerin önem derecesi.

Sürekli 1 yıllık verilerin KNN yöntemi ile tahmininde sürekli tüm verilerde olduğu gibi Stokastik %K, Larry Williams ve Stokastik %D indikatörleri ön plana çıkmıştır. Ağırlıklar açısından küçük farklarla beraber, genel olarak hem kesikli ve hem de sürekli yapıdaki veriler için tüm veri ile 1 yıllık veriler arası önem derecesi açısından paralellik bulunmaktadır. Sürekli yapıdaki verilerin kesikli yapıdaki gibi net değerler olmamasının KNN yöntemi ile tahminde önem derecesini etkisi fark edilmektedir.

Tablo 17. *Sürekli 1 Yıllık Veri KNN Yöntemi İndikatörlerin Önem Derecesi*

İndikatör	Önem Derecesi
StoK	0.098766
LW	0.092705
StoD	0.051383
upper	0.016672
atr14	0.016672
middle	0.015019
sma14	0.014468
ema14	0.013366
wma14	0.013366
RSI	0.012264
Mom	0.008408
GoogleTrends	0.007857
lower	0.004551
Tweets	0.001245



Şekil 30. *Sürekli 1 Yıllık Veri KNN Yöntemi İndikatörlerin Önem Oranı*

XGBoost Yöntemi İle Analiz

XGBoost Öğrenme Yönteminde en iyi doğruluk sonucuna ulaşabilmek için Rastgele Orman modelinde olduğu gibi 4 farklı derinlik ayarında 10-500 arası Gradyan Ağaç Sayısı, 20 katmanlı Çapraz Doğrulama Yöntemi ile Doğruluk değerlerinin ortalaması üzerinden test edilmiştir. Önem dereceleri hesaplanırken Feature Importance modülünden faydalanılmıştır.

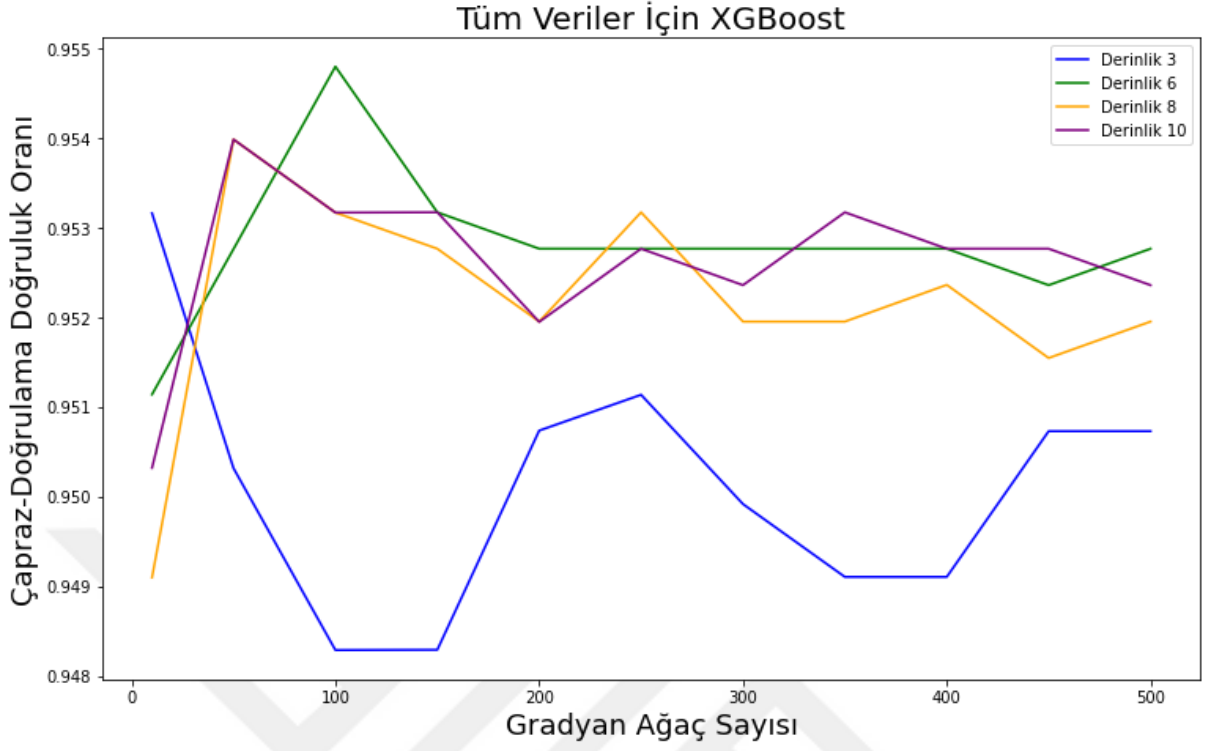
Kesikli verilerin analizi:

Kesikli tüm veri.

Kesikli yapıdaki tüm verilerden oluşan 1. veri seti için 100 gradyan ağaç ve 6 derinlikte en iyi sonuca ulaşılmıştır.

Tablo 18. Kesikli Tüm Veri XGBoost Optimizasyon Ayarları

Gradyan Ağaç Sayısı	Derinlik 3	Derinlik 6	Derinlik 8	Derinlik 10
10	0.953162	0.951140	0.949097	0.950320
50	0.950317	0.952762	0.953985	0.953985
100	0.948287	0.954798	0.953169	0.953169
150	0.948291	0.953172	0.952766	0.953172
200	0.950736	0.952766	0.951953	0.951949
250	0.951136	0.952766	0.953172	0.952766
300	0.949917	0.952766	0.951953	0.952359
350	0.949104	0.952766	0.951953	0.953172
400	0.949104	0.952766	0.952362	0.952766
450	0.950730	0.952359	0.951546	0.952766
500	0.950730	0.952766	0.951953	0.952359



Şekil 31. Kesikli Tüm Veri XGBoost Optimizasyon Ayarları

En iyi sonucu XGBoost'un standart düzen ayarları olan 100 Gradyan ağaç ve 6 derinlik vermiştir.

Ağaç Sayısı = 100, derinlik 6, çapraz doğrulama 20 katman için:

Karışıklık Matrisi:

	Pozitif	Negatif
Pozitif	1057	63
Negatif	48	1287

Sınıflandırma Raporu:

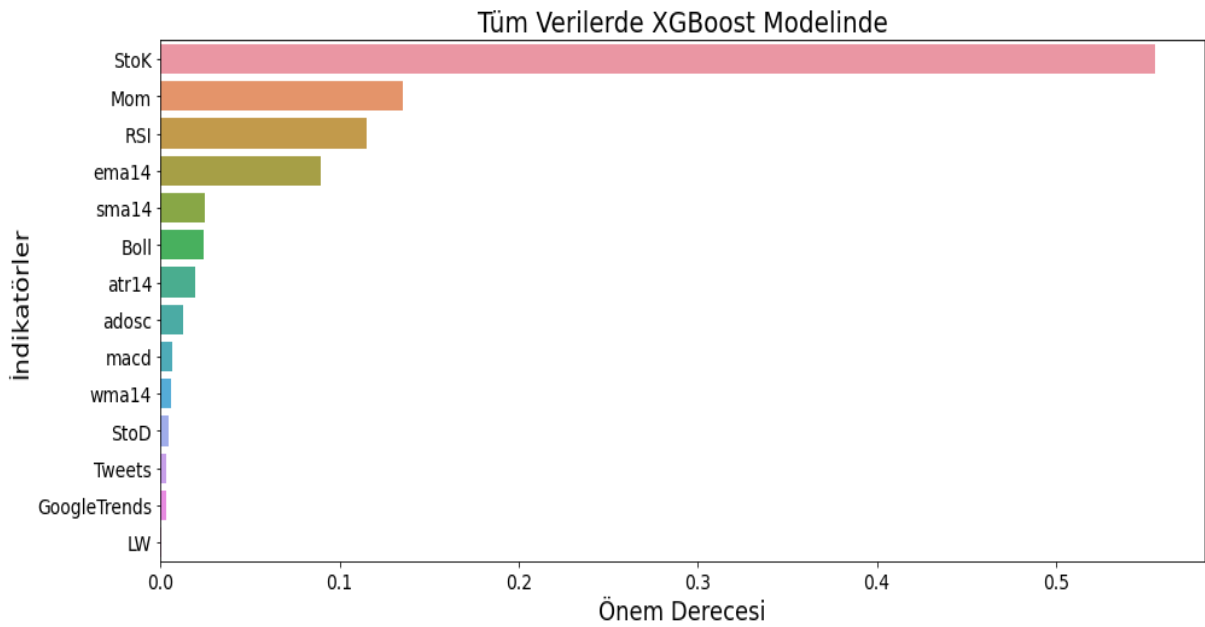
	Kesinlik	Hatırlama	F Skoru	Destek
0	0.96	0.94	0.95	1120
1	0.95	0.96	0.96	1335
Doğruluk			0.95	2455
Makro Ort.	0.95	0.95	0.95	2455
Ağırlıklı Ort.	0.95	0.95	0.95	2455

Kesikli tüm veriler için indikatörlerin önem derecesi.

Kesikli tüm veriler için XGBoost yönteminde Stokastik %K indikatörünün belirgin olarak tahminde ön plana çıktığı görülmektedir. Momentum ve RSI indikatörleri de yine önem açısından ön plandadır.

Tablo 19. *Kesikli Tüm Veri XGBoost Yöntemi İndikatörlerin Önem Derecesi*

İndikatör	Önem Derecesi
StoK	0.555295
Mom	0.135433
RSI	0.115190
ema14	0.089557
sma14	0.024964
Boll	0.023839
atr14	0.019035
adosc	0.012869
macd	0.006381
wma14	0.006077
StoD	0.004638
Tweets	0.003413
GoogleTrends	0.002870
LW	0.000438



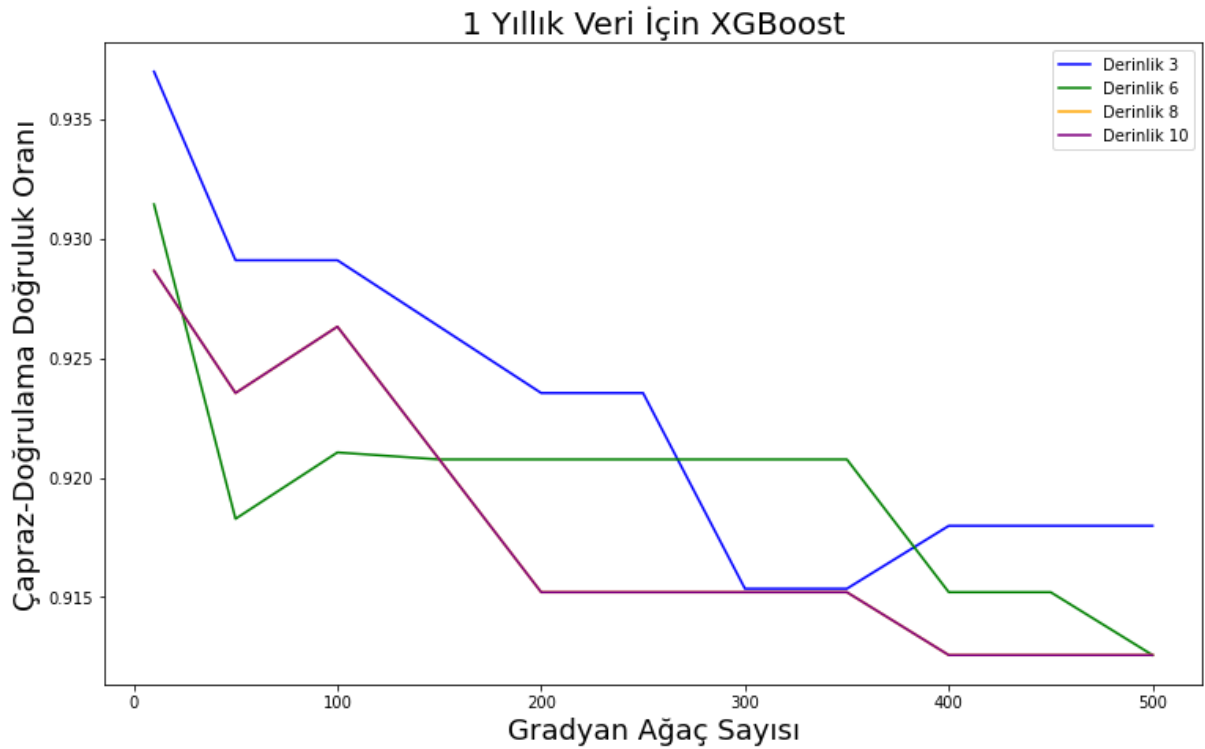
Şekil 32. Kesikli Tüm Veri XGBoost Yöntemi İndikatörlerin Önem Oranları

Kesikli 1 yıllık veri.

Kesikli yapıdaki 1 yıllık verilerden oluşan 2. veri seti için 10 ağaç ve 3 derinlikte en iyi sonuca ulaşılmıştır.

Tablo 20. Kesikli 1 Yıllık Veriler İçin XGBoost Optimizasyon Ayarları

Gradyan Ağaç Sayısı	Derinlik 3	Derinlik 6	Derinlik 8	Derinlik 10
10	0.936988	0.931433	0.928655	0.928655
50	0.929094	0.918275	0.923538	0.923538
100	0.929094	0.921053	0.926316	0.926316
150	0.926316	0.920760	0.920760	0.920760
200	0.923538	0.920760	0.915205	0.915205
250	0.923538	0.920760	0.915205	0.915205
300	0.915351	0.920760	0.915205	0.915205
350	0.915351	0.920760	0.915205	0.915205
400	0.917982	0.915205	0.912573	0.912573
450	0.917982	0.915205	0.912573	0.912573
500	0.917982	0.912573	0.912573	0.912573



Şekil 33. Kesikli 1 Yıllık Veriler İçin XGBoost Optimizasyon Sonuçları

Ağaç Sayısı = 10, derinlik 3, çapraz doğrulama 20 katman için:

Karışıklık Matrisi:

	Pozitif	Negatif
Pozitif	146	20
Negatif	12	185

Sınıflandırma Raporu:

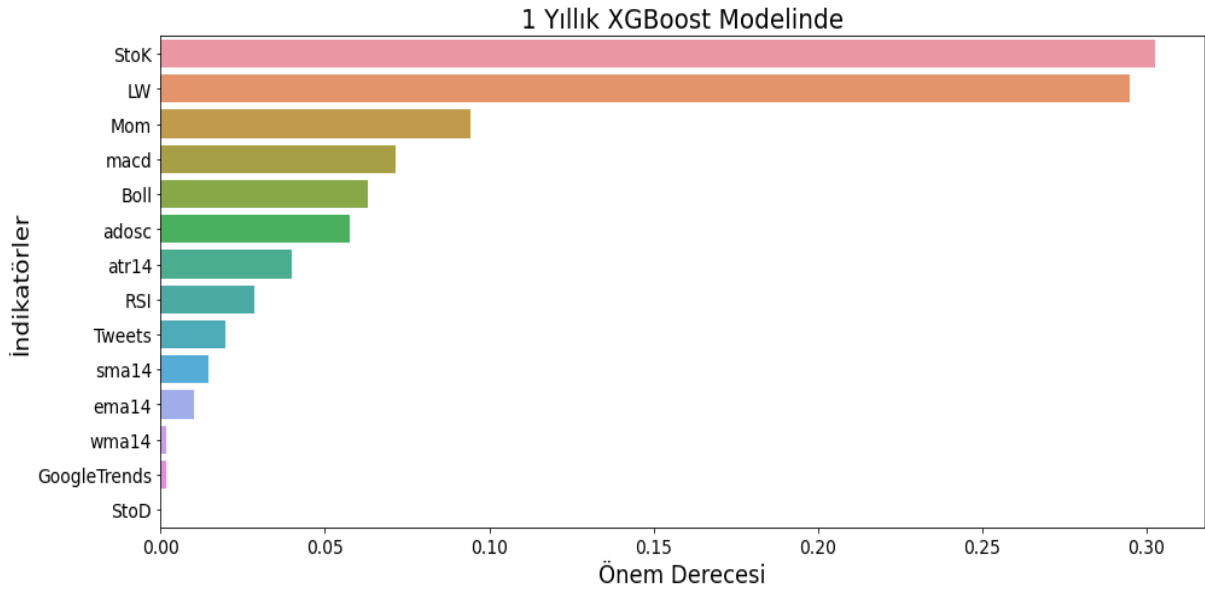
	Kesinlik	Hatırlama	F Skoru	Destek
0	0.92	0.88	0.90	166
1	0.90	0.94	0.92	197
Doğruluk			0.91	363
Makro Ort.	0.91	0.91	0.91	363
Ağırlıklı Ort.	0.91	0.91	0.91	363

Kesikli 1 yıllık veriler için indikatörlerin önem derecesi.

Kesikli 1 yıllık veriler için XGBoost yönteminde tahmin sonuçlarında, kesikli tüm verilerde olduğu gibi Stokastik %K indikatörü en önemli çıkarken, Larry Williams indikatörünün de neredeyse aynı oranda ön plana çıktığı görülmüştür. Kesikli yapıdaki verilerin önem derecesi XGBoost ve Rastgele Orman yöntemleri açısından benzerlik göstermektedir.

Tablo 21. *Kesikli 1 Yıllık Veri XGBoost Yöntemi İndikatörlerin Önem Derecesi*

İndikatör	Önem Derecesi
StoK	0.302629
LW	0.294975
Mom	0.094289
macd	0.071473
Boll	0.063235
adosc	0.057677
atr14	0.039966
RSI	0.028385
Tweets	0.019672
sma14	0.014504
ema14	0.010006
wma14	0.001612
GoogleTrends	0.001578
StoD	0.000000



Şekil 34. Kesikli 1 Yıllık Veri XGBoost Yöntemi İndikatörlerin Önem Oranları

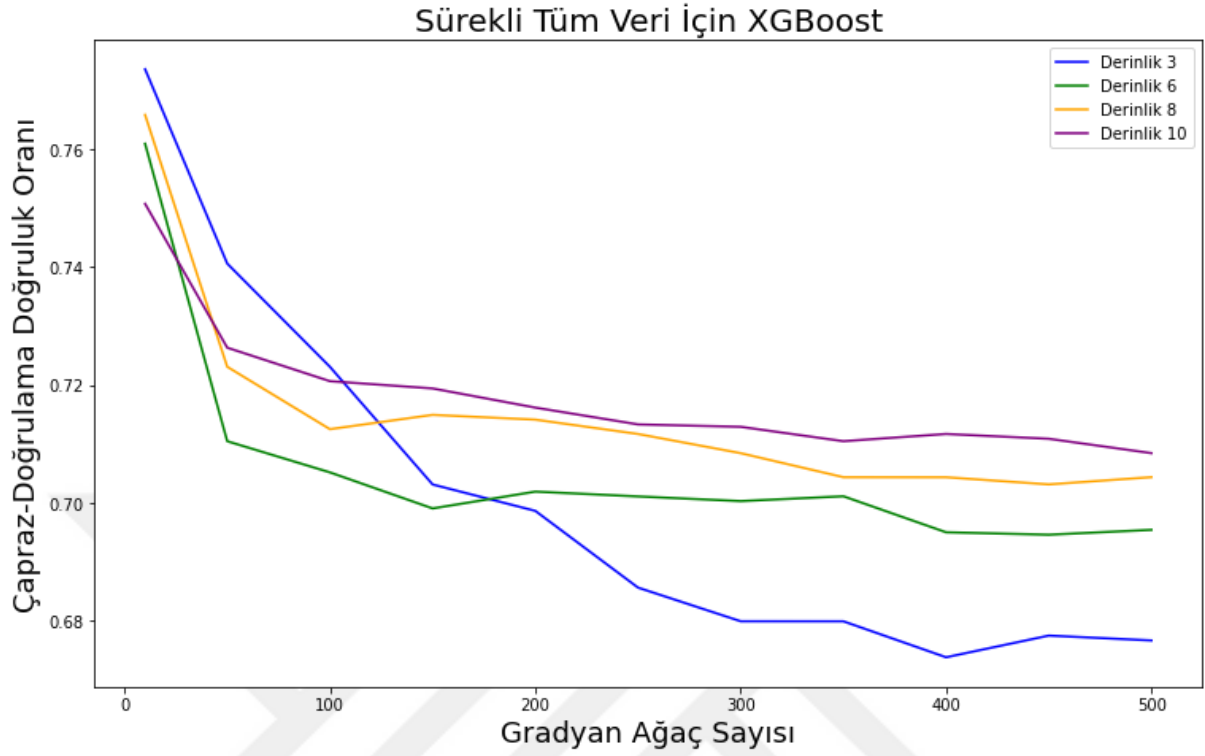
Sürekli verilerin analizi:

Sürekli tüm veri.

Sürekli yapıdaki tüm verilerden oluşan 3. veri seti için 10 ağaç ve 3 derinlikte en iyi sonuca ulaşılmıştır.

Tablo 22. Sürekli Tüm Veri İçin XGBoost Optimizasyon Ayarları

Gradyan Ağaç Sayısı	Derinlik 3	Derinlik 6	Derinlik 8	Derinlik 10
10	0.773547	0.760912	0.765807	0.750736
50	0.740600	0.710489	0.723107	0.726343
100	0.723101	0.705208	0.712528	0.720665
150	0.703159	0.699097	0.714957	0.719436
200	0.698684	0.701949	0.714148	0.716187
250	0.685679	0.701133	0.711715	0.713338
300	0.679975	0.700340	0.708463	0.712931
350	0.679968	0.701160	0.704392	0.710492
400	0.673870	0.695055	0.704385	0.711719
450	0.677546	0.694645	0.703172	0.710906
500	0.676723	0.695468	0.704398	0.708467



Şekil 35. Sürekli Tüm Veri İçin XGBoost Optimizasyon Sonuçları

Ağaç Sayısı = 10, derinlik 3, çapraz doğrulama 20 katman için:

Karışıklık Matrisi:

	Pozitif	Negatif
Pozitif	836	284
Negatif	272	1063

Sınıflandırma Raporu:

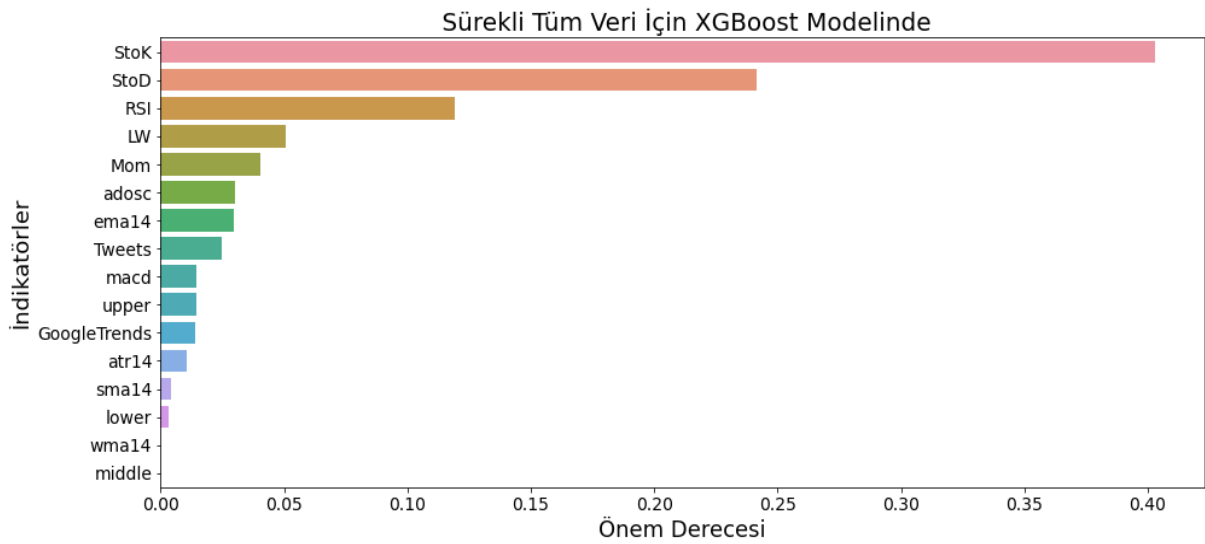
	Kesinlik	Hatırlama	F Skoru	Destek
0	0.75	0.75	0.75	1120
1	0.79	0.80	0.79	1335
Doğruluk			0.77	2455
Makro Ort.	0.77	0.77	0.77	2455
Ağırlıklı Ort.	0.77	0.77	0.77	2455

Sürekli tüm veriler için indikatörlerin önem derecesi.

Sürekli tüm veriler için XGBoost yönteminde Stokastik %K, Stokastik %D ve RSI indikatörlerinin diğer verilere göre tahmin için daha önemli çıktığı görülmüştür. Bu durum kesikli yapıdaki verilerle benzerlik göstermektedir.

Tablo 23. *Sürekli Tüm Veri XGBoost Yöntemi İndikatörlerin Önem Derecesi*

İndikatör	Önem Derecesi
StoK	0.402958
StoD	0.241380
RSI	0.119034
LW	0.050682
Mom	0.040460
adosc	0.030172
ema14	0.029741
Tweets	0.024812
macd	0.014647
upper	0.014313
GoogleTrends	0.014009
atr14	0.010601
sma14	0.004123
lower	0.003067



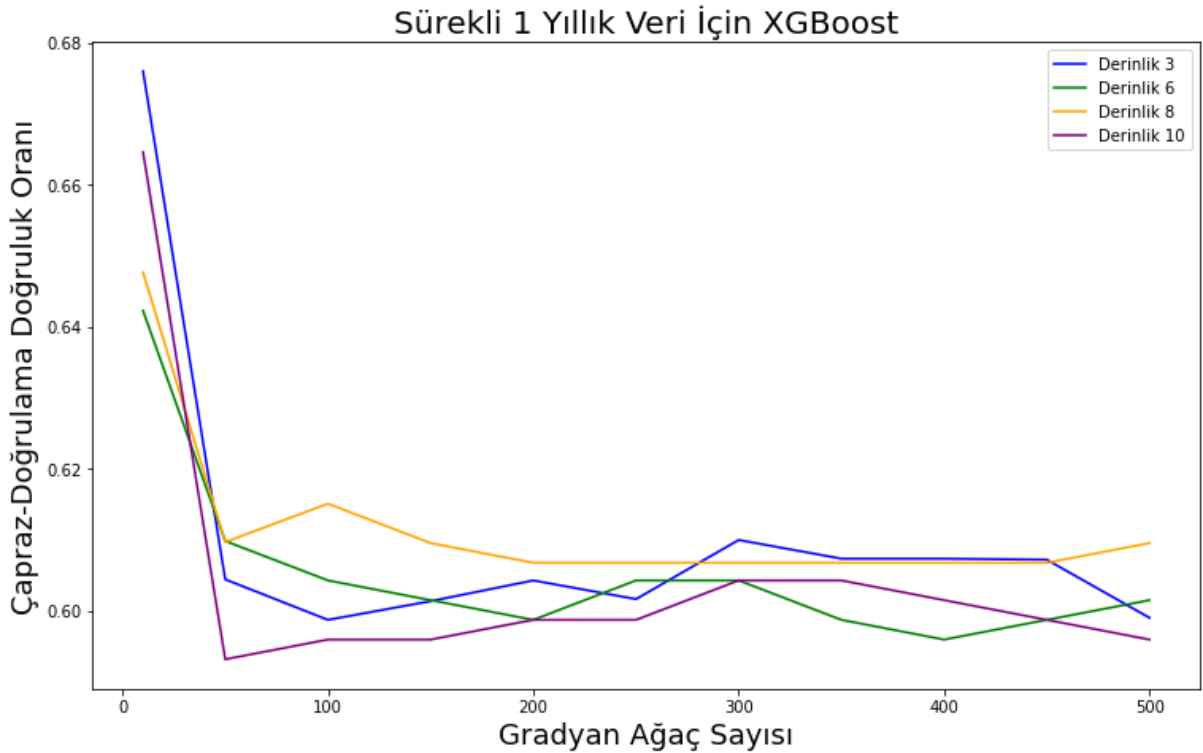
Şekil 36. Sürekli Tüm Veri XGBoost Yöntemi İndikatörlerin Önem Oranları

Sürekli 1 yıllık veri.

Sürekli yapıdaki 1 yıllık verilerden oluşan 4. veri seti için 10 ağaç ve 3 derinlikte en iyi sonuca ulaşılmıştır.

Tablo 24. Sürekli 1 Yıllık Veriler İçin XGBoost Optimizasyon Ayarları

Gradyan Ağaç Sayısı	Derinlik 3	Derinlik 6	Derinlik 8	Derinlik 10
10	0.676023	0.642251	0.647661	0.664620
50	0.604386	0.609795	0.609649	0.593129
100	0.598684	0.604240	0.615058	0.595906
150	0.601316	0.601462	0.609503	0.595906
200	0.604240	0.598684	0.606725	0.598684
250	0.601608	0.604240	0.606725	0.598684
300	0.609942	0.604240	0.606725	0.604240
350	0.607310	0.598684	0.606725	0.604240
400	0.607310	0.595906	0.606725	0.601462
450	0.607164	0.598684	0.606725	0.598684
500	0.598977	0.601462	0.609503	0.595906



Şekil 37. Sürekli tüm veriler için XGBoost optimizasyon ayarları

Ağaç Sayısı = 10, derinlik 3, çapraz doğrulama 20 katman için:

Karışıklık Matrisi:

	Pozitif	Negatif
Pozitif	109	57
Negatif	61	136

Sınıflandırma Raporu:

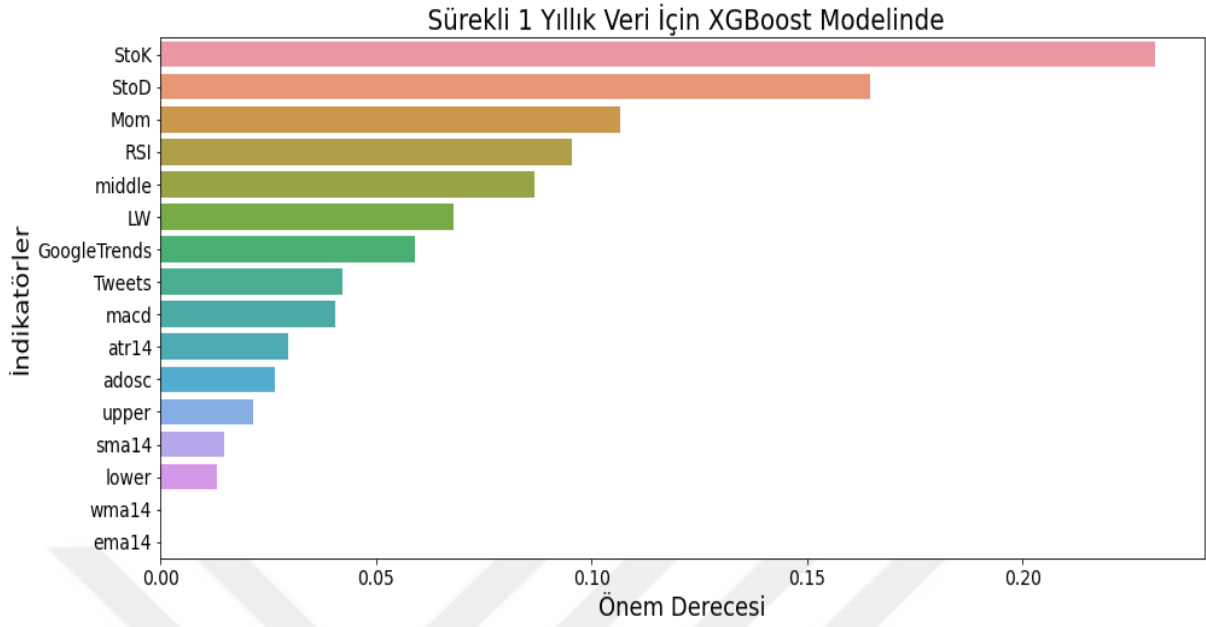
	Kesinlik	Hatırlama	F Skoru	Destek
0	0.64	0.66	0.65	166
1	0.70	0.69	0.70	197
Doğruluk			0.67	363
Makro Ort.	0.67	0.67	0.67	363
Ağırlıklı Ort.	0.68	0.67	0.68	363

Sürekli 1 yıllık veriler için indikatörlerin önem derecesi.

Sürekli 1 yıllık veriler için XGBoost yönteminde sürekli 1 yıllık verilerde olduğu gibi Stokastik %K, Stokastik %D, Momentum ve RSI indikatörlerinin diğer verilere göre tahmin için daha önemli çıktığı görülmüştür. Tüm veriye göre daha dengeli dağılım olan 1 yıllık veriler içinde kesikli yapıdaki gibi Stokastik %K indikatörünün en önemli gösterge çıkması, XGBoost algoritmasının verileri regresyon problemi olarak ele aldığını göstermektedir.

Tablo 25. *Sürekli 1 Yıllık Veri XGBoost Yöntemi İndikatörlerin Önem Derecesi*

İndikatör	Önem Derecesi
StoK	0.230813
StoD	0.164657
Mom	0.106808
RSI	0.095322
middle	0.086882
LW	0.067959
GoogleTrends	0.059107
Tweets	0.042308
macd	0.040471
atr14	0.029719
adosc	0.026562
upper	0.021515
sma14	0.014690
lower	0.013187



Şekil 38. Sürekli 1 yıllık veri XGBoost yöntemi indikatörlerin önem oranları

Destek Vektör Makinası (SVM) Yöntemi ile Analiz

Destek Vektör Makinası Öğrenme Yönteminde en iyi doğruluk sonucuna ulaşabilmek için 6 farklı Polinom Derecesine 0,025-100 arası regülasyon parametresi (C); ve Radyan Bazlı Fonksiyon metoduyla 0.1-5 arası Gamma değeri 20 katmanlı Çapraz Doğrulama Yöntemi ile Doğruluk değerlerinin ortalaması üzerinden test edilmiştir. Önem dereceleri yalnızca doğrusal modellerde hesaplanabildiğinden dolayı, sadece 1 yıllık sürekli veride hesaplanabilmektedir.

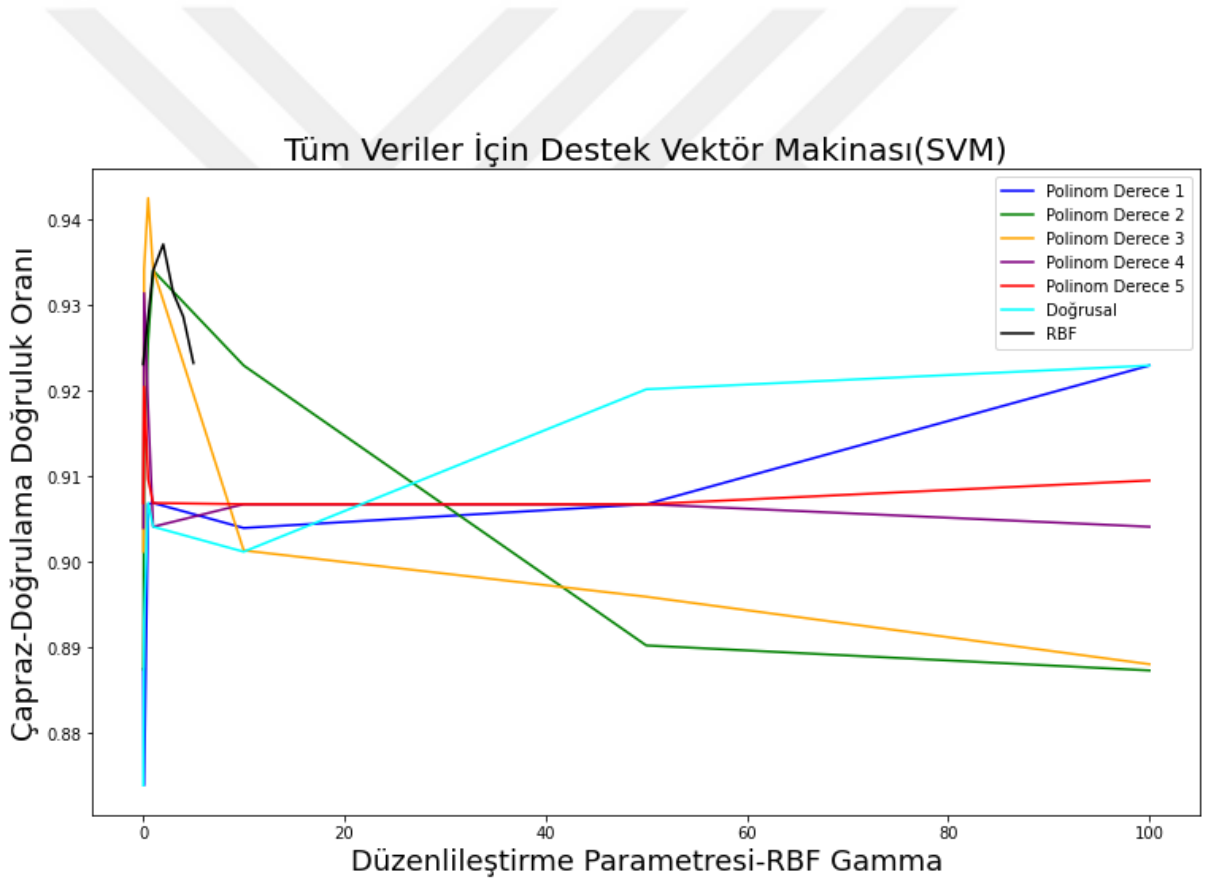
Kesikli verilerin analizi:

Kesikli tüm veri.

Kesikli yapıdaki tüm verilerden oluşan 1. veri seti için 300 ağaç ve 10 derinlikte en iyi sonuca ulaşılmıştır.

Tablo 26. Kesikli Tüm Veriler İçin SVM Optimizasyon Ayarları

C	Derece 1	Derece 2	Derece 3	Derece 4	Derece 5	Doğrusal	RBF	Gamma
0,025	0.923017	0.939294	0.954378	0.950720	0.950720	0.938078	0.953162	0.1
0,1	0.937672	0.949480	0.955604	0.956014	0.952356	0.940121	0.949920	0.5
0,5	0.940121	0.955191	0.956421	0.952759	0.951123	0.940121	0.949510	1
1	0.940121	0.955998	0.957234	0.951120	0.950307	0.940121	0.947471	2
10	0.940121	0.953169	0.945025	0.945015	0.949494	0.942153	0.940550	3
50	0.942153	0.951133	0.945848	0.945425	0.945832	0.942153	0.914074	4
100	0.942153	0.951136	0.947878	0.947065	0.945019	0.942153	0.881887	5



Şekil 39. Kesikli tüm veriler için SVM optimizasyon sonuçları

En iyi sonucu Polinom Derece 3 ve düzenleştirme Parametresi 1 olarak tespit edilmiştir.

Düzenleme = 1, Polinom Derecesi 3, çapraz doğrulama 20 katman için:

Karışıklık Matrisi:

	Pozitif	Negatif
Pozitif	1073	47
Negatif	58	1277

Sınıflandırma Raporu:

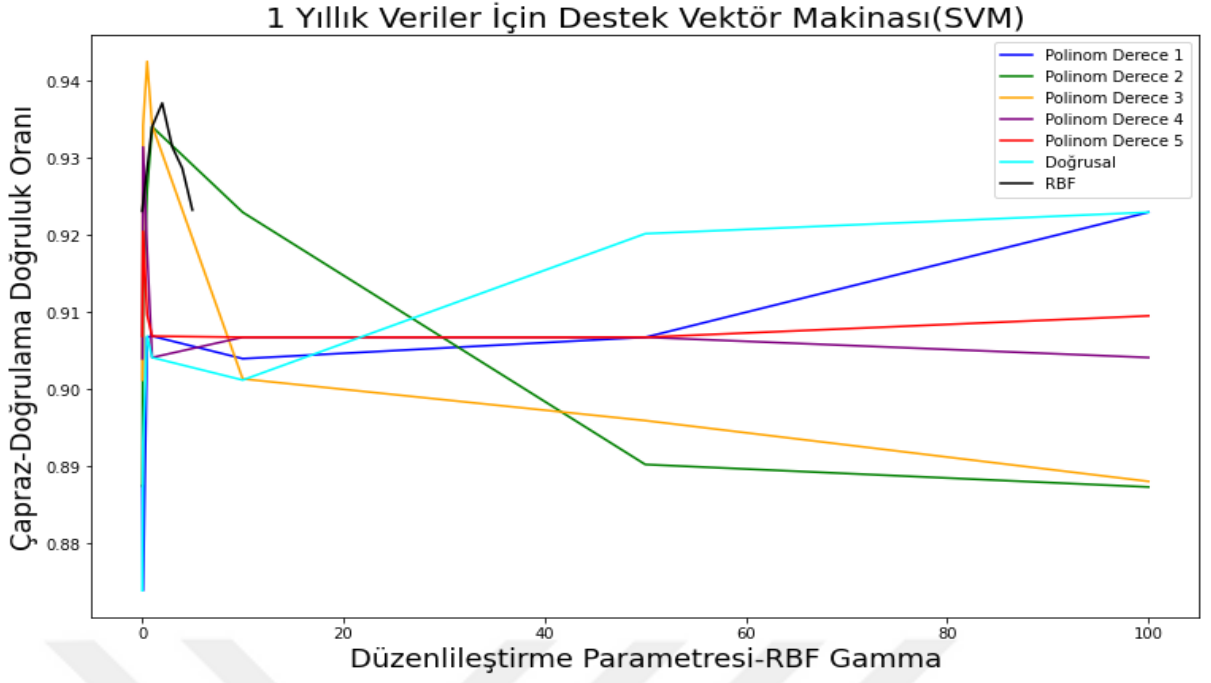
	Keskinlik	Hatırlama	F Skoru	Destek
0	0.95	0.96	0.95	1120
1	0.96	0.96	0.96	1335
Doğruluk			0.95	2455
Makro Ort.	0.96	0.96	0.95	2455
Ağırlıklı Ort.	0.96	0.96	0.95	2455

En iyi oran polinomda yakalandığı için polinom için önem derecesi hesaplanamamaktadır.

Kesikli 1 yıllık veri.

Tablo 27. Kesikli 1 Yıllık Veriler İçin SVM Optimizasyon Ayarları

C	Derece 1	Derece 2	Derece 3	Derece 4	Derece 5	Doğrusal	RBF	Gamma
0,025	0.887427	0.887427	0.901170	0.903947	0.906871	0.873830	0.923099	0,1
0,1	0.873830	0.912281	0.934211	0.931433	0.920468	0.893129	0.934064	0,2
0,5	0.906725	0.925877	0.942544	0.917836	0.909649	0.906871	0.937135	0,5
1	0.906871	0.934064	0.934211	0.904094	0.906871	0.904094	0.931433	0,8
10	0.903947	0.922953	0.901316	0.906725	0.906725	0.901170	0.928655	1
50	0.906725	0.890205	0.895906	0.906725	0.906725	0.920175	0.923246	1,5
100	0.922953	0.887281	0.888012	0.904094	0.909503	0.922953	0.901608	2



Şekil 40. Kesikli 1 yıllık veriler için SVM optimizasyon sonuçları

En iyi sonucu Polinom 3. Derece ve Düzenleştirme Parametresi 0,5 olarak tespit edilmiştir.

Düzenleştirme = 0,5, Polinom Derecesi 3, çapraz doğrulama 20 katman için:

Karışıklık Matrisi:

	Pozitif	Negatif
Pozitif	160	6
Negatif	15	182

Sınıflandırma Raporu:

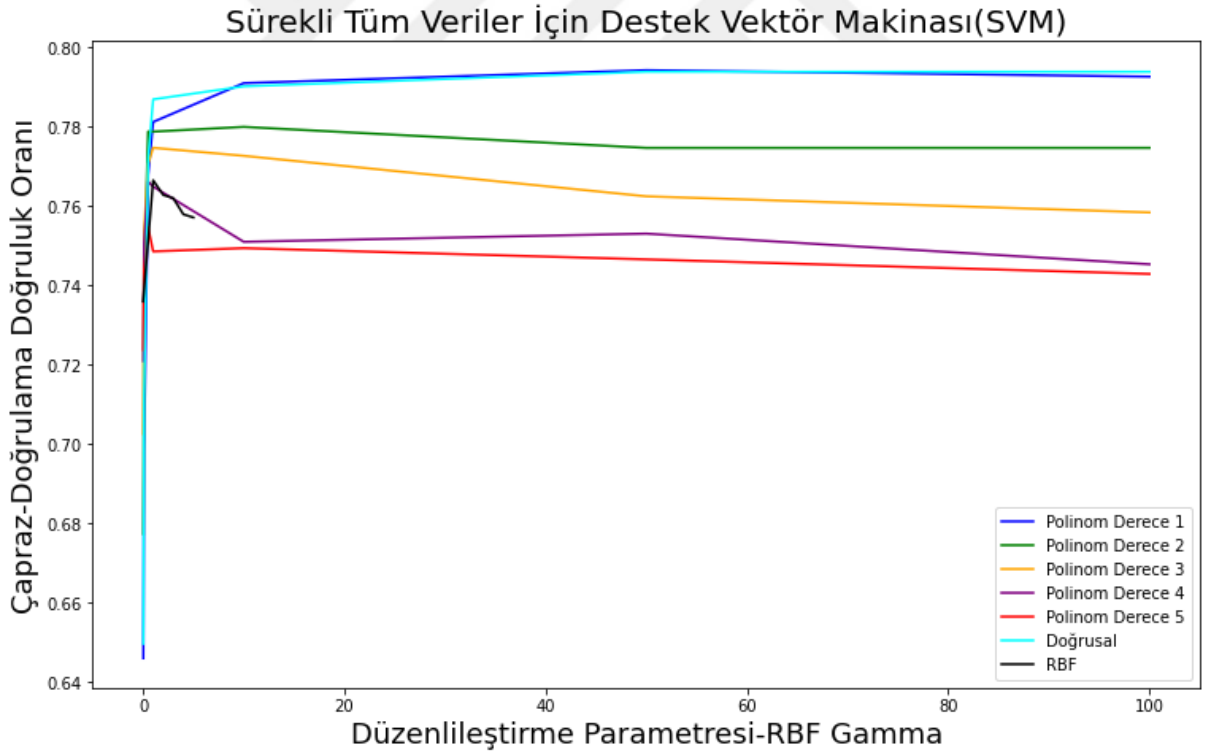
	Kesinlik	Hatırlama	F Skoru	Destek
0	0.91	0.96	0.94	166
1	0.97	0.92	0.95	197
Doğruluk			0.94	363
Makro Ort.	0.94	0.94	0.94	363
Ağırlıklı Ort.	0.94	0.94	0.94	363

Sürekli verilerin analizi.

Sürekli tüm veri.

Tablo 28. Sürekli Tüm Veriler İçin SVM Optimizasyon Ayarları

C	Derece 1	Derece 2	Derece 3	Derece 4	Derece 5	Doğrusal	RBF	Gamma
0,025	0.646018	0.677339	0.702599	0.720935	0.723787	0.649687	0.735946	0.1
0,1	0.698564	0.742490	0.753529	0.750666	0.743733	0.712385	0.766497	0.5
0,5	0.767783	0.778778	0.770602	0.766127	0.754685	0.771835	0.762808	1
1	0.781201	0.778758	0.774690	0.764907	0.748571	0.786899	0.761989	2
10	0.790987	0.779961	0.772668	0.750993	0.749397	0.790174	0.757927	3
50	0.794256	0.774667	0.762458	0.753052	0.746561	0.793852	0.757127	4
100	0.792623	0.774670	0.758397	0.745335	0.742903	0.793846	0.753865	5



Şekil 41. Sürekli tüm veriler için SVM optimizasyon sonuçları

En iyi sonucu Polinom 1. Derece ve Düzenleştirme Parametresi 50 olarak tespit edilmiştir.

Düzenleştirme = 50, Polinom Derecesi 1, çapraz doğrulama 20 katman için:

Karışıklık Matrisi:

	Pozitif	Negatif
Pozitif	822	298
Negatif	207	1128

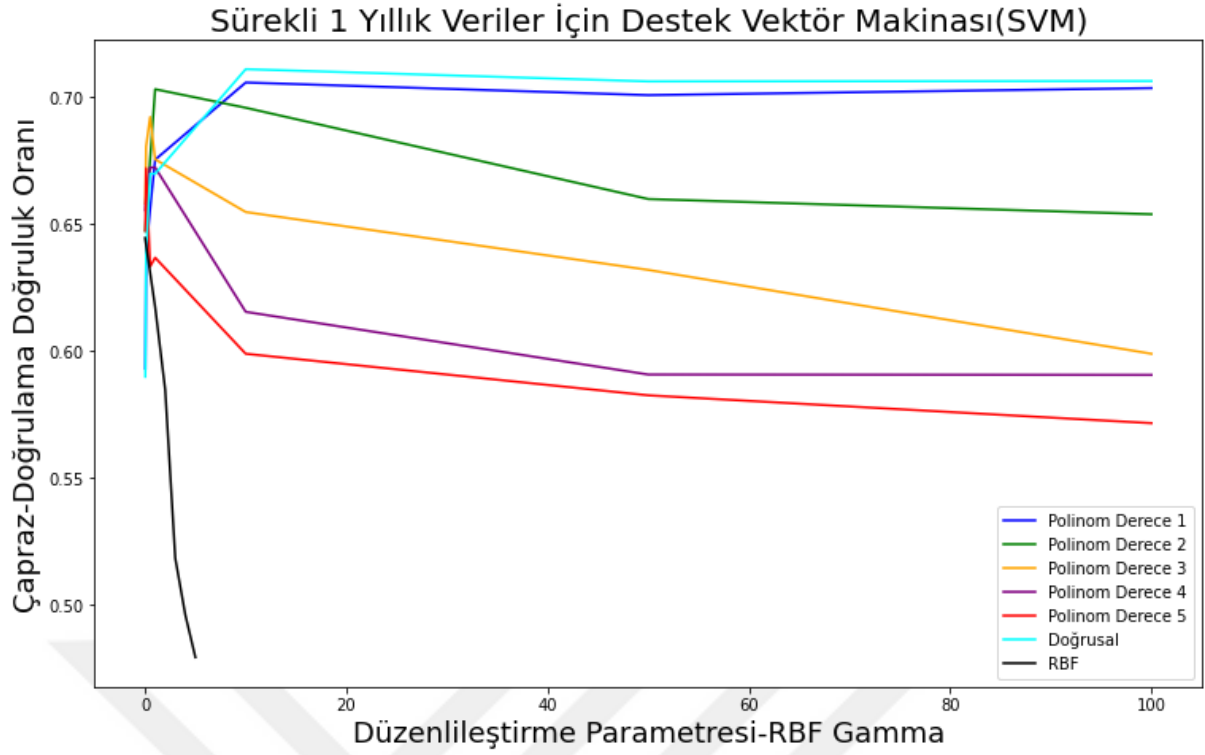
Sınıflandırma Raporu:

	Kesinlik	Hatırlama	F Skoru	Destek
0	0.80	0.73	0.77	1120
1	0.79	0.84	0.82	1335
Doğruluk			0.79	2455
Makro Ort.	0.79	0.79	0.79	2455
Ağırlıklı Ort.	0.79	0.79	0.79	2455

Sürekli 1 yıllık veri.

Tablo 29. Sürekli 1 Yıllık Veriler İçin SVM Optimizasyon Ayarları

C	Derece 1	Derece 2	Derece 3	Derece 4	Derece 5	Doğrusal	RBF	Gamma
0,025	0.592982	0.647222	0.658041	0.655409	0.647222	0.589766	0.644444	0.1
0,1	0.638889	0.652485	0.680409	0.661111	0.671930	0.638889	0.617251	0.5
0,5	0.655848	0.675292	0.692105	0.672222	0.633187	0.669737	0.584649	1
1	0.675292	0.703070	0.675439	0.672222	0.636696	0.669737	0.518275	2
10	0.705702	0.695760	0.654678	0.615351	0.598830	0.710965	0.495906	3
50	0.700731	0.659795	0.631871	0.590643	0.582456	0.706140	0.479240	4
100	0.703509	0.653801	0.598830	0.590497	0.571491	0.706287	0.479240	5



Şekil 42. Sürekli 1 yıllık veriler için SVM optimizasyon sonuçları

En iyi sonucu Doğrusal ve düzenleştirme Parametresi 10 olarak tespit edilmiştir.

Düzenleştirme = 10, Kernel= Doğrusal, çapraz doğrulama 20 katman için:

Karışıklık Matrisi:

	Pozitif	Negatif
Pozitif	105	61
Negatif	44	153

Sınıflandırma Raporu:

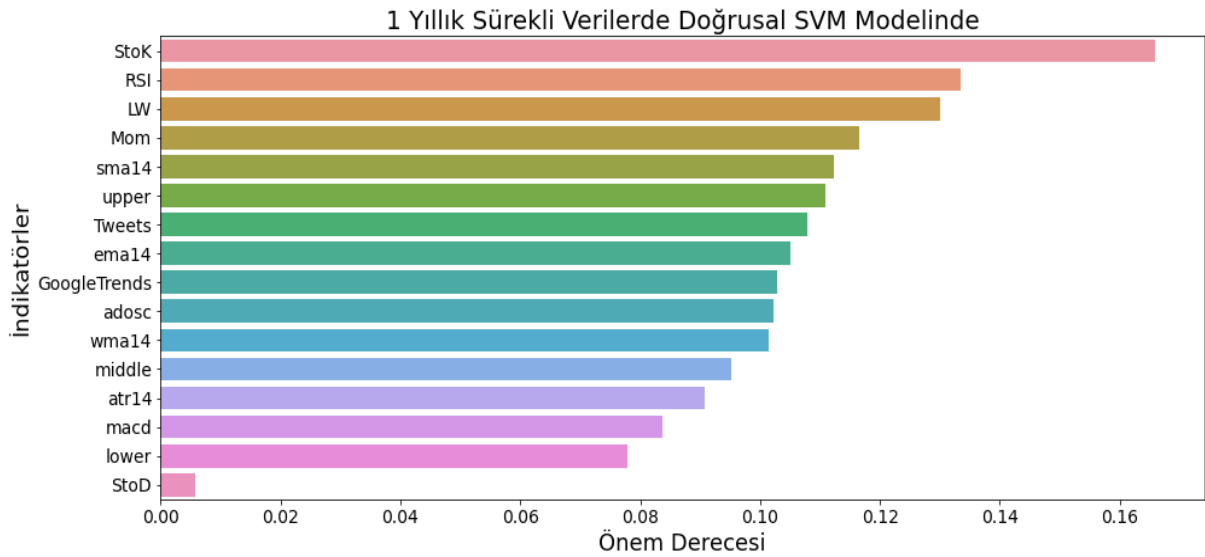
	Kesinlik	Hatırlama	F Skoru	Destek
0	0.70	0.63	0.67	166
1	0.71	0.78	0.74	197
Doğruluk			0.71	363
Makro Ort.	0.71	0.70	0.71	363
Ağırlıklı Ort.	0.71	0.71	0.71	363

Sürekli 1 yıllık veriler için indikatörlerin önem derecesi.

Çözüm doğrusal olduğundan dolayı Asal Problemdeki katsayılar(coef_) metodu ile önem derecesi hesaplanabilmektedir. Bu sonuçlara göre SVM yönteminde RF ve XGBoost yöntemlerine ile benzer sonuçlar elde edilmiştir.

Tablo 30. *Sürekli 1 Yıllık Veri SVM Yöntemi İndikatörlerin Önem Derecesi*

İndikatör	Önem Derecesi
StoK	0.165901
RSI	0.133501
LW	0.130048
Mom	0.116582
sma14	0.112231
upper	0.110891
Tweets	0.107836
ema14	0.104986
GoogleTrends	0.102863
adosc	0.102344
wma14	0.101458
middle	0.095202
atr14	0.090792
macd	0.083644
lower	0.077951
StoD	0.005767



Şekil 43. Sürekli 1 yıllık veri SVM yöntemi indikatörlerin önem oranları

Naive Bayes (NB) Yöntemi İle Analiz

Naive Bayes Makine Öğrenme Yönteminde en iyi doğruluk sonucuna ulaşabilmek için 4 farklı Bayes teoremi yöntemi 20 katmanlı Çapraz Doğrulama Yöntemi ile Doğruluk değerlerinin ortalaması üzerinden test edilmiştir. Önem dereceleri hesaplanırken Bernoulli yöntemi için coef_, Gaussian modeli için theta_ hesaplama yönteminden faydalanılmıştır.

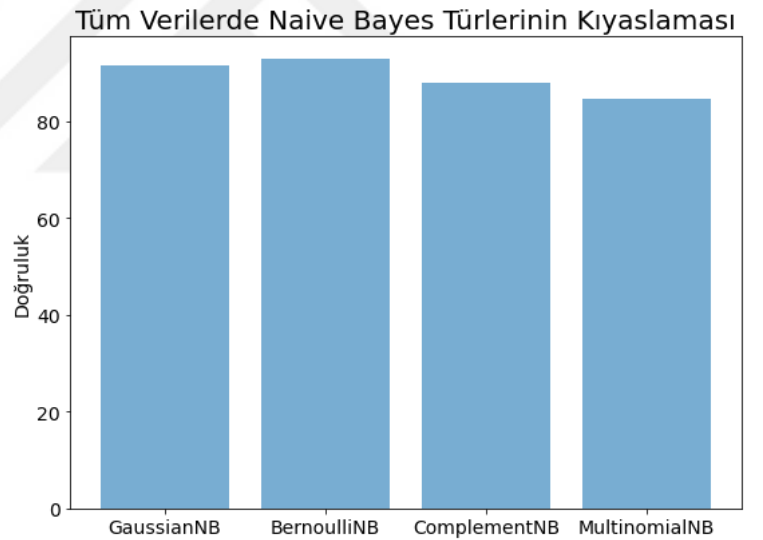
Kesikli verilerin analizi:

Kesikli tüm veriler.

Kesikli yapıdaki tüm verilerden oluşan 1. veri seti için BernoulliNB yönteminde %92,83 doğruluk ile en iyi sonuca ulaşılmıştır.

Tablo 31. Kesikli Tüm Veriler için en iyi NB Yöntemi

Yöntem	Doğruluk
GaussianNB	0.916077
BernoulliNB	0.928312
ComplementNB	0.878225
MultinomialNB	0.846052



En iyi doğruluk oranı %92,8 ile Bernoulli yönteminde görülmüştür. Bu yöneme göre 20 katmanlı çapraz doğruluma için:

Karışıklık Matrisi:

	Pozitif	Negatif
Pozitif	1019	101
Negatif	75	1260

Sınıflandırma Raporu:

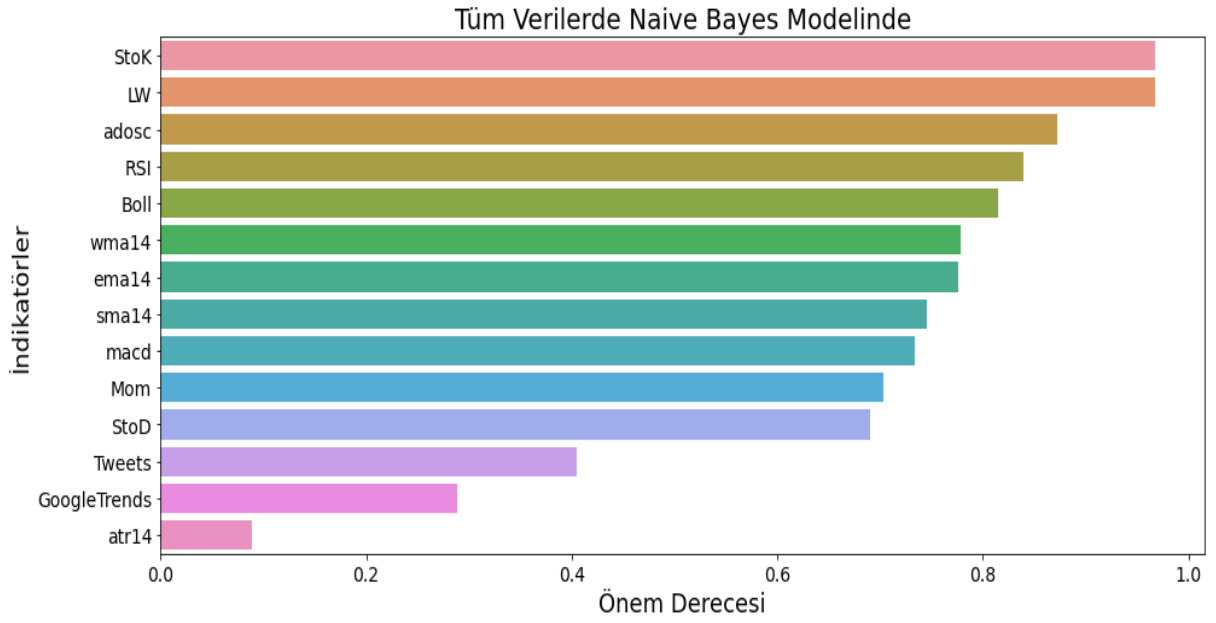
	Kesinlik	Hatırlama	F Skoru	Destek
0	0.93	0.91	0.92	1120
1	0.93	0.94	0.93	1335
Doğruluk			0.93	2455
Makro Ort.	0.93	0.93	0.93	2455
Ağırlıklı Ort.	0.93	0.93	0.93	2455

Kesikli Tüm Veriler İçin İndikatörlerin Önem Derecesi.

Bernoulli yönteminde önem derecesini hesaplayabilmek için coef_ değerlerinden faydalanılmıştır. Negatif değerin yansması alınabilmesi için sonuçlara +1.1 değer eklenmiştir. Kesikli tüm veriler için Naive Bayes yöntemi, diğer yöntemlerden farklı olarak dengeli şekilde önem derecesi dağıtmıştır. Ancak doğruluk oranı diğer yöntemlere göre daha düşük çıkmıştır.

Tablo 32. Kesikli Tüm Veriler NB Yöntemi İndikatörlerin Önem Derecesi

İndikatör	Önem Derecesi
StoK	0.967592
LW	0.967142
adosc	0.872911
RSI	0.839759
Boll	0.814939
wma14	0.778683
ema14	0.776499
sma14	0.746107
macd	0.734280
Mom	0.703125
StoD	0.690192
Tweets	0.405277
GoogleTrends	0.289055
atr14	0.088793



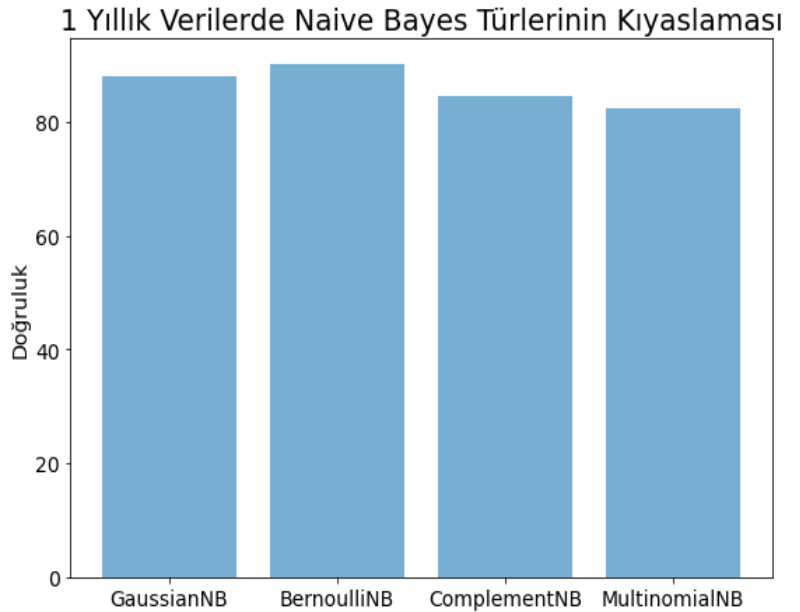
Şekil 44. Kesikli tüm veriler NB yöntemi indikatörlerin önem oranları

Kesikli 1 yıllık veriler.

Kesikli yapıdaki 1 yıllık verilerden oluşan 2. veri seti için BernoulliNB yönteminde %90,13 doğruluk ile en iyi sonuca ulaşılmıştır.

Tablo 33. Kesikli 1 Yıllık Veriler için en iyi NB Yöntemi

Yöntem	Doğruluk
GaussianNB	0.879678
BernoulliNB	0.901316
ComplementNB	0.846637
MultinomialNB	0.824708



Bernoulli yöntemine göre 20 katmanlı çapraz doğruluma için:

Karışıklık Matrisi:

	Pozitif	Negatif
Pozitif	149	17
Negatif	19	178

Sınıflandırma Raporu:

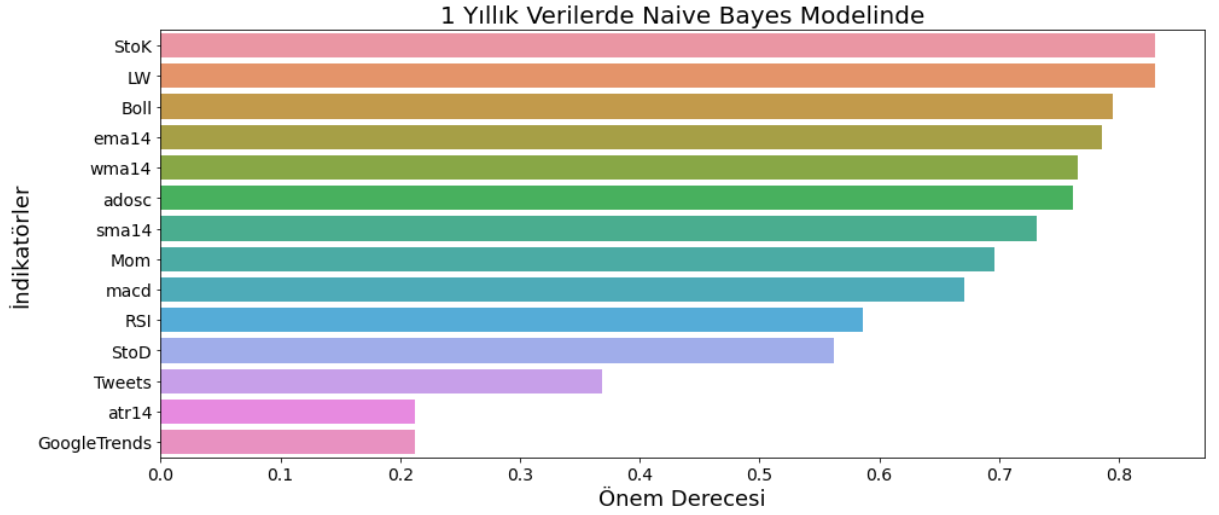
	Kesinlik	Hatırlama	F Skoru	Destek
0	0.89	0.90	0.89	166
1	0.91	0.90	0.91	197
Doğruluk			0.90	363
Makro Ort.	0.90	0.90	0.90	363
Ağırlıklı Ort.	0.90	0.90	0.90	363

Kesikli 1 yıllık veriler için indikatörlerin önem derecesi.

Bernoulli yönteminde önem derecesini hesaplayabilmek için coef_ değerlerinden faydalanılmıştır. Negatif değer yansıması alınabilmesi için sonuçlara +1 değer eklenmiştir. Kesikli 1 yıllık veri için Naive Bayes yöntemi yine kesikli tüm verilerde olduğu gibi önem derecelerini dengeli şekilde dağıtmıştır.

Tablo 34. Kesikli 1 Yıllık Veriler NB Yöntemi İndikatörlerin Önem Derecesi

İndikatör	Önem Derecesi
StoK	0.830253
LW	0.830253
Boll	0.795121
ema14	0.785412
wma14	0.765544
adosc	0.762109
sma14	0.731527
Mom	0.696157
macd	0.670665
RSI	0.586567
StoD	0.562240
Tweets	0.368398
atr14	0.212533
GoogleTrends	0.212533



Şekil 45. Kesikli 1 yıllık veriler NB yöntemi indikatörlerin önem oranları

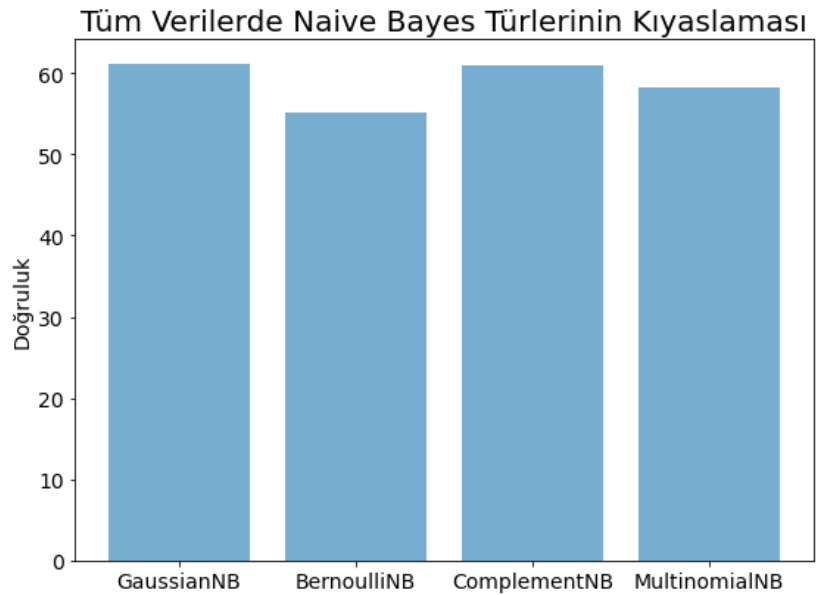
Sürekli verilerin analizi:

Sürekli Tüm Veriler.

Sürekli yapıdaki tüm verilerden oluşan 3. veri seti için GaussianNB yönteminde %61,09 doğruluk ile en iyi sonuca ulaşılmıştır.

Tablo 35. *Sürekli Tüm Veriler için en iyi NB Yöntemi*

Yöntem	Doğruluk
GaussianNB	0.610926
BernoulliNB	0.551916
ComplementNB	0.609303
MultinomialNB	0.581684



Gaussian yöntemine göre 20 katmanlı çapraz doğruluma için:

Karışıklık Matrisi:

	Pozitif	Negatif
Pozitif	591	529
Negatif	426	909

Sınıflandırma Raporu:

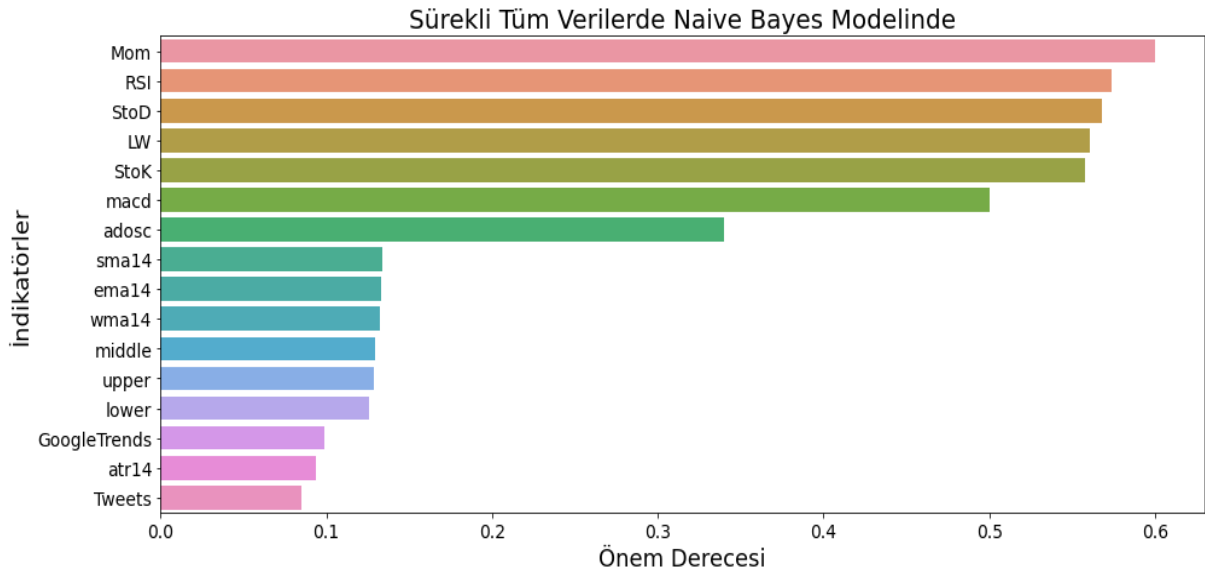
	Kesinlik	Hatırlama	F Skoru	Destek
0	0.58	0.53	0.55	1120
1	0.63	0.68	0.66	1335
Doğruluk			0.61	2455
Makro Ort.	0.61	0.60	0.60	2455
Ağırlıklı Ort.	0.61	0.61	0.61	2455

Sürekli Tüm Veriler İçin İndikatörlerin Önem Derecesi.

Gaussian yönteminde önem derecesini hesaplayabilmek için theta_ değerlerinden faydalanılmıştır. Sürekli yapıdaki verilerde kesikli yapıdan farklı olarak Momentum, RSI, Larry Williams, A/D ve Stokastik %K ve %D indikatörleri ön plana çıkmışken, diğerleri daha geri planda kalmıştır.

Tablo 36. *Sürekli Tüm Veriler NB Yöntemi İndikatörlerin Önem Derecesi*

İndikatör	Önem Derecesi
Mom	0.599976
RSI	0.573419
StoD	0.567845
LW	0.560544
StoK	0.557603
macd	0.500243
adosc	0.340012
sma14	0.133564
ema14	0.132731
wma14	0.132206
middle	0.129677
upper	0.129006
lower	0.125615
GoogleTrends	0.098695
atr14	0.093954
Tweets	0.085114



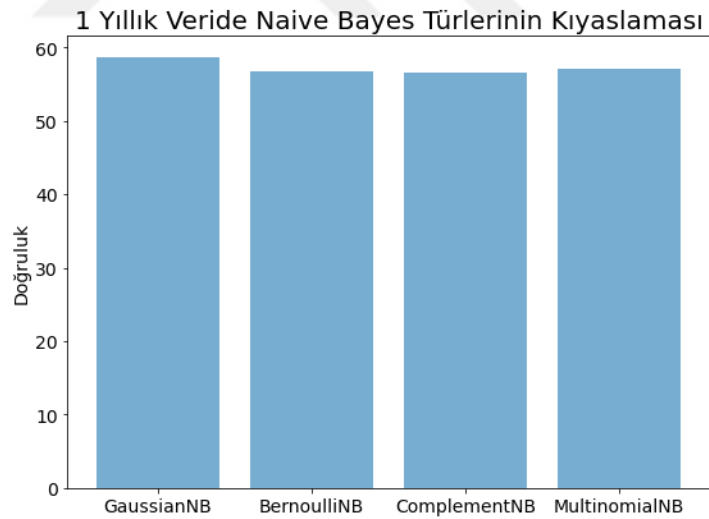
Şekil 46. Sürekli tüm veriler NB yöntemi indikatörlerin önem oranları

Sürekli 1 yıllık veri.

Sürekli yapıdaki 1 Yıllık verilerden oluşan 4. veri seti için GaussianNB yönteminde %58,61 doğruluk ile en iyi sonuca ulaşılmıştır.

Tablo 37. Sürekli 1 Yıllık Veriler için en iyi NB Yöntemi

Yöntem	Doğruluk
GaussianNB	0.586111
BernoulliNB	0.566667
ComplementNB	0.565058
MultinomialNB	0.570614



Gaussian yöntemine göre 20 katmanlı çapraz doğrulama için:

Karışıklık Matrisi:

	Pozitif	Negatif
Pozitif	91	75
Negatif	75	122

Sınıflandırma Raporu:

	Kesinlik	Hatırlama	F Skoru	Destek
0	0.55	0.55	0.55	166
1	0.62	0.62	0.62	197
Doğruluk			0.59	363
Makro Ort.	0.58	0.58	0.58	363
Ağırlıklı Ort.	0.59	0.59	0.59	363

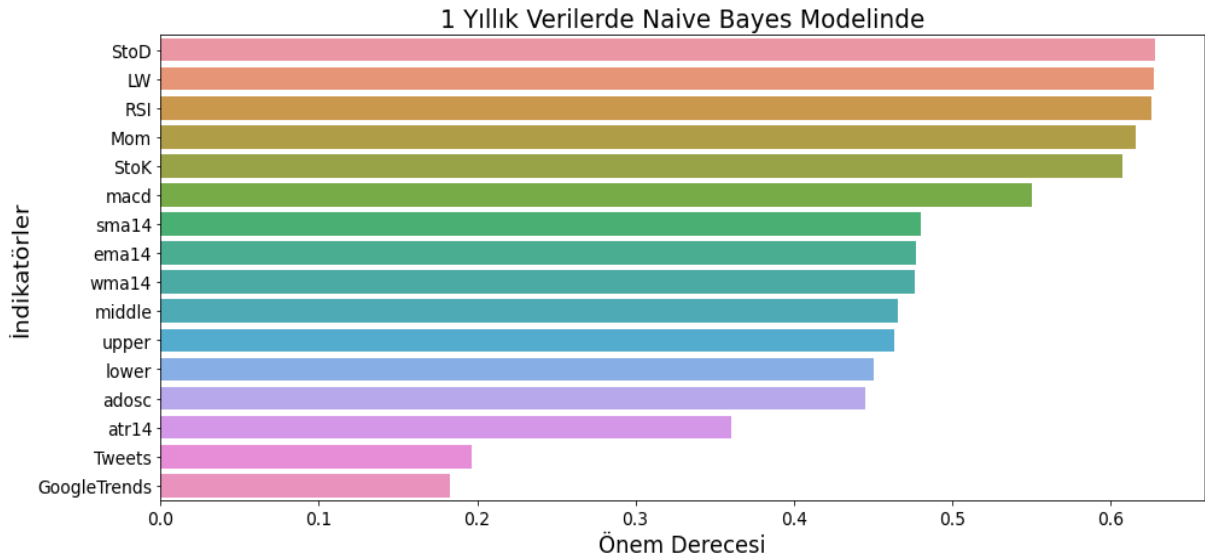
Yapılan önem testlerinin genel sonucu olarak kesikli yapıdaki verilerde Bernoulli yöntemi iyi sonuç verirken, bulanık yapıdaki sürekli veriler için Gaussian yönteminin daha iyi sonuçlar verdiği gözlenmiştir.

Sürekli 1 yıllık veriler için indikatörlerin önem derecesi.

Gaussian yönteminde önem derecesini hesaplayabilmek için theta_ değerlerinden faydalanılmıştır. Sürekli 1 yıllık veriler için önem derecelerinin, sürekli tüm veri ile paralellik gösterdiği ve önem derecesi hesaplamalarında Bernoulli yönteminde dengeli dağılım olduğu, Gaussian yönteminde ise bazı verilerin daha önemli çıktığı sonucuna ulaşılmıştır.

Tablo 38. *Sürekli 1 Yıllık Veriler NB Yöntemi İndikatörlerin Önem Derecesi*

İndikatör	Önem Derecesi
LW	0.171031
StoK	0.137362
StoD	0.106142
RSI	0.075713
Tweets	0.060442
Mom	0.056121
adosc	0.055710
macd	0.055572
sma14	0.046063
middle	0.042231
GoogleTrends	0.041141
ema14	0.034770
atr14	0.030601
wma14	0.030545



Şekil 47. Sürekli 1 yıllık veriler NB yöntemi indikatörlerin önem oranları

Yapay Sinir Ağları (ANN) Yöntemi İle Analiz

Yapay Sinir Ağlarında Katmanlı Çapraz Doğrulama yöntemleriyle ilgili çalışmalar devam ettiğinden dolayı ve yapay sinir ağları ile zaman serisi problemlerinin çözümünde katmanlı çapraz doğrulamaya ilişkin tartışmalardan dolayı; veri seti train_test_split yöntemiyle %75 eğitim %25 test seti olarak ayrılarak kullanılmıştır. Validation(Doğrulama) için test setlerinin %75'i yine train_test_split yöntemiyle rastgele ayrılarak validation için kullanılmıştır.

Yapay sinir ağlarında kaçınılması gereken en büyük problem aşırı uyum gösterme (overfitting) olarak karşımıza çıkmaktadır. Python keras ile oluşturulan formül test edilirken modelin her seferinde tekrar yüklenmediğinde kazandığı öğrenmenin sıfırlanmadığı, eğer bu işlemi her testten önce yapmazsa testi yapan kişinin fark edememesi durumunda modelin sürekli daha iyi sonuç verdiği hatasına sebep olabilecektir. Bu yüzden YSA için en iyi parametreler belirlenirken bu hatadan kaçınmak için işlemler her bir model için sıfırlanarak sonuçlar elde edilmiş, biriktirilen sonuçlar üzerinden tablo oluşturulmuştur. Modeller oluşturulurken batch size 32, Giriş katmanı nöron sayısı girdi sayısı x 2 olarak belirlenmiştir, gizli katman aktivasyon fonksiyonu olarak relu yöntemi seçilmiş ve 20 nöronlu 1 gizli katman kullanılmıştır. Çıkış katmanında sparse_categorical_crossentropy fonksiyonu için 2 katmanlı softmax, binary_crossentropy fonksiyonu için tek katmanlı relu kullanılmıştır.

En iyi optimizer için doğruluk:

En iyi optimizer belirlenirken tüm optimizerlar için sparse_categorical_crossentropy ve binary_crossentropy loss yöntemleri ayrı ayrı test edilmiştir. Test sırasında Optimizerların tamamında epoch 20, learning rate 0,001, epsilon=1e-07 olarak belirlenmiştir. Adam, Adamax, Nadam için beta_1=0.9, beta_2=0.999, RMSprop momentum 0.0 olarak belirlenmiştir.

Tablo 39. ANN Yöntemi En İyi Optimizer Ayarları

	Tüm Kesikli Verilerde	1 Yıllık Kesikli Verilerde	Tüm Sürekli Verilerde	1 Yıllık Sürekli Verilerde
Adadelta_bc	0,445471	0,460784	0,468116	0,397059
Adadelta_scc	0,598913	0,548529	0,436739	0,566667
Adagrad_bc	0,855652	0,708088	0,468116	0,595098
Adagrad_scc	0,699601	0,603676	0,505399	0,562500
Adam_bc	0,452174	0,788971	0,708152	0,639706
Adam_scc	0,932065	0,845098	0,723043	0,459314
Adamax_bc	0,935036	0,840441	0,665399	0,653922
Adamax_scc	0,895507	0,757843	0,659167	0,593873
Ftrl_bc	0,583188	0,480147	0,468116	0,438725
Ftrl_scc	0,530399	0,547794	0,543225	0,442892
Nadam_bc	0,923370	0,848775	0,727210	0,666422
Nadam_scc	0,939457	0,864951	0,716159	0,437990
RMSprop_bc	0,946993	0,889706	0,682609	0,689706
RMSprop_scc	0,930652	0,852451	0,732645	0,485294
SGD_bc	0,860797	0,829167	0,594130	0,663480
SGD_scc	0,896087	0,796324	0,620507	0,447794

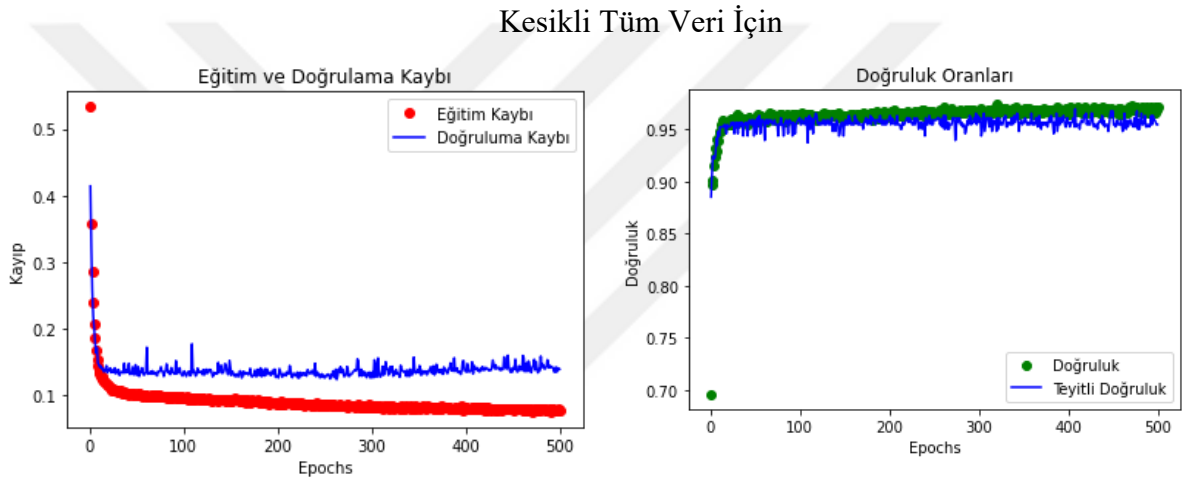
Sonraki aşamalarda Kesikli yapı için RMSprop optimizerı binary_crossentropy loss fonksiyonu ile, Sürekli verilerde yine RMSprop optimizerı en iyi sonucu vermiş, tüm sürekli veriler için sparse_categorical_crossentropy loss fonksiyonu ile, 1 Yıllık Sürekli Veride ise binary_crossentropy loss fonksiyonu ile kullanılmıştır.

En iyi epoch için doğruluk:

Tablo 40. ANN Yöntemi En İyi Epoch Ayarları

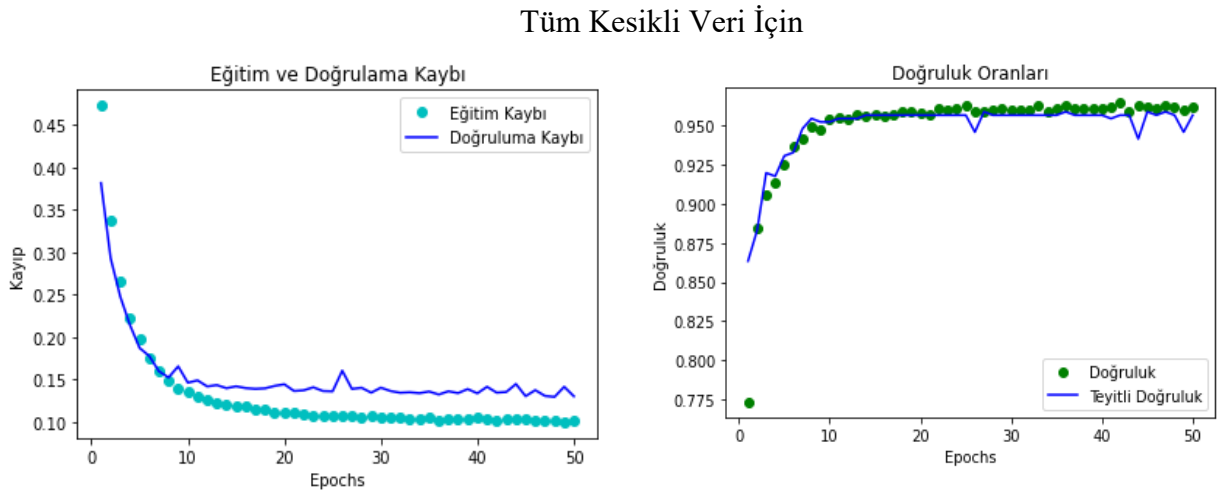
	Tüm Kesikli Verilerde	1 Yıllık Kesikli Verilerde	Tüm Sürekli Verilerde	1 Yıllık Sürekli Verilerde
20	0,949529	0,824510	0,732572	0,625245
50	0,956551	0,906961	0,770072	0,697451
100	0,957413	0,955441	0,786899	0,694118
200	0,964036	0,966765	0,800217	0,744828
500	0,968293	0,980529	0,811632	0,715657

Genel olarak epoch arttıkça doğruluk oranı artmış gibi görünmekte, ancak bu noktadan sonra aşırı uyumun(overfitting) başladığı noktayı gözlemlemek gerekecektir.



Şekil 48. Kesikli tüm veri için kayıp ve doğruluk oranları (500 Ölçek)

Tüm kesikli veri seti için için 50 epoch'tan sonra eğitim ve doğrulama kaybı belirgin şekilde ayrılmaya başlamaktadır. Ancak doğrulukta büyük bir değişim yaşanmamıştır.



Şekil 49. Kesikli tüm veri için kayıp ve doğruluk oranları (50 Ölçek)

50 epochtan sonra teyitlenmiş doğruluk oranlarında dalgalanma artmaya başlamıştır. Bu sebeple 50 Epoch ideal belirlenmiştir.

Optimizer = RMSprop, loss: binary_crossentropy, epoch= 50 için :

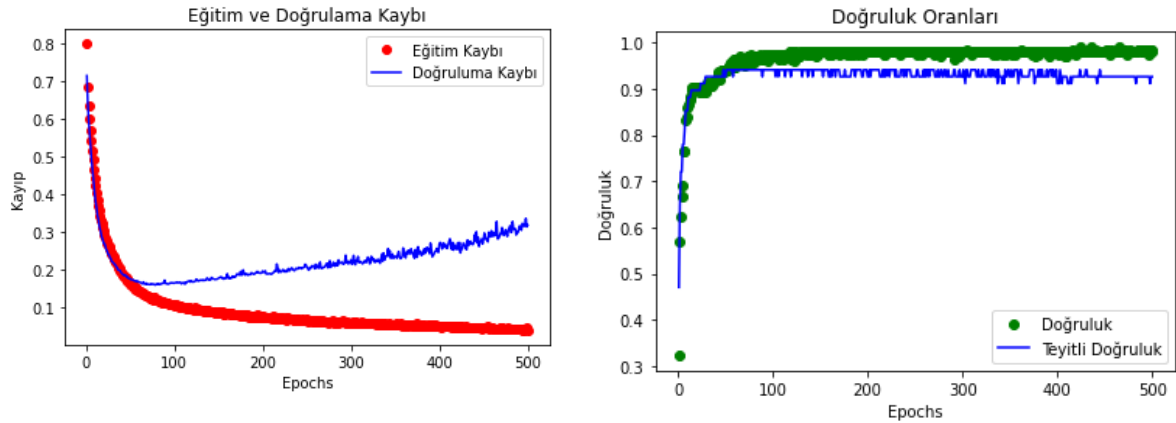
Karışıklık Matrisi:

	Pozitif	Negatif
Pozitif	255	15
Negatif	17	327

Sınıflandırma Raporu:

	Kesinlik	Hatırlama	F Skoru	Destek
0	0.94	0.94	0.94	270
1	0.96	0.95	0.95	344
Doğruluk			0.95	614
Makro Ort.	0.95	0.95	0.95	614
Ağırlıklı Ort.	0.95	0.95	0.95	614

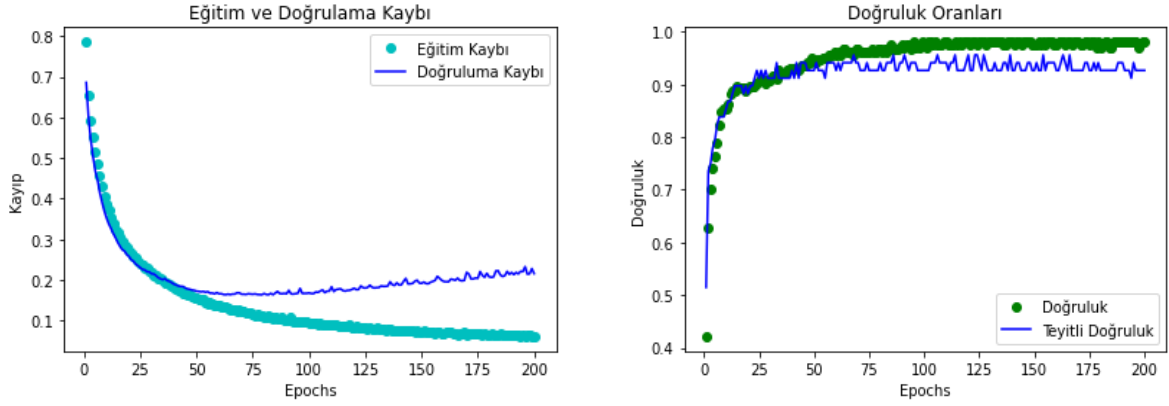
Kesikli 1 Yıllık Veri İçin



Şekil 50. Kesikli 1 yıllık veri için kayıp ve doğruluk oranları (500 Ölçek)

Tüm kesikli veri seti için için 100 epoch'tan sonra eğitim ve doğrulama kaybı belirgin şekilde ayrılmaya başlamaktadır.

Kesikli 1 Yıllık Veri İçin



Şekil 51. Kesikli 1 yıllık veri için kayıp ve doğruluk oranları (200 Ölçek)

Daha net görülebilmesi adına 200 epoch üzerinden bakıldığında, ayrışmanın 100 Epochdan sonra yön değiştirdiği görülmektedir. 200 epoch için doğrulukta azda olsa düşme meydana gelmiştir. Bu şekilde 100 Epoch ideal belirlenmiştir.

Optimizer = RMSprop, loss: binary_crossentropy, epoch= 100 için :

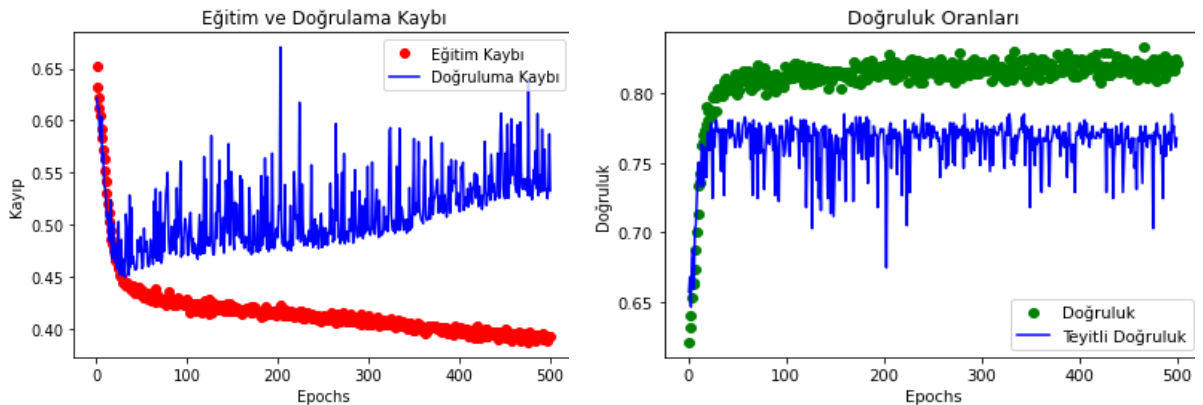
Karışıklık Matrisi:

	Pozitif	Negatif
Pozitif	34	7
Negatif	4	47

Sınıflandırma Raporu:

	Keskinlik	Hatırlama	F Skoru	Destek
0	0.92	0.83	0.87	41
1	0.87	0.94	0.90	50
Doğruluk			0.89	91
Makro Ort.	0.89	0.88	0.89	91
Ağırlıklı Ort.	0.89	0.89	0.89	91

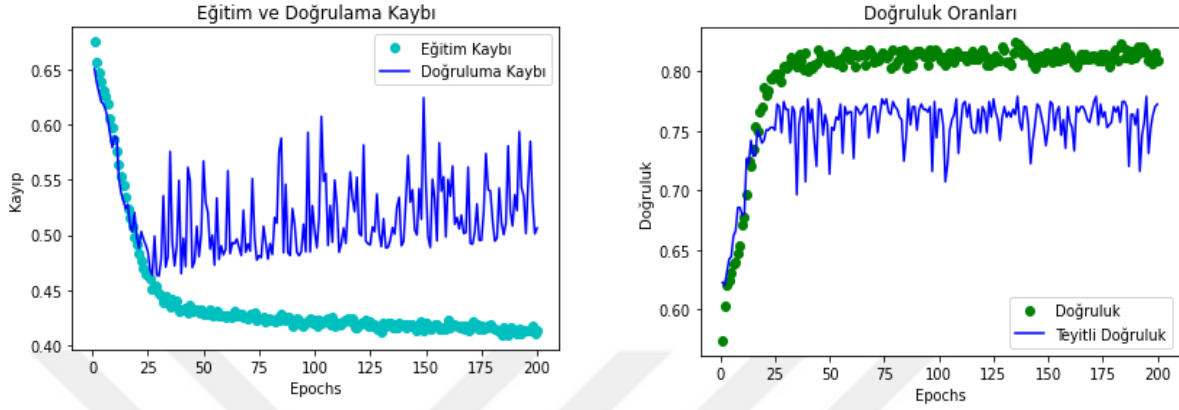
Sürekli Tüm Veri İçin



Şekil 52. Sürekli tüm veri için kayıp ve doğruluk oranları (500 Ölçek)

Sürekli verilerde bulanık olan veri yapısı değişikliği grafiğe de yansımıştır. Doğruluk oranı üst epochlarda en üst seviyeye çıkmış olsa da, 50 epoch sonrasında ayrışma oranı artmıştır, doğruluk yükselmezken teyitli doğruluktaki sapma oranları artmıştır. Bu sebeple 50 epoch tercih edilmiştir.

Sürekli Tüm Veri İçin



Şekil 53. Sürekli tüm veri için kayıp ve doğruluk oranları(200 Ölçek)

Optimizer = RMSprop, loss: sparse_categorical_crossentropy, epoch= 50 için :

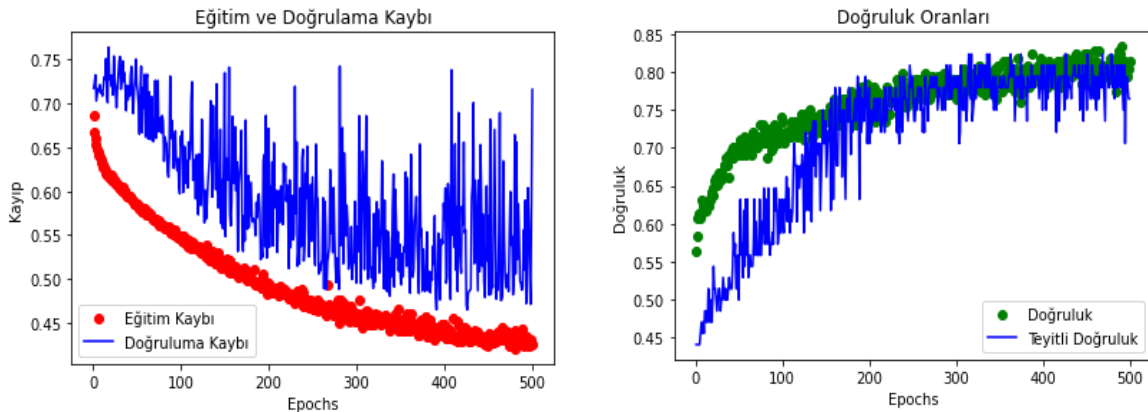
Karışıklık Matrisi:

	Pozitif	Negatif
Pozitif	195	82
Negatif	38	299

Sınıflandırma Raporu:

	Kesinlik	Hatırlama	F Skoru	Destek
0	0.84	0.70	0.76	277
1	0.78	0.89	0.83	337
Doğruluk			0.80	614
Makro Ort.	0.81	0.80	0.80	614
Ağırlıklı Ort.	0.81	0.80	0.80	614

Sürekli 1 Yıllık Veri İçin



Şekil 54. Sürekli 1 yıllık veri için kayıp ve doğruluk oranları (500 Ölçek)

1 Yıllık Sürekli veriler içinde 500 epoch’a kadar bir ayrışma gözlenmemiştir. Bu sebeple 500 epoch tercih edilmiştir.

Optimizer = RMSprop, loss: binary_crossentropy, epoch= 500 için :

Karışıklık Matrisi:

	Pozitif	Negatif
Pozitif	31	12
Negatif	8	40

Sınıflandırma Raporu:

	Kesinlik	Hatırlama	F Skoru	Destek
0	0.79	0.72	0.76	43
1	0.77	0.83	0.80	48
Doğruluk			0.78	91
Makro Ort.	0.78	0.78	0.78	91
Ağırlıklı Ort.	0.78	0.78	0.78	91

Uzun-Kısa Vadeli Bellek (LSTM) Yöntemi İle Analiz

Uzun-Kısa Vadeli Bellek(LSTM) yöntemi için veri seti train_test_split yöntemiyle %75 eğitim %25 test seti olarak ayrılarak kullanılmıştır. Doğrulama seti validation_split yöntemiyle veri setinin %25’lik kısmı olarak belirlenmiştir. LSYM yönteminde ANN’de olduğu gibi en iyi parametreler belirlenirken bu hatadan kaçınmak için işlemler her bir model için sıfırlanarak sonuçlar elde edilmiş, biriktirilen sonuçlar üzerinden tablo oluşturulmuştur. Modeller oluşturulurken batch size 32, Toplamda 4 katmanlı kurulan modelin giriş katmanı 50 unit olarak belirlenmiştir ve her katmandan sonra 0.2 dropout regülasyonu uygulanmıştır. Ara katmanların tamamında yine Unit değeri 50 olarak belirlenmiştir. Çıkış katmanında sparse_categorical_crossentropy fonksiyonu için 2 katmanlı softmax, binary_crossentropy fonksiyonu için tek katmanlı relu kullanılmıştır.

En iyi optimizer için doğruluk:

En iyi optimizer belirlenirken tüm optimizerlar için sparse_categorical_crossentropy ve binary_crossentropy loss yöntemleri ayrı ayrı test edilmiştir. Test sırasında Optimizerların tamamında epoch 20, learning rate 0,01, epsilon=1e-07 olarak belirlenmiştir. Adam, Adamax, Nadam için beta_1=0.9, beta_2=0.999, RMSprop momentum 0.0 olarak belirlenmiştir.

Tablo 41. LSTM Yöntemi En İyi Optimizer Ayarları

	Tüm Kesikli Verilerde	1 Yıllık Kesikli Verilerde	Tüm Sürekli Verilerde	1 Yıllık Sürekli Verilerde
Adadelta_bc	0,450000	0,426471	0,453623	0,485294
Adadelta_scc	0,547790	0,486520	0,540217	0,507353
Adagrad_bc	0,450000	0,426471	0,453623	0,485294
Adagrad_scc	0,549964	0,574510	0,546304	0,506863
Adam_bc	0,878696	0,702451	0,596667	0,521078
Adam_scc	0,916558	0,741422	0,669058	0,544118
Adamax_bc	0,823007	0,497059	0,529348	0,455882
Adamax_scc	0,867717	0,562010	0,609058	0,542647
Ftrl_bc	0,450000	0,426471	0,453623	0,485294
Ftrl_scc	0,549891	0,569853	0,546123	0,509559
Nadam_bc	0,912935	0,673529	0,598768	0,502451
Nadam_scc	0,933623	0,781373	0,718659	0,583333
RMSprop_bc	0,910507	0,749755	0,616920	0,520588
RMSprop_scc	0,922029	0,795588	0,636812	0,554902
SGD_bc	0,549928	0,572059	0,546304	0,512745
SGD_scc	0,549601	0,571324	0,546268	0,513726

Sonraki aşamalarda Kesikli tüm verilerden oluşan set için için Nadam optimizörü sparse_categorical_crossentropy loss fonksiyonu, 1 yıllık kesikli veriler için ise RMSprop optimizörü sparse_categorical_crossentropy loss fonksiyonu ile, Sürekli verilerde ise Nadam optimizörü optimizörü sparse_categorical_crossentropy loss fonksiyonu en iyi sonucu verdiği için bu metodlar kullanılmıştır.

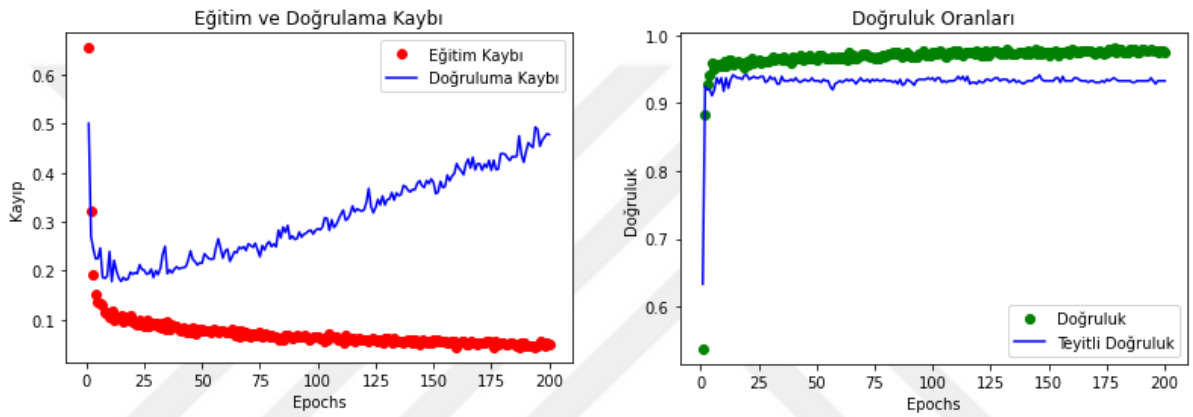
En iyi epoch için doğruluk:

Tablo 42. LSTM Yöntemi En İyi Epoch Ayarları

	Tüm Kesikli Verilerde	1 Yıllık Kesikli Verilerde	Tüm Sürekli Verilerde	1 Yıllık Sürekli Verilerde
20	0,924022	0,783578	0,691159	0,570588
50	0,942957	0,881471	0,741681	0,635784
100	0,960732	0,910931	0,788370	0,664804
200	0,968199	0,933333	0,804743	0,706838
500	0,964004	0,961892	0,798568	0,776608

Epoch değeri arttıkça doğruluk oranı artmış gibi görünmekte, ancak bu noktadan sonra aşırı uyum (overfitting) hesaplamak gerekecektir.

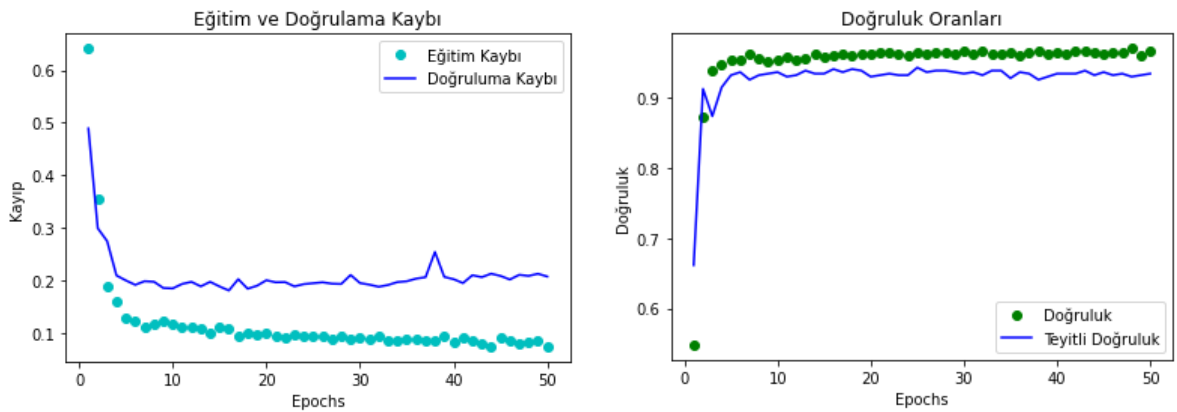
Kesikli Tüm Veriler İçin



Şekil 55. Kesikli Tüm Veri için kayıp ve doğruluk oranları(500 Ölçek)

Tüm kesikli veri seti için 20 epoch'tan sonra eğitim ve doğrulama kaybı belirgin şekilde ayrılmaya başlamaktadır. Ancak doğrulukta büyük bir değişim yaşanmamıştır.

Kesikli Tüm Veri İçin



Şekil 56. Kesikli Tüm Veri için kayıp ve doğruluk oranları(50 Ölçek)

100 epoch'tan sonra teyit edilmiş doğruluk oranlarında dalgalanma artmaya başlamıştır. Bu sebeple 20 epoch ideal belirlenmiştir.

Optimizer = Nadam, loss: sparse_categorical_crossentropy,, epoch= 20 için :

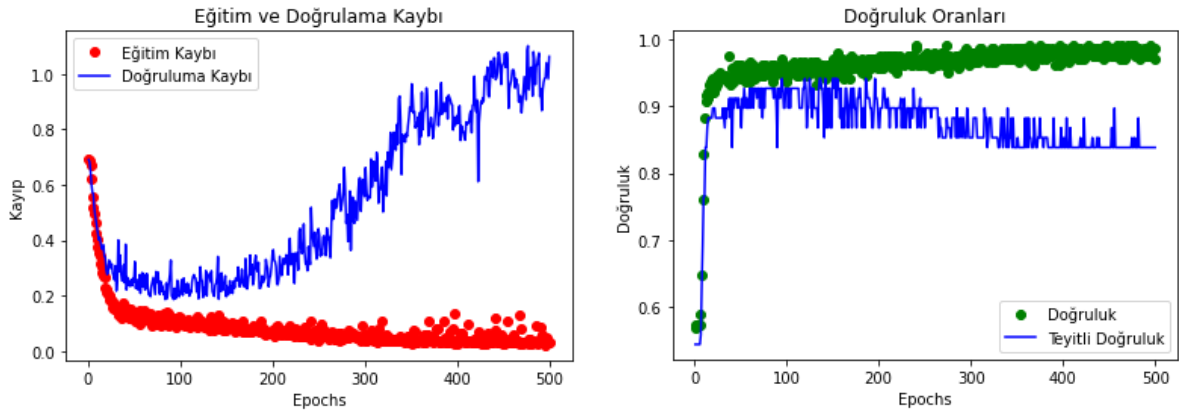
Karışıklık Matrisi:

	Pozitif	Negatif
Pozitif	273	10
Negatif	15	316

Sınıflandırma Raporu:

	Kesinlik	Hatırlama	F Skoru	Destek
0	0.95	0.96	0.96	283
1	0.97	0.95	0.96	331
Doğruluk			0.96	614
Makro Ort.	0.96	0.96	0.96	614
Ağırlıklı Ort.	0.96	0.96	0.96	614

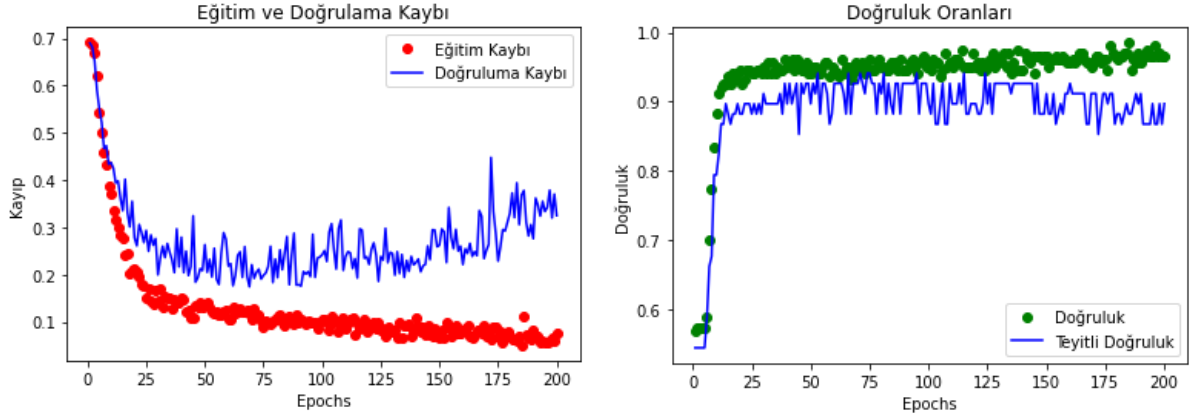
Kesikli 1 Yıllık Veri İçin



Şekil 57. Kesikli 1 yıllık veri için kayıp ve doğruluk oranları (500 Ölçek)

Yukarıda ki tablodan anlaşılacağı üzere tüm kesikli veri seti için için 100 epoch'tan sonra eğitim ve doğrulama kaybı belirgin şekilde ayrılmaya başlamaktadır.

Kesikli 1 Yıllık Veri İçin



Şekil 58. Kesikli 1 yıllık veri için kayıp ve doğruluk oranları (200 Ölçek)

Daha net görülebilmesi için 200 epoch üzerinden bakıldığında, ayrışmanın 100 epoch'tan sonra yön değiştirdiği görülmektedir. Şekil 58'de görüldüğü üzere 200 epoch için doğrulukta azda olsa düşme meydana gelmiştir. Bu şekilde 100 epoch ideal belirlenmiştir.

Optimizer = RMSprop, loss: sparse_categorical_crossentropy, epoch= 100 için :

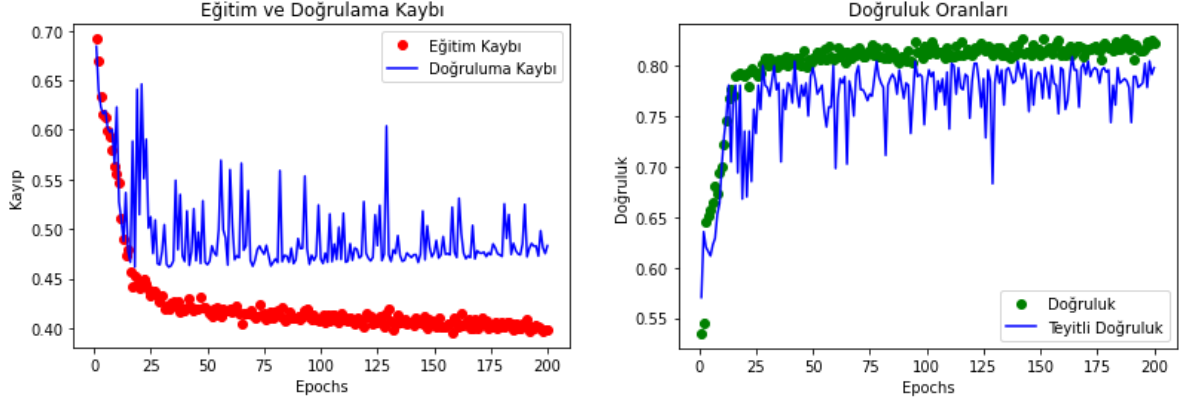
Karışıklık Matrisi:

	Pozitif	Negatif
Pozitif	41	4
Negatif	3	43

Sınıflandırma Raporu:

	Keskinlik	Hatırlama	F Skoru	Destek
0	0.93	0.91	0.92	45
1	0.91	0.93	0.92	46
Doğruluk			0.92	91
Makro Ort.	0.92	0.92	0.92	91
Ağırlıklı Ort.	0.92	0.92	0.92	91

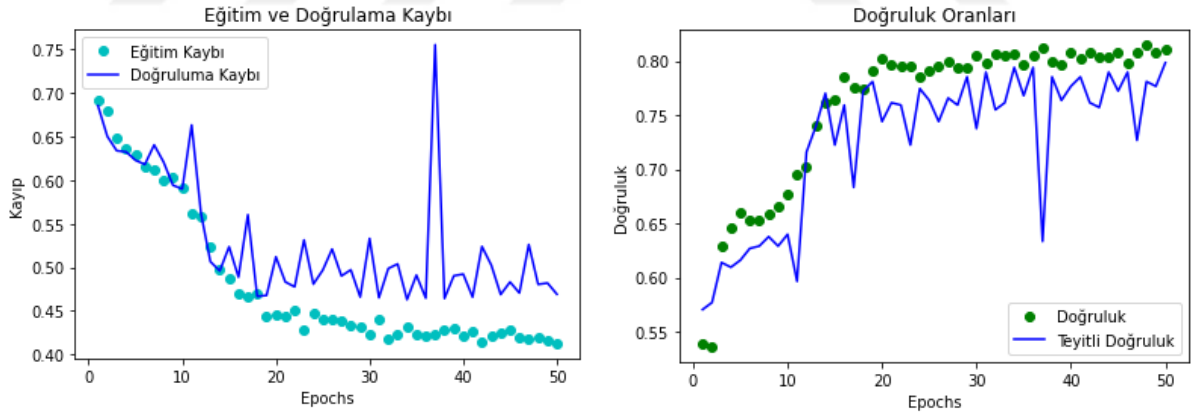
Sürekli Tüm Veri İçin



Şekil 59. Sürekli tüm veri için kayıp ve doğruluk oranları (500 Ölçek)

Sürekli verilerde veri yapısı değişikliği grafiğe de yansımıştır. Doğruluk oranı belli epochlarda en üst seviyeye çıkmış olsa da, 50 epoch sonrasında ayrışma oranı artmıştır, doğruluk yükselmezken teyitli doğruluktaki sapma oranları artmıştır. Bu sebeple 20 epoch tercih edilmiştir.

Sürekli Tüm Veriler İçin



Şekil 60. Sürekli tüm veri için kayıp ve doğruluk oranları (50 Ölçek)

Optimizer = Nadam, loss: sparse_categorical_crossentropy, epoch= 20 için :

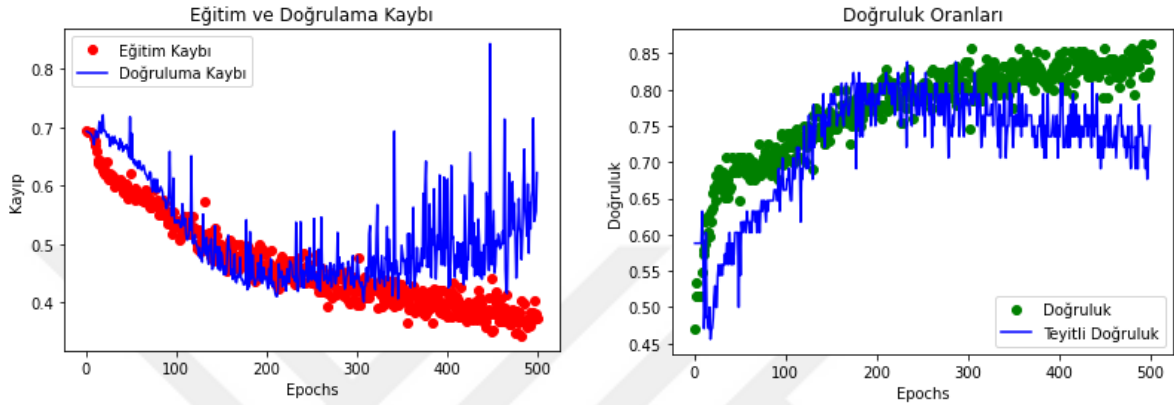
Karışıklık Matrisi:

	Pozitif	Negatif
Pozitif	158	124
Negatif	34	298

Sınıflandırma Raporu:

	Kesinlik	Hatırlama	F Skoru	Destek
0	0.82	0.56	0.67	282
1	0.71	0.90	0.79	332
Doğruluk			0.73	614
Makro Ort.	0.76	0.73	0.73	614
Ağırlıklı Ort.	0.76	0.74	0.73	614

Sürekli 1 Yıllık Veriler İçin



Şekil 61. Sürekli 1 yıllık veri için kayıp ve doğruluk oranları (500 Ölçek)

1 Yıllık Sürekli veriler içinde 200 epocha kadar bir ayrışma gözlenmemiştir. 200 epoch sonrası ayrışma başladığı için epoch 200 olarak belirlenmiştir.

Optimizer = Nadam, loss: sparse_categorical_crossentropy, epoch= 200 için :

Karışıklık Matrisi:

	Pozitif	Negatif
Pozitif	31	10
Negatif	22	28

Sınıflandırma Raporu:

	Kesinlik	Hatırlama	F Skoru	Destek
0	0.58	0.76	0.66	41
1	0.74	0.56	0.64	50
Doğruluk			0.65	91
Makro Ort.	0.66	0.66	0.65	91
Ağırlıklı Ort.	0.67	0.65	0.65	91

DÖRDÜNCÜ BÖLÜM

Bulgular ve Yorumlama

Bu bölümde analiz kısmında elde edilen metriklere göre veri setlerinin farklı makine öğrenme tekniklerinin optimize ayarlarına göre elde edilen sonuçların karşılaştırması ve yorumlanması yapılacaktır.

Elde edilen sonuçlar sonrasında en iyi performans veren Makine öğrenme yöntemlerinin tahmin performanslarını test edebilmek için verilerimizin bittiği noktadan başlayarak 11/06/2021-11/07/2021 tarihleri arası Bitcoin verileri ile 30 günlük Test adında yeni bir veri seti oluşturulmuştur. İndikatörleri doğru hesaplayabilmesi için geçmiş 2 aylık veri üzerinden son 30 günü bölünerek bu veri seti oluşturulmuştur. Makine öğrenme yöntemlerine en iyi performans verdikleri ayarlarda mevcut verilerinin tamamı öğrenme(fit) için yüklenecek ve 30 günlük verinin fiyat yönü sonuçlarının tahmin(predict) performansları kıyaslanacaktır.

Kesikli Verilerin Analizi

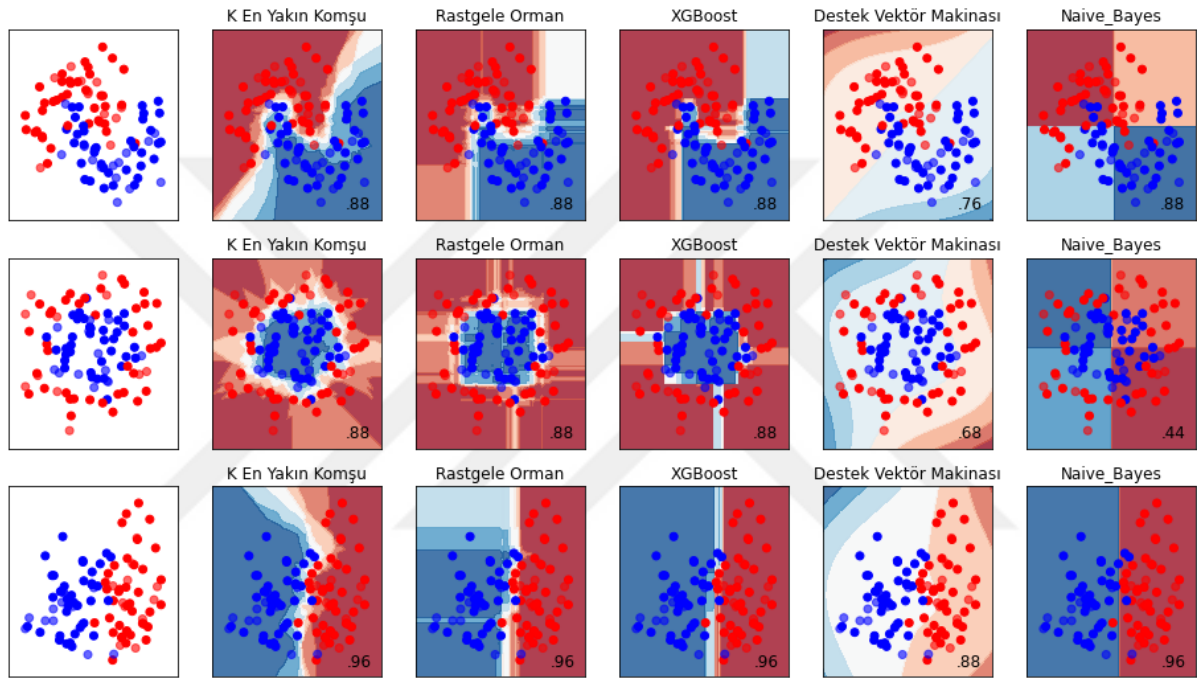
Kesikli Tüm Verilerin Genel Analizi

Tablo 43. Kesikli Tüm Veriler İçin Makine Öğrenme Yöntemleri Kıyaslaması

	Doğruluk	RMSE	ROC/AUC	MAE	F1
Rastgele Orman	0.954786	0.212635	0.954329	0.045214	0.958474
K En Yakın Komşu	0.944196	0.236230	0.944951	0.055804	0.948047
XGBoost	0.954786	0.212635	0.953897	0.045214	0.958659
Destek Vektör Makinası	0.957230	0.206809	0.957295	0.042770	0.960511
Naive Bayes	0.928310	0.267751	0.926821	0.071690	0.934718
Yapay Sinir Ağları	0.951140	0.221043	0.947513	0.052117	0.953846
Uzun-Kısa Süreli Bellek	0.959283	0.201784	0.959674	0.040717	0.961948

Kesikli tüm veriler için tablo incelendiğinde en iyi performans Uzun-Kısa Süreli Bellek (LSTM) makine öğrenme yöntemi Optimizer = Nadam, loss: sparse_categorical_crossentropy, epoch= 20 ayarında %95,93'lük doğrulukla elde edilmiştir. İkinci sırada %95,72'lik doğrulukla Destek Vektör Makinası ve %95,48 doğrulukla XGBoost ve Rastgele Orman makine öğrenme yöntemleri takip etmektedir. En düşük performansı %92,83'le Naive Bayes yöntemi vermiştir.

Değerler bu kadar yakın olduğundan dolayı yöntemlerin tahmin testleri ve eğitim zamanları üzerinden de değerlendirmeler yapılacaktır.



Şekil 62. Kesikli tüm verilerin algoritmalarında doğruluğa göre ayrışması

Kesikli tüm veri için tahmin testi.

Kesikli yapıdaki tüm veri ile eğitilen makine öğrenme yöntemlerinin Test Veri Setini tahmininin %100 çıkması üzerine, Test Veri Seti incelenmiş ve RSI indikatörünün 30 günlük kesikli yapıdaki sonuçlarının kapanış fiyat yönüyle aynı olduğu gözlenmiştir. Geriye dönük yapılan incelemede son 48 günde %100, son 80 günlük veride %91 benzerlik yakalanmıştır. Makine öğrenme yöntemlerinde RSI değerini kapanış olarak belirleyerek doğruluğu yakaladığı ve aşırı uyma(overfitting) sebep olduğu gözlenmiştir. Bu tespit üzerine RSI indikatörü hem eğitim hem de Test veri setlerinden çıkarılarak Makine öğrenme yöntemlerine Test Veri Seti için tahmin işlemi yaptırılmıştır. Elde edilen sonuçlar aşağıdaki gibi yer almıştır.

Tablo 44. Kesikli Tüm Veri Seti ile Eğitilmiş Yöntemlerin Tahmin Sonuçları

	Doğruluk	RMSE	ROC/AUC	MAE	F1
Rastgele Orman	0,967742	0,179605	0,970588	0,032258	0,969697
K En Yakın Komşu	0,967742	0,179605	0,970588	0,032258	0,969697
XGBoost	0,967742	0,179605	0,970588	0,032258	0,969697
Destek Vektör Makinası	0,967742	0,179605	0,970588	0,032258	0,969697
Naive Bayes	0,967742	0,179605	0,970588	0,032258	0,969697
Yapay Sinir Ağları	0,967742	0,179605	0,970588	0,032258	0,969697
Uzun-Kısa Süreli Bellek	0,967742	0,179605	0,970588	0,032258	0,969697

Kesikli Tüm Veri ile eğitilen makinelerin tamamı aynı tahmin performansını göstermişlerdir.



Şekil 63. Kesikli tüm veri seti ile eğitim ve tahmin süreleri (saniye)

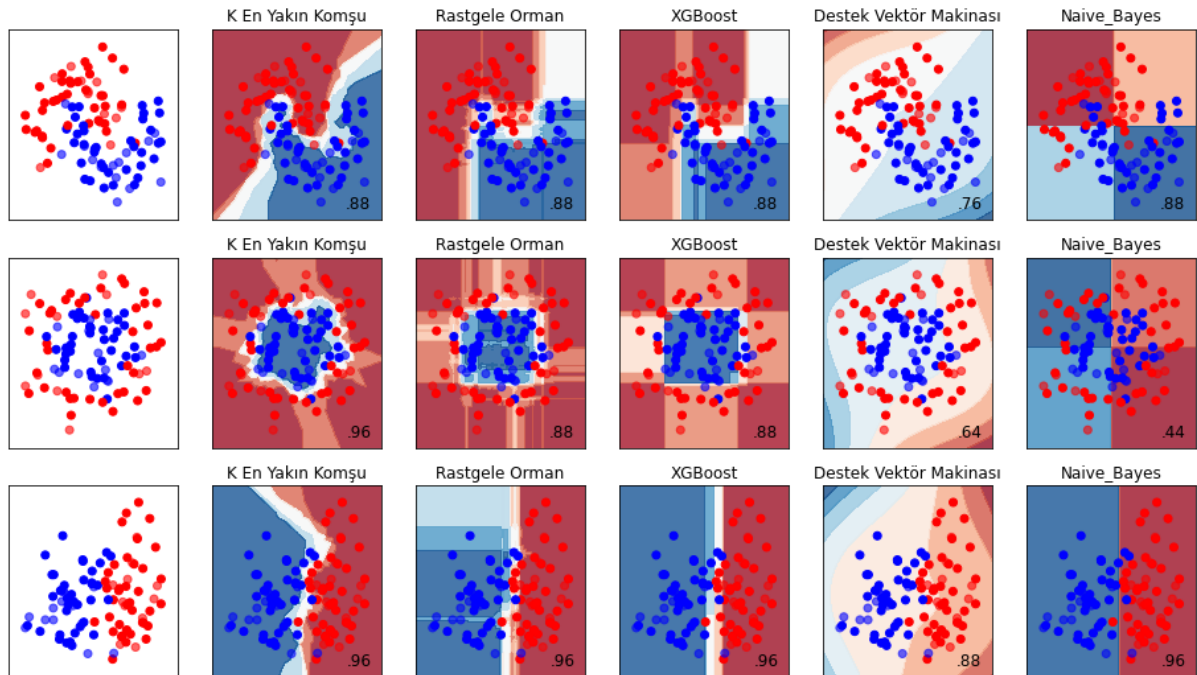
Tüm makine öğrenme yöntemlerinin Test Veri Seti tahmini için aynı doğruluk sonucunu verdiği Kesikli Tüm Veriler için, doğruluk testleri, eğitim ve tahmin süreleri incelendiğinde; **Destek Vektör Makinası** yönteminin bu büyüklükte veri seti için diğer yöntemlere göre daha verimli performans gösterdiği düşünülmektedir.

Kesikli 1 Yıllık Verilerin Genel Analizi

Tablo 45. Kesikli 1 Yıllık Veriler İçin Makine Öğrenme Yöntemleri Kıyaslaması

	Doğruluk	RMSE	ROC/AUC	MAE	F1
Rastgele Orman	0.933884	0.257130	0.934347	0.066116	0.938462
K En Yakın Komşu	0.911846	0.296908	0.912146	0.088154	0.917949
XGBoost	0.936639	0.251716	0.938307	0.063361	0.940260
Destek Vektör Makinası	0.942149	0.240523	0.943857	0.057851	0.945455
Naive Bayes	0.900826	0.314918	0.900572	0.099174	0.908163
Yapay Sinir Ağları	0.890110	0.331497	0.884634	0.109890	0.903846
Uzun-Kısa Süreli Bellek	0.923077	0.277350	0.922947	0.076923	0.924731

Kesikli 1 yıllık Veriler için tablo incelendiğinde en iyi performans Destek Vektör Makinası (SVM) makine öğrenme yönteminin Düzenleştirme = 0,5, Polinom Derecesi 3, çapraz doğrulama 20 katman ayarında %94,21 doğruluk ile elde edilmiştir. XGBoost makine öğrenme yöntemi %93,66 doğruluk ile İkinci sırada yer alırken, %93,39 doğruluk ile Rastgele Orman yöntemi üçüncü sırada yer almıştır. Yapay Sinir Ağları %89,01 doğruluk ile en düşük doğruluk performansını göstermiştir.



Şekil 64. Kesikli 1 yıllık verilerin algoritmalarında doğruluğa göre ayrışması

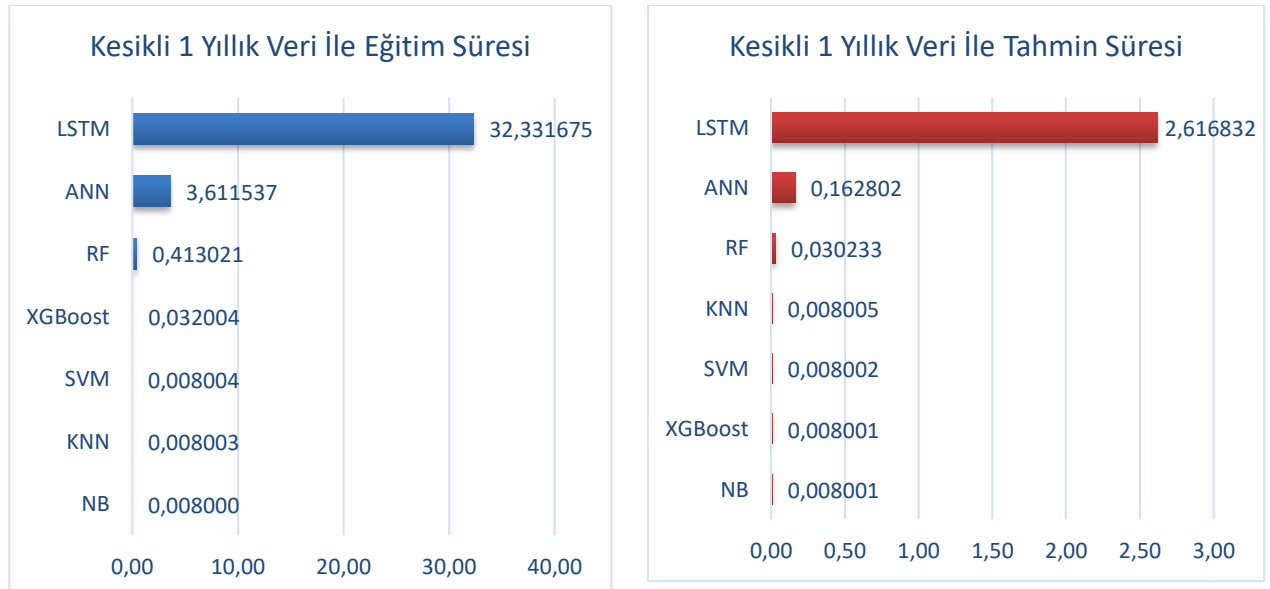
Kesikli 1 yıllık veri için tahmin testi.

Kesikli 1 yıllık verilerle eğitilen makine öğrenme yöntemlerinin de kesikli tüm verilerde olduğu gibi Test veri seti tahminin %100 doğrulukta olması üzerine aşırı uyumdan kaçınmak için RSI indikatörü eğitim ve test setlerinden çıkarılmıştır. Bu şekilde eğitilen Makine öğrenme yöntemlerinin Test Veri Seti için tahmin oranları aşağıdaki gibi yer almıştır.

Tablo 46. Kesikli 1 Yıllık Veri Seti ile Eğitilmiş Yöntemlerin Tahmin Sonuçları

	Doğruluk	RMSE	ROC/AUC	MAE	F1
Rastgele Orman	0,967742	0,179605	0,970588	0,032258	0,969697
K En Yakın Komşu	0,935484	0,254000	0,941176	0,064516	0,937500
XGBoost	0,967742	0,179605	0,970588	0,032258	0,969697
Destek Vektör Makinası	0,935484	0,254000	0,941176	0,064516	0,937500
Naive Bayes	0,903226	0,311086	0,911765	0,096774	0,903226
Yapay Sinir Ağları	0,903226	0,311086	0,905462	0,096774	0,909091
Uzun-Kısa Süreli Bellek	0,967742	0,179605	0,970588	0,032258	0,969697

Rastgele Orman, XGBoost ve Uzun-Kısa Süreli Bellek 1 Yıllık Kesikli Veri seti eğitiminde Kesikli Tüm Veri Seti ile aynı sonucu vererek %96,7'lik doğru tahmin elde etmişlerdir. K En Yakın Komşu ve Destek Vektör Makinasında %93,5 sonuç elde edilirken Naive Bayes ve Yapay Sinir Ağları %90,3'lük sonuç vermişlerdir. Test seti tahmin sonuçları performans analizleri ile büyük ölçüde paralellik göstermiştir.



Şekil 65. Kesikli 1 yıllık veri seti ile eğitim ve tahmin süreleri (saniye)

Kesikli 1 Yıllık Veri Seti ile eğitim ve tahmin sonuçları incelendiğinde, veri setinin iç testlerinde %93,6 doğruluk, %96,7 tahmin doğruluğu ve süre açısından çok hızlı olan **XGBoost** makine öğrenme yönteminin bu büyüklükte veri setlerinin tahminde daha verimli performans gösterdiği düşünülmektedir.

Tablo 47. Kesikli Veriler İçin T Testi Sonuçları

Veri Seti	Yöntem	Ortalama (Doğruluk)	N	Std. Sap.	t
Kesikli Tüm	Rastgele Orman	0,9667	30	0,18257	-1,000
	K En Yakın Komşu	0,9667	30	0,18257	-1,000
	XGBoost	0,9667	30	0,18257	-1,000
	Destek Vektör Makinası	0,9667	30	0,18257	-1,000
	Naive Bayes	0,9667	30	0,18257	-1,000
	Yapay Sinir Ağları	0,9667	30	0,18257	-1,000
	Uzun-Kısa Süreli Bellek	0,9667	30	0,18257	-1,00
Kesikli 1 Yıllık	Rastgele Orman	0,9667	30	0,18257	-1,000
	K En Yakın Komşu	0,9333	30	0,25371	-1,439
	XGBoost	0,9667	30	0,18257	-1,000
	Destek Vektör Makinası	0,9333	30	0,25371	-1,439
	Naive Bayes	0,9000	30	0,30513	-1,795
	Yapay Sinir Ağları	0,9000	30	0,30513	-1,795
	Uzun-Kısa Süreli Bellek	0,9667	30	0,18257	-1,000

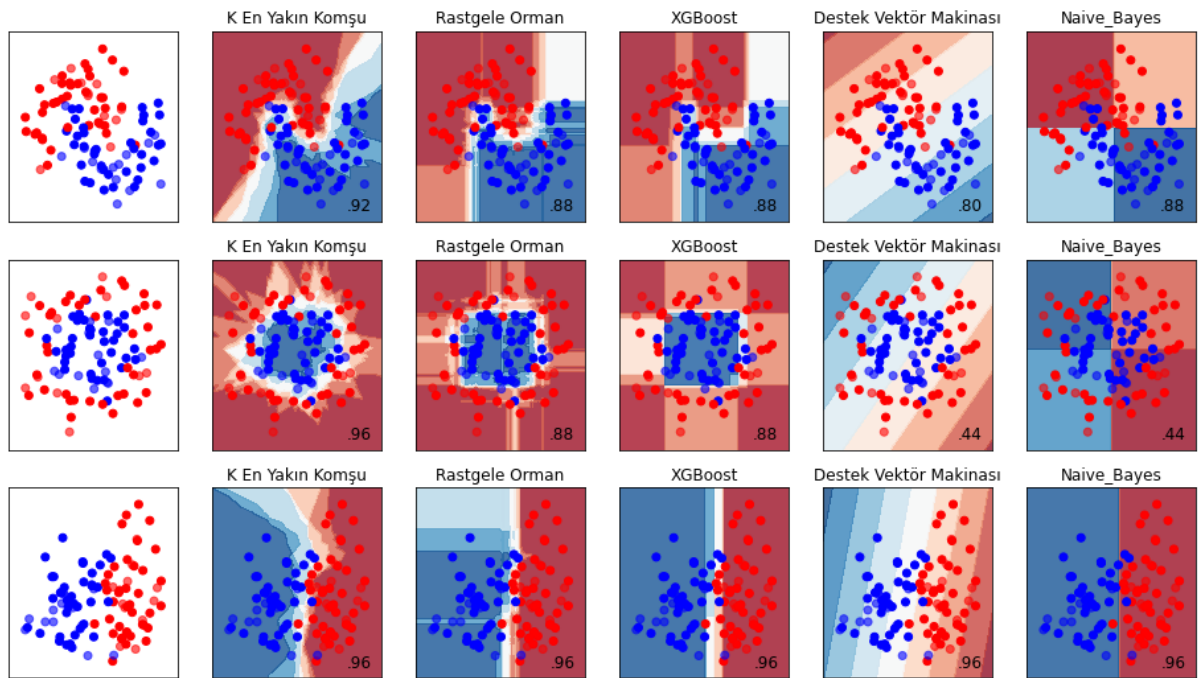
Sürekli Verilerin Analizi

Sürekli Tüm Verilerin Genel Analizi

Tablo 48. Sürekli Tüm Veriler İçin Makine Öğrenme Yöntemleri Kıyaslaması

	Doğruluk	RMSE	ROC/AUC	MAE	F1
Rastgele Orman	0.665988	0.577938	0.667720	0.334012	0.678431
K En Yakın Komşu	0.738086	0.511776	0.730345	0.261914	0.772711
XGBoost	0.773523	0.475896	0.771342	0.226477	0.792692
Destek Vektör Makinası	0.794297	0.453545	0.789436	0.205703	0.817095
Naive Bayes	0.610998	0.623700	0.604289	0.389002	0.655608
Yapay Sinir Ağları	0.804560	0.442086	0.795606	0.195440	0.832869
Uzun-Kısa Süreli Bellek	0.742671	0.507276	0.728937	0.257329	0.790451

Sürekli Tüm Veriler için tablo incelendiğinde en iyi performans Yapay Sinir Ağları (ANN) makine öğrenme yönteminin Optimizer = RMSprop, loss: sparse_categorical_crossentropy, epoch = 50 ayarında %80,46 doğruluk oranı ile elde edilmiştir. İkinci sırada %79,43 doğruluk ile Destek Vektör Makinası, üçüncü sırada ise %77,35 ile XGBoost makine öğrenme yöntemi en iyi doğruluk sonucunu vermiştir. En düşük performansı ise %61,10 doğruluk oranını Naive Bayes makine öğrenme yöntemi göstermiştir.



Şekil 66. Sürekli tüm verilerin algoritmalarında doğruluğa göre ayrışması

Sürekli tüm veri için tahmin testi.

Sürekli tüm veri test edilirken veriler sürekli yapıda olduğundan dolayı kesikli veride yer alan %100 benzerlik gibi durumlara rastlanmadığından dolayı herhangi bir veri setlerden çıkarılmamıştır. Test veri seti, eğitim için kullanılan Sürekli tüm veri setinde olduğu gibi 0-1 aralığına normalize edilmiştir. Sonuç değerleri için kesikli veri setlerinin kapanış sonucu kullanılmıştır.

Tablo 49. *Sürekli Tüm Veri Seti ile Eğitilmiş Yöntemlerin Tahmin Sonuçları*

	Doğruluk	RMSE	ROC/AUC	MAE	F1
Rastgele Orman	0,838710	0,401610	0,846639	0,161290	0,838710
K En Yakın Komşu	0,733333	0,516398	0,736607	0,266667	0,733333
XGBoost	0,774194	0,475191	0,781513	0,225806	0,774194
Destek Vektör Makinası	0,838710	0,401610	0,840336	0,161290	0,848485
Naive Bayes	0,548387	0,672022	0,500000	0,451613	0,708333
Yapay Sinir Ağları	0,870968	0,359211	0,869748	0,129032	0,882353
Uzun-Kısa Süreli Bellek	0,838710	0,401610	0,834034	0,161290	0,857143

Sürekli tüm veri setiyle eğitilmiş makine öğrenme yöntemlerinden Test Veri Setinin sonuçlarını %87,10 doğruluk ile en iyi Yapay Sinir Ağları tahmin etmiştir. Destek Vektör Makinası, Rastgele Orman ve XGBoost Makine öğrenme yöntemleri %83,88’lik doğruluk tahmin oranıyla ikinci sırada yer almıştır. %54,84 doğruluk tahmini ile Naive Bayes yöntemi en düşük performansı göstermiştir. Bu sonuçlarla en iyi ve en kötü tahmin oranları performans sonuçları ile paralellik göstermektedir.



Şekil 67. Sürekli tüm veri seti ile eğitim ve tahmin süreleri (saniye)

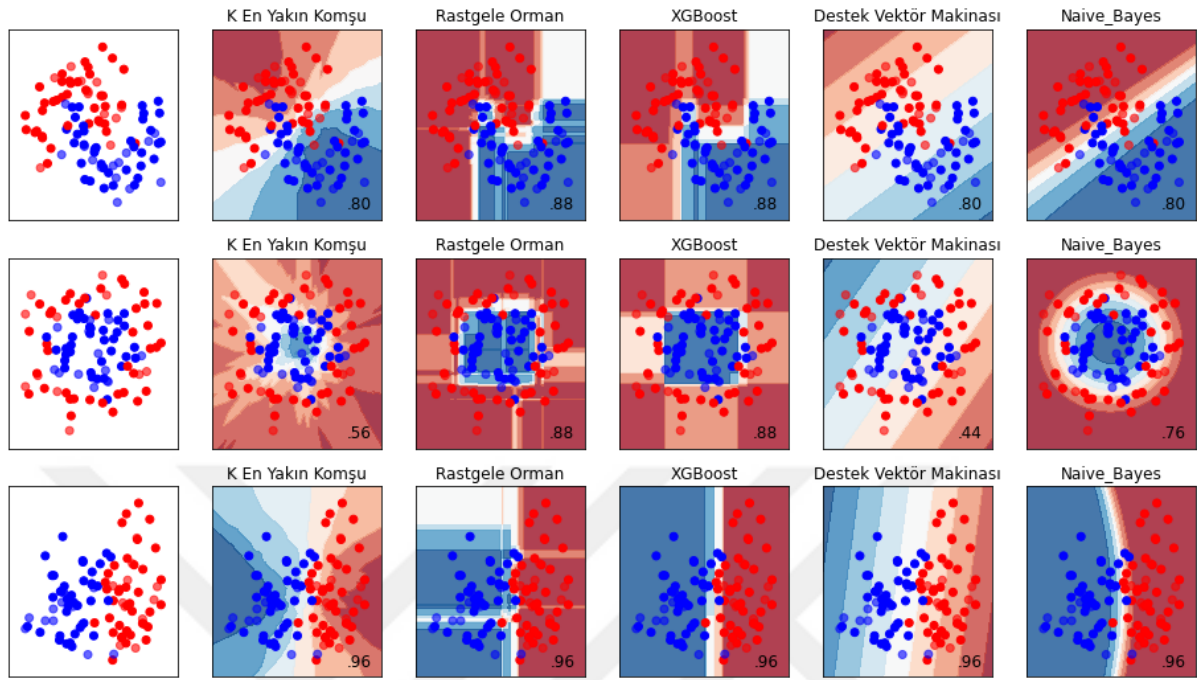
Sürekli tüm veri seti ile eğitim ve tahmin sonuçları incelendiğinde, veri setinin iç tahmin testlerinde %79,4 doğruluk, %83,87 tahmin doğruluğu ve süreler açısından çok hızlı olan **Destek Vektör Makinası** makine öğrenme yönteminin bu büyüklükte veri setlerinin tahminde daha verimli performans gösterdiği düşünülmektedir.

Sürekli 1 Yıllık Verilerin Genel Analizi

Tablo 50. Sürekli 1 Yıllık Veriler İçin Makine Öğrenme Yöntemleri Kıyaslaması

	Doğruluk	RMSE	ROC/AUC	MAE	F1
Rastgele Orman	0.575758	0.651339	0.573589	0.424242	0.605128
K En Yakın Komşu	0.608815	0.625447	0.604046	0.391185	0.646766
XGBoost	0.674931	0.570148	0.673491	0.325069	0.697436
Destek Vektör Makinası	0.710744	0.537825	0.704590	0.289256	0.744526
Naive Bayes	0.586777	0.642824	0.583741	0.413223	0.619289
Yapay Sinir Ağları	0.780220	0.468807	0.777132	0.219780	0.800000
Uzun-Kısa Süreli Bellek	0.681319	0.564519	0.676114	0.318681	0.632911

Sürekli tüm veriler için tablo incelendiğinde en iyi performans Yapay Sinir Ağları (ANN) makine öğrenme yönteminin Optimizer = RMSprop, loss: binary_crossentropy, epoch=500 ile elde edilmiştir.



Şekil 68. Sürekli 1 yıllık verilerin algoritmalarında doğruluğa göre ayrışması

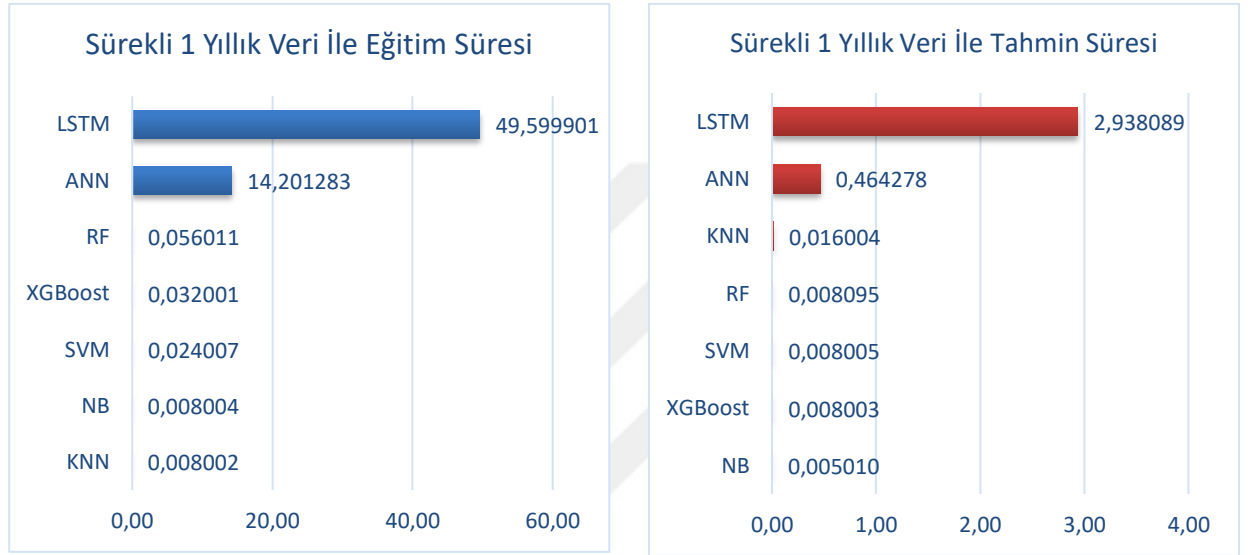
Sürekli 1 yıllık veri için tahmin testi.

Sürekli 1 Yıllık Verinin Tahmin testlerinde Sürekli tüm veride olduğu gibi verinin tamamı eğitim için kullanılmıştır. Test Veri Seti yine 0-1 aralığına normalize edilmiştir. Sonuç değerleri için kesikli veri setlerinin kapanış sonucu kullanılmıştır.

Tablo 51. Sürekli 1 Yıllık Veri Setiyle Eğitilmiş Yöntemlerin Tahmin Sonuçları

	Doğruluk	RMSE	ROC/AUC	MAE	F1
Rastgele Orman	0,645161	0,595683	0,638655	0,354839	0,685714
K En Yakın Komşu	0,612903	0,622171	0,615546	0,387097	0,625000
XGBoost	0,741935	0,508001	0,752101	0,258065	0,733333
Destek Vektör Makinası	0,806452	0,439941	0,810924	0,193548	0,812500
Naive Bayes	0,645161	0,595683	0,657563	0,354839	0,620690
Yapay Sinir Ağları	0,838710	0,401610	0,834034	0,161290	0,857143
Uzun-Kısa Süreli Bellek	0,800000	0,447213	0,799178	0,200000	0,833333

Sürekli 1 Yıllık Veri setiyle eğitilmiş makine öğrenme yöntemlerinden Test Veri Setinin sonuçlarını %83,87 doğruluk ile en iyi Yapay Sinir Ağları tahmin etmiştir. Destek Vektör Makinası %80,65 doğruluk ile 2. , Uzun Kısa Süreli Bellek %80 performans ile 3. Sırada yer alırken XGBoost Makine öğrenme yöntemi %74,19 doğruluk sonucu ile 4. Sırada yer almıştır. %61,29 doğruluk tahmini ile K En Yakın Komşu yöntemleri en düşük performansı göstermiştir. Bu sonuçlarla en iyi tahmin oranları performans sonuçları ile paralellik göstermiştir.



Şekil 69. Sürekli 1 yıllık veri ile eğitim ve tahmin süreleri (saniye)

Sürekli 1 Yıllık Veri Seti ile eğitim ve tahmin sonuçları incelendiğinde, veri setinin iç testlerinde %78 doğruluk, %83,9 tahmin doğruluğu ile Yapay Sinir Ağları performansı diğer Makine öğrenme türleri arasından belirgin olarak sıyrılıyor. Ancak süre açısından kendisine en yakın iç performans olarak %71, ve %80 tahmin doğruluğuna ulaşmış destek vektör makinasından 500 kat yavaş kalıyor. Yine de Sürekli Yapıdaki 1 Yıllık Verinin İşlenmesinde Yapay Sinir Ağları bu büyüklükte veri setlerinin tahminde daha verimli performans gösterdiği düşünülmektedir.

Tablo 52. Sürekli Veriler İçin T Testi Sonuçları

Veri Seti	Yöntem	Ortalama (Doğruluk)	N	Std. Sap.	t
Sürekli Tüm	Rastgele Orman	0,8333	30	0,37905	-2,408
	K En Yakın Komşu	0,6000	30	0,49827	-4,397
	XGBoost	0,7667	30	0,43018	-2,971
	Destek Vektör Makinası	0,8333	30	0,37905	-2,408
	Naive Bayes	0,5333	30	0,50742	-5,037
	Yapay Sinir Ağları	0,8667	30	0,34575	-2,112
	Uzun-Kısa Süreli Bellek	0,8333	30	0,37905	-2,408
Sürekli 1 Yıllık	Rastgele Orman	0,6333	30	0,49013	-2,408
	K En Yakın Komşu	0,6000	30	0,49827	-4,397
	XGBoost	0,7333	30	0,44978	-2,971
	Destek Vektör Makinası	0,8000	30	0,40684	-2,408
	Naive Bayes	0,6333	30	0,49013	-5,037
	Yapay Sinir Ağları	0,8333	30	0,37905	-2,112
	Uzun-Kısa Süreli Bellek	0,6000	30	0,49827	-2,408

Tablo 53. Makine Öğrenme Yöntemlerinin Genel Sonuçları

Makine Öğrenme Yöntemi	Kesikli Tüm Veri		Kesikli 1 Yıllık Veri		Sürekli Tüm Veri		Sürekli 1 Yıllık Veri	
	Test	Tahmin	Test	Tahmin	Test	Tahmin	Test	Tahmin
Rastgele Orman	0,9548	0,9677	0,9339	0,9677	0,6660	0,8387	0,5758	0,6452
K En Yakın Komşu	0,9442	0,9677	0,9118	0,9355	0,7381	0,7333	0,6088	0,6129
XGBoost	0,9548	0,9677	0,9366	0,9677	0,7735	0,7742	0,6749	0,7419
Destek Vektör Makinası	0,9572	0,9677	0,9421	0,9355	0,7943	0,8387	0,7107	0,8065
Naive Bayes	0,9283	0,9677	0,9008	0,9032	0,6110	0,5484	0,5868	0,6452
Yapay Sinir Ağları	0,9511	0,9677	0,8901	0,9032	0,8046	0,8710	0,7802	0,8387
Uzun-Kısa Süreli Bellek	0,9593	0,9677	0,9231	0,9677	0,7427	0,8387	0,6813	0,8000

BEŞİNCİ BÖLÜM

Sonuç ve Öneriler

20. yüzyılın ikinci yarısından itibaren geliştirilmeye başlanan makine öğrenme olgusu gelişen bilgisayar teknolojileri ile birlikte uygulama alanları sürekli genişletmektedir. İnternetinde yaygınlaşmasıyla birlikte kolaylaşan verilere ulaşım bu alanda yapılan çalışmaların artmasına sebep olmuştur. İnternet yalnızca kolayca verilere ulaşım sağlamamış, bunun yanında hızlı bir şekilde insanları birbirine bağlayarak zaman ve mekanlardan bağımsız sanal bir evren oluşturmuştur. Özellikle gelişen mobil teknolojisi ve uygulamalar ile tüm gelir düzeylerinden bu sanal evrene girişi hızlanmış; ticaret ve borsaya katılımı arttırarak paranın dijitalleşmesini hızlandırmıştır. Sanal evren kendi içinde bağımsızlığını destekleyecek sosyal ve finansal araçlarda geliştirmiştir. Bu yeni düzenin ürünü olan Kripto para geçtiğimiz 10 yıl içinde hayatımıza girmiş ve sanallaştığımız ölçüde değerini arttırarak ticaret ve yatırım aracı haline gelerek finans enstrümanlarından biri haline gelmiştir.

Yatırım dünyasında fiyat yönünün tahmini geçmişten günümüze önemini korumaktadır. Kripto paraların ilki olan Bitcoin, ticaret hacmi sayesinde yatırımcılar kadar araştırmacılar tarafından da sıkça incelenmektedir. Daha önce emtia ve borsa üzerine uygulanan makine öğrenme algoritmalarının Bitcoin fiyat tahmininde kullanıldığı örnekler dışında, farklı makine öğrenme algoritmalarının birlikte kullanıldığı veya yapay zeka uygulamaları ile birleştirildiği modellere de literatürde rastlanmaktadır.

Bu çalışmada Bitcoin fiyat tahmini için finans piyasasında sıkça kullanılan indikatörlerin yanında, Bitcoin konusunda günlük Google Arama Trendleri ve günlük atılan Tweet sayıları da tahmin algoritmasına dahil edilerek çeşitli makine öğrenme algoritmaları ile kesikli ve sürekli yapıda yer alan veriler için test edilmiştir.

Testler sırasında makine öğrenme yöntemleri optimize edilmek üzere tüm veri setleri ile farklı parametreler ile sınanmış ve elde edilen sonuçlar hata metrikleri ve performans ölçümleri açısından kıyaslanmıştır. Modellerde genel olarak Kesikli ve Sürekli yapıdaki veriler için optimize ayarlar her veri setinde hemen hemen farklılık göstermiştir. Bu yüzden genel bir ayar yerine elde edilen verinin yapısına göre yöntemin optimize edilmesinin önemi analizlerde bir kez daha ortaya çıkmıştır. Elde edilen sonuçlar içinde en iyi performansı veren modellerin tahmin sırasında girdilerin önem ağırlıkları sıralanmıştır. Genel olarak makine öğrenme tekniklerinin eğitim sırasında model oluştururken ağırlıklandırmalarını en iyi doğruluk üzerine

yaptıkları gözlenmiştir. Girdilerden Stochastic K ve D, RSI, LW ve Momentum indikatörlerinin diğerlerine oranla tahmin sırasında daha ön plana çıktığı gözlenmiştir. Özellikle kesikli yapıda tüm verilerde RSI, 1 yıllık tahminlerde ise LW indikatörü ön plana çıkmaktadır. Benzer çalışmalar farklı olarak tahmin algoritmaları eğitim ve tahmin süreleri açısından da kıyaslanarak hızlı tahmin gerektiren durumlarda model seçiminde yol gösterebilmesi hedeflenmiştir.

Günlük Bitcoine ilişkin atılan Tweet sayılarının ve Google Arama Trendlerinin önceki güne göre değişiminin etkisinin indikatörler kadar belirleyici olmadığı önem derecesi grafiklerinde gözlenmiştir. Bu sebeple bu verilere ilişkin hareketli ortamlar gibi indikatörler ile oluşturulan yeni veriler diğer teknik indikatörlerin yanına eklenebilecektir. Daha güçlü işlemcilerle metin sınıflandırma makine öğrenme yöntemlerini kullanarak anlık olarak olumlu/olumsuz şeklinde ayrıştırılan Tweet verileri ve duyarlılık analizi yapılmış Google Trend arama verilerine, teknik indikatörler ve piyasa toplam hacmi gibi veriler eklenerek fiyat yönünün ve fiyat bantlarının belirlenmesi bu alanda ihtiyaç duyulan çalışmalardandır. Bu sayede günümüzce sıkça dile getirilen kripto borsa manipülasyonunun gerçekleşmesi halinde makine tarafından öğrenilerek tekrar gerçekleşmesi halinde sonuçları doğru yönlü tahmin edebilecektir.

Veri setlerinden kesikli tüm veriler için makine öğrenme algoritmalarının iç testlerde ortalama %95'e yakın doğruluk yakaladığı ve bağımsız tahmin testlerinde yine tam doğruluğa yakın performans verdikleri gözlenmiştir. Verilerin neredeyse yedi de bir oranında azaltıldığı 1 yıllık kesikli veri setinde ise yine genel olarak iç testlerde %93 üzerinde doğruluk ve bağımsız tahmin testinde %90 üzerinde doğruluk yakaladığı gözlenmiştir. Kesikli verilerden elde edilen bu sonuçlardan son 1 yıllık fiyat hareketlerinde gerçekleşen hızlı değişimlerin model oluşturmaya yetecek kadar veri sağladığı düşünülmektedir. Alınan bu sonuçlar tahmin probleminin yalnızca zaman serisi çözümü olmadığını, aynı zamanda sınıflandırma ve regresyon yöntemlerinde verimli sonuç verdiği göstermiştir. Destek Vektör Makinası ve XGBoost modelleri hız ve performans açısından oldukça verimli sonuçlar göstermiştir. Kesikli verilerin iç testlerinde aşırı uyum problemi olmazken, test setinde aşırı uyum problemi oluşmuştur. Bunun sebebinin iç testlerde eğitim-test ayrışması ve çapraz doğrulama yöntemlerinde veri parçalar halinde alınırken, test setinde ise veri bir bütün alındığından dolayı zaman serisi tahmininde olduğu gibi sekansın içinde oluşan sebep-sonucun makine tarafından algılanarak serinin devamının aynı şekilde sürdürülmesi olduğu sonucuna ulaşılmıştır.

Sürekli verilerin performans testlerinde ise modellerin sonuçları arasında büyük farklar olduğu gözlenmiştir. Sürekli tüm veri için modeller içinde iç testlerde en yüksek %80 doğruluk yakalanabilmişken, Naive Bayes gibi modellerde doğruluk %60'a kadar düşmüştür. Sürekli 1 yıllık veri setinde ise bu oranlar %78 en iyi doğruluk elde edilirken, en düşük modelde %58 doğruluğa ulaşılmıştır. Sürekli yapıda tüm ve 1 yıllık verilerin tahmin performansları arasında hem iç testlerde hem de bağımsız tahmin testlerinde paralellik bulunmaktadır. Yalnızca Rastgele Orman modelinde eğitim verilerinin 1 yıllık olması durumunda tahmin performansının %20 azaldığı gözlenmiştir. Sürekli verilerde en iyi tahmini Yapay Sinir Ağları algoritması göstermiştir. Ancak Yapay Sinir Ağları ve eğitim ve tahmin için süre ve tüketimi Kısa-Uzun Süreli Bellek dışındaki modellere göre oldukça yüksektir.

Yapay Sinir Ağları ve Kısa-Uzun Süreli Bellek gibi sinir ağ modellerinde eğitim süreleri, çalıştıkları cihazın gücüne bağlı olarak diğer modellere göre daha uzun zaman almaktadır. Bu sürenin büyük kısmı veri setinin büyüklüğünden bağımsız modelin oluşturulması sırasında meydana gelmektedir. Literatürde fiyat hareket tahmini yanında doğrudan fiyat tahmini üzerine de sinir ağları modellerinin iyi performans gösterdiği bilinmektedir. Bu sebeple bu yöntemlerin durağan tekil tahminler yerine oluşturulan modellerin Pipeline ile sürekli beslemeli tahmin modelinde daha iyi performans verebileceği düşünülmektedir.

Kripto para piyasalarının hacminin her geçen gün arttığı günümüzde fiyat ve fiyat hareket tahminine yönelik modeller gelişmeye devam edecektir. Günlük fiyat yönü tahmini açısından iyi performans veren modeller tez çalışmamızda tespit edilebilmiştir. Ancak kripto para borsalarının 24 saat esasına göre süreklilik gösterdiği düşünüldüğünde günlük kapanıştan çok anlık alım satım daha büyük öneme sahip olmaktadır. Gelecek çalışmalarda bu tez çalışmasında makine öğrenme performanslarına yönelik elde edilen bilgiler ile saat, dakika ve 15 saniye gibi zaman periyotlarında sınanarak, elde edilen verilere özellikle futures piyasalarda yatırımcıların kısa(short) ve uzun(long) pozisyon alabilmelerini, ayrıca al-bekle-sat sinyali üretimi şeklinde modellerinde geliştirilebileceği düşünülmektedir. Kısa periyotlarda daha büyük veriler oluşacağından dolayı, LSTM gibi tekrarlayan sinir ağları modellerinden Kapılı Tekrarlayan Üniteler (GRU) ve Çoklu Katman Algılayıcı (MLP) gibi derin öğrenme modellerinin de araştırmalara dahil edilmesinin çalışmaları daha ileri seviyelere taşıyacaktır.

KAYNAKÇA

- Aery, M., & Ram, C. (2017). A Review on Machine Learning: Trends and Future. *Research Cell: An International Journal of Engineering Sciences*(25), s. 89-96.
- Afşar, M., Özdemir, B., & Arslan, B. G. (2012). *Para Teorisi*. Eskişehir: Anadolu Üniversitesi.
- Atakan, C., & Karabulut, İ. (2003). Derinliğe Dayalı Diskriminasyon. *Selçuk Üniversitesi Fen Fakültesi Fen Dergisi*, 53-64.
- Bitcoin Price Index* . (2021). Cointelegraph: <https://cointelegraph.com/bitcoin-price-index> adresinden edinilmiştir.
- Blockchain.com. (2021). <https://financesonline.com/number-of-blockchain-wallet-users/> adresinden edinilmiştir.
- Cervantesa, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., & Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, 189-215. doi:10.1016/j.neucom.2019.10.118
- Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *International Conference on Knowledge Discovery and Data Mining*, (s. 785–794). doi:<https://doi.org/10.1145/2939672.2939785>
- Chen, W., Xu, H., Jia, L., & Gao, Y. (2020, 2 8). Machine learning model for Bitcoin exchange rate prediction using economic and technology determinants. *International Journal of Forecasting*. doi:10.1016/j.ijforecast.2020.02.008
- Chen, Z., Li, C., & Sun, W. (2019). Bitcoin price prediction using machine learning: An approach to sample dimension engineering. *Journal of Computational and Applied Mathematics*(365). doi:10.1016/j.cam.2019.112395
- CoinDance. (2021). *Coin Dance*. Coin Dance: <https://coin.dance/> adresinden edinilmiştir.
- coinsspent.com. (2021). <https://coinsspent.com/best-bitcoin-cold-storage-wallets/> adresinden edinilmiştir.
- Cunningham, P., & Delany, S. (2021). k-Nearest Neighbour Classifiers - A Tutorial. *ACM Computing Surveys*, 54(6), 1-25. doi:10.1145/3459665
- Halaburda, H., & Sarvary, M. (2016). *Beyond Bitcoin*. Palgrave Macmillan. doi:10.1057/9781137506429

- Hochreiter, S., & Schmidhuber, J. (1997). Long Short-term Memory. *Neural Computation*, 1735-1780. doi:10.1162/neco.1997.9.8.1735
- Huang, J.-Z., Huang, W., & Ni, J. (2018, 11 22). Predicting bitcoin returns using high-dimensional technical. *The Journal of Finance and Data Science*, s. 140-155. doi:10.1016/j.jfds.2018.10.001
- IBM. (2021). *Doğrusal Regresyon*. IBM: <https://www.ibm.com/tr-tr/analytics/learn/linear-regression> adresinden edinilmiştir.
- Ibrahim, A., Kashef, R., & Corrigan, L. (2020, 11 19). Predicting market movement direction for bitcoin: A comparison of time series modeling methods. *Computers and Electrical Engineering*(89). doi:10.1016/j.compeleceng.2020.106905
- Ibrahim, A., Kashef, R., & Corrigan, L. (2021, 1). Predicting market movement direction for bitcoin: A comparison of time series modeling methods. *Computers & Electrical Engineering*. doi:10.1016/j.compeleceng.2020.106905
- Jeffares, A. (2019). *K-means: A Complete Introduction*. 1 28, 2021 tarihinde Towards Datascience. adresinden edinilmiştir.
- Karasu, S., Aytac, A., Sarac, Z., & Hacıoglu, R. (2018). Prediction of Bitcoin prices with machine learning methods using time series data. *2018 26th Signal Processing and Communications Applications Conference*. doi:10.1109/SIU.2018.8404760
- Karpuz, E. (2012). *Ödeme sistemleri ve araçlarının artan kullanımı: kredi kartı kullanımının para politikası etkinliğine etkisi*. Ankara: Türkiye Cumhuriyet Merkez Bankası.
- Keş, Y., & Turgut, B. F. (2015). Kültür ve Sanatın Yansıması Para Tasarımı"Osmanlı Son Dönem Kaimeleri ve Cumhuriyet'in İlk Emisyonları". *Medeniyet Sanat, İMÜ Sanat ve Tasarım Fakültesi Dergisi*, 31-47.
- Lahmiri, S., & Bekiros, S. (2020). Intelligent forecasting with machine learning trading systems in chaotic intraday Bitcoin market. *Chaos, Solitons and Fractals*(133). doi:10.1016/j.chaos.2020.109641
- Langs, G., Röhrich, S., Hofmanninger, J., Prayer, F., Pan, J., Herold, C., & Prosch, H. (2018). Machine learning: from radiomics to discovery and routine. *Der Radiologe*. doi:10.1007/s00117-018-0407-3
- Madan, I., Saluja, S., & Zhao, A. (2015). *Automated Bitcoin Trading via Machine Learning Algorithms*. Stanford University.
- Mallqui, D., & Fernandes, R. (2018). Predicting the direction, maximum, minimum and closing prices of daily Bitcoin exchange rate using machine learning techniques. *Applied Soft Computing*(75), s. 596-606. doi:10.1016/j.asoc.2018.11.038
- Mangla, N., Bhat, A., Avabratha, G., & Bhat, N. (2019). Bitcoin Price Prediction Using Machine Learning. *International Journal of Information and Computing Science*, 6(5), s. 318-320.

- McNally, S. (2016). *Predicting the price of Bitcoin using Machine*. Yüksek Lisans Tezi, National College of Ireland, Dublin. <http://norma.ncirl.ie/2496/1/seanmcnally.pdf> adresinden edinilmiştir.
- Mudassir, M., Bennbaia, S., Unal, D., & Hammoudeh, M. (2020). Time-series forecasting of Bitcoin prices using high-dimensional features: a machine learning approach. *Neural Computing and Applications*, s. 1-15. doi:10.1007/s00521-020-05129-6
- Nakamoto, S. (2008). *Bitcoin: A Peer-to-Peer Electronic Cash System*. <https://bitcoin.org> adresinden edinilmiştir.
- Öztemel, E. (2003). *Yapay Sinir Ağları*. İstanbul: Papatya Bilim.
- Öztürk, N., & Koç, A. (2006). Elektronik Para, Diğer Para Türleriyle Karşılaştırılması ve Olası Etkileri. *SÜ İİBF Sosyal ve Ekonomik Araştırmalar Dergisi*, 207 - 243.
- Pabuçcu, H., & Yakıcı Ayan, T. (2017). The Development of an Alternative Method for the Sovereign Credit Rating System Based on Adaptive Neuro-Fuzzy Inference System. *American Journal of Operations Research*, 7(1), 41-55. doi:10.4236/ajor.2017.71003
- Pabuçcu, H., Ongan, S., & Ongan, A. (2020, 11 16). Forecasting the movements of Bitcoin prices: an application of machine learning algorithms. *Quantitative Finance and Economics*, 4(4), s. 679–692. doi:10.3934/QFE.2020031
- Pawlak, Z. (2003). A rough set view on Bayes' theorem. *International Journal of Intelligent Systems*, 487 - 498. doi:10.1002/int.10099
- Sarı, Y. (1998). *Borsada Sistemli Teknik Analiz (Cilt 3)*. İstanbul: Alfa Yayıncılık.
- Sel, B. (2020). *Forecasting With a Large Number Of Predictors: A Tree Based Moving Average Approach*. Yüksek Lisans Tezi, Yükseköğretim Kurulu Ulusal Tez Merkezi'nden edinilmiştir. (Tez No:650510)
- Shah, D., & Zhang, K. (2014). Bayesian regression and Bitcoin. Allerton: Annual Allerton Conference on Communication, Control, and Computing. doi:10.1109/ALLERTON.2014.7028484
- Shivaprasa, P. (2020). *Understanding Confusion Matrix, Precision-Recall, and F1-Score*. Towards Datascience. adresinden edinilmiştir.
- Şıklar, İ. (2004). *Para Teorisi ve Politikası*. Eskişehir: Anadolu Üniversitesi Yayınları.
- Teke, Ç., Çolak, M., Kiraz, A., & İpek, M. (2019, 1). Prediction of Critical Fraction of Solid in Low Pressure Die Casting of Aluminum Alloys Using Artificial Neural Network. *Scientia Iranica*, 26(6), 3304–3312. doi:10.24200/sci.2019.50819.1881
- Tekinay, M. (2021). *Geleceğe Bakış; Bitcoin'in Derin Öğrenme Yöntemiyle Modellenmesi ve Tahmini*. Yüksek Lisans Tezi, Yükseköğretim Kurulu Ulusal Tez Merkezi'nden edinilmiştir. (Tez No:664975)

Türkiye'nin İlk Bitcoin ATM'si Hizmete Girdi. (2018, 12 19). Haberler: <https://www.haberler.com/bitcoin-islemlerinde-atm-donemi-11553089-haberi/> adresinden edinilmiştir.

VeriBilimcisi. (2017). *Makine Öğrenmesi.* Veri Bilimcisi: <https://veribilimcisi.com/2017/07/14/mse-rmse-mae-mape-metrikleri-nedir/> adresinden edinilmiştir.

Vranken, H. (2017). Sustainability of bitcoin and blockchains. *Current Opinion in Environmental Sustainability*, 1-9. doi:10.1016/j.cosust.2017.04.011

(2020). *World Payments Report.* Capgemini Research Institute.

Yavuz, U., Özen, Ü., Taş, K., & Çağlar, B. (2020). Yapay Sinir Ağları ile Blockchain Verilerine Dayalı Bitcoin Fiyat Tahmini. *Journal of Information Systems and Management Research*, 2(1), s. 1-9.



ÖZGEÇMİŞ

.....’de doğdu. Lisans eğitimini Anadolu Üniversitesi İşletme ve Ahmet Yesevi Üniversitesi Bilgisayar Mühendisliği Bölümlerinde tamamladı. 2010 yılından bu yana Bayburt Üniversitesinde çalışmakta olan Hulusi Mehmet TANRIKULU Evli ve 1 çocuk babasıdır. Halen Bayburt Üniversitesi İşletme Programında Yüksek Lisans eğitimi görmektedir.

