

**T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**MAKİNE ÖĞRENMESİ YÖNTEMLERİNİ KULLANARAK BİR
PETROKİMYA FİRMASININ HİSSE SENEDİ FİYAT TAHMİNİ**

YÜKSEK LİSANS TEZİ

Şevval TOPRAK

Endüstri Mühendisliği Anabilim Dalı

Endüstri Mühendisliği Bilim Dalı

ŞUBAT 2023

**T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**MAKİNE ÖĞRENMESİ YÖNTEMLERİNİ KULLANARAK BİR
PETROKİMYA FİRMASININ HİSSE SENEDİ FİYAT TAHMİNİ**

YÜKSEK LİSANS TEZİ

Şevval TOPRAK

Endüstri Mühendisliği Anabilim Dalı

Endüstri Mühendisliği Bilim Dalı

Tez Danışmanı: Doç. Dr. Gültekin ÇAĞIL

ŞUBAT 2023

Şevval TOPRAK tarafından hazırlanan “Makine Öğrenmesi Yöntemlerini Kullanarak Bir Petrokimya Firmasının Hisse Senedi Fiyat Tahmini” adlı tez çalışması 10.02.2023 tarihinde aşağıdaki jüri tarafından oy birliği ile Sakarya Üniversitesi Fen Bilimleri Enstitüsü Endüstri Mühendisliği Anabilim Dalı Endüstri Mühendisliği Bilim Dalı’nda Yüksek Lisans tezi olarak kabul edilmiştir.

Tez Jürisi

Jüri Başkanı :

Jüri Üyesi :

Jüri Üyesi :



ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ

Sakarya Üniversitesi Fen Bilimleri Enstitüsü Lisansüstü Eğitim-Öğretim Yönetmeliğine ve Yükseköğretim Kurumları Bilimsel Araştırma ve Yayın Etiği Yönergesine uygun olarak hazırlamış olduğum “MAKİNE ÖĞRENMESİ YÖNTEMLERİNİ KULLANARAK BİR PETROKİMYA FİRMASININ HİSSE SENEDİ FİYAT TAHMİNİ” başlıklı tezin bana ait, özgün bir çalışma olduğunu; çalışmamın tüm aşamalarında yukarıda belirtilen yönetmelik ve yönergeye uygun davrandığımı, tezin içerdiği yenilik ve sonuçları başka bir yerden almadığımı, tezde kullandığım eserleri usulüne göre kaynak olarak gösterdiğimi, bu tezi başka bir bilim kuruluna akademik amaç ve unvan almak amacıyla vermediğimi ve 20.04.2016 tarihli Resmi Gazete’de yayımlanan Lisansüstü Eğitim ve Öğretim Yönetmeliğinin 9/2 ve 22/2 maddeleri gereğince Sakarya Üniversitesi’nin abonesi olduğu intihal yazılım programı kullanılarak Enstitü tarafından belirlenmiş ölçütlere uygun rapor alındığını, çalışmamla ilgili yaptığım bu beyana aykırı bir durumun ortaya çıkması halinde doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi beyan ederim.

(01/06/2022)

Şevval Toprak





Aileme...



TEŞEKKÜR

Tez çalışmamın hazırlanmasında desteklerini ve görüşlerini esirgemeyen sayın danışman hocam Doç. Dr. Gültekin Çağıl'a, makale sürecindeki desteklerinden dolayı sayın hocam Dr. Öğr. Üyesi Abdullah Hulusi Kökçam'a ve tüm eğitim hayatım süresince üzerimde emeği olan öğretmenlerime/hocalarıma teşekkürlerimi sunarım.

Bugüne kadar benden maddi ve manevi desteklerini ve fedakarlıklarını esirgemeyen babam Murat Hüseyin Toprak'a, annem Şengün Toprak'a ve ağabeyim Ahmet Serhat Toprak'a; varlıklarıyla mutlu eden canım yeğenlerim Bulut Toprak'a ve Alp Toprak'a; tez hazırlama sürecinde hep benimle birlikte olan manevi desteklerini her zaman hissettiğim arkadaşlarıma çok teşekkür ederim.

Şevval Toprak



İÇİNDEKİLER

Sayfa

ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ	v
TEŞEKKÜR	ix
İÇİNDEKİLER	xi
KISALTMALAR	xiii
SİMGELER	xv
TABLO LİSTESİ	xvii
ŞEKİL LİSTESİ	xix
ÖZET	xxi
SUMMARY	xxiii
1. GİRİŞ	1
1.1. Literatür Taraması	2
2. MATERYAL VE YÖNTEM	9
2.1. Materyal	9
2.1.1. Veri seti	9
2.1.2. Modeller	22
2.2. Tahmin Metodu	23
2.3. Veri Ön İşleme	24
2.3.1. Veri temizleme yöntemleri	25
2.3.1.1. Eksik verinin veri setinden kaldırılması	25
2.3.1.2. Eksik verinin 0 olarak doldurulması	26
2.3.1.3. Aritmetik ortalama	26
2.3.1.4. Doğrusal enterpolasyon	26
2.3.2. Min-max normalizasyonu	26
2.4. Makine Öğrenmesi	27
2.4.2. Karar ağaçları	30
2.4.3. Rassal orman regresyonu	32
2.5. Yapay Sinir Ağları	33
2.5.1. Yapay sinir ağlarının sınıflandırılması	37
2.5.1.1. Katman sayılarına göre yapay sinir ağları	37
2.5.1.2. Bağlantı şekillerine göre yapay sinir ağları	38
2.6. Derin Öğrenme	39
2.6.1. Evrişimli sinir ağı (CNN)	40
2.6.2. Tekrarlayan sinir ağı (RNN)	42
2.6.3. Uzun kısa süreli bellek (LSTM)	44
2.6.4. CNN+LSTM	47
2.7. Hiperparametre Optimizasyonu	47
2.8. İstatistiksel Analiz	49
3. UYGULAMA	51
3.1. Eksik Verinin Doldurulma Yönteminin Bulunması	52

3.2. Model 1.....	53
3.3. Model 2.....	66
3.4. Model 3.....	77
3.5. Model 4.....	88
4. BULGULAR VE TARTIřMA	99
5. SONUÇ VE ÖNERİLER.....	123
KAYNAKLAR.....	127
ÖZGEÇMİř.....	131



KISALTMALAR

ANN	: Yapay Sinir Ağı
ANN-GARCH	: Yapay Sinir Ağları-Genelleştirilmiş Otoregresif Koşullu Değişken Varyans
ANN-LSTM	: Yapay Sinir Ağları – Uzun Kısa Süreli Bellek
ANOVA	: Tek Yönlü Varyans Analizi
AO	: Aritmetik Ortalama
ARMA	: Otoregresif Hareketli Ortalama
ARIMA	: Otoregresif Entegre Hareketli Ortalama
CART	: Sınıflandırma Regresyon Ağaçları
CMSB	: China Minsheng Bank İndeksi
CNN	: Evrişimli Sinir Ağı
CNN+LSTM	: Evrişimli Sinir Ağı+Uzun Kısa Süreli Bellek
DJI	: Dow Jones Industrial Average İndeksi
DL	: Derin Öğrenme
DNN	: Derin Sinir Ağları
DT	: Karar Ağaçları
EWI	: Ampirik Dalgack Dönüşümü
GA-CNN	: Genetik Algoritma-Evrişimli Sinir Ağı
GA-SVM	: Genetik Algoritma-Destek Vektör Makineleri
GARCH	: Genelleştirilmiş Otoregresif Koşullu Değişken Varyans
GBR	: Gradient Boosting Regresyonu
GRU-LSTM	: Kapılı Yinelemeli Uniteler-Uzun Kısa Süreli Bellek
LKOH	: Lukoil Hisse Senedi
LSTM	: Uzun Kısa Süreli Bellek
MAE	: Ortalama Mutlak Hata
MAPE	: Ortalama Mutlak Yüzde Hata
ML	: Makine Öğrenmesi
MLP	: Çok Katmanlı Algılayıcılar
MOEXOG	: MOEX Oil and Gas İndeksi

MPE	: Ortalama Yüzde Hata
MSE	: Ortalama Karesel Hata
NIFTY50	: NIFTY 50 İndeksi
NSE	: Ulusal Menkul Kıymetler Borsası (Hindistan)
QECON	: QE Consumer Goods&Services Endeksi
QFLS	: Qatar Fuel Co Hisse Senedi
PETKM	: Petkim Petrokimya Holding A. Ş. Hisse Senedi
R²	: Determinasyon Katsayısı
PSO	: Parçacık Sürü Optimizasyonu
ReLU	: Rectified Linear Unit Aktivasyon Fonksiyonu
RF	: Rassal Orman
RFR	: Rassal Orman Regresyonu
RMSE	: Kök Ortalama Karesel Hata
RNN	: Tekrarlayan Sinir Ağı
RTSOG	: RTS Oil & Gas Endeksi
SLNN	: Tek Katmanlı Yapay Sinir Ağı
S&P500	: Standard and Poor's 500 İndeksi
SVM	: Destek Vektör Makineleri
USD/TRY	: Dolar Birim Fiyatı
Tanh	: Hiperbolik Tanjant Aktivasyon Fonksiyonu
TUPRAS	: Türkiye Petrol Rafinerileri A.Ş. Hisse Senedi
XKMYA	: BİST Kimyasal, Petrol & Plastik İndeksi

SİMGELER

σ : Sigmoid Aktivasyon Fonksiyonu





TABLO LİSTESİ

Sayfa

Tablo 1.1. Literatür taraması.	6
Tablo 2.1. Veri setindeki değerlerin değer aralıkları.....	22
Tablo 2.2. Modellerin girdi ve çıktı şablonu.	22
Tablo 3.1. Eksik veri doldurma yöntemlerinin tahmin sonuçlarına etkisi	53
Tablo 3.2. Model 1 için RFR algoritmasının hiperparametre değerleri ve hata metrik değerleri.....	55
Tablo 3.3. Model 1 için LSTM algoritmasının hiperparametre değerleri ve hata metrik değerleri.....	57
Tablo 3.4. Model 1 için CNN+LSTM algoritmasının hiperparametre değerleri ve hata metrik değerleri.	61
Tablo 3.5. Model 2 için RFR algoritmasının hiperparametre değerleri ve hata metrik değerleri.....	68
Tablo 3.6. Model 2 için LSTM algoritmasının hiperparametre değerleri ve hata metrik değerleri.....	69
Tablo 3.7. Model 2 için CNN+LSTM algoritmasının hiperparametre değerleri ve hata metrik değerleri.	72
Tablo 3.8. Model 3 için RFR algoritmasının hiperparametre değerleri ve hata metrik değerleri.....	79
Tablo 3.9. Model 3 için LSTM algoritmasının hiperparametre değerleri ve hata metrik değerleri.....	80
Tablo 3.10. Model 3 için CNN+LSTM algoritmasının hiperparametre değerleri ve hata metrik değerleri.	83
Tablo 3.11. Model 4 için RFR algoritmasının hiperparametre değerleri ve hata metrik değerleri.....	90
Tablo 3.12. Model 4 için LSTM algoritmasının hiperparametre değerleri ve hata metrik değerleri.	91
Tablo 3.13. Model 4 için CNN+LSTM algoritmasının hiperparametre değerleri ve hata metrik değerleri.	94
Tablo 4.1. Tüm modeller için RFR algoritmasındaki eğitim ve test veri seti boyutları hiperparametrelerinin tahmin sonucuna etkisi.	100
Tablo 4.2. Tüm modeller için RFR algoritmasındaki maksimum özellik sayısı hiperparametresinin tahmin sonucuna etkisi.	101
Tablo 4.3. Tüm modeller için RFR algoritmasındaki ağaç sayısı hiperparametresinin tahmin sonucuna etkisi.	102
Tablo 4.4. Tüm modeller için LSTM algoritmasındaki eğitim, doğrulama ve test veri seti boyutu hiperparametrelerinin tahmin sonucuna etkisi.....	104
Tablo 4.5. Model2 için LSTM algoritmasındaki rastgele öğrenme durumu hiperparametresinin tahmin sonucuna etkisi.	105

Tablo 4.6.	Modeller için LSTM algoritmasındaki aktivasyon fonksiyonu hiperparametresinin tahmin sonucuna etkisi.	106
Tablo 4.7.	Tüm modeller için LSTM algoritmasındaki öğrenme aralığı hiperparametresinin tahmin sonucuna etkisi.	107
Tablo 4.8.	Modeller İçin LSTM algoritmasındaki devir sayısı hiperparametresinin tahmin sonucuna etkisi.	109
Tablo 4.9.	Modeller İçin LSTM algoritmasındaki birim sayısı hiperparametresinin tahmin sonucuna etkisi.	110
Tablo 4.10.	Modeller için LSTM algoritmasındaki eşik değeri hiperparametresinin tahmin sonucuna etkisi.	111
Tablo 4.11.	Modeller için LSTM algoritmasındaki parti boyutu hiperparametresinin tahmin sonucuna etkisi.	112
Tablo 4.12.	Modeller için CNN+LSTM algoritmasındaki filtre sayısı hiperparametresinin tahmin sonucuna etkisi.	113
Tablo 4.13.	Modeller için CNN+LSTM algoritmasındaki havuz boyutu hiperparametresinin tahmin sonucuna etkisi.	114
Tablo 4.14.	Modeller için CNN+LSTM algoritmasındaki birim sayısı hiperparametresinin tahmin sonucuna etkisi.	115
Tablo 4.15.	Modeller için CNN+LSTM algoritmasındaki eşik değeri hiperparametresinin tahmin sonucuna etkisi.	116
Tablo 4.16.	Modeller için CNN+LSTM algoritmasındaki eğitim, doğrulama ve test veri seti boyutu hiperparametresinin tahmin sonucuna etkisi.	117
Tablo 4.17.	Modeller için CNN+LSTM algoritmasındaki çekirdek boyutu hiperparametresinin tahmin sonucuna etkisi.	118
Tablo 4.18.	Modeller için CNN+LSTM algoritmasındaki öğrenme aralığı hiperparametresinin tahmin sonucuna etkisi.	119
Tablo 4.19.	Modeller için CNN+LSTM algoritmasındaki aktivasyon fonksiyonu hiperparametresinin tahmin sonucuna etkisi.	120
Tablo 4.20.	Modeller için CNN+LSTM algoritmasındaki parti boyutu hiperparametresinin tahmin sonucuna etkisi.	121
Tablo 4.21.	Modeller için CNN+LSTM algoritmasındaki devir sayısı hiperparametresinin tahmin sonucuna etkisi.	122

ŞEKİL LİSTESİ

Sayfa

Şekil 2.1. PETKM hisse senedinin açılış ve kapanış fiyatları.	10
Şekil 2.2. PETKM hisse senedinin açılış fiyatlarının histogram grafiği.	10
Şekil 2.3. PETKM hisse senedinin kapanış fiyatlarının histogram grafiği.	10
Şekil 2.4. USD/TRY açılış ve ortalama alış/satış kur fiyatı	11
Şekil 2.5. USD/TRY açılış alış satış kur fiyatının histogram grafiği.	11
Şekil 2.6. USD/TRY ortalama alış/satış kur fiyatının histogram grafiği.	12
Şekil 2.7. XKMYA endeksinin açılış ve kapanış fiyatları.	12
Şekil 2.8. XKMYA endeksinin açılış fiyatlarının histogram grafiği.	13
Şekil 2.9. XKMYA endeksinin kapanış fiyatlarının histogram grafiği.	13
Şekil 2.10. TUPRS hisse senedinin açılış ve kapanış fiyatları.	14
Şekil 2.11. TUPRS hisse senedinin açılış fiyatlarının histogram grafiği.	14
Şekil 2.12. TUPRS hisse senedinin kapanış fiyatlarının histogram grafiği.	15
Şekil 2.13. QECON endeksinin açılış ve kapanış fiyatları.	15
Şekil 2.14. QECON endeksinin açılış fiyatlarının histogram grafiği.	16
Şekil 2.15. QECON endeksinin kapanış fiyatlarının histogram grafiği.	16
Şekil 2.16. QFLS hisse senedinin açılış ve kapanış fiyatları.	17
Şekil 2.17. QFLS hisse senedinin açılış fiyatlarının histogram grafiği.	17
Şekil 2.18. QFLS hisse senedinin kapanış fiyatlarının histogram grafiği.	17
Şekil 2.19. MOEXOG endeksinin açılış ve kapanış fiyatları.	18
Şekil 2.20. MOEXOG endeksinin açılış fiyatlarının histogram grafiği.	18
Şekil 2.21. MOEXOG endeksinin kapanış fiyatlarının histogram grafiği.	19
Şekil 2.22. RTSOG endeksinin açılış ve kapanış fiyatları.	19
Şekil 2.23. RTSOG endeksinin açılış fiyatlarının histogram grafiği.	20
Şekil 2.24. RTSOG endeksinin kapanış fiyatlarının histogram grafiği.	20
Şekil 2.25. LKOH hisse senedinin açılış ve kapanış fiyatları.	21
Şekil 2.26. LKOH hisse senedinin açılış fiyatlarının histogram grafiği.	21
Şekil 2.27. LKOH hisse senedinin kapanış fiyatlarının histogram grafiği.	21
Şekil 2.28. Tahmin metodu.	24
Şekil 2.29. Veri ön işleme adımları.	24
Şekil 2.30. AI, ML, ANN ve DL arasındaki ilişki.	28
Şekil 2.31. ML yöntem grupları.	28
Şekil 2.32. Karar ağacı algoritmalarının yapısı.	30
Şekil 2.33. RFR algoritmasının yapısı.	32
Şekil 2.34. ANN algoritmasının katmanları ve genel yapısı.	34
Şekil 2.35. Bir sinir hücresinin yapısı.	35
Şekil 2.36. Sigmoid fonksiyonunun grafiği.	36
Şekil 2.37. Hiperbolik tanjant fonksiyonunun grafiği.	36

Şekil 2.38. ReLU fonksiyonunun grafiği.	37
Şekil 2.39. Bir ANN algoritmasının yapısı.	38
Şekil 2.40. Bir MLP algoritmasının yapısı.	38
Şekil 2.41. İleri beslemeli bir sinir ağının yapısı.	39
Şekil 2.42. Geri beslemeli bir sinir ağının yapısı.	39
Şekil 2.43. CNN algoritma yapısı.	41
Şekil 2.44. RNN algoritmasının çalışma prensibi.	43
Şekil 2.45. Bir RNN hücresinin yapısı.	43
Şekil 2.46. LSTM algoritmasının çalışma prensibi.	45
Şekil 2.47. CNN+LSTM algoritmasının yapısı.	47
Şekil 4.1. (a) R1.2 (b) R1.5 (c) R.1.11 örneklerinin tahmin grafikleri.	100
Şekil 4.2. (a) R4.12, (b)R4.14 (c) R4.15 örneklerinin tahmin grafikleri.	102
Şekil 4.3. (a)R3.12 (a) R3.14 (c) R3,15 örneklerinin tahmin grafikleri.	103
Şekil 4.4. (a) L1.3 (b) L1.4 örneklerinin tahmin grafikleri.	104
Şekil 4.5. (a) L1.3 (b) L1.4 örneklerinin kayıp fonksiyon grafikleri.	105
Şekil 4.6. (a) L2.2 (b) L2.3 örneklerinin tahmin grafikleri.	105
Şekil 4.7. (a) L1.4 (b) L1.5 (c) L1.6 örneklerinin tahmin grafikleri.	106
Şekil 4.8. (a) L1.4 (b) L1.5 (c) L1.6 örneklerinin kayıp fonksiyon grafikleri.	107
Şekil 4.9. (a) L3.4 (b) L3.5 örneklerinin tahmin grafikleri.	108
Şekil 4.10. (a) L3.4 (b) L3.5 örneklerinin kayıp fonksiyon grafikleri.	108
Şekil 4.11. (a) L4.4 (b) L4.5 örneklerinin kayıp fonksiyon grafikleri.	109
Şekil 4.12. (a) L4.4 (b) L4.5 örneklerinin kayıp fonksiyon grafikleri.	109
Şekil 4.13. (a) L3.11 (b) L3.12 örneklerinin tahmin grafikleri.	110
Şekil 4.14. (a) L3.11, (b) L3.12 örneklerinin kayıp fonksiyon grafikleri.	110
Şekil 4.15. (a) L4.2 (b) L4.3 örneklerinin tahmin grafikleri.	111
Şekil 4.16. (a) L4.2 (b) L4.3 örneklerinin kayıp fonksiyon grafikleri.	111
Şekil 4.17. (a) CL2.6 (b) CL2.7 örneklerinin tahmin grafikleri.	113
Şekil 4.18. (a) CL2.6 (b) CL2.7 örneklerinin kayıp fonksiyon grafiği.	113
Şekil 4.19. (a) CL1.11 (b) CL1.12 örneklerinin tahmin grafikleri.	114
Şekil 4.20. (a) CL1.11 (b) CL2.12 örneklerinin kayıp fonksiyon grafikleri.	114
Şekil 4.21. (a) CL4.7 (b) CL4.8 örneklerinin tahmin grafikleri.	115
Şekil 4.22. (a) CL4.7 (b) CL4.8 örneklerinin kayıp fonksiyon grafikleri.	115
Şekil 4.23. (a) CL2.13 (b) CL2.14 örneklerinin tahmin grafikleri.	116
Şekil 4.24. (a) CL2.13 (b) CL2.14 örneklerinin kayıp fonksiyon grafikleri.	116
Şekil 4.25. (a) CL2.9 (b) CL2.10 örneklerinin tahmin grafikleri.	117
Şekil 4.26. (a) CL2.9 (b) CL2.10 örneklerinin kayıp fonksiyon grafikleri.	117
Şekil 4.27. (a) CL1.10 (b) CL1.11 örneklerinin tahmin grafikleri.	118
Şekil 4.28. (a) CL1.10 (b) CL1.11 örneklerinin kayıp fonksiyon grafikleri.	118
Şekil 4.29. (a) CL1.1 (b) CL1.3 örneklerinin tahmin grafikleri.	119
Şekil 4.30. (a) CL1.1 (b) CL1.3 örneklerinin kayıp fonksiyon grafikleri.	119
Şekil 4.31. (a) CL4.13 (b) CL4.14 örneklerinin tahmin grafikleri.	120
Şekil 4.32. (a) CL4.13 (b) CL4.14 örneklerinin kayıp fonksiyon grafikleri.	121
Şekil 4.33. (a) CL2.11 (b) CL2.12 örneklerinin tahmin grafikleri.	122
Şekil 4.34. (a) CL2.11 (b) CL2.12 örneklerinin kayıp fonksiyon grafikleri.	122

MAKİNE ÖĞRENMESİ YÖNTEMLERİNİ KULLANARAK BİR PETROKİMYA FİRMASININ HİSSE SENEDİ FİYAT TAHMİNİ

ÖZET

Günümüzde borsalar önemli bir yatırım aracı halini almıştır. Yatırımcılar deneyimlerine ve sezgilerine güvenerek yatırım portföylerini oluşturup gelir elde etmeye çalışmaktadırlar. Ancak borsadaki hisse senedi ve endekslere etki eden pek çok etken bulunmaktadır. Bu da onların tahminini zorlaştırmaktadır. Bu nedenle literatürde zaman serisi tahminlerinde kullanılan yöntemler hisse senedi ve endeks değerlerinin tahmininde de kullanılmaya başlanmıştır. Özellikle son yıllarda yapılan hisse senedi tahmin çalışmalarında geleneksel yöntemlerinden çok yapay zeka yöntemlerinin önem kazandığı görülmektedir.

Bu çalışmada, Türkiye, Rusya ve Katar'daki önemli petrol endeksi ve hisse senedi fiyatlarını içeren bir veri seti kullanılarak Petkim Petrokimya Holding A.Ş.'nin hisse senedi (PETKM) kapanış fiyatı tahmini yapılmıştır. Rassal Orman Regresyonu (RFR), Uzun Kısa Süreli Bellek (LSTM) ve Evrişimli Sinir Ağı+Uzun Kısa Süreli Bellek (CNN+LSTM) yöntemleriyle tahmin edilerek sonuçları karşılaştırılmıştır. Tahmin sonuçlarının başarısını ölçmek için Ortalama Karesel Hata (MSE), Kök Ortalama Karesel Hata (RMSE), Ortalama Mutlak Hata (MAE), Ortalama Mutlak Yüzde Hatası (MAPE) gibi hata metrik yöntemleri ve Determinasyon Katsayısı (R^2) kullanılmıştır. Çalışmada kullanılan bu 3 tahmin yönteminin kullanılıp sonuçlarının karşılaştırıldığı bir çalışmaya literatürde rastlanmamıştır.

Tahmin işlemi için 4 farklı girdi modeli oluşturulmuştur. 1. girdi veri setinde (Model 1), Dolar açılış ve ortalama alış/satış kur fiyatı; Türkiye'nin hisse senedi ve endeks fiyatları, 2. girdi veri setinde (Model 2), Dolar fiyatı açılış ve ortalama alış/satış kur fiyatı, Türkiye'deki ve Katar'daki hisse senedi-endeks fiyatları, 3. girdi veri setinde (Model 3), Dolar fiyatı açılış ve ortalama alış/satış kur fiyatı, Türkiye'deki ve Rusya'daki hisse senedi ve endeks fiyatları, 4. girdi veri setinde (Model 4) Dolar açılış ve ortalama alış/satış kur fiyatı, Türkiye, Katar ve Rusya'daki hisse senedi-endeks fiyatları kullanılmıştır.

Çalışmada farklı ülkelerden veriler bulunduğu borsaların işlem günleri farklılık göstermektedir. Bu nedenle eksik verilerin ön işleme için 4 farklı yöntem kullanılmıştır. Bunlardan 1'inci eksik verinin bulunduğu tarihin veri setinden çıkarılması, 2'ncisi eksik verilerin yerine 0 yazılması, 3'üncüsü eksik verinin ortalama ile doldurulması ve 4'üncüsü eksik verinin doğrusal enterpolasyon yöntemi ile doldurulmasıdır. Model 1'deki veriler bu 4 farklı yöntemle doldurulmuş, veri setleriyle tahmin işlemi yapılmıştır. Çalışmadaki veriler için en iyi sonucu veren yöntemin doğrusal enterpolasyon yöntemi olduğu görülmüştür. En iyi sonucu veren eksik veri doldurma yöntemi diğer modeller için de uygulanmıştır.

Çalışmanın amacı çalışmada kullanılan veriler için en doğru sonucu veren algoritmayı, algoritmalarda kullanılan hiperparametrelerin tahmin başarısına etkisini ve Türkiye'nin petrol temin ettiği ülkelerin PETKM hisse senedi fiyatı tahmininde bir etkisi olup olmadığını gözlemlemektir.

Çalışmanın sonucunda RFR algoritmasında tahmin sonucuna en çok etki eden hiperparametrenin eğitim ve test veri seti boyutu olduğu, LSTM algoritmasında tahmin sonucuna en çok etkisi olan hiperparametrenin rastgele öğrenme durumu (Shuffle) olduğu, CNN+LSTM algoritmasında tahmin sonucuna en çok etkisi olan hiperparametrelerin havuzlama boyutu, öğrenme aralığı ve çekirdek boyutu olduğu görülmüştür. Oluşturulan bu 4 farklı girdi modeli karşılaştırıldığında Rusya ve Katar'daki hisse senedi ve endeks fiyatlarının PETKM hisse senedi kapanış fiyatını tahmin etmede anlamlı bir etkisinin olmadığı sonucuna varılmıştır. Algoritmaların başarısı karşılaştırıldığında genellikle en iyi sonuçları veren algoritmanın LSTM olduğu görülmüştür.

Çalışmanın Bulgular ve Tartışma bölümünde tüm karşılaştırmalar detaylı bir şekilde yapılmış ve incelenmiştir. Elde edilen sonuçlar ve ileride yapılacak çalışmalar için öneriler sonuç bölümünde paylaşılmıştır.

STOCK PRICE PREDICTION OF A PETROCHEMICAL COMPANY USING MACHINE LEARNING METHODS

SUMMARY

Today, stock markets have become an important investment tool. Investors rely on their experience and intuition to create their investment portfolios and try to generate income. However, there are many factors that affect stocks and indexes in the stock market. This makes their prediction difficult. For this reason, the methods used in time series estimations in the literature have also started to be used in the estimation of stock and index values. While estimations in this area were made with traditional methods in the early days, it is seen that artificial intelligence methods have gained importance rather than traditional methods in stock estimation studies carried out in recent years. This is because artificial intelligence methods are intelligent algorithms that can think, make decisions, interpret and make inferences.

In this study, the closing price of Petkim Petrokimya Holding A.Ş.'s stock (PETKM) closing price is estimated by using a data set including dollar price, important oil index and stock prices in Turkey, Russia and Qatar. For the estimation process, 4 different input data sets were created. In the 1st input dataset (Model 1), USD/TRY opening and average buying/selling exchange rate; the opening and closing prices of Turkish stocks (PETKM, TUPRS-Türkiye Petrol Rafinerileri A.Ş.-) and the opening and closing prices of Turkey's sectoral petroleum index (XKMYA -BIST Chem Petrol Plastic-) are used. USD/TRY opening and average buying/selling exchange rate; the opening and closing prices of Sectoral petroleum index-stocks in Turkey (XKMYA, PETKM, TUPRAS), the opening and closing prices of sectoral petroleum index in Qatar (QECON- QE Consumer Goods & Services-) and the opening and closing prices of stock in Qatar (QFLS- Qatar Fuel Co-) are contained in 2nd input dataset (Model 2). 3rd Input data set (Model 3) consists of the USD/TRY opening and average buying/selling exchange rate; the opening/closing prices of sectoral oil index-stocks in Turkey (PETKM, TUPRS, XKMYA), the opening/ closing prices of Russia's sectoral oil indexes (MOEXOG-MOEX Oil & Gas and RTSOG -RTS Oil & Gas-) and the opening/ closing stock prices of Russia's leading oil company (LKOI - Lukoil-). The USD/TRY opening and average buying/selling exchange rate; the opening/closing prices of Sectoral petroleum index-stocks in Turkey (XKMYA, PETKM, TUPRAS), the opening/ closing prices of sectoral petroleum index-stock in Qatar (QECON, QFLS) and the opening/ closing prices of sectoral petroleum index-stock in Russia (MOEXOG, RTSOG, LKOI), are contained in 4th input dataset (Model 4). The data of these features between 03.01.2010-31.12.2020 were used. Since the data set includes data from foreign countries, the days when stock markets do not trade are changing. Four different methods were applied to the data set to fill in the missing data. One of them is to replace the missing data with 0, another is to remove the missing

data from the data set, another is to use the average of the data 1 day before and after the missing data, and the last one is to fill in the missing data by linear interpolation method. In this way, estimation results were observed in Model 1 by applying these methods to the data set. In general, the data filled with the linear interpolation method gave the most successful results. For this reason, in order to avoid data loss, data filling was performed with the Linear Interpolation method for the dates except for the dates in which there is no data in all data properties.

In addition, since not all feature values in the data set are within a certain value range, all features are normalized. Max-Min Normalization method, which is the most used normalization method, was used. With normalization, the feature values are set to 0-1 range. Failure to do this will adversely affect the prediction algorithms and algorithm results.

The data sets created are estimated with Random Forest Regression (RFR) which is a powerful machine learning method, Long Short Term Memory (LSTM) algorithm which is the most used method in time series estimations and Convolutional Neural Network + Long Short Term Memory (CNN+LSTM) which one of the deep learning hybrid methods that have been used recently. These 3 methods were applied to all 4 data sets. Error metric methods such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Coefficient of Determination (R2) were used to evaluate the estimation values. The Hyperparameter Tuning method was used for the optimization of the error results. The values of the hyperparameters provide different results in the data. For this reason, in the tests with high error results, experiments were made by changing the hyperparameter of the algorithm. Since these algorithms learn different features each time they run, they give different results. Algorithms were run at least 10 times in order to stabilize the results. Generally, while there are large differences between the evaluation results in the first runs, similar values are obtained in the next runs.

In the implementation of the study, Python 3.7 writing language and Spyder 3 editor were used. Python's Pandas, SciKit Learn, Keras, Numpy, Math and Matplotlib libraries were used to create the algorithms. The machine on which the algorithms used are run has 8.00GB RAM, Intel(R) Core(TM) i7-4510U CPU processor and Windows 10 Pro 64 bit operating system. The data were kept in MS Excel 2013 program. The excel format used in the study is csv.

The aims of this study are: 1- To observe whether the sectoral oil indexes and stock prices of the countries from which Turkey supplies oil have an effect on the estimation of the stock (PETKM) of Petkim Petrokimya Holding A.Ş., which has an important role in the petrochemical industry of our country, 2- Comparing the methods used in estimation, finding and interpreting the most appropriate estimation method, 3- To find the effect of the hyperparameter of these algorithms on the prediction results.

In the study, data set splitting (training data set size, test data set size), number of trees, maximum number of features hyperparameters for the RFR algorithm were tested and compared. It has been seen that the most influential hyperparameter on this algorithm is the size of the training and test data sets. If the training data set contains less data and the test data set contains more data, the algorithm has been largely unsuccessful.

Other hyperparameters showed less effect on estimation results compared to training and test dataset size.

For LSTM algorithm, dataset splitting (training dataset size, validation dataset size, test dataset size), epoch, number of layers, number of units in layers and threshold value, batch size, activation function, random learning situation, optimization function and learning range hyperparameters are used. Among these hyperparameters, it has been observed that the most influential hyperparameter on the prediction values is the random learning situation. This hyper parameter provides random learning of the given data in the training data set while running the algorithm. The values of this parameter are selected as Yes and No. If it is selected as Yes, it has been observed that it gives mostly accurate and consistent results compared to the case of No in all tests.

Dataset splitting (training dataset size, validation dataset size, test dataset size), epoch, number of layers, number of units in layers, threshold value and pool size, kernel size, optimization function, learning interval, batch in CNN+LSTM algorithm dimension, random learning situation hyperparameters were used. It was observed that the most influential hyperparameters among these hyperparameters were the maxpooling size, learning rate and kernel size hyperparameter. The success of this algorithm is generally similar to that of LSTM.

When the estimation results obtained from the algorithm for the 4 data sets created were evaluated, it was seen that Model 1 was the model that gave the best result according to most of the results in the tests. Model 2 and Model 3's prediction success rates are close to Model 1. However, Model 4 performed less well than the other 3 data sets. For this reason, it is seen that the oil indexes and stocks related to Qatar and Russia used in the models do not make a difference in the algorithm estimation for PETKM stock closing price.

In the Findings and Discussion section of the study, all comparisons were made and examined in detail. The results obtained and suggestions for future studies are shared in the conclusion section.



1. GİRİŞ

Zaman serileri belirli bir zaman dilimi içerisinde, zamana bağılı olarak değişim gösteren sürekli dizilerdir. Gerçek hayatta görebileceğimiz zaman serilerine örnek olarak finansal veriler, sensör verileri, hava kalitesi verileri, su seviyesi verileri ve hava durumu verileri gibi veriler gösterilebilmektedir. Bu veriler birçok çevresel faktörden etkilendiklerinden doğrusal bir yapıdan çok dalgalı ve gürültülü bir yapıya sahiptirler [1-3]. Bu nedenle diziler tek başlarına bir anlam ifade etmemektedirler ancak veriler üzerinde yapılabilecek zaman serisi analizleriyle anlamlı hale getirilebilmektedirler. Zaman serisi analizleri verilerin anlamlı istatistiklerinin ve gizli özelliklerinin ortaya çıkmasını sağlamaktadır [2]. Zaman serilerinin tahmini ise bir dizinin zaman içindeki gelişimini, gelişim yönünü takip ederek dizinin gelecek verilerinin eğilimlerinin tahmini için yapılır. Zaman serisi tahmini 3 aşamada yapılmaktadır. Bunlar: 1- Verilerin toplanması ve düzenlenmesi, 2-Zaman serisi analizi ve tahmin modelinin oluşturulması, 3-Oluşturulan tahmin modelleriyle tahmin işleminin yapılması ve sonuçların değerlendirilerek yorumlanmasıdır [4]. Tahminde kullanılan veri sayısının fazla miktarda olması ve modelin doğru bir şekilde oluşturulması sonuçların güvenilirliği açısından son derece önemlidir [5].

Günümüzde zaman serilerine bağlı tahmin ihtiyaçları büyük oranda artmıştır. Zaman serisi modelleri içerisinde en çok tahmin edilmek istenen finansal zaman serileridir. [5]. Akademik ve gerçek hayat tahminlerinde de finansal zaman serileri içerisinde borsadaki endeks ve hisse senedi fiyatı değerleri büyük oranda kullanılmaktadır. Borsa verileri ülke içerisindeki üretimden, tüketimden, politik faktörlerden, diğer ülkelerle olan ilişkilerinden yatırımcı duyarlılığına kadar pek çok kontrol edilemez dış faktörlerden ve borsa içerisindeki hareketler, endeks değerleri gibi birçok iç faktörden büyük oranda etkilenmektedir [6]. Etkilendiği bu geniş yelpazedeki faktörlerden dolayı borsa endeksleri ve fiyatları dalgalı bir yapıya sahiptir. Bu sebeple borsa tahmin yöntemleri tahminin güvenilirliği için, gürültülere; trendlere, aykırı değerlere karşı dayanıklı ve doğrusal olmayan tahmin yöntemleri olmalıdır [1].

Ekonominin önemli bir parçası olan borsalardaki hisse senedi fiyatlarının ve endeks değerlerinin doğru tahmini kurumlar ve yatırımcılar için doğru bir portföy oluşturmada çok önemli bir yere sahiptir. Borsa ve diğer zaman serisi verilerinin tahmini için geçmiş yıllardan günümüze Otoregresif Hareketli Ortalama (ARMA), Otoregresif Entegre Hareketli Ortalama (ARIMA), Genelleştirilmiş Otoregresif Koşullu Değişken Varyans (GARCH) vb. gibi geleneksel tahmin yöntemleri kullanılmaktadır. Ancak bu yöntemler borsa verilerinin hareketli gürültülü yapısına yeterince uyum sağlayamamaktadır. Son yıllarda ise tahmin için daha doğru sonuçların elde edildiği kanıtlanmış olan yapay zeka yöntemleri kullanılmaktadır [7-9]. Yapay Sinir Ağları(ANN), Destek Vektör Makineleri (SVM); Karar Ağaçları (DT); Rassal Orman Regresyonu (RFR) vb. makine öğrenmesi, Tekrarlayan Sinir Ağı (RNN); Uzun Kısa Süreli Bellek (LSTM) vb. derin öğrenme gibi yapay zeka yöntemleri bir insan beyni özelliklerini göstererek doğrusal olmayan karmaşık verinin içindeki gizli özellikleri bularak gelecekle ilgili daha doğru tahminlerde bulunabilmektedirler. Özellikle son yıllarda bu yapay zeka yöntemlerine Tekrarlayan Sinir Ağı+Uzun Kısa Süreli Bellek (RNN+LSTM), Yapay Sinir Ağları+Uzun Kısa Süreli Bellek (ANN+LSTM), Evrişimli Sinir Ağı+Uzun Kısa Süreli Bellek (CNN+LSTM) gibi hibrit derin öğrenme yöntemleri de eklenmiştir. Son yapılan çalışmalarda hibrit yapay zeka yöntemlerinin daha da iyi sonuçlar verdiği görülmektedir [9,10].

Çalışmada Borsa İstanbul'da bulunan Petkim Petrokimya Holding A.Ş.'nin hisse senedi (PETKM) kapanış fiyatı RFR, LSTM ve CNN+LSTM yöntemleriyle tahmin edilmiştir. Bu bölümden sonra hisse senedi ve çalışma içerisinde kullanılan yöntemlerle ilgili literatür taraması bilgilerinin verildiği 2. bölüm, veri setinin tanıtıldığı; veri ön işleme yöntemlerini ve algoritmaların adımlarının anlatıldığı 3. bölüm, uygulamanın verilerinin paylaşıldığı 4. Bölüm, uygulama sonuçlarının karşılaştırıldığı/tartışıldığı 5. Bölüm ve uygulama sonuçlarının paylaşıldığı 6. Bölüm bulunmaktadır.

1.1. Literatür Taraması

Çalışmada kullanılan Rassal Orman Regresyonu (RFR) algoritmasının, Uzun Kısa Süreli Bellek (LSTM) algoritmasının, Evrişimli Sinir Ağı+Uzun Kısa Süreli Bellek

(CNN+LSTM) yöntemleri ile oluşturulmuş hibrit algoritmanın ve hisse senedi fiyat-borsa indeksi tahminlerinin yapıldığı çalışmaların literatürdeki özeti bu bölümde sunulmuştur.

Karakoyun [11] 2018 yılında Bitcoin fiyatını ve Apple şirketinin hisse senedi fiyatını tahmin etmek için geleneksel istatistikî tahmin yöntemi olan Birleştirilmiş Otoregresif Hareketli Ortalama (ARIMA) algoritmasını ve derin öğrenme (DL) yöntemleri arasında zaman serisi tahminlerinde büyük oranda kullanılan LSTM algoritmasını kullanmıştır. ARIMA modellerini oluşturmada R ve LSTM modellerini oluşturmada Python yazılım dillerini kullanmıştır. Oluşturduğu modelleri Kök Ortalama Karesel Hata (RMSE), Ortalama Mutlak Hata (MAE), Ortalama Yüzde Hata (MPE) ve Ortalama Mutlak Yüzde Hata (MAPE) metrik değerleri ile karşılaştırmıştır. Apple şirketinin ARIMA modeliyle tahmininde RMSE değerini 34.04762, MAE değerini 27.01132, MPE değerini -15.38245 ve MAPE değerini 15.6444 olarak; LSTM modeliyle tahmininde RMSE değerini 9.61, MAE değerini 6.48, MPE değerini 2.75 ve MAPE değerini 5.78 olarak bulmuştur. Bitcoin fiyatının ARIMA modeli ile tahmininde RMSE değerini 1146.067, MAE değerini 939.5819, MPE değerini 10.86483 ve MAPE değerini 11.86484 olarak; LSTM modeli ile tahmininde RMSE değeri 93.27, MAE değerini 81.56, MPE değerini 1.40 ve MAPE değerini 1.40 olarak bulmuştur. ARIMA ve LSTM modelleri arasında en düşük hata değerlerini LSTM algoritmasının verdiğini kanıtlamıştır.

Manjula ve Karthikeyan [12] 2019 yılında ham petrol fiyatı, rupi dolar kuru, enflasyon ve faiz oranlarını kullanarak Ocak 2000-Aralık 2018 tarihleri arasında altın fiyatını analiz etmek için Lineer Regresyon, Rassal Orman Regresyonu (RFR) ve Gradient Boosting Regresyonu (GBR) yöntemlerini kullanmıştır. Veri analizi için Ocak 2000-Aralık 2018, verilen tarih aralığı içerisindeki verileri altın fiyatının yükselme eğilimi gösterdiği Ocak 2000-Ekim 2011 ve altın fiyatının yatay bir eğilim gösterdiği Kasım 2011-Aralık 2018 olarak 3 farklı veri seti hazırlamıştır. Tarih aralığı Ocak 2000-Aralık 2018 olan veri seti için en başarılı sonucu Determinasyon Katsayısı (R^2) değeri 0.9802, MAE değeri 2808.44, MSE değeri 16294903.01 ve RMSE değeri 4036.69 olarak Rassal Orman Regresyonu vermiştir. Tarih aralığı Ocak 2000-Ekim 2011 olan veri seti için R^2 değeri 0.979, MAE değeri 1529.11, MSE değeri 5831754.475, RMSE değeri

2414.9 olarak ve tarih aralığı Kasım 2011-Aralık 2018 olan veri seti için R2 değeri 0.6288, MAE değeri 2728.56, MSE değeri 12264286.33, RMSE değeri 3502.04 olarak en başarılı sonuçları GBR vermiştir. Sonuçları karşılaştırdıklarında uzun dönemli veri setlerinde RFR ve kısa dönemli veri setlerinde GBR yöntemlerinin iyi sonuçlar verdiğini savunmuşlardır.

2019 yılında Raşo ve Demirci [13] Türkiye Menkul Kıymetler Borsası'nın Borsa İstanbul A.Ş.(BIST) endekslerinden BIST 30 endeksinin 01.01.2016-11.04.2018 tarihleri arasındaki fiyat tahminini Yapay Sinir Ağları (ANN), Destek Vektör Makineleri Regresyonu (SVR) ve Derin Sinir Ağları (DNN) yöntemleri ile yapmışlardır. Çalışmalarında oluşturdukları DNN modelinin ANN ve SVR modelleriyle 5 günlük tahminin yapmışlardır. İlk 4 gün DNN modelinin diğer iki modele göre daha iyi sonuçlar verdiği görülürken 5. gün de en iyi sonucu SVR modeli vermiştir.

Liu ve Long [8] 2020 yılında Standard and Poor's 500 (S&P 500) endeksinden 17 Aralık 2010-17 Ocak 2013 tarihleri arasında, China Minsheng Bank (CMSB) endeksinden 18 Aralık 2013-18 Ocak 2016 tarihleri arasında ve Dow Jones Industrial Average (DJI) endeksinden 17 Aralık 2014-17 Ocak 2017 tarihleri arasında endekslerin kapanış fiyatlarından oluşan 3 farklı veri seti oluşturmuşlardır. Tahminler için veri işleme aşamasında Ampirik Dalgacık Dönüşümü (EWT), tahmin aşamasında LSTM ağı ve algoritmanın optimizasyonu için Parçacık Sürü Optimizasyonu (PSO) yöntemlerini kullanmışlardır. Yeni combine ettikleri metotla literatürdeki benzer yöntemleri karşılaştıracak çalışma içerisinde önerdikleri yeni yaklaşımın diğerlerine göre çok daha iyi sonuçlar verdiğini kanıtlamışlardır.

2020 yılında Yadav ve ark. [14] Bir Hindistan Borsası olan Ulusal Menkul Kıymetler Borsası'nın (NSE) NIFTY 50 endeksi içerisindeki farklı sektörlerden şirketlerin 29.12.2008-24.05.2019 tarihleri arasındaki hisse senedi fiyat tahminlerini LSTM yöntemiyle gerçekleştirmişlerdir. Bu veri setleri için durum bilgisinin tutulduğu ve her partide unutulduğu iki tür LSTM türü uygulamışlardır. Veriler arasındaki standart sapma ve yayılım gibi değerlerin az olduğu sayısal veri setlerinde durum bilgisi tutulmayan (stateless) LSTM'in tercih edilmesi gerektiği, içerisinde metin ifadelerinin de bulunduğu modeller oluşturuluyorsa durum bilgisinin tutulduğu (stateful) LSTM

türünün tercih edilmesi gerektiği sonucuna varmışlardır. Katman sayısı hiperparametresi üzerinde değişiklikler yaparak LSTM modelini optimize etmeye çalışmışlardır. Oluşturdukları modelde en iyi sonunu tek katmanlı LSTM vermiştir. Sonucun doğruluğunu test etmek için kullandıkları tek yönlü ANOVA testinde de aynı sonuca ulaşarak doğruluğunu kanıtlamışlardır.

Vidal ve Kristjanpoller [15] 2020 yılında altın oynaklığının tahmininde CNN+LSTM hibrit yöntemini kullanmışlardır. Bu hibrit model, altın fiyatının hareketini etkileyen statik ve dinamik özellikleri girdi olarak kabul edebilmektedir. Hibrit modelin sonuçları Genelleştirilmiş Otoregresif Koşullu Değişken Varyans (GARCH), Destek Vektör Makineleri (SVM), ANN, ANN-GARCH, CNN, LSTM modelleriyle MSE metrik hata değeriyle karşılaştırılmıştır. CNN+LSTM hibrit modelinin MSE değeri GARCH modeline göre %37, SVR modeline göre %44, ANN modeline göre %26.09, ANN-GARCH modeline göre %25.99, CNN modeline göre %26.27 ve LSTM modeline göre %18.10 daha iyi bir sonuç vermiştir.

Jing ve ark. [16] 2021 yılında derin öğrenme algoritmalarından Evrişimli Sinir Ağları (CNN) ve LSTM ile duyarlılık analizini birleştirerek yeni bir yaklaşım sunmuşlardır. Bir hisse senedi forum sayfasından yatırımcıların gizli duygularını sınıflandırmada CNN algoritması kullanılmıştır. Borsa teknik göstergeleri ve yatırımcı duyarlılığı sonuçlarını LSTM yöntemi ile analiz etmişlerdir. Geliştirilen yeni yaklaşımın kullanılabilirliğinin ölçülebilmesi için Şangay Menkul Kıymetler Borsası'ndaki (SSE) altı önemli sektörden elde edilen verileri kullanmışlardır. Verilerin analizi için SVM, Genetik Algoritma+Destek Vektör Makineleri (GA-SVM), CNN, Genetik Algoritma+Evrişimli Sinir Ağları (GA-CNN), LSTM ve oluşturdukları hibrit model olan CNN+LSTM'i kullanmışlardır. Analiz sonucu MAPE hata metriği ile değerlendirilmiştir. Üç dönem ve bu üç dönemin ortalamasının MAPE değerleri sırasıyla 0.0413, 0.0405, 0.0528 ve 0.0449 ile en iyi sonucu geliştirdikleri hibrit sistemin elde ettiğini göstermişlerdir.

Hsu [17] 2021 yılında doğru portföy oluşturmanın önemini araştırmak amacıyla borsa endekslerinde performans iyileştirmek için bir hisse seçim modeli geliştirmiştir. Modelin arka planında Karar ağaçları (DT) ve Rassal Orman (RF) modelleri bulunmaktadır ve uygulama ABD Borsası'ndan S&P 500 ve İngiltere Borsası'ndan

FTSE 100 endekslerinin 01.01.2014-31.12.2018 tarihleri arasındaki verileri ile yapılmıştır. Daha etkin bir portföy oluşturmak için geliştirdiği bu Makine Öğrenmesi (ML) tabanlı yöntemi ile oluşturulan portföyleri diğeri ile karşılaştırdığında geliştimenin kullanıldığı portföylerin çok daha başarılı sonuçlar elde ettiğini öne sürmüştür.

2021 yılında Demirel ve ark. [18] İstanbul Menkul Kıymetler Borsası'nın BİST 100 endeksindeki 42 şirketin hisse senetlerinin 01.01.2010-01.01.2019 tarihleri arasındaki borsa açılış ve kapanış fiyatlarını Çok Katmanlı Algılayıcılar (MLP), SVM ve LSTM yöntemleriyle tahmin etmeye çalışmışlardır. Modeller tüm şirketlere uygulanarak RMSE, MSE ve Determinasyon Katsayısı (R^2) değerleri ile karşılaştırmışlardır. Karşılaştırmalar sonucunda MLP ve LSTM modellerinin SVM modellerinden daha iyi sonuçlar verdiğini ortaya koymuşlardır.

Literatür taramasının özet tablosu Tablo 1.1'de verilmiştir.

Tablo 1.1. Literatür taraması.

Yazarlar (Yıl)	Çalışmanın Adı	Veriler	Algoritmalar
Ebru Şeyma Karakoyun (2018)	Derin öğrenme ile zaman serilerinin gerçek zamanlı tahmini	- Bitcoin fiyatı - Apple şirketinin hisse senedi fiyatı	- ARIMA - LSTM
K. A. Manjula P. Karthikeyan (2019)	Gold Price Prediction using Ensemble based Machine Learning Techniques	- Ham petrol fiyatı - Rupi Dolar Kuru - Enflasyon ve Faiz oranları - Altın fiyatı	- Lineer Regresyon - RFR - GBR
Halil Raşo Mehmet Demirci (2019)	Predicting the Turkish Stock Market BIST 30 Index using Deep Learning	- BİST30 endeks fiyatı	- Deep Model - SVR - ANN
Hui Liu Zhiho Long (2020)	An improved deep learning model for predicting stock market price time series	- S&P 500 endeksi - CMBS İndeksi - DJI İndeksi	- Ampirik Dalgacık Dönüşümü (EWT) - LSTM - PRçacık Sürü Optimizasyonu(PSO)
Anita Yadav C. K. Jha Aditi Sharan (2020)	Optimizing LSTM for time series prediction in Indian stock market	- NIFTY 50	- LSTM - ANOVA

Tablo 1.1. (Devamı): Literatür taraması.

Yazarlar (Yıl)	Çalışmanın Adı	Veriler	Algoritmalar
Andres Vidal Werner Kristjanpoller (2020)	Gold volatility prediction using a CNN-LSTM approach	- Altın fiyatları	- CNN+LSTM - GARCH - SVM - ANN - ANN-GARCH - CNN - LSTM
Nan Jing Zhao Wu Hefei Wang (2021)	A hybrid model integrating deep learning with investor sentiment analysis for stock price prediction	- SSE'den 6 farklı sektördeki şirketlerin hisse senedi fiyatları	- CNN+LSTM - SVM - GA-SVM - GA-CNN - LSTM
Tien-Yu Hsu (2021)	Machine learning applied to stock index performance enhancement	- S&P 500 - FTSE 100	- Karar Ağaçları - Rassal Orman
Uğur Demirel Handan Çam Ramazan Ünlü (2021)	Predicting Stock Prices Using Machine Learning Methods and Deep Learning Algorithms: The Sample of the Istanbul Stock Exchange	- BİST100	- MLP - SVM - LSTM



2. MATERYAL VE YÖNTEM

2.1. Materyal

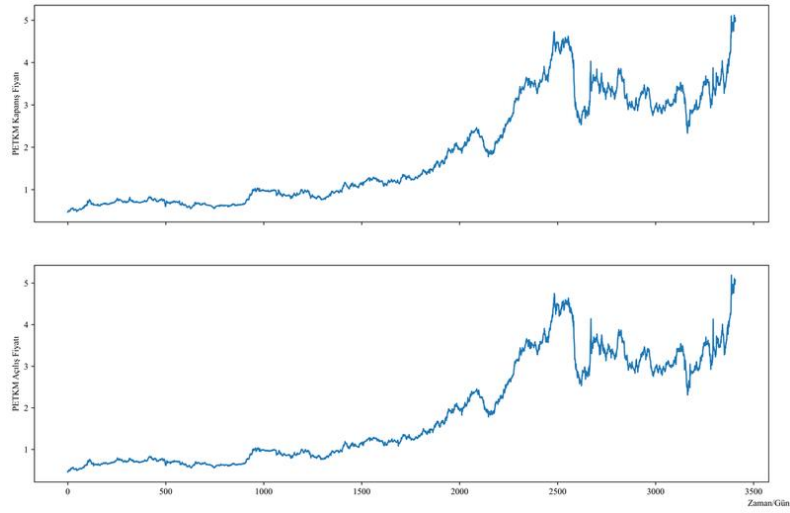
Çalışmada, Petkim Petrokimya Holding A.Ş (PETKM); Türkiye Petrol Rafinerileri A.Ş. (TUPRS); Lukoil (LKOH); Qatar Fuel Co (QFLS) hisse senedi açılış ve kapanış fiyatları, Amerikan Doları Türk Lirası alış/satı kur değeri (USD/TRY), BİST Kimyasal, Petrol & Plastik (XKMYA); MOEX Oil and Gas (MOEXOG); RTS Oil & Gas (RTSOG) ve QE Consumer Goods & Services (QECON) endekslerinin açılış ve kapanış değerlerinin 03.01.2010-31.12.2020 tarihleri arasındaki verileri kullanılmaktadır. Veriler <https://tr.investing.com/> internet adresinden alınmıştır. Bu başlık altında veri seti içerisindeki verilerle ve çalışma içerisinde kullanılacak modellerle ilgili bilgi verilecektir.

2.1.1. Veri seti

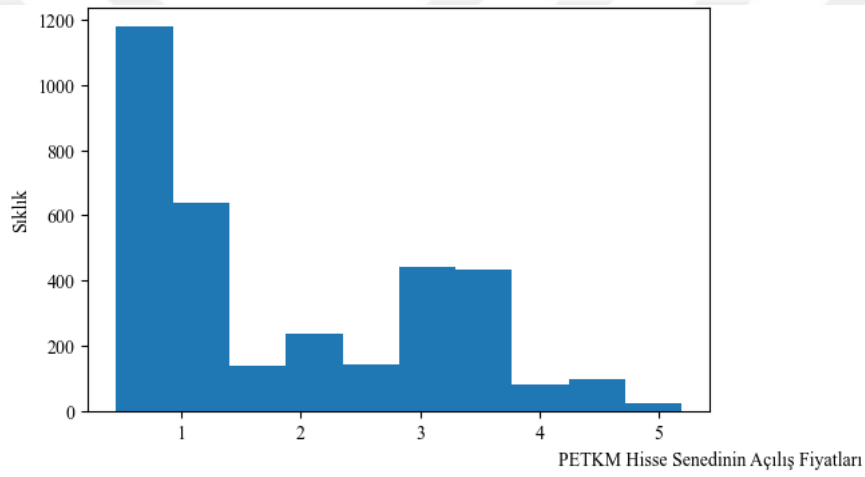
PETKM: Petkim Petrokimya Holding A.Ş. 03.04.1965 yılında Türkiye Petrolleri Anonim Ortaklığı (TPAO) liderliğinde kurulmuştur. Toplam 1.9 milyon kapasitesiyle ana ve ara petrokimyasal hammaddeler üretmektedir. Ürünleri dünya genelinde Avrupa, Asya, Birleşik Devletler, Orta Doğu ve Uzak Doğu'ya ihraç edilmektedir. 09.07.1990 tarihinden beri İstanbul Menkul Kıymetler Borsası'nda (İMKB) işlem görmektedir. BİST 100 (XU100), BİST 50 (XU050), BİST 30 (XU030), XKMYA ve bunlar gibi daha pek çok İMKB endeksinin bileşeni olarak görülmektedir [19,20].

Çalışmada tahmin edilmesi için PETKM hisse senedinin seçilmiş olmasının nedeni petrol madeninin önemli bir maden olması, gün geçtikçe daha da önem kazanması ve gelişmesidir. Çalışmanın akademik katkısının dışında yatırımcıya da yol göstereceği düşünülmektedir.

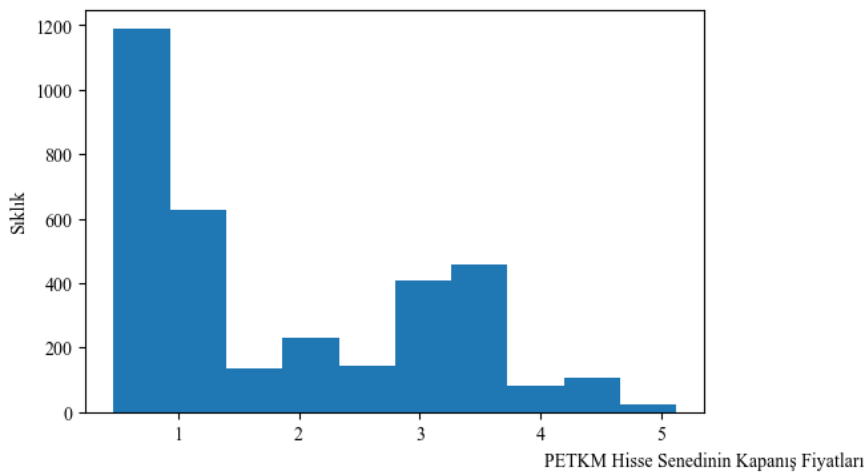
Şekil 2.1'de PETKM hisse senedinin açılış ve kapanış fiyatlarının grafiği verilmektedir. PETKM hisse senetlerinin Şekil 2.2'de açılış fiyatlarının ve Şekil 2.3'te kapanış fiyatlarının değer dağılımını göstermek için histogram grafikleri paylaşılmıştır.



Şekil 2.1. PETKM hisse senedinin açılış ve kapanış fiyatları.

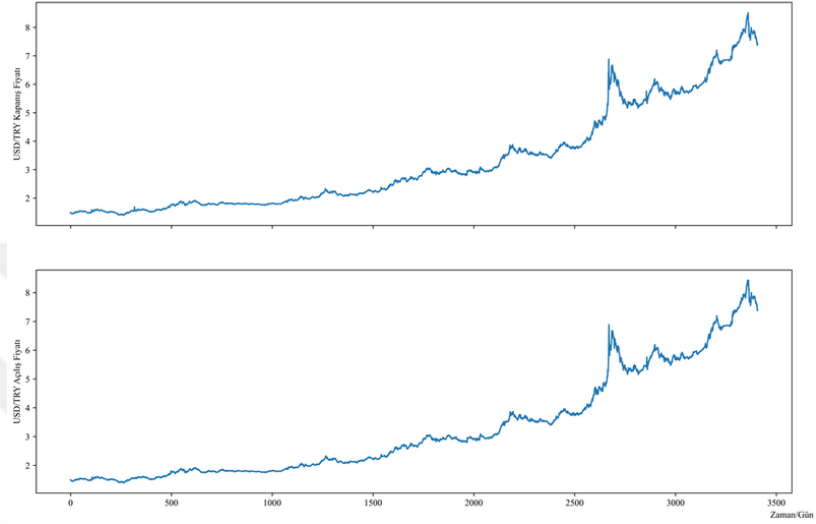


Şekil 2.2. PETKM hisse senedinin açılış fiyatlarının histogram grafiği.

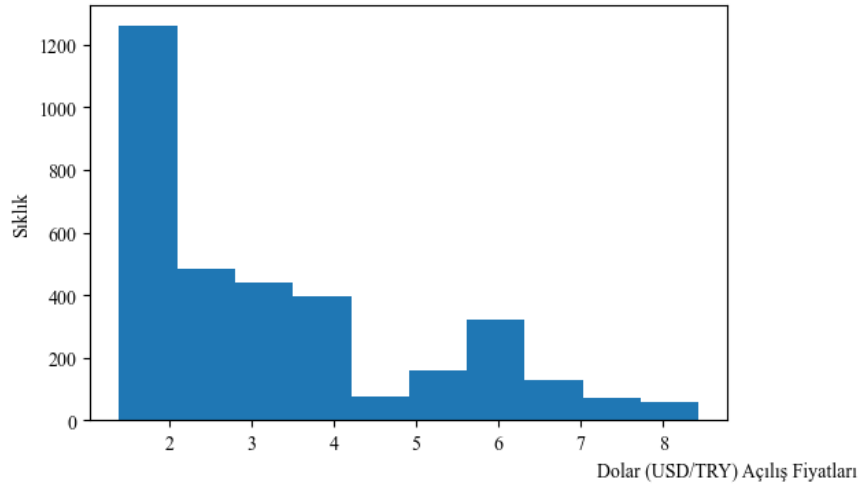


Şekil 2.3. PETKM hisse senedinin kapanış fiyatlarının histogram grafiği.

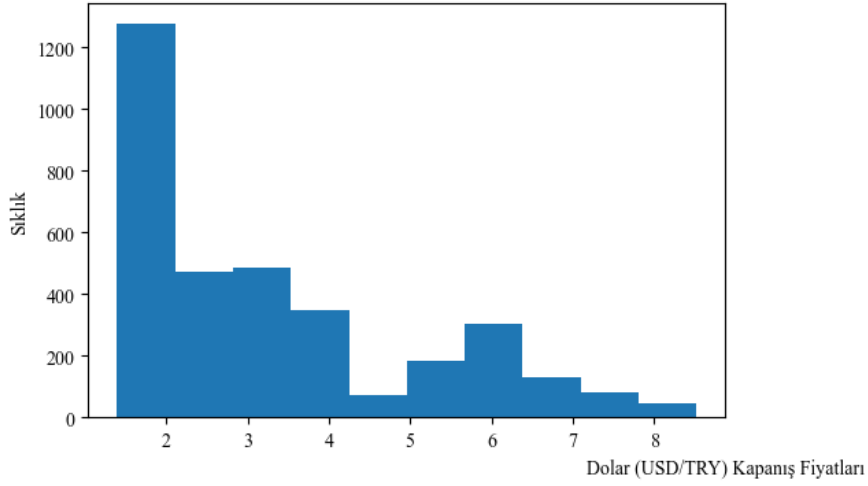
USD/TRY: Amerikan Doları (USD) dünya genelinde en yaygın kullanılan para birimidir. Türkiye'nin ithalat işlemlerinde yaygın olarak kullanılmaktadır. USD/TRY fiyatı 1 USD'nin (\$) karşılık geldiği Türk Lirası (₺) değerini göstermektedir [19]. Şekil 2.4'te USD/TRY'nin açılış ve ortalama alış/satış kur fiyat grafiği verilmektedir. USD/TRY fiyatlarının Şekil 2.5'te açılış ve Şekil 2.6'da kapanış değerlerinin dağılımını göstermek için histogram grafikleri paylaşılmıştır.



Şekil 2.4. USD/TRY açılış ve ortalama alış/satış kur fiyatı

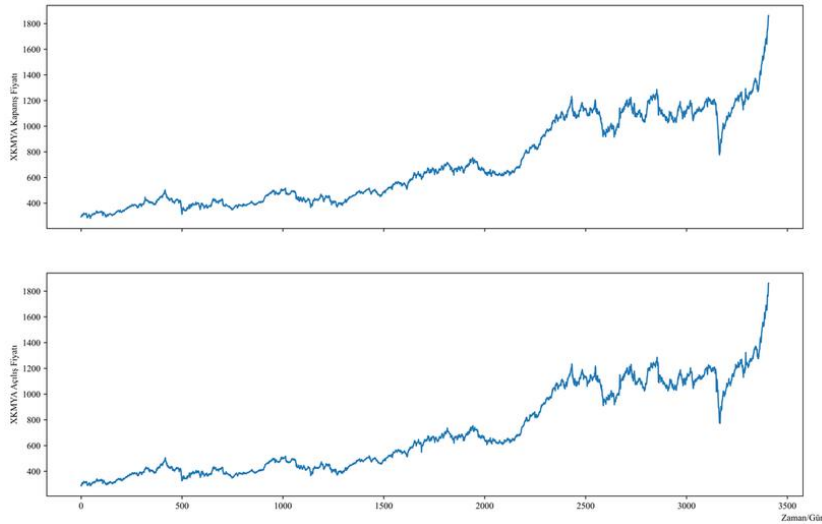


Şekil 2.5. USD/TRY açılış alış satış kur fiyatının histogram grafiği.

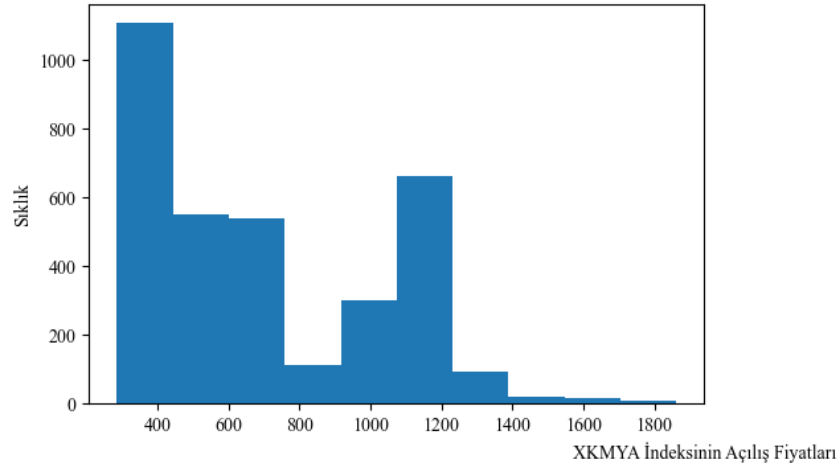


Şekil 2.6. USD/TRY ortalama alış/satış kur fiyatının histogram grafiği.

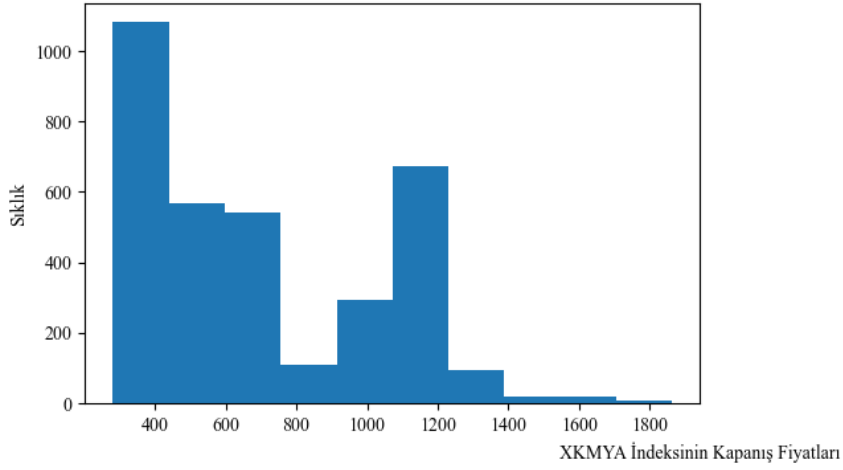
XKMYA: Borsa İstanbul (BİST) içerisindeki kimyasal, petrol ve türevleri için hammadde sektöründe bulunan şirketlerin sektörel endeksidir. İndeks 03.01.1997 tarihinden beri BİST’te işlem görmektedir. İndeksin bileşenleri arasında Aygaz A.Ş., Tüpraş, Petkim, Türk İlaç, Deva Holding A.Ş., Marshall Boya ve Vernik Sanayi A.Ş. gibi pek çok şirket bulunmaktadır [19]. Şekil 2.7’de XKMYA endeksinin açılış ve kapanış fiyatlarının grafiği verilmektedir. XKMYA endeksinin Şekil 2.8’de açılış fiyatlarının ve Şekil 2.9’da kapanış fiyatlarının değer dağılımını göstermek için histogram grafikleri paylaşılmıştır.



Şekil 2.7. XKMYA endeksinin açılış ve kapanış fiyatları.



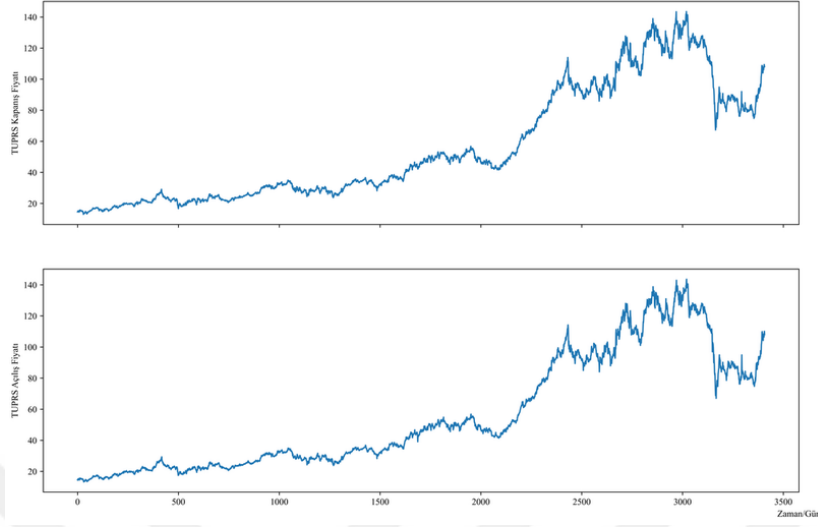
Şekil 2.8. XKMYA endeksinin açılış fiyatlarının histogram grafiği.



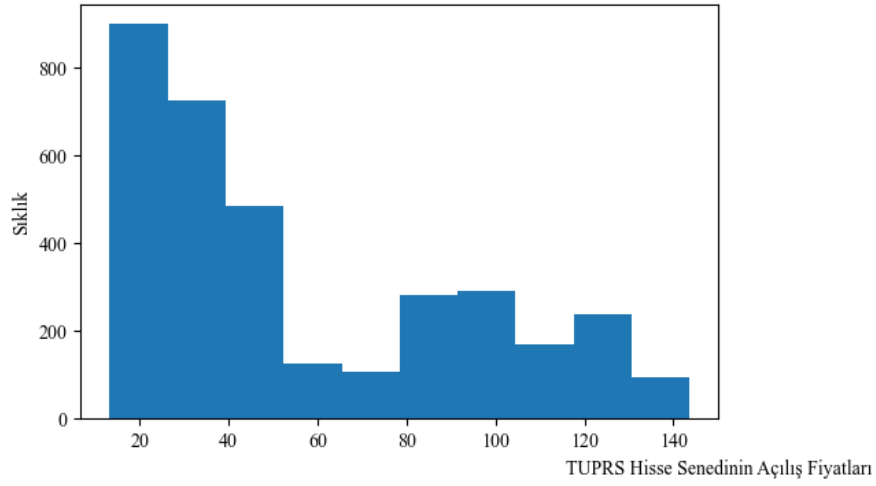
Şekil 2.9. XKMYA endeksinin kapanış fiyatlarının histogram grafiği.

TUPRS: 1955 yılında 1.1 milyon ton rafineri kapasitesiyle Türkiye'nin ilk rafinerisi olan Batman Rafinerisi ile üretime başlayan Türkiye Petrol Rafinelerileri A.Ş. günümüzde 28.1 milyon tonluk rafinaj kapasitesi ve ürünlerinin yarattığı katma değer ile Türkiye'nin en büyük sanayi şirkettir. Batman, Kocaeli, İzmir ve Kırıkkale illerinde kurduğu rafineri tesislerinde 37 farklı petrol ürünü üretmektedir. 31.05.1991 tarihinden beri İMKB'de işlem görmektedir. Bist Ulusal Tüm (XUTUM), Bist Sınai (XUSIN), XKIMYA, BİST 100 (XU100), BİST 50 (XU050), BİST 30 (XU030) vb. pek çok İMKB endeksinin ve MIEF00000PUS küresel endeksinin bileşenidir [19,21]. Şekil 2.10'da TUPRS hisse senedinin açılış ve kapanış fiyatlarının grafiği verilmektedir. TUPRS hisse senedinin Şekil 2.11'de açılış fiyatlarının ve Şekil 2.12'de

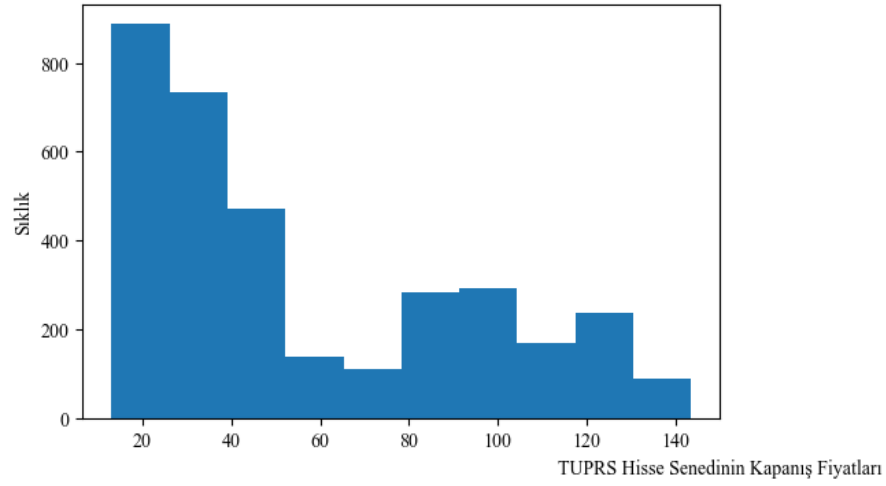
kapanış fiyatlarının değer dağılımını göstermek için histogram grafikleri paylaşılmıştır.



Şekil 2.10. TUPRS hisse senedinin açılış ve kapanış fiyatları.

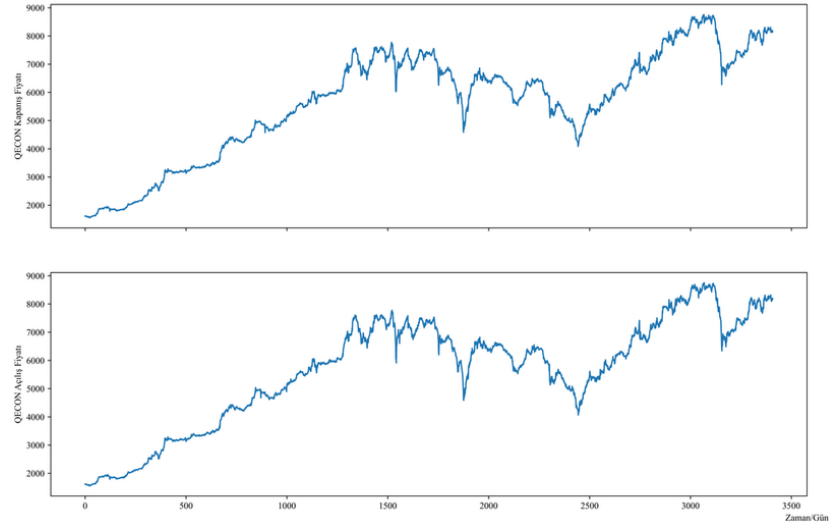


Şekil 2.11. TUPRS hisse senedinin açılış fiyatlarının histogram grafiği.

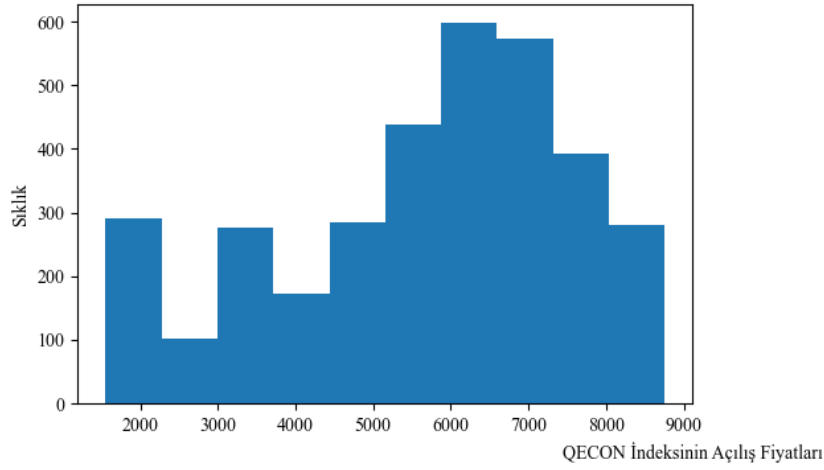


Şekil 2.12. TUPRS hisse senedinin kapanış fiyatlarının histogram grafiği.

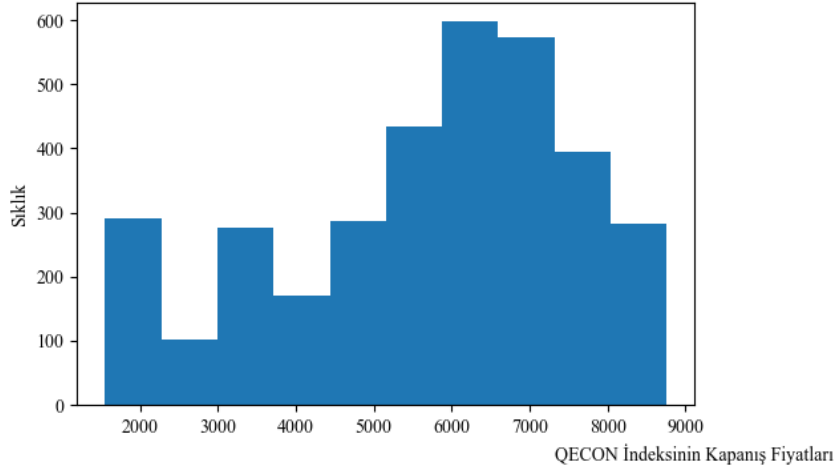
QECON: Doha Menkul Kıymetler Borsası (QSI)'nın Tüketici Malları & Hizmetler (QE Consumer Goods & Services - QECON) endeksidir. İndeks 04.01.2007 tarihinden beri QSI'da işlem görmektedir. İndeks bileşenleri arasında Medicare Group (MCGS), Baladna Food Industries Co (BLDN), QFLS gibi pek çok farklı sektörden şirket bulunmaktadır [19]. Şekil 2.13'te QECON endeksinin açılış ve kapanış fiyatlarının grafiği verilmektedir. QECON endeksinin Şekil 2.14'te açılış fiyatlarının ve Şekil 2.15'te kapanış fiyatlarının değer dağılımını göstermek için histogram grafikleri paylaşılmıştır.



Şekil 2.13. QECON endeksinin açılış ve kapanış fiyatları.

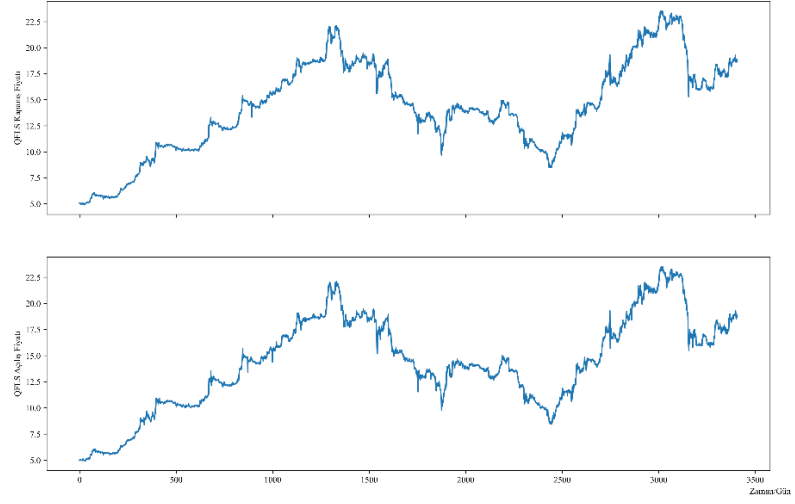


Şekil 2.14. QECON endeksinin açılış fiyatlarının histogram grafiği.

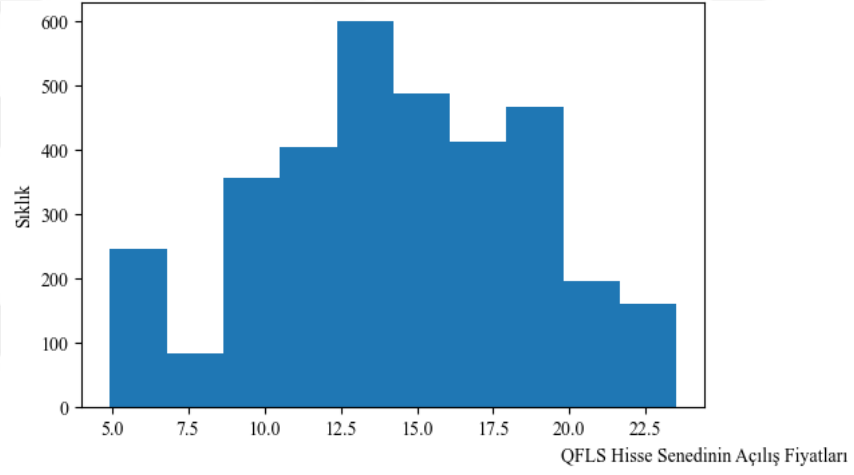


Şekil 2.15. QECON endeksinin kapanış fiyatlarının histogram grafiği.

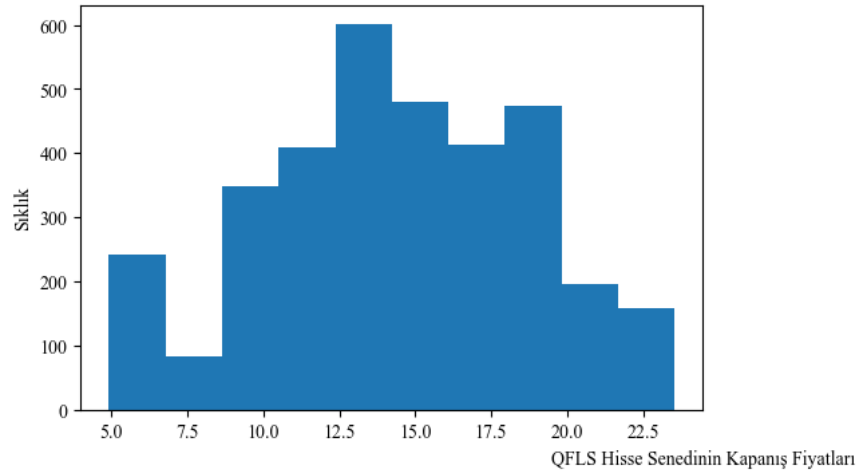
QFLS: Qatar Fuel (WOQOD) şirketi, Qatar Petroleum şirketi tarafından üretilen rafine petrolün ülke genelinde, ticari, endüstriyel alanlarda akaryakıt dağıtım ve pazarlanmasında tek imtiyaz sahibi şirkettir. 2002 yılından beri hizmet vermektedir. 16.12.2002 tarihinden beri QSI’da işlem görmektedir [19, 22]. Şekil 2.16’da QFLS hisse senedinin açılış ve kapanış fiyatlarının grafiği verilmektedir. QFLS hisse senedinin Şekil 2.17’de açılış fiyatlarının ve Şekil 2.18’de kapanış fiyatlarının değer dağılımını göstermek için histogram grafikleri paylaşılmıştır.



Şekil 2.16. QFLS hisse senedinin açılış ve kapanış fiyatları.

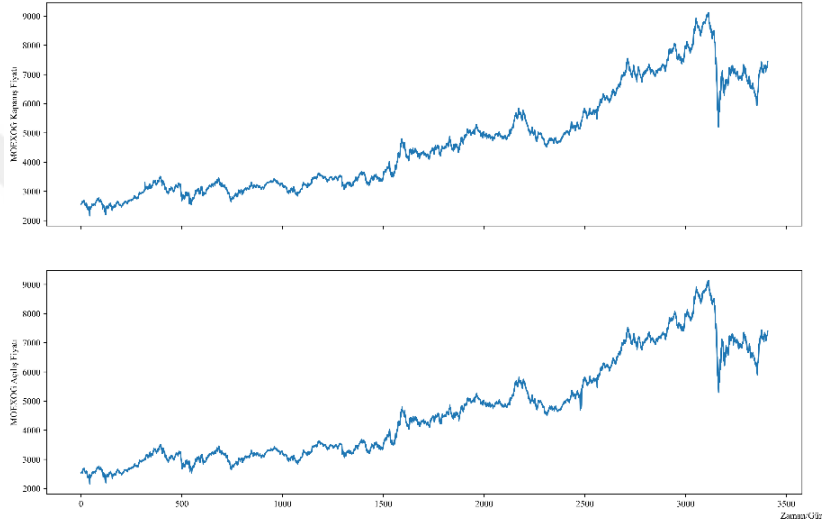


Şekil 2.17. QFLS hisse senedinin açılış fiyatlarının histogram grafiği.

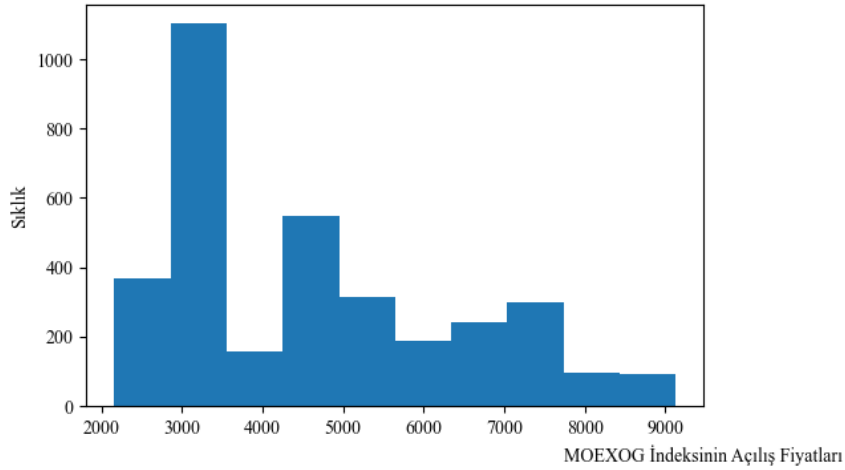


Şekil 2.18. QFLS hisse senedinin kapanış fiyatlarının histogram grafiği.

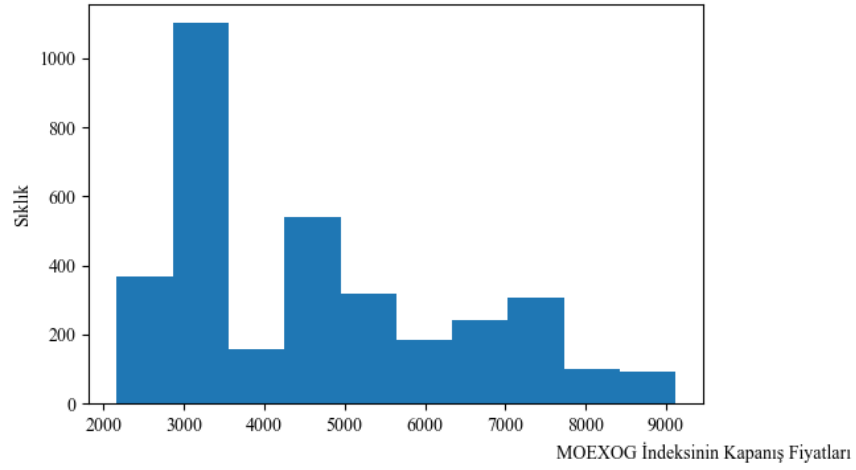
MOEXOG: Moskova Menkul Kıymetler Borsası (MOEX) içerisindeki bir petrol ve gaz sektörel endekslerinden biridir. İndeks 23.04.2007 tarihinden beri MOEX’te işlem görmektedir. İndeks bileşenleri arasında NK Lukoil PAO (LKOH), NK RussNeft PAO (RNFT), NK Rosneft PAO (ROSN) gibi doğal gaz ve petrol hisseleri bulunmaktadır. Şekil 2.19’da MOEXOG endeksinin açılış ve kapanış değerlerinin grafiği verilmektedir. MOEXOG endeksinin Şekil 2.20’de açılış fiyatlarının ve Şekil 2.21’de kapanış fiyatlarının değer dağılımını göstermek için histogram grafikleri paylaşılmıştır.



Şekil 2.19. MOEXOG endeksinin açılış ve kapanış fiyatları.

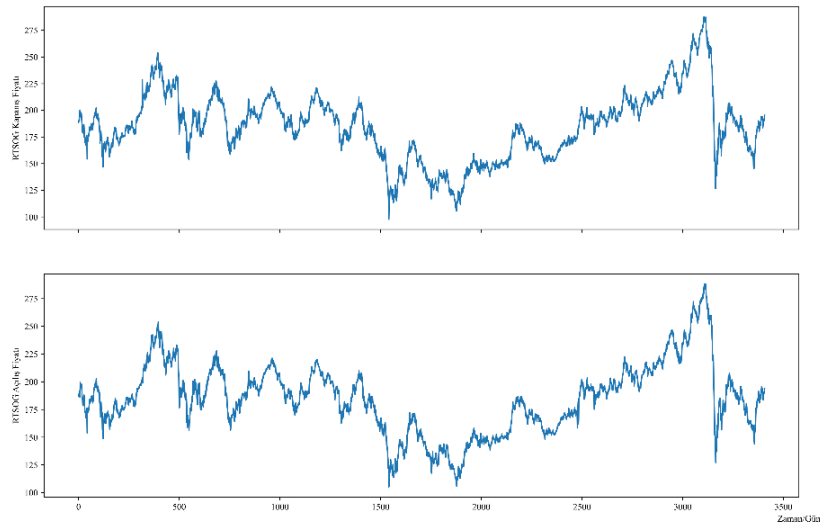


Şekil 2.20. MOEXOG endeksinin açılış fiyatlarının histogram grafiği.

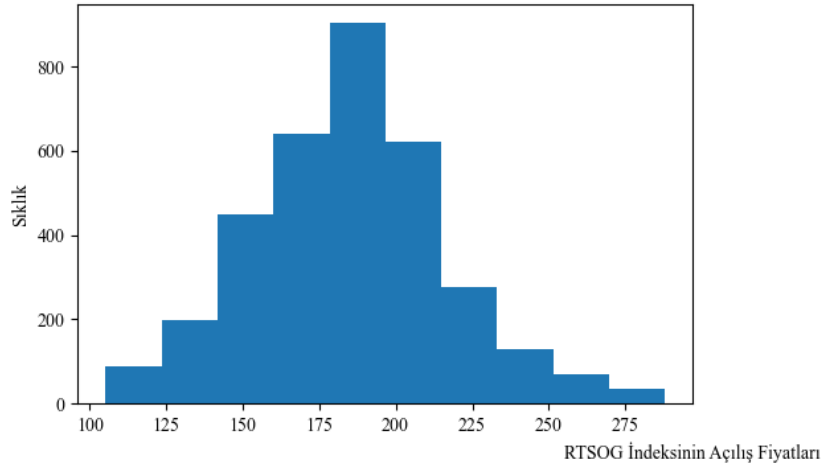


Şekil 2.21. MOEXOG endeksinin kapanış fiyatlarının histogram grafiği.

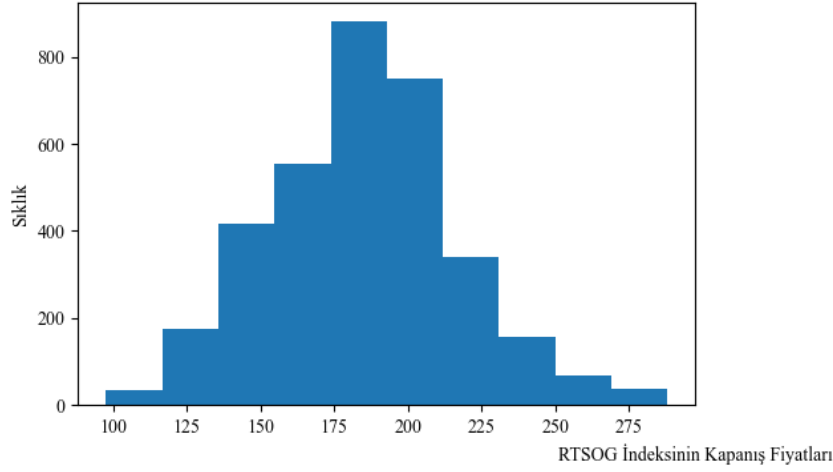
RTSOG: MOEX içerisindeki bir petrol ve gaz sektörel endekslerinden biridir. İndeks 06.01.2000 tarihinden beri MOEX'te işlem görmektedir. Endeks bileşenleri arasında NK Lukoil PAO (LKOH), NK RussNeft PAO (RNFT), NK Rosneft PAO (ROSN) gibi doğal gaz ve petrol hissetleri bulunmaktadır. Şekil 2.22'de RTSOG endeksinin açılış ve kapanış değerlerinin grafiği verilmektedir. RTSOG endeksinin Şekil 2.23'te açılış fiyatlarının ve Şekil 2.24'te kapanış fiyatlarının değer dağılımını göstermek için histogram grafikleri paylaşılmıştır.



Şekil 2.22. RTSOG endeksinin açılış ve kapanış fiyatları.

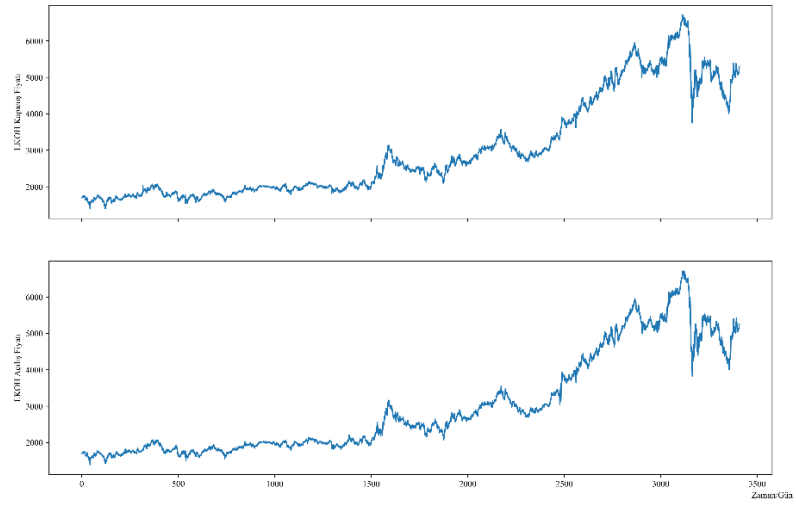


Şekil 2.23. RTSOG endeksinin açılış fiyatlarının histogram grafiği.

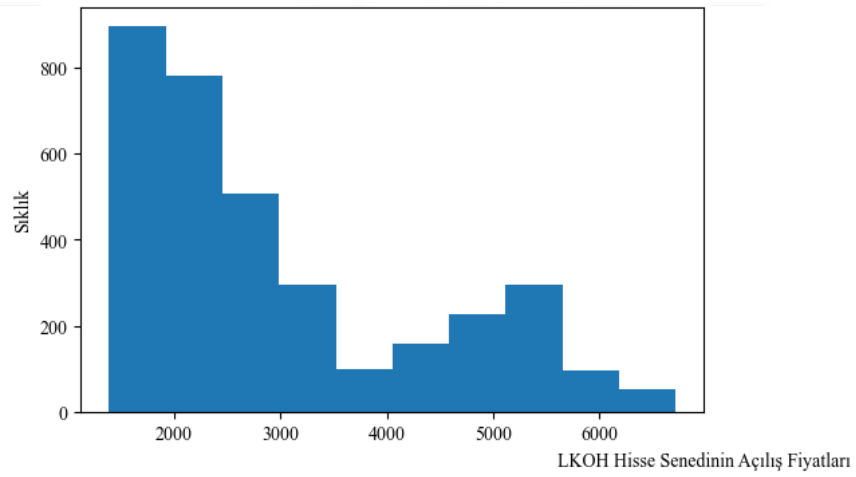


Şekil 2.24. RTSOG endeksinin kapanış fiyatlarının histogram grafiği.

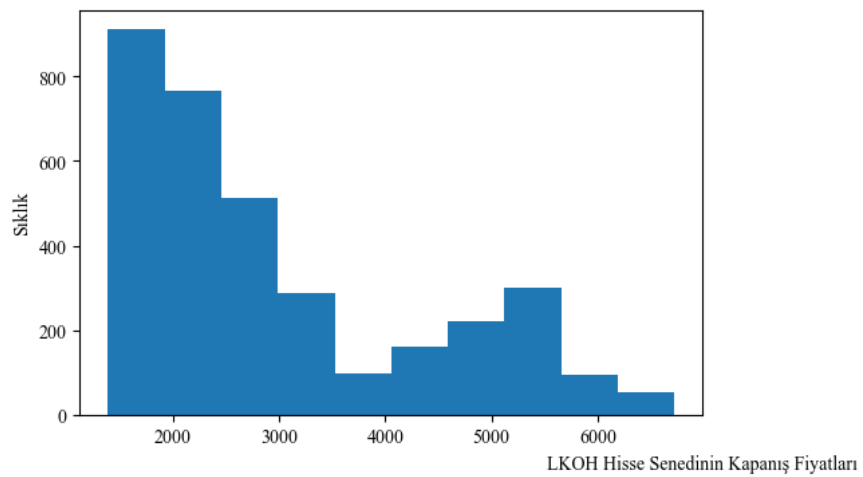
LKOH: LangepasUrayKogalym (LUKOIL) petrol endüstri grubu 25 Kasım 1991 tarihinde kuruldu. Rusya merkezli en büyük petrol şirketleri arasında yer almaktadır. MOEX, MOEX 10, MOEX Oil and Gas (MOEXOG), RTS Oil & Gas (RTSOG) gibi Rusya ve Russian USD (RTX), RTX Oil & Gas USD (RXOILUSD), RTX Oil & Gas EUR (RXOILEUR) gibi Avusturya endekslerinin bileşenidir [19, 23]. Şekil 2.25'te LKOH hisse senedinin açılış ve kapanış fiyatlarının grafiği verilmektedir. QFLS hisse senedinin Şekil 2.26'da açılış fiyatlarının ve Şekil 2.27'de kapanış fiyatlarının değer dağılımını göstermek için histogram grafikleri paylaşılmıştır.



Şekil 2.25. LKOH hisse senedinin açılış ve kapanış fiyatları.



Şekil 2.26. LKOH hisse senedinin açılış fiyatlarının histogram grafiği.



Şekil 2.27. LKOH hisse senedinin kapanış fiyatlarının histogram grafiği.

Tablo 2.1’de veri seti değişkenlerinin normalize edilmemiş değer aralıkları verilmiştir.

Tablo 2.1. Veri setindeki değerlerin değer aralıkları.

Veri Seti Değişkenleri	Aralıklar	
	Alt Değer	Üst Değer
Tarih	01.0.1.2010	31.12.2020
USD/TRY Kapanış Fiyatı	1,395	8,520
USD/TRY Açılış Fiyatı	1,392	8,432
XKMYA İndeksi Kapanış Değeri	281,410	1862,440
XKMYA İndeksi Açılış Değeri	286,270	1861,610
PETKM Hisse Senedi Kapanış Fiyatı	0,469	5,120
PETKM Hisse Senedi Açılış Fiyatı	0,453	5,190
TUPRS Hisse Senedi Kapanış Fiyatı	12,990	143,500
TUPRS Hisse Senedi Açılış Fiyatı	13,230	143,600
MOEXOG İndeksi Kapanış Değeri	2164,880	9118,110
MOEXOG İndeksi Açılış Değeri	2162,089	9127,120
RTSOG İndeksi Kapanış Değeri	97,680	287,920
RTSOG İndeksi Açılış Değeri	105,180	287,970
LKOH Hisse Senedi Kapanış Fiyatı	1393,333	6717,000
LKOH Hisse Senedi Açılış Fiyatı	1388,522	6719,000
QECON İndeksi Kapanış Değeri	1554,110	8756,770
QECON İndeksi Açılış Değeri	1554,110	8753,000
QFLS Hisse Senedi Kapanış Fiyatı	4,911	23,540
QFLS Hisse Senedi Açılış Fiyatı	4,942	23,500

2.1.2. Modeller

Çalışmada kullanılan verinin amacı, ülkemizdeki ve/veya ülkemizin yurtdışından en çok petrol ithalatı yaptığı ülkelerdeki petrol/petrokimyasal borsa endekslerinin ve önemli petrol şirketlerinin hisse senetlerinin ülkemizin petrokimya sanayisinde önemli bir yeri olan Petkim Petrokimya Holding A.Ş.’nin hisse senedi (PETKM) kapanış fiyatının üzerinde bir etkisi bulunup bulunmadığını gözlemlemektir. Tüm veri seti içerisinde dört farklı model oluşturulup tahmin işlemi gerçekleştirilmiştir. Tablo 2.2’de oluşturulan modellere ilişkin girdi ve çıktı bilgileri verilmektedir.

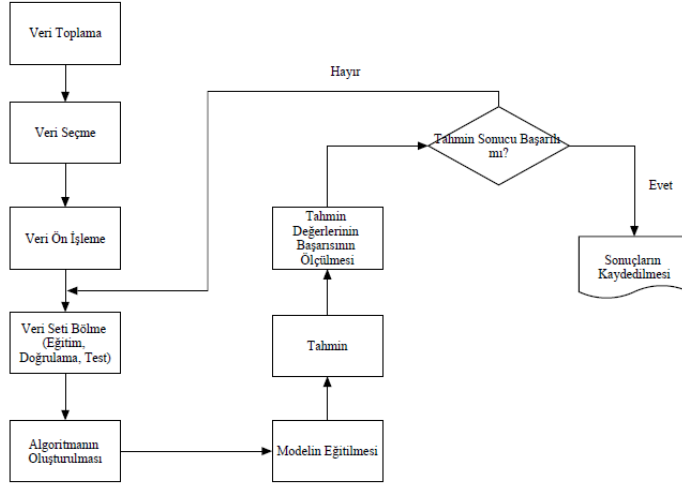
Tablo 2.2. Modellerin girdi ve çıktı şablonu.

Modeller	Girdi	Çıktı
M1	USD/TRY	PETKM
	XKMYA	
	TUPRS, PETKM	
M2	USD/TRY	PETKM
	QECON	
	QFLS, PETKM	
M3	USD/TRY	PETKM
	MOEXOG, RTSOG	
	LKOH, PETKM	
M4	USD/TRY	PETKM
	XKMYA, QECON, MOEXOG, RTSOG	
	TUPRS, QFLS, LKOH, PETKM	

- M1 Modeli – BİST Kimyasal, Petrol & Plastik (XKMYA) endeksi, Türkiye Petrol Rafinelerileri A.Ş. (TUPRS); PETKM hisse senedi fiyatlarının ve USD/TRY (dolar) açılış ve ortalama alış/satış kur değerleri girdi olarak kullanılarak PETKM hisse senedi kapanış fiyatı tahmin edilmektedir.
- M2 Modeli - QE Consumer Goods & Services (-Katar'ın tüketici malları endeksi-QECON); XKMYA endeksi, Qatar Fuel (QFLS); PETKM hisse senedi fiyatları ve dolar açılış ve ortalama alış/satış kur değerleri girdi olarak kullanılarak PETKM hisse senedi kapanış fiyatı tahmin edilmektedir.
- M3 Modeli - XKMYA; MOEX Oil and Gas (-Rusya'nın petrol endeksi-MOEXOG); RTS Oil & Gas (-Rusya'nın petrol endeksi- RTSOG), LUKOIL (LKOH); PETKM hisse senedi fiyatları ve dolar açılış ve ortalama alış/satış kur değerleri girdi olarak kullanılarak PETKM hisse senedi kapanış fiyatı tahmin edilmektedir.
- M4 Modeli - XKMYA; QECON; MOEXOG; RTSOG endeksleri, TUPRS; QFLS; LUKOIL; PETKM hisse senedi, dolar açılış ve ortalama alış/satış kur değerleri girdi olarak kullanılarak PETKM hisse senedi kapanış fiyatı tahmin edilmektedir.

2.2. Tahmin Metodu

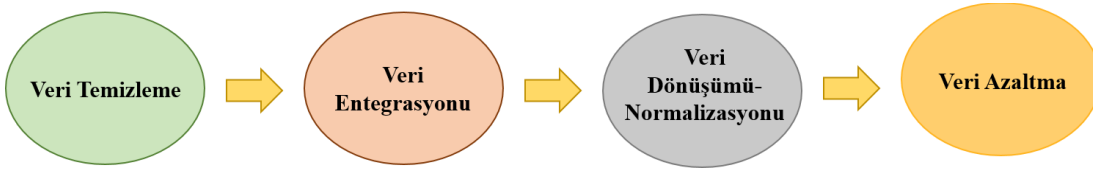
Tahmin çalışmalarında verilerin toplanması ve hazırlanması, modellerin oluşturulması, tahmin işleminin gerçekleştirilmesi ve sonuçların yorumlanması gerekmektedir. Bu çalışmadaki tahmin metodu Şekil 2.28'de verilmiştir.



Şekil 2.28. Tahmin metodu.

2.3. Veri Ön İşleme

Diğer tüm çalışmalarda olduğu gibi tahmin çalışmalarında da doğru, güvenilir ve anlaşılır sonuçlar elde edebilmek için verilerin kalitesi son derece önemlidir. Ham bir veriden kaliteli bir veri elde edebilmek için veri, çalışmanın alanına ve verinin durumuna göre veri ön işleme denilen; uygulamaya özel olarak belirlenen bir dizi süreçten geçmektedir. Genellikle ihmal edilen bu adım, bir çalışmadan elde edilecek sonuçların daha doğru ve kabul sınırları içinde olmasını sağlamaktadır [24]. Veri ön işleme adımları Şekil 2.29’da verilmektedir.



Şekil 2.29. Veri ön işleme adımları.

Veri ön işlemenin adımları aşağıda açıklanmıştır.

- Veri Temizleme: Bir veri seti içerisindeki gürültülü, eksik ve gereksiz veri üzerinde düzenleme yapılan aşamadır [25]. Veri setindeki hatalı değerlere gürültülü veri adı verilmektedir. Gürültülü verilerin tespit edilerek düzenlenmesi, düzenlenemeyecek bir durumda ise veri setinden çıkartılması veya çalışma içerisinde gürültüye duyarlılığı az olan algoritmaların

kullanılması gerekmektedir [24]. Verinin toplanması, iletilmesi ve birleştirilmesi aşamalarında oluşabilen eksik veri sorununun düzenlenmesi için herhangi bir geçmiş veriden faydalanılamıyorsa bu veri, veri setinden çıkarılabilmekte veya ilgili veri özelliğinin ortalamasını değiştirmeyecek hesaplama yöntemleriyle bu değer doldurulabilmektedir.

- Veri Entegrasyonu: Bir veri seti birden fazla sayıdaki veri deposundan toplanarak oluşturulabilmektedir. Farklı veri depolarından veri setlerinde aynı veriden birden fazla veya tutarsız veri elde edilebilmektedir. Veri seti içerisinde tutarsızlığa neden olmayan birleştirme işlemleri veri setinin genişletilmesini sağlamakta ancak maliyetlerin artmasına da neden olabilmektedir [24,25].
- Veri Dönüşümü - Normalizasyonu: Bazı sistemlerin ham verilerinin öznitelik değerleri, çalışmalarda kullanılmaya uygun olmamaktadır. Bu sebeple orjinal özniteliklerden modellerin tahmin gücünü arttırmak amacıyla, yeni öznitelikler meydana getirilmeden uygun özelliklere sahip yeni bir değer dizisi oluşturulmaktadır. Bu veri dönüşümüne normalizasyon denilmektedir [24,25].
- Veri Azaltma: Veri setinin etkinliğini artıracak bir veri ön işleme özelliği de veri azaltmadır. Çok büyük hacimdeki veri kümeleri, veri analizlerinin süresini büyük oranda uzatmaktadır. Bu adımda oluşturulacak veri seti, ham veri setinin temel özelliklerini ve bütünlüğünü koruyarak veri miktarı azaltılmaktadır.

2.3.1. Veri temizleme yöntemleri

Çalışmada kullanılan ve eksik verilerin tamamlamasında sıklıkla kullanılan 4 yöntem bu başlık altında verilmiştir.

2.3.1.1. Eksik verinin veri setinden kaldırılması

Bir zaman serisi veri setinde birden fazla özellik bulunduğunda bu özelliklerin bazılarında ortak tarihte/günde/saatte vb. veri bulunmayabilir. Bu durum gerçek hayat problemlerinde karşılaşılan bir problemdir. Bu problemin çözümü için kullanılan yöntemlerden biri eksik verinin bulunduğu tarihin veri setinden kaldırılmasıdır [25, 26]. Ancak bu yöntem ilgili zamanda bulunan diğer verilerin kaybına yol açmaktadır.

2.3.1.2. Eksik verinin 0 olarak doldurulması

Zaman serisi verilerinde eksik verilerin yerine 0 yazılması bir diğer eksik veri doldurma yöntemlerinden biridir. Eksik veri miktarı az sayıda olduğunda bu yöntem ile başarılı sonuçlar elde edilebilmektedir [17].

2.3.1.3. Aritmetik ortalama

Aritmetik Ortalama (AO) bir değer kümesindeki tüm elemanlarının toplamının eleman sayısına bölümü ile bulunur. AO Denklem 2.1’de verilmiştir.

$$\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i \quad (2.1)$$

$\bar{X}$ tüm elemanların ortalama değeri, N eleman sayısı ve X_i değer kümesinde bulunan elemanların her birini ifade etmektedir.

Kitle ve örneklem ortalama hesaplamaları için aritmetik ortalama en çok kullanılan merkezi eğilim ölçülerinden biridir. Aynı zamanda eksik veriyi tamamlama ve gürültülü veriyi temizleme de kullanılabilecek yöntemlerdendir. Bu yöntemin dezavantajı kitle veya örneklemelerde, bulunan verinin içerisinde bulunabilecek aykırı verilere yüksek oranda duyarlı olmasıdır [27].

2.3.1.4. Doğrusal enterpolasyon

Veri setindeki eksik verilerin doldurulmasında kullanılan yöntemlerden biri de doğrusal enterpolasyon yöntemidir. Enterpolasyon yöntemi ile bilinen iki nota arasındaki bir değer bulunmasını sağlamaktadır [28] Lineer enterpolasyon yöntemi Denklem 2.2’de verilmiştir.

$$y_0 = \frac{y_2 - y_1}{x_2 - x_1} (x_2 - x_0) + y_1 \quad (2.2)$$

(x_1, y_1) ve (x_2, y_2) bilinen iki noktayı, (x_0, y_0) ise bilinen iki nokta arasında bulunan bir noktayı temsil etmektedir.

2.3.2. Min-max normalizasyonu

Gerçek hayat problemlerinde kullanılan veri setlerindeki tüm özelliklerin değer aralıkları genellikle birbirine benzer olmamaktadır. Tahmin çalışmalarında bu veri setleri kullanıldığında değerleri diğer özelliklere göre büyük olan özellikler, eğitim

aşamasında ağırlık hesaplamalarında daha büyük bir etkiye sahip olmaktadır. Aynı zamanda tahmin sonuçlarının değerlendirilme aşamasında hata değerler ölçümelerini olumsuz yönde etkilemektedir. Bu nedenle veriler belirli bir sayı aralığına çekilerek bu olumsuz sonuçlar ortadan kaldırılmaya çalışılmaktadır. Bu yöntem genel olarak normalizasyon işlemi denmektedir. Normalizasyon yöntemleri arasında en sık kullanılan yöntem ise min-max normalizasyonudur. Min-max normalizasyonunda tüm özellikler kendi içlerinde [0, 1] aralığına çekilmektedir. Bir özellik içerisindeki en küçük değer 0 ve en büyük değer 1 olacak şekilde konumlandırılmaktadır [24, 28]. Min-max normalizasyon işlemi Denklem 2.3'te verilmiştir.

$$X' = \frac{(X - \min(X))}{\max(X) - \min(X)} \quad (2.3)$$

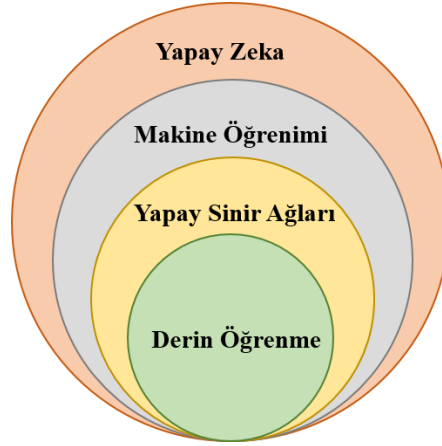
X' normalize edilmiş değeri, X orjinal değeri, $\min(X)$ özelliğin değer kümesindeki en küçük değeri ve $\max(X)$ özelliğin değer kümesindeki en büyük değeri ifade etmektedir.

2.4. Makine Öğrenmesi

1990'lı yıllarda önemi fark edilen Makine Öğrenmesi (ML), gelecekle ilgili olay ve durumları tahmin edebilmek için geçmiş verileri bilgiye dönüştüren türlü algoritmalar uygulayan bir Yapay Zeka (AI) alanıdır. Fark edilmesinden sonra büyük bir hızla ilerleyerek günümüze kadar popülerliğini korumuş ve bir o kadar da ilerlemiş disiplinler arası bir bilim dalıdır. Bir ML modeli tanımlayabilmek için istatistik, matematik, ekonomi, fizik, bilgisayar bilimi, psikoloji ve sinirbilimi gibi pek çok farklı disiplin kullanılarak geçmiş verilere göre model hiperparametreleri optimize edilir. ML yöntemleri, büyük hacimli verilerin içerisindeki temel ilişkileri anlamlandırıp karakterize edebilme yeteneği sayesinde yeni bir modelin kurulmasında, var olan bir sürecin değişen koşullarını kategorize etmede, mevcut bir sistemin sorunlarını belirlemede kullanılabilir ve var olan sorunları gidermek için yeni bilgilerle kendisini yeniden tasarlayabilmektedir [29].

ML, kendisi gibi veri tanıma, kategorize etme, veriden tahminler yapma gibi alanlarda başarılı sonuçlar veren Yapay Sinir Ağları (ANN) ve Derin Öğrenme (DL)

algortitmalarını da kapsamaktadır [29]. Şekil 2.30’da AI, ML, ANN ve DL arasındaki ilişki şeması verilmiştir.



Şekil 2.30. AI, ML, ANN ve DL arasındaki ilişki.

2.4.1. Makine öğrenmesi türleri

Makine Öğrenmesi (ML) algoritmaları kendi içlerinde 4 ana gruba ayrılmaktadırlar. ML yöntem grupları Şekil 2.31’de verilmiştir.



Şekil 2.31. ML yöntem grupları.

- Gözetimli (Denetimli) Öğrenme: Gözetimli öğrenme, bir modelin oluşturulmasında eğitim sürecinin gerçekleştirildiği algoritmaları içermektedir. Bu tür algoritmalar mevcutta bulunan girdi-çıkı eşleşmeleri ile

bir model oluřturmakta ve bu modelden öęrendiklerini kullanarak modelde olmayan eřleşmeleri tahmin etmektedirler. Bu modeller veri ierisindeki bilinen ve bilinmeyen nitelik deęerlerinin kurallarından oluřmaktadır. Eęitim verileri, belli bir ön iřleme ařamasından gemiř veriden modeli oluřturacak tüm veri özelliklerini ieren bir örneklem alınarak oluřturulmaktadır. Modellerin ıktısı, sonlu-kategorik ve belli bir sayı deęeri aralıęında sonsuz sayıdan oluřabilmektedir. Bu nedenle gözetimli öęrenme, sınıflandırma ve regresyon algoritmaları olmak üzere iki tür problem tipine sahiptir. Sınıflandırma algoritmalarının ıktıları sonlu-kategorik, regresyon algoritmalarının sonuçları ise sonsuz-sayısaldır. Yaygın olarak kullanılan gözetimli öęrenme algoritmalarına örnek olarak, Karar Aęaçları (DT), Rassal Orman Regresyonu (RFR), Destek Vektör Makineleri (SVM), ANN, LSTM vb. algoritmalar verilebilmektedir. Cinsiyet tahmini, tahlil ve hastalık teřhisleri, hava durumu tahminleri, hisse senedi ve endeks tahminleri gözetimli öęrenme problemlerine örnek olarak gösterilebilmektedir [24, 29].

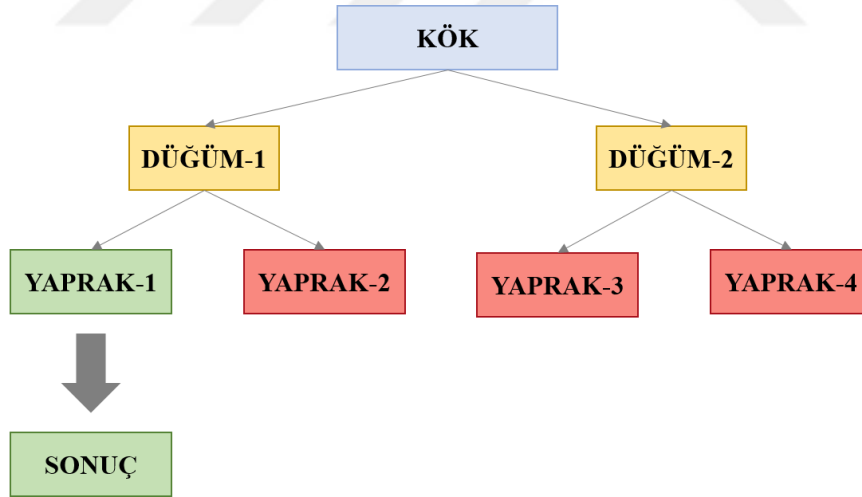
- Gözetimsiz (Denetimsiz) Öęrenme: Gözetimsiz öęrenme algoritmaları yalnızca girdi verilerini kullanmaktadır. Bu algortimaların amacı, girdi verisinde bulunan düzenlilik, düzensizlik, aykırılık ve benzerlikleri bulmaktır [24]. Gözetimsiz öęrenme algoritmaları, var olan verilerle tahmine dayalı kümeleme iřlemi ile oluřturulmuř kümelere yeni veriler ekleneceęinde ilgili model için olan kümeleme modelini öęrenebilmektedirler. Kümeleme, Boyutsal Küçölme, Anomali Tespiti ve Birliktelik Kuralları algoritmaları gözetimsiz öęrenme algoritmaları arasında gösterilmektedir. Maęazalarda-marketlerde reyon düzenleme, internet alışverişlerinde-kredi kartı kullanımında sahtekarlık tespiti vb. gibi problemlerde gözetimsiz öęrenme algoritmaları kullanılmaktadır [29].
- Takviyeli Öęrenme: Takviyeli Öęrenme algoritmaları, bir hedefe ulaşmak için yapılan eylemlere dayalı algoritmalarlardır. Algoritmanın başarısı tek bir eyleme baęlı olmamaktadır. Hedefe ulaşmak için yapılan her eylemden yeni bir bilgi öęrenilmektedir. Yeniden bařlayan her öęrenme sürecinden gemiř deneyimler göz önüne alınarak en iyi eylem strajesi oluřturulmaya alışılmaktadır.

Robotik kodlama ve oyun geliřtirmeleri bu öğrenme tipine örnek olarak verilebilmektedir [29].

- Örneęe Dayalı Öğrenme: Örneęe Dayalı Öğrenme, bir veri setiyle yeni veriyi test etmek için eğitimli bir model oluşturmak yerine yeni ve görölmemiş verilerin etkilerini gözlemlemek için ham veri noktalarını kullanan algoritmaları içerir. Verilerin özellikleri, boyutları ve uzaklıkları dikkate alınarak yeni veri noktasının bulunması sağlanmaktadır [29].

2.4.2. Karar ağaçları

Karar Ağacı (DT) algoritmaları, eğitim ve test verisi olarak ayrılmış iki kümeden eğitim kümesinde veri setinin kuralını yansıtan tutarlı bir ağaç yapısı oluşturup bu yapıyı test verilerine uygulayan gözetimli öğrenme yöntemleridir. Sınıflandırma ve regresyon olmak üzere iki tür karar ağacı bulunmaktadır. Ağacın, hedef deęişkeninin ayrık küme deęerlerinden oluşması sınıflandırma, sürekli küme deęişkenlerden oluşması regresyon ağacı olduğunu göstermektedir. Veri setindeki yinelemelerin bulunduğu verilerde her adımda farklı kararlar verilmektedir [17, 30]. Şekil 2.32’de DT algoritmalarının yapısı verilmektedir.



Şekil 2.32. Karar ağacı algoritmalarının yapısı.

DT algoritmaları kendi aralarında, Entropiye Dayalı Algoritmaları ve Sınıflandırma - Regresyon Ağaçları (CART) olmak üzere 2’ye ayrılmaktadır. Entropiye Dayalı Algoritmalarda karar ağacı dallanmaları, veri setini içeren özelliklerin entropisi ve

bilgi kazanımları ile bulunmaktadır. Entropi veri setindeki belirsizliği temsil etmektedir. Entropi değeri $[0, 1]$ aralığında değişmektedir. Bu değer 1'e yaklaştıkça belirsizlik artmaktadır. Bilgi kazanımı, entropi değerini göz önüne alarak veri setindeki özelliklerin sağlayacağı bilgi kazanımını tespit etmektedir. Bilgi kazanımı her özellik için uygulanmakta ve en yüksek bilgi kazanımına sahip değer düğüm oluşturmak için seçilen değer olmaktadır. Entropi değeri Denklem 2.4'te verilmiştir [30].

$$H(T) = - \sum_{i=1}^n p_i \log_2(p_i) \quad (2.4)$$

$H(T)$ hedef değişkenin entropisini, i bir özellik içerisindeki her bir durumunu, p_i özellik içerisindeki tüm durumların olasılıklarını temsil etmektedir.

Bir veri seti içerisindeki özelliklerin entropileri Denklem 2.5'te verilmiştir.

$$H(X_k) = - \sum_{i=1}^n \frac{|T_i|}{|X_k|} \log_2 \frac{|T_i|}{|X_k|} \quad (2.5)$$

$H(X_k)$ her bir özelliğin entropi değerini, T_i özelliğin içerisindeki sınıflarını ifade etmektedir.

Özelliklerin hedef değişken ile olan entropileri Denklem 2.6 ile hesaplanmaktadır.

$$H(X, T) = \sum_{i=1}^n \frac{|X_k|}{|X|} H(X_k) \quad (2.6)$$

$H(X, T)$ bir özellik içerisindeki sınıfların hedef değişkenine bağımlı olan entropisini, $\frac{|X_k|}{|X|}$ değeri bir özelliğin içerisindeki sınıfın özellikteki tüm değerlere oranını ifade etmektedir.

Bir özellikten elde edilen bilgi kazancı Denklem 2.7'de verilmektedir.

$$\text{Bilgi Kazancı}(X, T) = H(T) - H(X, T) \quad (2.7)$$

Bilgi kazancı tüm özellikler için hesaplanmaktadır. Hesaplamalar sonucunda en yüksek bilgi kazanına sahip özellik karar ağacının düğümünü oluşturmaktadır. Karar ağacına düğüm olarak seçilmiş özellik/özellikler dışındaki tüm özellikler için aynı adımlar tekrarlanmakta ve bir karar ağacı oluşturulmaktadır [30].

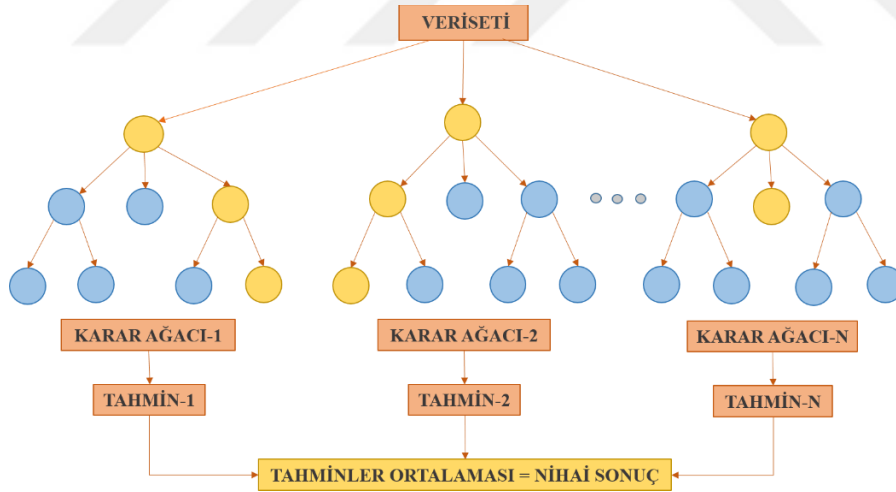
CART algoritmalarından en yaygın kullanılan yöntem ise Gini algoritmasıdır. Bir karar ağacını oluşturacak dallanmalar ikiye bölünmektedir. Her bir dal için Denklem 2.8'deki indeks hesaplanmaktadır.

$$\text{Gini İndeks} = 1 - \sum_{i=1}^n p_i^2 \quad (2.8)$$

P_i değeri bir daldaki özelliğin içerisindeki sınıfın oranını temsil etmektedir. Gini indeks değeri tüm özellikler için hesaplanmaktadır. En düşük indeks değerine sahip olan özellik karar ağacının düğümünü oluşturmaktadır [30].

2.4.3. Rassal orman regresyonu

Rasgele Orman Regresyonu (RFR) tahmin konusunda Makine Öğrenmesi (ML) yöntemleri arasında hem sınıflandırma hem de regresyon problemlerinde en iyi sonuçları veren topluluk (ensemble) algoritmaları arasındadır. Veri seti eğitim ve test olarak ikiye bölünüp, algoritmanın tahmin için gerekli öğrenimleri kazanması sağlanır. RFR algoritması tahmin çalışmalarının başarısının yanında veri seti içerisindeki en önemli özellikleri bulma konusunda da son derece başarılıdır. Şekil 2.33'te RFR algoritmasının genel yapısının şeması verilmiştir [29, 30].



Şekil 2.33. RFR algoritmasının yapısı.

Şekil 2.33'te görüldüğü gibi yöntem birçok karar ağacından elde edilen sonuçların ortalamasını alarak nihai sonuca ulaşmaktadır. Oluşan tüm karar ağaçları için veri setindeki tüm özelliklerden rastgele seçimler yapılmasıyla, bir tahmin probleminde kullanılan verilerin eğitimi aşamasında meydana gelen aşırı uydurma (overfitting)

sorununun önüne geçebilmektedir. Veri seti içerisinde bulunan aykırı ve gürültülü değerlere karşı tolerans gösterebilmektedir. Lineer problemlerde olduğu kadar lineer olmayan veride de son derece başarılı sonuçlar vermektedir. RFR algoritması, birçok alandaki ve özellikteki veri setlerine iyi bir şekilde uyum sağlayabilmekte ve modelleri basit bir şekilde kurulabilmektedir. Bu sayede RFR algoritması zaman serisi tahmin çalışmalarında güvenli sonuçlar elde edilebilecek algoritmalar arasında gösterilmektedir [12, 30, 31]. Bu nedenle bu yöntem çalışmada kullanılmıştır.

RFR algoritması bir sınıflandırma problemi için kullanıldığında verileri belirli sınıflara ayıracak etiketler, regresyon problemi için kullanıldığında Θ olarak rassal değerler içeren bir vektör kullanılmaktadır. k adet karar ağacının $\Theta_1, \Theta_2, \dots, \Theta_k$ adet tahmin vektörü bulunmaktadır. Tüm tahmin vektörleri kendisinden önceki tahmin vektörlerine bağımlı olarak oluşturulmaktadır. x girdi vektörü ve Θ vektörüne bağlı olarak birçok karar ağacından oluşan karar modeli, $h(x, \Theta)$ olarak ifade edilmektedir. K adet ağaç topluluğunun tahmin modeli Denklem 2.9'daki gibi oluşturulmaktadır [31, 32].

$$\text{Sequence} = \{h(X; \Theta_1), h(X; \Theta_2), \dots, h(X; \Theta_k)\} \quad (2.9)$$

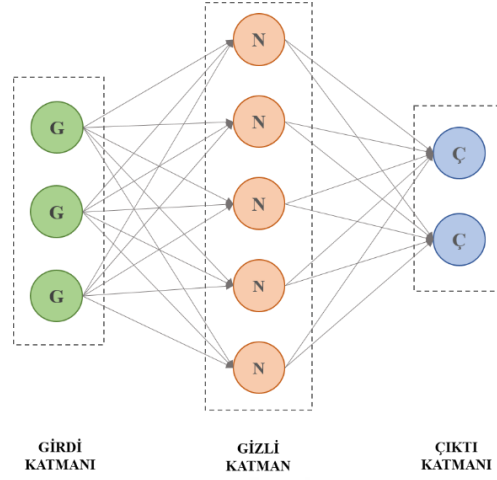
K adet ağacın tahmin değerlerinin ortalamasıyla modelin nihai tahmin sonucunu vermektedir. Modelin tahmin sonucu Denklem 2.10'daki gibi ifade edilmektedir.

$$\bar{h}(X) = (1/K) \sum_{k=1}^K h(X; \Theta_k) \quad (2.10)$$

2.5. Yapay Sinir Ağları

Gerçek hayat problemlerinin çözümünde kullanılan veri setleri genellikle lineer bir düzene sahip değildir. Bu lineer olmayan veriler içeren problemler için tahmin çalışmalarında kullanılan geleneksel tahmin yöntemleri yeterli düzeyde doğruluk başarısına ulaşamadığı için bilim insanları tarafından insan beyninin öğrenme yeteneğinden esinlenerek bir matematiksel model geliştirilmiştir. Yeni bilgileri öğrenebilme, keşfedebilme, öğrendiklerinden yeni bilgiler türetebilme, öğrendiği bilgileri analiz edebilme vb. yetenekleri olan bu uygulamaya Yapay Sinir Ağı (ANN) denmektedir [33-36].

ANN algoritmaları temelde yapay sinir hücreleri (nöronlar) ve katmanlardan oluşmaktadır. Şekil 2.34'te ANN algoritmasının katmanları verilmektedir.

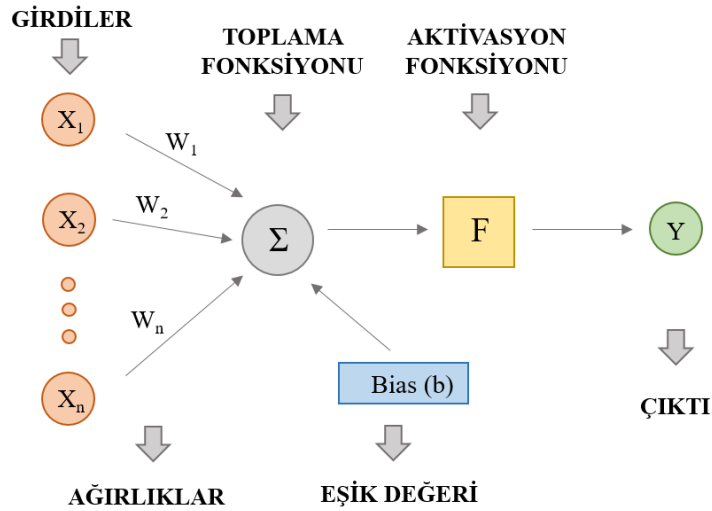


Şekil 2.34. ANN algoritmasının katmanları ve genel yapısı.

Bu katmanlar [33]

- Girdi Katmanı: Bu katmanın görevi modele girişi yapılacak yeni bilgilerin alınması ve gizli katmanlara iletilmesidir. Bu katmanda bilgi üzerinde herhangi bir işlem gerçekleştirilmemektedir. Bir ANN modelinde yalnızca bir adet girdi katmanı bulunmaktadır.
- Gizli Katman (lar): Bilgilerin işleme prosesinin uygulandığı adımdır. Prosesten geçen bilgiler çıktı katmanına iletilmektedir. Bir ANN modelinde gizli katmanların miktarı problemin karmaşıklığına göre değişiklik göstermektedir.
- Çıktı Katmanı: Bir dizi işlemde geçen bilgilerin çıkışının yapıldığı katmandır. Bir ANN modelinde yalnızca bir adet çıktı katmanı bulunmaktadır.

Bir sinir hücresi, girdiler; ağırlıklar; toplama (birleştirme) fonksiyonu; aktivasyon fonksiyonu ve çıktılarından oluşmaktadır.



Şekil 2.35. Bir sinir hücresinin yapısı.

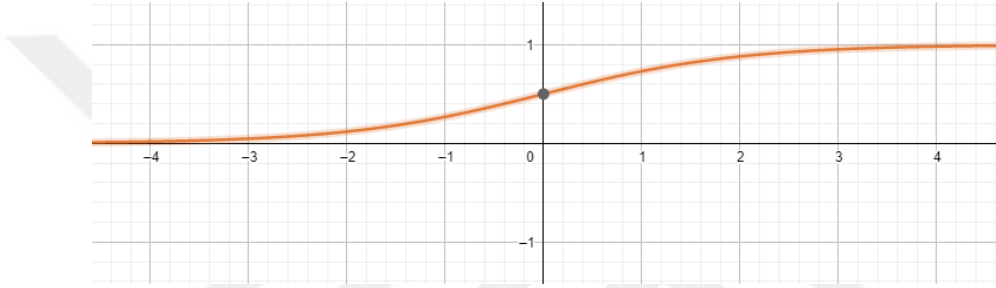
Şekil 2.35’te bir sinir hücresinin yapısı verilmiştir. Bilşenler [28-34],

- Girdiler: Bir sinir ağı hücresine dışarıdan veya başka bir sinir ağı hücresinden gelen bilgilerdir.
- Ağırlıklar: Bir sinir ağı modeli oluştururken kullanılan girdilerin model üzerinde belli etkileri vardır. Bu etki değerlerinin belirlenmesinde ağırlıklar kullanılmaktadır. Toplama fonksiyon prosesine girecek veriler öncelikle sabit veya değişken olabilen ağırlık değerleriyle çarpılmaktadır.
- Toplama Fonksiyonu: Bir sinir ağı hücresine gelen net girdi burada hesaplanmaktadır. Net girdi değeri bulunurken kullanılan Bias değeri aktivasyon fonksiyonunun yönünü belirlemede önemli rol oynar. Girdilerin oluşturdukları sinyallerin toplam değeri 0 olduğunda öğrenme gerçekleşmemektedir. Bias değerinin her durumda 1 olması, oluşan sinyallerin 0’dan farklı olmasını ve bu şekilde sinir ağının devamlılığını sağlamaktadır.
- Aktivasyon Fonksiyonu: Net girdi kullanılarak çıktının belirlenmesini gerçekleştirmektedir. Farklı karakteristik özelliklere sahip olan birçok aktivasyon fonksiyonu bulunmaktadır. Model içinde kullanılan aktivasyon fonksiyonu lineer bir fonksiyon olduğunda, modelin gizli katman içermesi ve katmanın nöron sayısı lineer olmayan bir hesaplama yapılmasını sağlamayacaktır. Bu nedenle lineer olmayan bir veri seti için kullanılan

aktivasyon fonksiyonunun da lineer olmaması modelin başarısı açısından önemlidir. Sık kullanılan aktivasyon fonksiyonlarından bazıları aşağıdaki gibidir [28, 36, 38].

1. Sigmoid Fonksiyonu: 0 ve 1 arasında değer alan, lineer olmayan ve türevlenebilen bir fonksiyondur. Bir fonksiyonun türevlenebilir olması, oluşturulan modelin öğrenebilen bir model olduğunu göstermektedir. Denklem 11'de sigmoid fonksiyonu ve Şekil 2.36'da Sigmoid fonksiyonunun grafiği vermektedir.

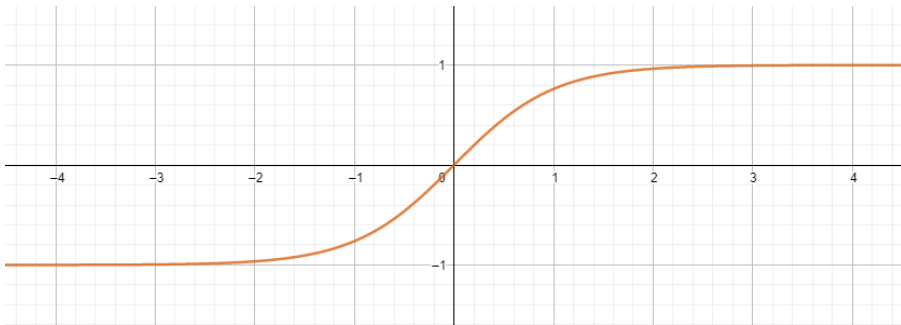
$$\sigma(z) = \frac{1}{1+e^{-z}} \quad (2.11)$$



Şekil 2.36. Sigmoid fonksiyonunun grafiği.

2. Hiperbolik Tanjant Fonksiyonu: -1 ve 1 arasında değer alan, lineer olmayan ve türevlenebilen bir fonksiyondur. Negatif değerleri de eşleyebilmektedir. Denklem 2.12'de Hiperbolik Tanjant fonksiyonu ve Şekil 2.37'de Hiperbolik Tanjant fonksiyonunun grafiği vermektedir.

$$\sigma(z) = \frac{1}{1+e^{-z}} \quad (2.12)$$

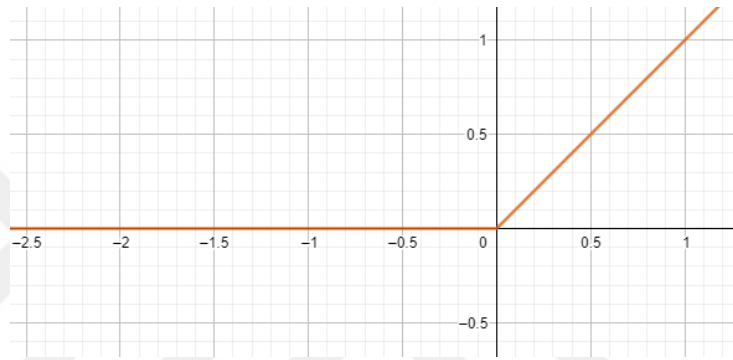


Şekil 2.37. Hiperbolik tanjant fonksiyonunun grafiği.

3. ReLU Fonksiyonu: En çok kullanılan aktivasyon fonksiyonudur. Negatif deęerleri iin fonksiyon sonucu 0 dndrlr ve trevlenebilir bir fonksiyon deęildir. Deęer aralıęı 0 ve $+\infty$ arasındadır. Denklem 2.13a’da 0’dan kk deęerler iin, Denklem 2.13b’de 0 ve 0’dan byk deęerler iin ReLU fonksiyonu ve ekil 2.38’de ReLU fonksiyonunun grafięi vermektedir.

$$f(z) = 0, \quad z < 0 \quad (2.13a)$$

$$f(z) = z, \quad z \geq 0 \quad (2.13b)$$



ekil 2.38. ReLU fonksiyonunun grafięi.

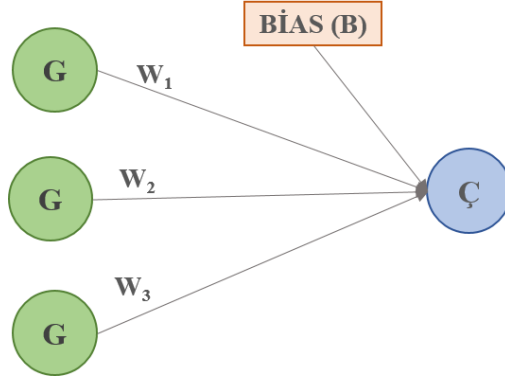
2.5.1. Yapay sinir aęlarının sınıflandırılması

Gerek hayat problemlerinde belirli bir modelin tm veri setleri iin uygun olma durumu bulunmamaktadır. Bařka bir deyiřle her veri iin farklı bir model uyarlanmalıdır. Bu ANN modelleri temelde katman sayılarına ve baęlantı řekillerine gre sınıflandırılmaktadır [34].

2.5.1.1. Katman sayılarına gre yapay sinir aęları

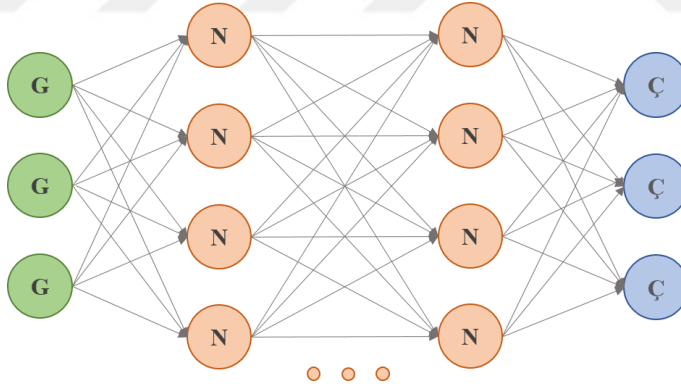
ANN modelleri katmanlardan oluřmaktadır. Bir modelin katman sayısı belirtilirken girdi katmanı dahil edilmemektedir. Yalnızca girdi ve ıktı katmanlarından oluřan ANN modellerine Tek Katmanlı Yapay Sinir Aęı (SLNN) denmektedir. Bu tr sinir aęlarında gizli katmanlar bulunmamaktadır. Bu nedenle doęrusal olmayan verilerin tahmininde yetersiz kalacaęından yalnızca doęrusal verilerin tahmininde kullanılmaktadır. ıktı fonksiyonu, 1 veya -1 olmak zere iki deęer alabilmektedir.

Çıktı değerinin 1 veya -1 olmasına göre veriler sınıflandırılmaktadır [34, 39]. Şekil 2.39'da tek katmanlı bir ANN modelinin yapısı verilmektedir.



Şekil 2.39. Bir ANN algoritmasının yapısı.

Girdi ve çıktı katmanlarının dışında gizli katmanları da içeren ANN türlerine Çok Katmanlı Sinir Ağı (MLP) denmektedir. Bu türden olan sinir ağları içeride bulunan doğrusal olmayan aktivasyon fonksiyonları sayesinde doğrusal olmayan ve karmaşık veri türlerine uyum sağlayabilmektedir. Verinin karmaşıklığına göre gizli katman sayısı artırılabilir [25, 37, 38]. Şekil 2.40'ta çok katmanlı bir ANN modelinin yapısı verilmektedir.

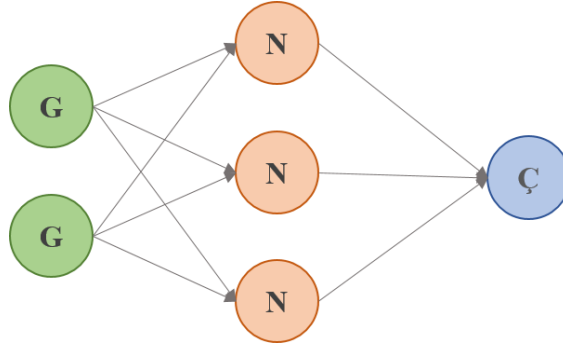


Şekil 2.40. Bir MLP algoritmasının yapısı.

2.5.1.2. Bağlantı şekillerine göre yapay sinir ağları

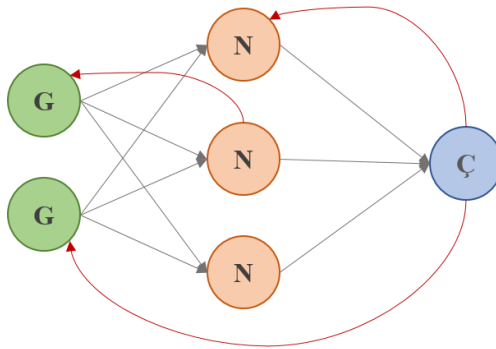
Yapay Sinir Ağları (ANN) bağlantı şekillerine göre birbirlerinden ayrılmaktadırlar. Bilgi akışının yalnızca girdi katmanı, gizli katmanlar ve çıktı katmanına doğru olarak tek yönlü aktığı ANNler'e İleri Beslemeli Sinir Ağı denmektedir. Bu tür ANNler'de yalnızca sıralı katmanlar arasında ileri doğru bağlantılar bulunmaktadır. Girdi

katmanında herhangi bir işlem yapılmamaktadır. Ağın gizli ve çıktı katmanlarında bilgi işlenerek çıktı belirlenir. [35, 39, 40]. Şekil 2.41’de ileri beslemeli bir sinir ağının yapısı verilmiştir.



Şekil 2.41. İleri beslemeli bir sinir ağının yapısı.

ANNler’in bir diğer bağlantı şekline verilen isim ise Geri Beslemeli Sinir Ağı’dır. Bu tür ANN’lerde her katman yalnızca kendisinden bir sonraki katmanla bağlantılı değildir. Model içerisinde en az bir katmanın kendisinden önceki bir katmanla bağlantısı bulunmaktadır. Bu durum ağın dinamik bir yapıya sahip olmasını sağlamaktadır. Hata miktarları gizli katmanlara geri dönmektedir. Bu şekilde katmanlardaki tüm ağırlıklar güncellenmekte ve ağın optimize edilmesi sağlanmaktadır. [28, 35, 40]. Şekil 2.42’de geri beslemeli bir sinir ağının yapısı verilmiştir.



Şekil 2.42. Geri beslemeli bir sinir ağının yapısı.

2.6. Derin Öğrenme

Zaman içinde verilerin tahmininde kullanılan yöntemlerin, güvenli ve doğru tahmin performansının beklenenden düşük olmasıyla Yapay Sinir Ağı (ANN) ilerletilerek

Derin Öğrenme (DL) yöntemleri bulunmuştur. DL algoritmaları aynı ANN algoritmalarında olduğu gibi girdi, çıktı ve gizli katmanlardan oluşmaktadır. Ancak ANN algoritmaları ile tahminde ağırlıklar, sapma değerleri vb. gibi birkaç hiperparametre bulunurken DL algoritmalarında bu hiperparametrelerden çok daha fazla bulunmaktadır. DL algoritmalarının bu veri çözümündeki çok yönlü hiperparametreleri ile diğer yöntemlerle başarılı sonuçlar elde etmede direnen verileri çok daha kolay bir şekilde tahmin edebilmektedir. Bu şekilde karmaşık ve büyük hacimdeki verinin işlenmesi kolaylaşmıştır [25, 29].

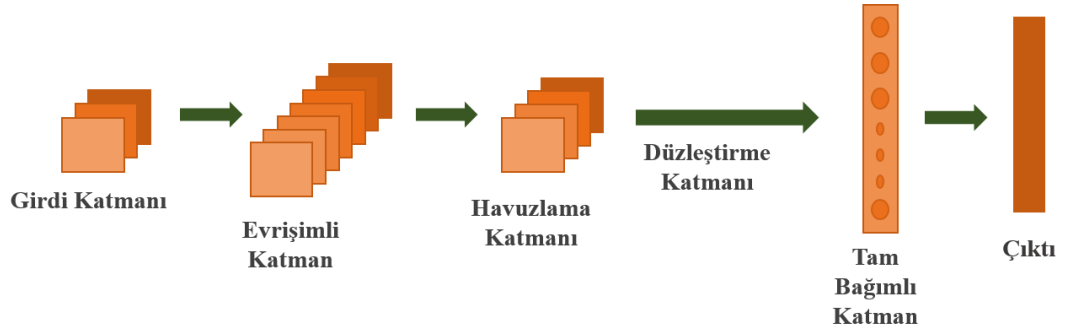
Verilerin öğrenilmesi, gereksiz verilerin unutulması, yeni öğrendiğini anlamlandırma, öğrenilmiş bilginin kullanılması, geçmiş ve gelecek bilgilerin ilişkilendirilmesi gibi özellikleri sayesinde bir insanın öğrenme yeteneklerine sahip olacak şekilde tasarlanmışlardır. Öğrenme işlemi, DL’de türlü katmanlar ve bu katmanlardaki nöronlar/birimler ile gerçekleşmektedir. Sistemde girdi olarak kullanılan her veriden katmanlarda bulunan birimler/nöronlar sayesinde yeni bilgiler ve özellikler elde edilmektedir. DL algoritmaları bu bilgi ve özelliklerin kullanılıp kullanılmayacağına karar verebilme yeteneğine sahiptirler. Bu sayede algoritmanın eğitimde kullanılan eğitim kümesinden elde edilen bilgilerle sisteme girilen yeni bilgilerin tahminindeki veya sınıflandırılmasındaki doğrulukta yüksek performansa ulaşabilmektedirler [29, 40].

DL algoritmaları, sosyal medyada, e-ticaret sitelerinde, anomali tespitinde, görsel; video ve ses çıkarımlarında, doğal dil işleme, görüntü işleme tibi pek çok alanda kullanılmakta ve başarılı sonuçlar elde etmektedir [40].

2.6.1. Evrişimli sinir ağı (CNN)

Evrişimli Sinir Ağı (CNN), görüntü sınıflandırma ve doğal dil işleme gibi tek ve iki boyutlu özellikler içeren verilerde en çok kullanılan DL yöntemlerindendir. Gizli katmanların amacı veriler içerisindeki doğrusal olmayan kalıpları bulmak ve bunları kullanarak doğru tahminlerde bulunabilmektir. Özellikle evrişim katmanında verilen filtre sayısı hiperparametresi verideki gizli özellikleri bulmayı kolaylaştırmaktadır. Aynı zamanda geleneksel ANN algoritmasında olduğu gibi gizli katmanlarındaki nöronların kendini optimize edebilme yeteneği bulunmaktadır. Son zamanlarda

yapılan çalışmalarda CNN yönteminin zaman serisi analizlerinde de etkin ve verimli değerler verdiği görülmüştür [29, 41-43].



Şekil 2.43. CNN algoritma yapısı.

Şekil 2.43'te bir CNN algoritmasının yapısı verilmiştir. CNN algoritmalarında 4 tür katman bulunmaktadır. Bu katmanlar [28, 41].

- Evrişim Katmanı (Convolutional Layer): Girdiler ve ağırlıkların evrişim katmanındaki nöron/birim çıkışı skaler çarpım ile belirlenmektedir. Katmanın içerisinde özellik haritalarını belirlemek için filtreleme davranışı gösteren birimler bulunmaktadır. Bu filtreleme işlemiyle girdinin 2 boyutlu bir aktivasyon haritası üretilir. Filtreleme işlemi sayesinde veri içerisindeki gizli özellikler bulunmaktadır. Evrişim Katmanı'nda girdi özellikleri satırlara yerleştirilir. Girdi verileri ağırlıklarla ilişkilendirilen küçük parçalara ayrılarak matris formuna girmektedir.

Parçalar haline getirilmiş girdi kümeleri Denklem 2.14'te verilen filtreleme işlemiyle çeşitli nitelik kümelerini oluşturmaktadır.

$$F_l^k = (A * K_l^k) \quad (2.14)$$

Denklem 14'te A girdi matrisini, K_l^k k. katmanın l. filtresini ve F_l^k ise katmanın özellik haritası değerini göstermektedir.

- Havuzlama Katmanı (Pooling Layer): Matris halindeki verinin boyutunu kademeli olarak azaltmayı sağlamaktadır. Modelin hiperparametre sayısı azaldığından hesaplama işlemi sırasında oluşabilecek karmaşıklıkların

azalması hedeflenmektedir. Bu durumda modelin baskın özellikleri ortaya çıkarılmaktadır.

Havuzlama katmanı, girdi setini oluşturan her parçanın içerisindeki baskın özelliği Denklem 2.15 yardımıyla bulmaktadır.

$$P_l = f_p(F^l) \quad (2.15)$$

Denklem 15'te f_p havuzlama işlemini, F^l l. girdiyi ve P_l ise l. girdinin özellik haritasını göstermektedir.

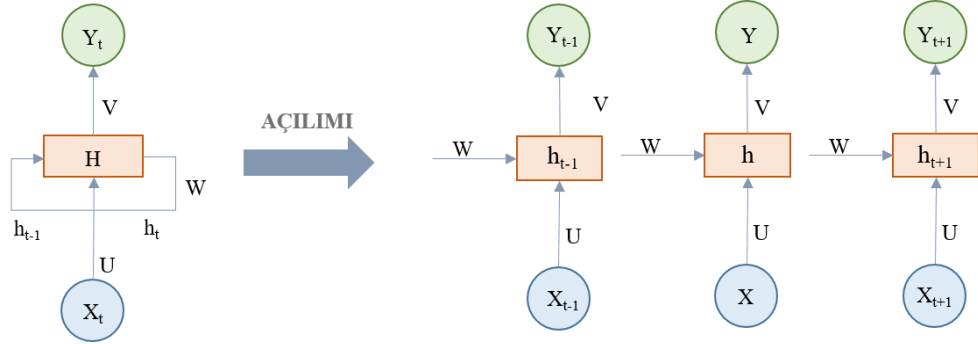
- Düzleştirme Katmanı (Flatten Layer): Matris şeklindeki girdinin Tam Bağımlı Katman'a geçmeden önce vektör durumuna gelmesini sağlamaktadır.
- Tam Bağımlı Katman (Fully-Connected Layer): Kendisinden önceki katman ve sonraki katman ile doğrudan bağlantılıdır. Bu katmanın yapısı geleneksel ANN'e benzemektedir. Katmana gelen bilgiler vektör boyutundadır ve katman çok sayıda hiperparametre içermektedir. Bu nedenle karmaşık hesaplamalar bu katmanda uzun bir zaman almaktadır. Önceki katmanların çıktılarını da göz önüne alarak en iyi tahmin sonunu elde edilmeye çalışılmaktadır. Sınıflandırma ve regresyon modellerinde ağı son katmanı olarak yer almaktadır.

2.6.2. Tekrarlayan sinir ağı (RNN)

Tekrarlayan Sinir Ağı (RNN) modelinin başlangıcı, ileri beslemeli bir sinir ağına benzemektedir. Ancak modele yeni girecek bilgilerin yanında daha önceki öğrenilen bilgilerin de bir girdi olarak kullanılması RNN'i ileri beslemeli sinir ağından ayırmaktadır. RNNler'in içerisinde bulunan bağlam birimleri çıktı katmanından çıkan çıktının bir kopyasını tutmaktadır. Bir önceki birimin çıktısı bir sonraki birimin girdisini oluşturmaktadır. Bu sayede sonucun yalnızca yeni bilgilere göre oluşması engellenmektedir [28, 44].

Önceki birimin çıktısındaki bilgiyi depolayan yapıya bellek adı verilmektedir. Karmaşık yapı verilerde İleri Beslemeli Sinir Ağı'nın yetersiz olması nedeniyle bu tür bellek içeren bir model oluşturulmuştur. RNN modellerini eğitmek için Geri Yayılımlı ve Gerçek Zamanlı Öğrenme algoritmaları kullanılmaktadır. Bu da RNN

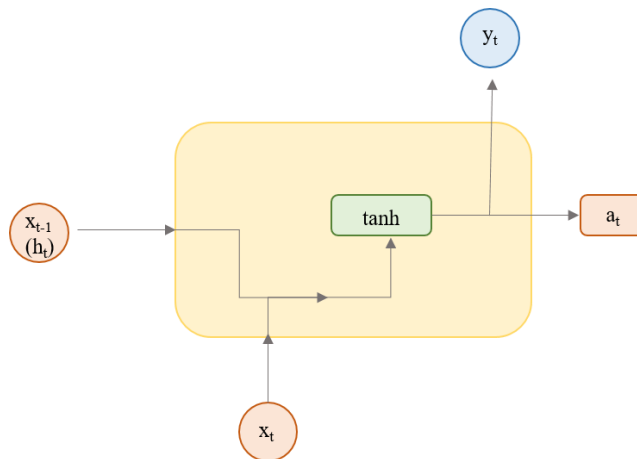
modellerinin dinamik bir hale gelmesini sağlamaktadır. Ancak RNN’de bulunan bellek kısa süreli bir bellektir. Bu nedenle kısa süreli tahminlerde uzun süreli tahmin problemlerine göre daha iyi sonuç vermektedirler [45, 46].



Şekil 2.44. RNN algoritmasının çalışma prensibi.

Şekil 2.44’te RNN Algoritmasının çalışma prensibi verilmiştir. H gizli hücrenin durumu, X_t ağ girdisini, Y_t ağın çıktısını, W, U, V birimler arası bağların ağırlıklarını temsil etmektedir. h mevcut hücre belleğinin durumunu h_{t-1} bir önceki hücrenin belleğini, h_{t+1} bir sonraki hücrenin bellek durumunu ifade etmektedir. Y mevcut hücrenin çıktısını Y_{t-1} bir önceki hücrenin çıktısını ve aynı zamanda mevcut hücrenin girdisini göstermektedir. X mevcut durumdaki girdiyi, X_{t-1} bir önceki adımın girdisini ve X_{t+1} bir sonraki adımın girdisini ifade etmektedir.

Bir RNN hücresi yalnızca bir tane tanh fonksiyonunu içermektedir [45]. Şekil 2.45’te bir RNN hücresinin yapısı verilmiştir.



Şekil 2.45. Bir RNN hücresinin yapısı.

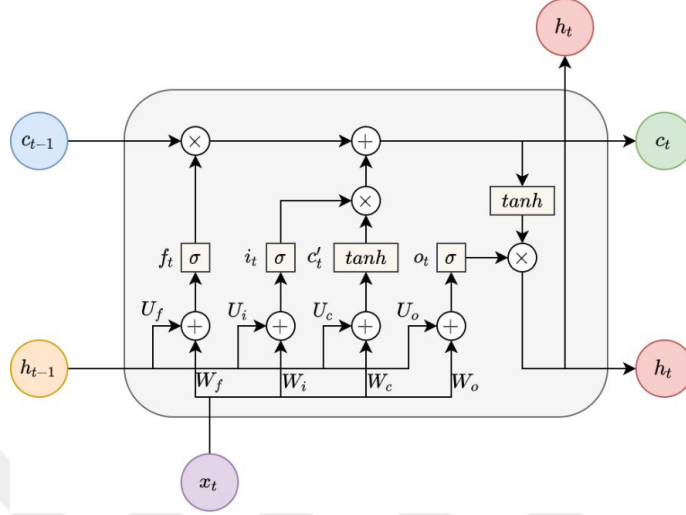
2.6.3. Uzun kısa süreli bellek (LSTM)

Tekrarlayan Sinir Ağları (RNN) zaman serilerinde sıklıkla kullanılan ve iyi sonuçlar veren bir yöntem olsa da genel olarak uzak geçmişî hatırlamada ve gereksiz verilerin unutulması konusunda eksiklikleri olduđunun saptanmasıyla RNN'in gelişmiş hali olan Uzun-Kısa Süreli Bellek (LSTM) algoritması oluşturulmuştur [47-50]. RNN hücrelerinde her gizli durumu bulmak için bir tane lineer olmayan fonksiyon kullanılmaktadır. LSTM hücrelerinde ise her gizli durum için birden fazla fonksiyon kullanılmasıyla algoritmanın belleđini süresini uzatmaktadır. Bu sayede LSTM algoritması uzak geçmişteki bilgileri hatırlamada daha başarılıdır [28]. Uzak geçmişî hatırlamadaki başarısı ve birçok çalışmada da en başarılı sonuçlar veren algoritma olması nedeniyle bu çalışmada kullanılmıştır.

Bir LSTM hücresinde, algoritmanın verilerin hücreye girişı, gereksiz bilgileri unutma, değerli bilgileri bir sonraki hücreye taşıma, yeni verideki bilgileri anlamlandırma gibi verilerin içerisindeki bilgileri öğrenmek için bulunan 1 hücre ve bu hücre içerisinde bulunan kapı adı verilen 3 bileşen bulunmaktadır. Bu bileşenler şunlardır [49, 51, 52]:

- Hücre: LSTM birimine gelen bir verinin bilgiye dönüşmesi için kullanılacak olan diğer hücre bileşenlerini saklayan yapıdır.
- Unut Kapısı (f_t): Önceki birimlerden geçerek mevcut birime ulaşan bilgilerin kullanılıp kullanılmayacağına karar veren kapıdır.
- Giriş Kapısı (i_t): İlk LSTM hücresi için verilerin hücreye ilk giriş yaptığı kapı, sonraki hücreler için ise yeni veri girişinin ve bir önceki LSTM hücresinde dönüştürülmüş bilgilerin yeni hücreye girdiđi kapıdır. Bu kapıdaki karar mekanizması iki adımdan oluşmaktadır. İlk adımda yeni bilginin durumuna karar verilmektedir. İkinci adımda ise giriş kapısının verdiđi karara bađlı olarak gerekiyorsa sisteme giren yeni veri ve bir önceki birimden gelen bilgiler birleştirilmektedir.
- Çıktı kapısı (o_t): Önceki birimlerden gelen bilgilerin bir sonraki birime iletilip ileilmeyeceđine karar veren unut kapısı ve yeni girdinin kullanılıp kullanılmayacağına karar veren girdi kapılarının kararları sonucunda bilgiler birleştirilmiş olarak çıktı kapısına gelmektedir. Çıktı kapısı ise mevcut

birimden bir sonraki hücreye bir veri aktarımı olup olmayacağına karar veren kapıdır. Bu kapıdan çıkan bilgiler bir sonraki LSTM hücresinin girdi kapısına ulaşmaktadır.



Şekil 2.46. LSTM algoritmasının çalışma prensibi.

Şekil 2.46’da bir LSTM algoritmasının çalışma prensibi verilmiştir. LSTM hücre girişi x_t , bellek durumu c_t ve hücrenin çıkışı h_t ve σ - $\tanh$ aktivasyon fonksiyonu olarak gösterilmiştir. Tüm hücrelerin girdisini bir önceki LSTM hücresinin çıktısı (h_{t-1}) ve yeni girdi (x_t) oluşturmaktadır. Bu girdi değerleri U ve W ağırlık kümeleriyle unut kapısı (f_t), girdi kapısı (i_t) ve çıktı kapısı (o_t) girişlerinde ilişkilendirilmektedir. Algoritmanın çalışma prensibi ile ilgili denklemler aşağıdaki gibidir [51].

Unut kapısının, bir önceki hücrenin çıktısından gelen h_{t-1} bilgisine ve x_t yeni girdisine dayalı bir karar mekanizması bulunmaktadır. Bir sonraki hücreye taşınmaması karar verilen bilgiler unutulmaktadır. Bu mekanizma Denklem 2.16’da gösterilmiştir.

$$f_t = \sigma(W_f x_t + U_f h_{t-1}) \quad (2.16)$$

c_{t-1} bir önceki LSTM hücresinin bellek durumunu gösteren bir vektördür. Bu vektördeki bilgilerin bir sonraki LSTM hücresine geçip geçmeyeceği vektörün her elemanına uygulanacak σ aktivasyon fonksiyonu ile belirlenmektedir. Bu fonksiyonun işlem sonucunda elde edilen değer 0 ise bilgiler silinmekte, 1 ise bilgilerin bir sonraki LSTM hücresine geçmesi gerektiğine karar verilmektedir.

Girdi kapısında LSTM hücresine girecek bilgilerle ilgili kararlar verilmektedir. Kararlar, x_t yeni girdisine ve bir önceki LSTM hücresinin çıktısı (h_{t-1}) verilerine göre σ aktivasyon fonksiyonu yardımıyla verilmektedir. Aktivasyon fonksiyonun çıktısı 0 olduğunda hücreye yeni bilgi girişi yapılamamakta, yalnızca bir önceki hücreden gelen bilgi kullanılmaktadır. 1 olduğunda ise yeni bilgiler hücreye girebilmektedir. Girdi kapısında yapılan bu proses Denklem 2.17’de verilmiştir.

$$i_t = \sigma(W_i x_t + U_i h_{t-1}) \quad (2.17)$$

Aday girdi vektörü (c'_t) bir önceki LSTM hücresinin çıktısıyla (h_{t-1}) ve mevcut hücrenin yeni girdisiyle (x_t) bağımlıdır. Tanh aktivasyon fonksiyonu c'_t ’yi gerekli formata dönüştürmektedir. Bu işlem Denklem 2.18’de verilmiştir.

$$c'_t = \tanh(W_c x_t + U_c h_{t-1}) \quad (2.18)$$

Hücrenin yeni bellek kominasyonu, girdi ve unut kapıları yardımı ile mevcut durumdaki bilgiler ve hücreye giren yeni verilerle oluşturulmaktadır. Bu karar işlemi Denklem 2.19’da verilmiştir.

$$c_t = f_t * c_{t-1} \oplus i_t * c'_t \quad (2.19)$$

Bir LSTM hücresindeki çıktı kapısı, mevcut hücrenin bir çıktısı olup olmayacağına ve bir sonraki hücreye geçmek üzere bir çıktı verilecekse bu çıktının hangi bilgilerin kalıtlarından oluşacağına karar vermektedir. Bu karar σ fonksiyonu yardımıyla verilmektedir. σ aktivasyon fonksiyonu ile çıktı değeri 0 veya 1 değerlerini alır. İşlem sonucu 0 olduğunda mevcut LSTM hücresinden bir sonraki hücreye bir bilgi geçişi olmamakta, 1 olduğunda bir sonraki hücreye bilgi aktarımı yapılmaktadır. Bu işlem Denklem 2.20’de gösterilmiştir.

$$o_t = \sigma(W_o x_t + U_o h_{t-1}) \quad (2.20)$$

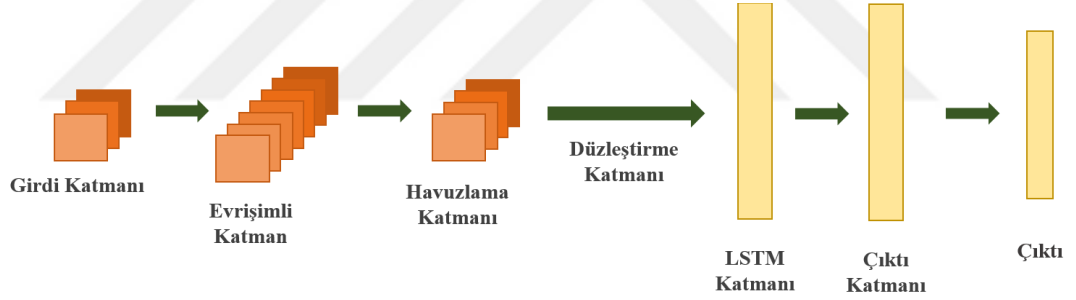
Sigmoid aktivasyon fonksiyonu çıktının oluşup oluşmayacağına karar verirken, bir çıktı üretilceğine dair bir karar verildiğinde Tanh aktivasyon fonksiyonu çıktı kapısında bir sonraki LSTM hücresine iletilecek bilgi kümesine karar verilmektedir. Bu işlem için Denklem 2.21 verilmiştir.

$$h_t = o_t * \tanh(c_t) \quad (2.21)$$

2.6.4. CNN+LSTM

Hisse senedi fiyat tahmininde doğru sonuçlara ulaşabilmek için modelde kullanılan veri setlerinin çok yönlü olması ve kullanılan yöntemin de veri seti içinde bulunan gizli özellikleri tespit etmede başarılı olması gerekmektedir. Bu nedenle CNN yönteminin, veri seti içerisindeki özellikleri bulma konusundaki başarısı ile LSTM yönteminin zaman serileri üzerindeki öğrenme başarısı bileştirilerek LSTM'den daha çok yönlü ve güvenilir yeni bir yaklaşım olarak CNN+LSTM yöntemi oluşturulmuştur. Bu yeni yaklaşım, hisse senedi fiyat tahmini, borsa endeks tahmini ve altın, gümüş gibi değerli madenlerin fiyatları ile ilgili tahminlerde LSTM algoritmasından daha doğru ve güvenilir sonuçlar verdiği için bu çalışmada kullanılmıştır [15, 52].

CNN+LSTM algoritması Şekil 2.47'deki gibi CNN algoritmasındaki Düzleştirme katmanından sonra LSTM modelinin eklenmesi ile oluşturulmaktadır [28, 53].



Şekil 2.47. CNN+LSTM algoritmasının yapısı.

2.7. Hiperparametre Optimizasyonu

Makine Öğrenmesi (ML) ve Derin Öğrenme (DL) modellerinde bir veri setinin doğruluğu ve ön işleme adımının yapılması dışında, oluşturulan tahmin modelinde kullanılan hiperparametrelerin de optimize edilmesi gerekmektedir. Bu hiperparametreler deneme yanılma yoluyla modellerde uygulanarak model hata değerlerinin iyileştirilmesi amaçlanır. Rassal Orman Regresyonu (RFR), Evrişimli Sinir Ağı (CNN) ve Uzun-Kısa Süreli Bellek (LSTM) modellerde en çok kullanılan hiperparametrelerden bazıları aşağıda verilmiştir [28, 54, 55].

Veri Bölme: ML ve DL yöntemlerinde veri eğitilmesi, doğrulanması ve test edilmesi için “Eğitim Kümesi” “Doğrulama Kümesi”ve “Test Kümesi” olmak üzere 3 gruba veya “Eğitim Kümesi” ve “Test Kümesi” olmak üzere 2 gruba ayrılabilir. Çalışmada RFR algoritmasında eğitim ve test kümesi olmak üzere 2 küme kullanılmaktadır. LSTM ve CNN+LSTM yöntemlerinde de bu kümeler eğitim ve test olmak üzere ikiye ayrılıp test kümesi aynı zamanda doğrulama kümesi olarak da kullanılabilir. Ancak doğrulama kümesinin ayrı olarak oluşturulduğu senaryolarda bu algoritmaların daha iyi ve tutarlı sonuçlar verdiği görülmüştür. Bu nedenle çalışmada LSTM ve CNN+LSTM algoritmaları için eğitim, doğrulama ve test olarak 3 küme oluşturulmuştur.

Ağaç Sayısı (n_estimator): RFR algoritmasında, oluşturulacak toplam karar ağacının sayısıdır.

Ağacın Maksimum Derinliği (max_depth): RFR algoritmasında oluşturulan ağaçların en fazla ilerleyebileceği düğüm sayısını göstermektedir. Herhangi bir değer girilmediğinde, ağaçlar, ağaçlardan elde edilebilecek en iyi tahmin değerine ulaşmaya kadar düğüm oluşturmaya devam etmektedir.

Özellik Sayısı (max_features): RFR algoritmasında oluşturulan karar ağaçlarında kullanılması gereken en fazla özellik sayısıdır.

Rasgelelik Eşik Değeri (random_state): RFR algoritmasında bu hiperparametreye bir değer verildiğinde modelin her çalıştığında farklı bir tahminde bulunması engellenir.

Veri Bölmede Rastgelelik Durumu (shuffle): RFR algoritmasında eğitim ve test veri setlerinin veri içinden seçilme yöntemi ile ilgili hiperparametredir. Hiperparametre değeri “Hayır” seçildiğinde veriden sırasıyla, “Evet” seçildiğinde veriden rastgele bir eğitim ve test veri setleri oluşturulur.

Devir Sayısı (Epoch): Bir DL modelinde algoritmanın kaç kez çalıştırılacağını gösteren hiperparametredir.

Birim Sayısı (Unit): Bir LSTM hücresinin gizli katmanlarında bulunan gizli birimlerin sayısıdır.

Katman Sayısı: Oluşturulan algoritmaların içinde bulunacak katman sayısıdır.

Aktivasyon Fonksiyonu (activation): Zaman serisi tahminlerinde genellikle kullanılan aktivasyon fonksiyonları, ReLU, Tanh ve Sigmoid'dir.

Optimizasyon Fonksiyonu (Optimize-Momentum): Algoritmanın optimize edilmesinde kullanılan fonksiyondur. En yaygın kullanılan optimizasyon fonksiyonu Adam'dır (Momentum).

Öğrenme Aralığı (learning_rate): Optimizasyon fonksiyonunun öğrenme oranındaki adım sayısını temsil etmektedir. Öğrenme aralığı değeri ağıın verileri öğrenmesinde kullanılan bir hiperparametredir. Büyük bir değer verildiğinde ağıın veri içerisindeki gizli bilgileri öğrenmesinde problem oluşmaktadır. Çok düşük bir değer verildiğinde öğrenme konusunda başarılı olmakta ancak algoritmanın çalışma süresi uzamaktadır.

Çekirdek Boyutu (kernel_size): CNN algoritmasındaki filtrelerin boyutunu gösteren matristir. Tam sayı formatında yazılmaktadır.

Filtre Sayısı (filters): CNN algoritmasındaki filtre sayısıdır. Bu hiperparametre girdide tespit edilen özelliklerin bir özellik haritasını oluşturmayı sağlamaktadır. Tam sayı formatında yazılmaktadır.

Kayıp Fonksiyonu (loss): Belirlenen verilerle yapılan eğitimin başarısını ölçmektedir. En yaygın kullanılan fonksiyonu Ortalama Karesel Hata'dır (MSE).

2.8. İstatistiksel Analiz

Çalışmada elde edilen tahmin sonuçlarının doğruluğunun ve başarısının kontrolü, literatürde hisse senedi fiyat tahmini çalışmalarında yaygın olarak kullanılan Ortalama Karesel Hata (MSE), Kök Ortalama Karesel Hata (RMSE), Ortalama Mutlak Hata (MAE), Ortalama Mutlak Yüzde Hata (MAPE) gibi hata metrik yöntemleri ve Determinasyon Katsayısı (R^2) ile yapılmıştır. Hata metrik değerlerinin 0'a ve R^2 değerinin 1'e yakın olması algoritmanın kullanılan verilerdeki gözlem değerlerine yakın tahminler yaptığını göstermektedir. R^2 değerinde negatif bir sonuç elde edilmesi oluşturulan algoritmanın veri setine uyum sağlamadığını göstermektedir [56, 57]. Bu durumda algoritmadaki hiperparametler ile yeni kombinasyonlar oluşturularak tahmin işlemleri tekrarlanmaktadır. MSE değeri Denklem 2.22'deki RMSE değeri

Denklem 2.23'teki, MAE değeri Denklem 2.24'teki, MAPE değeri Denklem 2.25'deki ve R^2 değeri Denklem 2.26'daki gibidir.

$$MSE = \frac{1}{n} \sum_{j=1}^n e_j^2 \quad (2.22)$$

$$RMSE = \sqrt{\frac{\sum_{j=1}^n e_j^2}{n}} \quad (2.23)$$

$$MAE = \frac{1}{n} \sum_{j=1}^n |e_j| \quad (2.24)$$

$$MAPE = 100 \times \frac{\sum_{i=1}^n \frac{|y_i - \bar{y}_i|}{y_i}}{n} \quad (2.25)$$

$$R^2 = 1 - \frac{\sum (y_i - \bar{y}_i)^2}{(\sum y_i - \bar{y})^2} \quad (2.26)$$

Tüm metrik hata denklemlerinde n işlem yapılacak veri sayısını, e_j j. gözlem değerini ve onun tahmin değeri arasındaki hata değerini (farkı) göstermektedir. R^2 denkleminde ise y_i gerçek değerleri, $\bar{y}_i$ tahmin değerlerini ve $\bar{y}$ ise ortalama değeri temsil etmektedir.

3. UYGULAMA

Bu çalışmada Petkim Petrokimya Holding A.Ş.'nin (PETKM) hisse senedi kapanış fiyat tahmini yapılmıştır. Bu tahmin işlemi için 4 farklı girdi veri seti oluşturulmuştur. Girdi için kullanılan veriler Türkiye'deki ve Türkiye'nin petrol temin ettiği ülkelerdeki ilgili olabilecek hisse senedi fiyatlarını ve borsa endeks fiyatlarını içermektedir. Bu verilere ayrı ayrı Rassal Orman Regresyonu (RFR) algoritması, Uzun-Kısa Süreli Bellek (LSTM) algoritması ve Evrişimli Sinir Ağı (CNN) ile LSTM (CNN+LSTM) hibrit algoritma uygulanarak algoritmaların tahmin başarıları bulunmuştur. Bu bölüm içerisinde oluşturulan 4 farklı veri seti ile ilgili bilgiler, modellerin kuruluşunda kullanılan fonksiyonlar, algoritmaların içerdiği hiperparametre aralıkları ve bu hiperparametrelerin sonuçlara etkisini gösteren örnek test değerlerini içeren tablolar verilmiştir.

Çalışmanın uygulamasında Python 3.7 yazım dili, Spyder 3 editörü kullanılmıştır. Algoritmaların oluşturulmasında Python'ın Pandas, SciKit Learn, Keras, Numpy, Math ve Matplotlib kütüphanelerinden yararlanılmıştır. Kullanılan algoritmaların çalıştırıldığı makine 8.00GB RAM, Intel(R) Core(TM) i7-4510U CPU işlemci ve Windows 10 Pro 64 bit işletim sistemi özelliklerine sahiptir. Veriler MS Excel 2013'te tutulmaktadır. Çalışmada kullanılan excel formatı csv türündedir.

Excel'de bulunan eksik veri doldurma işlemi uygulanan veriler, Pandas kütüphanesindeki read_csv fonksiyonu ile çağrılmaktadır. Tüm verilerin arasından girdi olarak kullanılacak özellikler iloc fonksiyonu ile seçilmekte ve DataFrame fonksiyonu ile matris formatına getirilmektedir. Aynı zamanda bu verilerin değer aralıklarının çok yüksek ve farklı aralıklarda olmasından kaynaklanacak problemleri ortadan kaldırmak amacıyla her veri seti özelliği için normalizasyon işlemi uygulanmıştır. Bu normalizasyon işlemiyle veri seti özelliklerinin değerleri 0-1 arasına çekilmiştir. Normalizasyon işlemi için SciKit Learn kütüphanesindeki preprocessing modülündeki MinMaxScaler fonksiyonu kullanılmıştır. Bu fonksiyonda normalize edilecek verilerin hangi sayı aralığına çekileceğinin bilgisi verilmektedir. fit_transform ve transform fonksiyonlarıyla normalizasyon işlemi verilere uygulanır.

Tüm algoritmalarda tüm modeller için veri seti ile oluşturulan matristen, geçmiş verileri kullanılarak mevcut günün tahmini için oluşturulmuş bir denetimli öğrenme fonksiyonu [58] kullanılmıştır. Denetimli öğrenme modelinden drop fonksiyonu ile tahmin edilmeyecek veriler silinmiştir.

Oluşturulan modellerin algoritma sonuçlarını değerlendirmek için Ortalama Karesel Hata (MSE), Kök Ortalama Karesel Hata (RMSE), Ortalama Mutlak Hata (MAE), Ortalama Mutlak Yüzde Hata (MAPE) gibi hata metrik yöntemleri ve Determinasyon Katsayısı (R^2) kullanılmıştır. Bu metrik değerleri için SciKit Learn kütüphanesinin `metrics` modülündeki `mean_squared_error`, `mean_absolute_error`, `r2_score` fonksiyonları ve Numpy kütüphanesindeki `sqrt` (karekök) fonksiyonu kullanılmıştır.

Test kümesi çıktılarındaki gözlem ve tahmin değerlerinin grafiklerde gösterilmesi için Matplotlib kütüphanesindeki, `subplots`, `scatter`, `plot`, `xlabel`, `ylabel` ve `legend` fonksiyonları kullanılmıştır. Grafiklerde turuncu ile gösterilen değerler gözlem değerleri, kırmızı ile gösterilen değerler tahmin değerlerini ifade etmektedir. Grafiklerdeki yazım şeklinin belirlenmesi için Matplotlib kütüphanesindeki `rcParams.update` fonksiyonu kullanılmıştır. Matplotlib kütüphanesindeki `savefig` fonksiyonu yardımıyla grafikler kaydedilmektedir.

3.1. Eksik Verinin Doldurulma Yönteminin Bulunması

Veri setinin içinde Türkiye, Rusya ve Katar ülkelerindeki hisse senedi ve endeks fiyatlarının bulunmasıyla veriler arasında farklı tarihlerdeki resmi tatil ve diğer nedenlerden dolayı eksik kalmış ve/veya hatalı veriler bulunmaktadır. Çalışmada 4 farklı eksik veri doldurma yöntemi kullanılarak tüm verilerin içinde bulunduğu ayrı ayrı 4 veri seti oluşturulmuştur. Bunlar: 1- Eksik verinin bulunduğu tarihin veri setinde çıkarılması, 2- Eksik verinin 0 ile doldurulması, 3- Eksik verinin ortalama değer ile doldurulması, 4- doğrusal enterpolasyon yöntemi ile eksik verinin doldurulmasıdır. Bu yöntemlerle oluşturulmuş 4 veri seti ile Model 1 için tahmin çalışması yapılarak, çalışmadaki veriler için en iyi sonucu veren eksik veri doldurma yöntemi bulunmaya çalışılmıştır. Tablo 3.1’de bu veri setleri ile Model 1 üzerinde yapılmış ön tahmin çalışmalarının değerlendirme sonuçları verilmiştir. En iyi sonuçları veren veri setinin

doğrusal enterpolasyon yöntemi kullanılarak oluşturulmuş veri seti olduğu görülmüştür.

Tablo 3.1. Eksik veri doldurma yöntemlerinin tahmin sonuçlarına etkisi

Örnek	Algoritma	Eksik Veri Doldurma Yöntemi	MSE	RMSE	MAE	MAPE	R ²
1	RFR	Eksik Veriyi Silme	0,048	0,220	0,161	4,85	0,757
2	RFR	Eksik Veriyi Silme	0,043	0,207	0,158	4,803	0,784
3	RFR	Eksik Veriyi 0 İle Doldurma	3,185	1,784	1,654	NaN	-0,711
4	RFR	Eksik Veriyi 0 ile Doldurma	2,436	1,561	1,414	NaN	-0,308
5	RFR	Aritmetik Ortalama	0,032	0,178	0,140	4,260	0,835
6	RFR	Aritmetik Ortalama	0,041	0,203	0,154	4,704	0,787
7	RFR	Doğrusal Enterpolasyon	0,044	0,209	0,159	4,856	0,776
8	RFR	Doğrusal Enterpolasyon	0,044	0,209	0,163	4,990	0,775
9	LSTM	Eksik Veriyi Silme	0,027	0,164	0,124	3,699	0,865
10	LSTM	Eksik Veriyi Silme	0,077	0,278	0,219	6,370	0,612
11	LSTM	Eksik Veriyi 0 İle Doldurma	2,724	1,650	1,134	Inf	-0,463
12	LSTM	Eksik Veriyi 0 ile Doldurma	2,872	1,695	1,103	Inf	-0,543
13	LSTM	Aritmetik Ortalama	0,024	0,156	0,131	0,875	4,007
14	LSTM	Aritmetik Ortalama	0,012	0,107	0,073	0,940	2,197
15	LSTM	Doğrusal Enterpolasyon	0,019	0,137	0,094	0,904	2,820
16	LSTM	Doğrusal Enterpolasyon	0,021	0,145	0,102	0,892	3,044
17	CNN+LSTM	Eksik Veriyi Silme	0,051	0,227	0,183	5,540	0,742
18	CNN+LSTM	Eksik Veriyi Silme	0,047	0,216	0,166	4,961	0,765
19	CNN+LSTM	Eksik Veriyi 0 İle Doldurma	3,281	1,811	1,701	Inf	-0,763
20	CNN+LSTM	Eksik Veriyi 0 ile Doldurma	3,221	1,795	1,697	Inf	-0,730
21	CNN+LSTM	Aritmetik Ortalama	0,044	0,20	0,174	5,439	0,773
22	CNN+LSTM	Aritmetik Ortalama	0,199	0,446	0,410	12,144	-0,029
23	CNN+LSTM	Doğrusal Enterpolasyon	0,044	0,211	0,148	4,262	0,773
24	CNN+LSTM	Doğrusal Enterpolasyon	0,038	0,196	0,156	4,909	0,805

3.2. Model 1

Model 1'deki veri setinde, BİST Kimya Petrol Plastik (XKMYA) endeks açılış/kapanış fiyatının, Türkiye Petrol Rafinelerileri A.Ş. (TUPRS); PETKM hisse senedi açılış/kapanış fiyatlarının ve dolar (USD/TRY) açılış ve ortalama alış/satış kur değerleri girdi, PETKM hisse senedi kapanış fiyatı çıktı olarak kullanılmaktadır.

RFR algoritmasında, eğitim ve test veri setlerini ayırmak için SciKit Learn kütüphanesinde `model_selection` modülündeki `train_test_split` fonksiyonu

kullanılmaktadır. Bu fonksiyonda kullanılan hiperparametreler veri setindeki girdi çıktı değerleri, test veri kümesinin boyutu (test_size) ve verilerin test-eğitim kümesi olarak ayrılırken verilerin rastgele seçilme durumudur (shuffle). Eğitim ve test kümelerinin oluşturulması diğer modellerin RFR algoritmalarında da aynı şekilde yapılmıştır. Fonksiyon sonucunda eğitim kümesinin girdisi, eğitim kümesinin çıktısı, test kümesinin girdisi ve test kümesinin çıktısı olmak üzere 4 küme elde edilmektedir. Test veri kümesinin boyutu tüm veri setinin %20-%40'ı arasında değerler arasından belirlenmiştir. SciKit Learn kütüphanesindeki ensemble metotlar arasından RandomForestRegressor kullanılmaktadır. Bu metodun içerisinde kullanılan hiperparametreler oluşturulacak ağaç sayısı (n_estimators) ve en fazla kullanılacak özellik miktarıdır (max_features). Ağaç sayısı 10-100 arasındaki değerler arasında ve kullanılacak özellik miktarı 0,6-Otomatik (auto) değerleri arasında değişmektedir.

fit fonksiyonu eğitim verileri ile modelin eğitilmesini sağlamaktadır. predict fonksiyonu ile test kümesi girdisi ile tahmin işlemi yapılmaktadır. Tahmin işlemi diğer modellerin RFR algoritmalarında da aynı şekilde yapılmıştır.

Tablo 3.2.'de Model 1 girdileri için RFR algoritmasında kullanılan hiperarametre değerleri, tahmin değerlerinin gözlem değerleriyle olan hata değerleri (MSE, RMSE, MAE) ve tahmin değerlerinin gözlem değerlerine uygunluğu (R^2) verilmiştir.

Belirlenmiş girdiler ve çıktıdan eğitim, doğrulama ve test kümeleri oluşturulmuştur. Eğitim ve doğrulama kümelerinin toplamı %60-%80 arasında belirlenmiştir. Doğrulama kümesindeki veri sayısı eğitim ve doğrulama kümesinin toplamının %20'si olarak belirlenmiştir. Burada belirtilen denetimli öğrenme fonksiyonu ve eğitim, doğrulama, test kümelerinin bölünme şekli diğer tüm derin öğrenme algoritmalarında aynı şekilde yapılmıştır.

Tablo 3.2. Model 1 için RFR algoritmasının hiperparametre değerleri ve hata metrik değerleri.

Model1	Eğitim Veri Seti Boyutu	Test Veri Seti Boyutu	Rastgele Veri Bölme Durumu	Ağaç Sayısı	Maksimum Özellik Sayısı	Değerlendirme				
						MSE	RMSE	MAE	MAPE	R ²
R1.1	2049 (%60)	1367 (%40)	Hayır	10	Auto	1,513	1,230	1,064	30,440	-2,573
R1.2	2049 (%60)	1367 (%40)	Hayır	50	Auto	1,573	1,254	1,088	31,182	-2,714
R1.3	2049 (%60)	1367 (%40)	Hayır	100	Auto	1,563	1,250	1,083	30,996	-2,691
R1.4	2391 (%70)	1025 (%30)	Hayır	10	Auto	0,224	0,473	0,377	10,904	0,237
R1.5	2391 (%70)	1025 (%30)	Hayır	50	Auto	0,225	0,475	0,368	10,621	0,233
R1.6	2391 (%70)	1025 (%30)	Hayır	100	Auto	0,218	0,467	0,362	10,455	0,257
R1.7	2391 (%70)	1025 (%30)	Hayır	50	0,8	0,234	0,484	0,383	11,130	0,202
R1.8	2391 (%70)	1025 (%30)	Hayır	100	0,8	0,215	0,464	0,361	10,390	0,267
R1.9	2391 (%70)	1025 (%30)	Hayır	100	0,6	0,224	0,473	0,370	10,673	0,238
R1.10	2732 (%80)	684 (%20)	Hayır	10	Auto	0,054	0,234	0,185	5,670	0,721
R1.11	2732 (%80)	684 (%20)	Hayır	50	Auto	0,042	0,206	0,162	4,934	0,782
R1.12	2732 (%80)	684 (%20)	Hayır	100	Auto	0,048	0,220	0,174	5,339	0,752
R1.13	2732 (%80)	684 (%20)	Hayır	50	0,8	0,054	0,233	0,192	5,963	0,721
R1.14	2732 (%80)	684 (%20)	Hayır	100	0,8	0,050	0,225	0,185	5,717	0,741
R1.15	2732 (%80)	684 (%20)	Hayır	100	0,6	0,069	0,263	0,225	6,954	0,644

Oluşturulan LSTM algoritması içerisindeki katmanlar sırası ile ilerlediği için Keras kütüphanesindeki models modülünden Sequential (sıralı) ve layers modülünden LSTM, Dropout ve Dense katmanları kullanılmaktadır. Bu model için oluşturulan algoritmada LSTM katmanında birim sayısı 50-100 arasında değişmektedir. LSTM katmanında Tanh, ReLU ve Sigmoid aktivasyon fonksiyonları kullanılmıştır. Girdi şekli bu katmanda belirtilmektedir. Tahmin algoritmalarındaki en büyük problemlerden biri aşırı öğrenme (overfitting) durumudur. Bu problemin önüne geçme için LSTM katmanına Dropout katmanı bağlanmıştır. Bu katmandaki eşik değeri (threshold) 0,01 ve 0,03 olarak verilmiştir. Bu katmanda belirlenen eşik değerinin altında kalan değerler silinmektedir. Bu şekilde aşırı öğrenme durumunun önüne geçilmektedir. Dense katmanı çıktı katmanıdır. Algoritmanın tek çıktısı olduğu için bu katmandaki birim sayısı 1'dir. Katman sayısı artırıldığında elde edilen sonuçların hata değerleri arttığı için genellikle 2 katmanlı modeller oluşturulmuştur.

Optimizasyon (optimizer) fonksiyonu olarak Adam (Momentum) kullanılmıştır. Optimizasyon fonksiyonunun öğrenme aralığı 0,01 ve 0,03 olarak verilmiştir. Bu değer ağırlık verileri öğrenme hızını temsil etmektedir. Kayıp fonksiyonu mean_squared_error olarak belirlenmiştir. Parti Boyutu (batch_size) hiperparametresi aynı anda öğrenme işleminin uygulanacağı veri sayısını göstermektedir. Model 1'in LSTM fonksiyonu için bu hiperparametre 20-100 arasında belirlenmiştir. Eğitim setinden alınan her parça aynı anda eğitilmemektedir. Bu eğitim işleminin her birine devir denmektedir. Devir sayısı (Epoch) 50-100 arasında değişmektedir. Her eğitim işleminde verilerin rastgele seçimini belirleyen hiperparametre shuffle'dır. Evet ve Hayır olmak üzere iki seçeneği bulunmaktadır.

Algoritma oluşturulduktan sonra predict fonksiyonuyla tahmin işlemi yapılır. inverse_transform fonksiyonu ile tahmin değerleri normalize değerlerden gerçek değerlere dönüştürülmektedir. Bu tahmin işlemi diğer tüm derin öğrenme algoritmalarında da aynı şekilde yapılmıştır.

Tablo 3.3'te Model 1 girdileri için LSTM algoritmasında kullanılan hiperparametre değerleri ve modellerin değerlendirme sonuçları verilmiştir.

Tablo 3.3. Model 1 için LSTM algoritmasının hiperparametre değerleri ve hata metrik değerleri.

Model1	Eğitim Veri Seti Boyutu	Doğrulama Veri Seti Boyutu	Test Veri Seti Boyutu	Devir Sayısı	Katman	Birim Sayısı/ Eşik Değeri	Parti Boyutu	Aktivasyon Fonksiyonu	Rastgele Öğrenme Durumu	Optimizasyon Fonksiyonu/ Öğrenme Aralığı	Değerlendirme
L1.1	1639 (%48)	341 (%12)	1367 (%40)	50	LSTM Dropout Dense	50 0,03 1	100	Tanh	Evet	Adam 0,01	MSE = 0,096 MAE = 0,255 RMSE = 0,310 MAPE = 7,777 R ² = 0,774
L1.2	1639 (%48)	341 (%12)	1367 (%40)	50	LSTM Dropout Dense	100 0,03 1	100	Tanh	Evet	Adam 0,01	MSE = 0,082 MAE = 0,234 RMSE = 0,287 MAPE = 7,167 R ² = 0,805
L1.3	1639 (%48)	341 (%12)	1367 (%40)	50	LSTM Dropout Dense	100 0,01 1	100	Tanh	Evet	Adam 0,01	MSE = 0,140 MAE = 0,313 RMSE = 0,374 MAPE = 9,625 R ² = 0,670
L1.4	1912 (%56)	479 (%14)	1025 (%30)	50	LSTM Dropout Dense	100 0,01 1	100	Tanh	Evet	Adam 0,01	MSE = 0,032 MAE = 0,145 RMSE = 0,180 MAPE = 4,416 R ² = 0,890
L1.5	1912 (%56)	479 (%14)	1025 (%30)	50	LSTM Dropout Dense	100 0,01 1	100	Sigmoid	Evet	Adam 0,01	MSE = 0,286 MAE = 0,423 RMSE = 0,535 MAPE = 11,418 R ² = 0,028

Tablo 3.3. (Devamı): Model 1 için LSTM algoritmasının hiperparametre değerleri ve hata metrik değerleri.

Model1	Eğitim Veri Seti Boyutu	Doğrulama Veri Seti Boyutu	Test Veri Seti Boyutu	Devir Sayısı	Katman	Birim Sayısı/ Eşik Değeri	Parti Boyutu	Aktivasyon Fonksiyonu	Rasgele Öğrenme Durumu	Optimizasyon Fonksiyonu/ Öğrenme Aralığı	Değerlendirme
L1.6	1912 (%56)	479 (%14)	1025 (%30)	50	LSTM Dropout Dense	100 0,01 1	100	ReLU	Evet	Adam 0,01	MSE = 0,101 MAE = 0,282 RMSE = 0,318 MAPE = 8,458 R ² = 0,656
L1.7	1912 (%56)	479 (%14)	1025 (%30)	50	LSTM Dropout Dense	100 0,01 1	100	Tanh	Evet	Adam 0,03	MSE = 0,218 MAE = 0,382 RMSE = 0,467 MAPE = 10,720 R ² = 0,260
L1.8	1912 (%56)	479 (%14)	1025 (%30)	50	LSTM Dropout Dense	50 0,01 1	100	Tanh	Evet	Adam 0,01	MSE = 0,021 MAE = 0,100 RMSE = 0,146 MAPE = 2,912 R ² = 0,927
L1.9	1912 (%56)	479 (%14)	1025 (%30)	100	LSTM Dropout Dense	50 0,01 1	100	Tanh	Evet	Adam 0,01	MSE = 0,231 MAE = 0,410 RMSE = 0,480 MAPE = 11,688 R ² = 0,216
L1.10	2185 (%64)	547 (%16)	684 (%20)	50	LSTM Dropout Dense	70 0,01 1	100	Tanh	Evet	Adam 0,01	MSE = 0,034 MAE = 0,147 RMSE = 0,184 MAPE = 4,447 R ² = 0,827
L1.11	2185 (%64)	547 (%16)	684 (%20)	70	LSTM Dropout Dense	70 0,01 1	100	Tanh	Hayır	Adam 0,01	MSE = 0,023 MAE = 0,102 RMSE = 0,150 MAPE = 3,065 R ² = 0,885

Tablo 3.3. (Devamı) Model 1 için LSTM algoritmasının hiperparametre değerleri ve hata metrik değerleri.

Model1	Eğitim Veri Seti Boyutu	Doğrulama Veri Seti Boyutu	Test Veri Seti Boyutu	Devir Sayısı	Katman	Birim Sayısı/ Eşik Değeri	Parti Boyutu	Aktivasyon Fonksiyonu	Rastgele Öğrenme Durumu	Optimizasyon Fonksiyonu/ Öğrenme Aralığı	Değerlendirme
L1.12	2185 (%64)	547 (%16)	684 (%20)	100	LSTM	70	100	Tanh	Evet	Adam 0,01	MSE = 0,019 MAE = 0,101 RMSE = 0,139 MAPE = 3,057 R ² = 0,902
					Dense	1					
L1.13	2185 (%64)	547 (%16)	684 (%20)	100	LSTM	70	80	Tanh	Evet	Adam 0,01	MSE = 0,018 MAE = 0,094 RMSE = 0,133 MAPE = 2,827 R ² = 0,909
					Dense	1					
L1.14	2185 (%64)	547 (%16)	684 (%20)	100	LSTM	70	50	Tanh	Evet	Adam 0,01	MSE = 0,015 MAE = 0,082 RMSE = 0,121 MAPE = 2,456 R ² = 0,926
					Dense	1					
L1.15	2185 (%64)	547 (%16)	684 (%20)	100	LSTM	70	20	Tanh	Hayır	Adam 0,01	MSE = 0,019 MAE = 0,096 RMSE = 0,137 MAPE = 2,874 R ² = 0,903
					Dense	1					

Oluşturulan CNN+LSTM algoritması içerisindeki katmanlar sırası ile ilerlediği için Keras kütüphanesindeki models modülünden Sequential (sıralı) ve layers modülünden TimeDistributed (TD), Conv1D, MaxPooling1D, Flatten, LSTM, Dropout, Dense katmanları kullanılmaktadır. TimeDistributed katmanı Conv1D katmanını LSTM katmanına bağlamak için kullanılmıştır. Conv1D katmanındaki filtre sayısı (filters) 32 ve 64 olarak belirlenmiştir. Çekirdek boyutu (kernel_size) 2 ve 3 olarak kullanılmıştır. Girdi şekli bu katmanda belirtilmektedir. MaxPooling1D katmanında havuz boyutu (pool_size) 2 ve 3 olarak belirtilmiştir. Flatten katmanı ile veriler düzleştirilmektedir. LSTM katmanından birim sayısı 50 olarak belirlenmiştir. Conv1D, LSTM katmanlarında Tanh, ReLU ve Sigmoid aktivasyon fonksiyonları kullanılmıştır. LSTM katmanına Dropout katmanı bağlanmıştır. Bu katmandaki eşik değeri 0,01 olarak verilmiştir. Dense katmanı çıktı katmanıdır. Algoritmanın tek çıktısı olduğu için bu katmandaki birim sayısı 1'dir.

Optimizasyon (optimizer) fonksiyonu olarak Adam (Momentum) kullanılmıştır. Optimizasyon fonksiyonunun öğrenme oranı 0,002-0,02 arasında verilmiştir. Kayıp fonksiyonu mean_squared_error olarak belirlenmiştir. Parti Boyutu (batch_size) değeri Model 1'in CNN+LSTM fonksiyonu için 50 ve 100 değerleri belirlenmiştir. Devir sayısı (Epoch) 50 ve 100 olarak belirlenmiştir. Her eğitim işleminde verilerin rastgele seçimini belirleyen hiperparametre shuffle'dır. Evet ve Hayır olmak üzere iki seçeneği bulunmaktadır. Bu algoritma için shuffle hiperparametresi Evet olarak seçilmiştir.

Tablo 3.4'te Model 1 girdileri için CNN+LSTM algoritmasında kullanılan hiperparametre değerleri ve modellerin değerlendirme sonuçları verilmiştir.

Tablo 3.4. Model 1 için CNN+LSTM algoritmasının hiperparametre değerleri ve hata metrik değerleri.

Tablo 3.4. Model 1 için CNN+LSTM algoritmasının hiperparametre değerleri ve hata metrik değerleri.

Tablo 3.4. (Devamı): Model 1 için CNN+LSTM algoritmasının hiperparametre değerleri ve hata metrik değerleri.

Model1	Eğitim Veri Seti Boyutu	Doğrulama Veri Seti Boyutu	Test Veri Seti Boyutu	Devir Sayısı	Katman	Birim/ Sayısı/Eşik Değeri/ Havuz Boyutu	Aktivasyon Fonksiyonu	Çekirdek Boyutu	Optimizasyon Fonksiyonu/ Öğrenme Aralığı	Parti Boyutu	Rastgele Öğrenme Durumu	Değerlendirme
CL1.4	1912 (%56)	479 (%14)	1025 (%30)	50	TD(Conv1D)	32	Tanh	2	Adam 0,002	100	Evet	MSE = 0,173 MAE = 0,262 RMSE = 0,416 MAPE = 6,858 R ² = 0,412
					MaxPooling1D	2	Tanh					
					Flatten							
					LSTM	50						
					Dropout	0,01						
					Dense	1						
CL1.5	1912 (%56)	479 (%14)	1025 (%30)	80	TD(Conv1D)	32	Tanh	2	Adam 0,002	100	Evet	MSE = 0,199 MAE = 0,279 RMSE = 0,446 MAPE = 7,289 R ² = 0,325
					MaxPooling1D	2	Tanh					
					Flatten							
					LSTM	50						
					Dropout	0,01						
					Dense	1						
CL1.6	1912 (%56)	479 (%14)	1025 (%30)	100	TD(Conv1D)	32	Tanh	2	Adam 0,002	100	Evet	MSE = 0,231 MAE = 0,308 RMSE = 0,481 MAPE = 8,120 R ² = 0,213
					MaxPooling1D	2	Tanh					
					Flatten							
					LSTM	50						
					Dropout	0,01						
					Dense	1						

Tablo 3.4. (Devamı): Model 1 için CNN+LSTM algoritmasının hiperparametre değerleri ve hata metrik değerleri.

Model1	Eğitim Veri Seti Boyutu	Doğrulama Veri Seti Boyutu	Test Veri Seti Boyutu	Devir Sayısı	Katman	Birim Sayısı/Eşik Değeri/Havuz Boyutu	Aktivasyon Fonksiyonu	Çekirdek Boyutu	Optimizasyon Fonksiyonu/ Öğrenme Aralığı	Parti Boyutu	Rastgele Öğrenme Durumu	Değerlendirme
CL1.7	1912 (%56)	479 (%14)	1025 (%30)	50	TD(Conv1D)	64	Tanh	2	Adam 0,002	100	Evet	MSE = 0,198 MAE = 0,281 RMSE = 0,445 MAPE = 7,432 R ² = 0,327
					MaxPooling1D	2						
					Flatten							
					LSTM	50	Tanh					
					Dropout	0,01						
CL1.8	1912 (%56)	479 (%14)	1025 (%30)	50	Dense	1			Adam 0,002	100	Evet	MSE = 0,208 MAE = 0,283 RMSE = 0,457 MAPE = 7,368 R ² = 0,292
					TD(Conv1D)	64	Tanh	2				
					MaxPooling1D	2						
					Flatten							
					LSTM	100	Tanh					
CL1.9	1912 (%56)	479 (%14)	1025 (%30)	50	Dropout	0,01			Adam 0,002	100	Evet	MSE = 0,225 MAE = 0,315 RMSE = 0,474 MAPE = 8,278 R ² = 0,236
					Dense	1						
					TD(Conv1D)	32	Tanh	3				
					MaxPooling1D	2						
					Flatten							
					LSTM	50	Tanh					
					Dropout	0,01						
					Dense	1						

Tablo 3.4. (Devamı): Model 1 için CNN+LSTM algoritmasının hiperparametre değerleri ve hata metrik değerleri.

Model1	Eğitim Veri Seti Boyutu	Doğrulama Veri Seti Boyutu	Test Veri Seti Boyutu	Devir Sayısı	Katman	Birim Sayısı/Eşik Değeri/Havuz Boyutu	Çekirdek Boyutu	Aktivasyon Fonksiyonu	Optimizasyon Fonksiyonu/ Öğrenme Aralığı	Parti Boyutu	Rastgele Öğrenme Durumu	Değerlendirme
CL1.10	2185 (%64)	547 (%16)	684 (%20)	50	TD(Conv1D)	32	3	Tanh	Adam 0,002	100	Evet	MSE = 0,051 MAE = 0,160 RMSE = 0,225 MAPE = 4,719 R ² = 0,742
					MaxPooling1D	2						
					Flatten							
					LSTM	50						
					Dropout	0,01						
Dense	1											
CL1.11	2185 (%64)	547 (%16)	684 (%20)	50	TD(Conv1D)	32	2	Tanh	Adam 0,002	100	Evet	MSE = 0,032 MAE = 0,130 RMSE = 0,180 MAPE = 3,894 R ² = 0,835
					MaxPooling1D	2						
					Flatten							
					LSTM	50						
					Dropout	0,01						
Dense	1											
CL1.12	2185 (%64)	547 (%16)	684 (%20)	50	TD(Conv1D)	32	2	Tanh	Adam 0,002	100	Evet	MSE = 0,325 MAE = 0,404 RMSE = 0,570 MAPE = 12,365 R ² = -0,662
					MaxPooling1D	3						
					Flatten							
					LSTM	50						
					Dropout	0,01						
Dense	1											

Tablo 3.4. (Devamı): Model 1 için CNN+LSTM algoritmasının hiperparametre değerleri ve hata metrik değerleri.

Model1	Eğitim Veri Seti Boyutu	Doğrulama Veri Seti Boyutu	Test Veri Seti Boyutu	Devir Sayısı	Katman	Birim Sayısı/Eşik Değeri/Havuz Boyutu	Aktivasyon Fonksiyonu	Çekirdek Boyutu	Optimizasyon Fonksiyonu/ Öğrenme Aralığı	Parti Boyutu	Rastgele Öğrenme Durumu	Değerlendirme
CL1.13	2185 (%64)	547 (%16)	684 (%20)	50	TD(Conv1D)	32	ReLU	2	Adam 0,002	50	Evet	MSE = 0,257 MAE = 0,435 RMSE = 0,507 MAPE = -0,315 R ² = 12,924
					MaxPooling1D	2						
					Flatten							
					LSTM	50	ReLU					
					Dropout	0,01						
CL1.14	2185 (%64)	547 (%16)	684 (%20)	50	Dense	1			Adam 0,005	50	Evet	MSE = 0,107 MAE = 0,255 RMSE = 0,328 MAPE = 7,702 R ² = 0,451
					TD(Conv1D)	32	Sigmoid	2				
					MaxPooling1D	2						
					Flatten							
					LSTM	50	Sigmoid					
CL1.15	1912 (%56)	479 (%14)	1025 (%30)	50	Dropout	0,01			Adam 0,005	50	Evet	MSE = 0,034 MAE = 0,145 RMSE = 0,183 MAPE = 4,474 R ² = 0,828
					Dense	1						
					TD(Conv1D)	32	Sigmoid	2				
					MaxPooling1D	2						
					Flatten							
					LSTM	50	Tanh					
					Dropout	0,01						
					Dense	1						
					TD(Conv1D)	32	Sigmoid	2				
					MaxPooling1D	2						

3.3. Model 2

Model 2'deki veri setinde BİST Kimya Petrol Plastik'in (XKMYA) ve QE Consumer Goods & Services'in (QECON) endeksinin açılı ve kapanış fiyatlarının, Türkiye Petrol Rafinelerileri A.Ş. (TUPRS), Petkim Petrokimya Holding A.Ş (PETKM) ve Qatar Fuel Co (QFLS) açılı ve kapanış hisse senedi fiyatlarının ve dolar (USD/TRY) açılış ve ortalama alış/satı kur değerleri girdi, PETKM hisse senedi kapanış fiyatı çıktı olarak kullanılmaktadır.

Model 2'nin RFR algoritmasında, Test veri kümesinin boyutu tüm veri setinin %20-%40'ı arasında değerler olarak belirlenmiştir. RandomForestRegressor metodunun içerisinde kullanılan hiperparametrelerden oluşturulacak ağaç sayısı 10-100 ve kullanılacak özellik miktarı 0,6-Otomatik (auto) değerleri arasında değişmektedir.

Tablo 3.5'te Model 2 için RFR algoritmasında kullanılan hiperparametre değerleri, tahmin değerlerinin gözlem değerleriyle olan hata değerleri (MSE, RMSE, MAE, MAPE) ve tahmin değerlerinin gözlem değerlerine uygunluğu (R^2) verilmiştir.

Model 2 için oluşturulan LSTM algoritması içerisindeki katmanlar sırası ile ilerlediği için Keras kütüphanesindeki models modülünden Sequential (sıralı) ve layers modülünden LSTM, Dropout ve Dense katmanları kullanılmaktadır Bu model için oluşturulan algorithmda LSTM katmanından birim sayısı 50 ve 100 olarak verilmiştir. LSTM katmanında Tanh, ReLU ve Sigmoid aktivasyon fonksiyonlar kullanılmıştır. LSTM katmanına bağlı olan Dropout katmanında eşik değer 0,02 ve 0,04 olarak verilmiştir. Dense katmanı tam bağlı çıktı katmanıdır. Algoritmanın tek çıktısı olduğu için bu katmandaki birim sayısı 1'dir. Katman sayısı artırıldığında elde edilen sonuçların hata değerleri arttığı için 2 katmanlı modeller oluşturulmuştur.

Optimizasyon (optimizer) fonksiyonu olarak Adam (Momentum) kullanılmıştır. Optimizasyon fonksiyonunun öğrenme oranı 0,001 ve 0,01 olarak belirlenmiştir. Kayıp fonksiyonu mean_squared_error olarak belirlenmiştir. Parti Boyutu (batch_size) 20-100 arasında belirlenmiştir. Devir sayısı (Epoch) 50 ve 70 olarak değişmektedir. Her eğitim işleminde verilerin rastgele seçimini belirleyen hiperparametre shuffle'dır. Buradaki testlerde genellikle shuffle hiperparametresi Evet olarak seçilmiştir.

Tablo 3.6’da Model 2 için LSTM algoritmasında kullanılan hiperarametre değerleri, tahmin değerlerinin gözlem değerleriyle olan hata değerleri (MSE, RMSE, MAE, MAPE) ve tahmin değerlerinin gözlem değerlerine uygunluğu (R^2) verilmiştir.

Oluşturulan CNN+LSTM algoritması içerisindeki katmanlar sırası ile ilerlediği için Keras kütüphanesindeki models modülünden Sequential (sıralı) ve layers modülünden TimeDistributed (TD), Conv1D, MaxPooling1D, Flatten, LSTM, Dropout, Dense katmanları kullanılmaktadır. TimeDistributed katmanı Conv1D katmanını LSTM katmanına bağlamak için kullanılmıştır. Conv1D katmanındaki filtre sayısı (filters) 16, 32, 64, 128 olarak belirlenmiştir. Çekirdek boyutu (kernel_size) 2 ve 3 olarak kullanılmıştır. MaxPooling1D katmanında havuz boyutu (pool_size) 2 ve 3 olarak belirtilmiştir. Flatten katmanı ile veriler düzleştirilmektedir. LSTM katmanında birim sayısı 50 ve 70 olarak değişmektedir. Conv1D, LSTM katmanlarında Tanh, ReLU ve Sigmoid aktivasyon fonksiyonları kullanılmıştır. LSTM katmanına bağlı Dropout katmanındaki eşik değeri 0,01 ve 0,03 olarak verilmiştir. Dense katmanı çıktı katmanıdır. Algoritmanın tek çıktısı olduğu için bu katmandaki birim sayısı 1’dir.

Optimizasyon (optimizer) fonksiyonu olarak Adam (Momentum) kullanılmıştır. Optimizasyon fonksiyonunun öğrenme oranı 0,001-0,005 arasında değişmektedir. Kayıp fonksiyonu mean_squared_error olarak belirlenmiştir. Parti Boyutu (batch_size) değeri 100 olarak belirlenmiştir. Devir sayısı (Epoch) 50-100 arasında değişmektedir. Rastgele Öğrenme Durumu hiperparametresi shuffle, Evet olarak seçilmiştir.

Tablo 3.7’de Model 2 için CNN+LSTM algoritmasında kullanılan hiperarametre değerleri ve modellerin değerlendirme sonuçları verilmiştir.

Tablo 3.5. Model 2 için RFR algoritmasının hiperparametre değerleri ve hata metrik değerleri.

Model2	Eğitim Veri Seti Boyutu	Test Veri Seti Boyutu	Rasgele Veri Bölme Durumu	Ağaç Sayısı	Maksimum Özellik Sayısı	Değerlendirme				
						MSE	MAE	RMSE	MAPE	R ²
R2.1	2049 (%60)	1367 (%40)	Hayır	10	Auto	1,589	1,098	1,260	31,482	-2,752
R2.2	2049 (%60)	1367 (%40)	Hayır	50	Auto	1,541	1,076	1,241	30,798	-2,640
R2.3	2049 (%60)	1367 (%40)	Hayır	100	Auto	1,533	1,072	1,238	30,668	-2,622
R2.4	2391 (%70)	1025 (%30)	Hayır	10	Auto	0,264	0,368	0,514	9,858	0,102
R2.5	2391 (%70)	1025 (%30)	Hayır	50	Auto	0,253	0,358	0,503	9,625	0,141
R2.6	2391 (%70)	1025 (%30)	Hayır	100	Auto	0,248	0,348	0,498	9,286	0,155
R2.7	2391 (%70)	1025 (%30)	Hayır	50	0,8	0,246	0,350	0,496	9,413	0,164
R2.8	2391 (%70)	1025 (%30)	Hayır	100	0,8	0,253	0,354	0,503	9,467	0,140
R2.9	2391 (%70)	1025 (%30)	Hayır	100	0,6	0,236	0,345	0,486	9,325	0,196
R.2.10	2732 (%80)	684 (%20)	Hayır	10	Auto	0,053	0,169	0,231	5,114	0,726
R2.11	2732 (%80)	684 (%20)	Hayır	50	Auto	0,046	0,167	0,215	5,083	0,763
R2.12	2732 (%80)	684 (%20)	Hayır	100	Auto	0,049	0,175	0,222	5,309	0,748
R2.13	2732 (%80)	684 (%20)	Hayır	50	0,8	0,055	0,193	0,235	5,915	0,716
R2.14	2732 (%80)	684 (%20)	Hayır	100	0,8	0,052	0,185	0,228	5,635	0,733
R2.15	2732 (%80)	684 (%20)	Hayır	100	0,6	0,078	0,232	0,280	7,132	0,598

Tablo 3.6. Model 2 için LSTM algoritmasının hiperparametre değerleri ve hata metrik değerleri.

Model2	Eğitim Veri Seti Boyutu	Doğrulama Veri Seti Boyutu	Test Veri Seti Boyutu	Devir Sayısı	Katman	Birim Sayısı/ Eşik Değeri	Parti Boyutu	Aktivasyon Fonksyonu	Rastgele Öğrenme Durumu	Optimizasyon Fonksiyonu / Öğrenme Aralığı	Değerlendirme
L2.1	1639 (%48)	341 (%12)	1367 (%40)	50	LSTM Dropout Dense	50 0,02 1	20	Tanh	Evet	Adam 0,01	MSE = 0,368 MAE = 0,497 RMSE = 0,607 MAPE = 14,260 R ² = 0,130
L2.2	1639 (%48)	341 (%12)	1367 (%40)	50	LSTM Dropout Dense	100 0,02 1	20	Tanh	Evet	Adam 0,01	MSE = 0,263 MAE = 0,436 RMSE = 0,513 MAPE = 12,625 R ² = 0,378
L2.3	1639 (%48)	341 (%12)	1367 (%40)	50	LSTM Dropout Dense	100 0,02 1	20	Tanh	Hayır	Adam 0,01	MSE = 3,435 MAE = 1,757 RMSE = 1,853 MAPE = 53.001 R ² = -7,111
L2.4	1912 (%56)	479 (%14)	1025 (%30)	50	LSTM Dropout Dense	50 0,02 1	20	Tanh	Evet	Adam 0,01	MSE = 0,117 MAE = 0,239 RMSE = 0,342 MAPE = 6,393 R ² = 0,603
L2.5	1912 (%56)	479 (%14)	1025 (%30)	50	LSTM Dropout Dense	50 0,02 1	20	Tanh	Evet	Adam 0,001	MSE = 0,020 MAE = 0,099 RMSE = 0,142 MAPE = 2,893 R ² = 0,931

Tablo 3.6. (Devamı): Model 2 için LSTM algoritmasının hiperparametre değerleri ve hata metrik değerleri.

Model2	Eğitim Veri Seti Boyutu	Doğrulama Veri Seti Boyutu	Test Veri Seti Boyutu	Devir Sayısı	Katman	Birim Sayısı/ Eşik Değeri	Parti Boyutu	Aktivasyon Fonksyonu	Rastgele Öğrenme Durumu	Optimizasyon Fonksiyonu/ Öğrenme Aralığı	Değerlendirme
L2.6	1912 (%56)	479 (%14)	1025 (%30)	70	LSTM Dropout Dense	50 0,02 1	50	Tanh	Evet	Adam 0,001	MSE = 0,126 MAE = 0,319 RMSE = 0,355 MAPE = 9,415 R ² = 0,573
L2.7	1912 (%56)	479 (%14)	1025 (%30)	70	LSTM Dropout Dense	50 0,02 1	100	Tanh	Evet	Adam 0,001	MSE = 0,074 MAE = 0,211 RMSE = 0,271 MAPE = 6,036 R ² = 0,750
L2.8	1912 (%56)	479 (%14)	1025 (%30)	70	LSTM Dropout Dense	50 0,02 1	20	Tanh	Evet	Adam 0,001	MSE = 0,032 MAE = 0,140 RMSE = 0,891 MAPE = 4,190 R ² = 0,891
L2.9	1912 (%56)	479 (%14)	1025 (%30)	70	LSTM Dropout Dense	50 0,04 1	20	Tanh	Evet	Adam 0,001	MSE = 0,036 MAE = 0,189 RMSE = 0,151 MAPE = 4,515 R ² = 0,878
L2.10	2185 (%64)	547 (%16)	684 (%20)	50	LSTM Dropout Dense	50 0,02 1	20	Tanh	Evet	Adam 0,01	MSE = 0,027 MAE = 0,127 RMSE = 0,164 MAPE = 3,915 R ² = 0,863

Tablo 3.6. (Devamı): Model 2 için LSTM algoritmasının hiperparametre değerleri ve hata metrik değerleri.

Model2	Eğitim Veri Seti Boyutu	Doğrulama Veri Seti Boyutu	Test Veri Seti Boyutu	Devir Sayısı	Katman	Birim Sayısı/ Eşik Değeri	Parti Boyutu	Aktivasyon Fonksiyonu	Rastgele Öğrenme Durumu	Optimizasyon Fonksiyonu/ Öğrenme Aralığı	Değerlendirme Hata Metrikleri
L2.11	2185 (%64)	547 (%16)	684 (%20)	50	LSTM Dropout Dense	50 0,02 1	20	Tanh	Evet	Adam 0,001	MSE = 0,101 MAE = 0,261 RMSE = 0,318 MAPE = 7,695 R ² = 0,482
L2.12	2185 (%64)	547 (%16)	684 (%20)	50	LSTM Dropout Dense	100 0,02 1	20	Tanh	Evet	Adam 0,001	MSE = 0,036 MAE = 0,141 RMSE = 0,190 MAPE = 4,175 R ² = 0,816
L2.13	2185 (%64)	547 (%16)	684 (%20)	70	LSTM Dropout Dense	100 0,02 1	20	Tanh	Evet	Adam 0,001	MSE = 0,076 MAE = 0,235 RMSE = 0,276 MAPE = 7,038 R ² = 0,610
L2.14	2185 (%64)	547 (%16)	684 (%20)	70	LSTM Dropout Dense	100 0,02 1	20	Sigmoid	Evet	Adam 0,001	MSE = 0,116 MAE = 0,306 RMSE = 0,340 MAPE = 9,637 R ² = 0,409
L2.15	2185 (%64)	547 (%16)	684 (%20)	70	LSTM Dropout Dense	100 0,02 1	20	ReLU	Evet	Adam 0,001	MSE = 0,469 MAE = 0,605 RMSE = 0,685 MAPE = 17,955 R ² = -1,396

Tablo 3.7. Model 2 için CNN+LSTM algoritmasının hiperparametre değerleri ve hata metrik değerleri.

Tablo 3.7. Model 2 için CNN+LSTM algoritmasının hiperparametre değerleri ve hata metrik değerleri.

Tablo 3.7. (Devamı): Model 2 için CNN+LSTM algoritmasının hiperparametre değerleri ve hata metrik değerleri.

Model2	Eğitim Veri Seti Boyutu	Doğrulama Veri Seti Boyutu	Test Veri Seti Boyutu	Devir Sayısı	Katman	Birim Sayısı/Eşik Değeri/Havuz Boyutu	Çekidek Boyutu	Aktivasyon Fonksiyonu	Optimizasyon Fonksiyonu/ Öğrenme Adımı	Parti Boyutu	Rastgele Öğrenme Durumu	Değerlendirme
CL2.4	1912 (%56)	479 (%14)	1025 (%30)	70	TD(Conv1D)	64	2	Tanh	Adam 0,001	100	Evet	MSE = 0,199 MAE = 0,274 RMSE = 0,446 MAPE = 7,087 R ² = 0,323
					MaxPooling1D	2						
					Flatten		Tanh					
					LSTM	70						
					Dropout	0,01						
					Dense	1						
CL2.5	1912 (%56)	479 (%14)	1025 (%30)	70	TD(Conv1D)	16	2	Tanh	Adam 0,001	100	Evet	MSE = 0,172 MAE = 0,246 RMSE = 0,415 MAPE = 6,390 R ² = 0,415
					MaxPooling1D	2						
					Flatten		Tanh					
					LSTM	70						
					Dropout	0,01						
					Dense	1						
CL2.6	1912 (%56)	479 (%14)	1025 (%30)	70	TD(Conv1D)	32	2	Tanh	Adam 0,001	100	Evet	MSE = 0,241 MAE = 0,329 RMSE = 0,491 MAPE = 8,601 R ² = 0,181
					MaxPooling1D	2						
					Flatten		Tanh					
					LSTM	50						
					Dropout	0,01						
					Dense	1						

Tablo 3.7. (Devamı): Model 2 için CNN+LSTM algoritmasının hiperparametre değerleri ve hata metrik değerleri.

Model2	Eğitim Veri Seti Boyutu	Doğrulama Veri Seti Boyutu	Test Veri Seti Boyutu	Devir Sayısı	Katman	Birim Sayısı/Eşik Değeri/Havuz Boyutu	Çekirdek Boyutu	Aktivasyon Fonksiyonu	Optimizasyon Fonksiyonu/ Öğrenme Adımı	Parti Boyutu	Rastgele Öğrenme Durumu	Değerlendirme
CL2.7	1912 (%56)	479 (%14)	1025 (%30)	70	Conv1D	16	2	Tanh	Adam 0,001	100	Evet	MSE = 0,267 MAE = 0,337 RMSE = 0,517 MAPE = 8,781 R ² = 0,092
					MaxPooling1D	2						
					Flatten		Tanh					
					LSTM	50						
					Dropout	0,01						
					Dense	1						
CL2.8	1912 (%56)	479 (%14)	1025 (%30)	70	Conv1D	16	2	ReLU	Adam 0,001	100	Evet	MSE = 0,387 MAE = 0,445 RMSE = 0,622 MAPE = 12,263 R ² = -0,313
					MaxPooling1D	2						
					Flatten		ReLU					
					LSTM	50						
					Dropout	0,01						
					Dense	1						
CL2.9	1912 (%56)	479 (%14)	1025 (%30)	70	Conv1D	16	2	Sigmoid	Adam 0,001	100	Evet	MSE = 1,275 MAE = 0,984 RMSE = 1,129 MAPE = 27,220 R ² = -3,333
					MaxPooling1D	2						
					Flatten		Sigmoid					
					LSTM	50						
					Dropout	0,01						
					Dense	1						

Model2	Eğitim Veri Seti Boyutu	Doğrulama Veri Seti Boyutu	Test Veri Seti Boytu	Devir Sayısı	Katman	Birim Sayısı/Eşik Değeri/Havuz Boyutu	Çekirdek Boyutu	Aktivasyon Fonksiyonu	Optimizasyon Fonksiyonu/ Öğrenme Adımı	Parti Boyutu	Rastgele Öğrenme Durumu	Değerlendirme
CL2.10	2185 (%64)	547 (%16)	684 (%20)	70	TD(Conv1D)	16	2	Sigmoid	Adam 0,001	100	Evet	MSE = 0,125 MAE = 0,262 RMSE = 0,353 MAPE = 7,503 R² = 0,364
					MaxPooling1D	2						
					Flatten		Sigmoid					
					LSTM	50						
					Dropout	0,01						
					Dense	1						
CL2.11	2185 (%64)	547 (%16)	684 (%20)	70	TD(Conv1D)	16	2	Tanh	Adam 0,001	100	Evet	MSE = 0,129 MAE = 0,296 RMSE = 0,359 MAPE = 8,849 R² = 0,340
					MaxPooling1D	2						
					Flatten		Tanh					
					LSTM	50						
					Dropout	0,01						
					Dense	1						
CL2.12	2185 (%64)	547 (%16)	684 (%20)	100	TD(Conv1D)	16	2	Tanh	Adam 0,001	100	Evet	MSE = 0,047 MAE = 0,170 RMSE = 0,216 MAPE = 5,074 R² = 0,762
					MaxPooling1D	2						
					Flatten		Tanh					
					LSTM	50						
					Dropout	0,01						
					Dense	1						

Model2	Eğitim Veri Seti Boyutu	Doğrulama Veri Seti Boyutu	Test Veri Seti Boytu	Devir Sayısı	Katman	Birim Sayısı/Eşik Değeri/Havuz Boyutu	Çekirdek Boyutu	Aktivasyon Fonksiyonu	Optimizasyon Fonksiyonu/ Öğrenme Adımı	Parti Boyutu	Rastgele Öğrenme Durumu	Değerlendirme
CL2.10	2185 (%64)	547 (%16)	684 (%20)	70	TD(Conv1D)	16	2	Sigmoid	Adam 0,001	100	Evet	MSE = 0,125 MAE = 0,262 RMSE = 0,353 MAPE = 7,503 R² = 0,364
					MaxPooling1D	2						
					Flatten		Sigmoid					
					LSTM	50						
					Dropout	0,01						
					Dense	1						
CL2.11	2185 (%64)	547 (%16)	684 (%20)	70	TD(Conv1D)	16	2	Tanh	Adam 0,001	100	Evet	MSE = 0,129 MAE = 0,296 RMSE = 0,359 MAPE = 8,849 R² = 0,340
					MaxPooling1D	2						
					Flatten		Tanh					
					LSTM	50						
					Dropout	0,01						
					Dense	1						
CL2.12	2185 (%64)	547 (%16)	684 (%20)	100	TD(Conv1D)	16	2	Tanh	Adam 0,001	100	Evet	MSE = 0,047 MAE = 0,170 RMSE = 0,216 MAPE = 5,074 R² = 0,762
					MaxPooling1D	2						
					Flatten		Tanh					
					LSTM	50						
					Dropout	0,01						
					Dense	1						

Tablo 3.7. (Devamı): Model 2 için CNN+LSTM algoritmasının hiperparametre değerleri ve hata metrik değerleri.

Model2	Eğitim Veri Seti Boyutu	Doğrulama Veri Seti Boutu	Test Veri Seti Boyutu	Devir Sayısı	Katman	Birim Sayısı/Eşik Değeri/Havuz Boyutu	Aktivasyon Fonksiyonu	Çekirdek Boyutu	Optimizasyon Fonksiyonu /Öğrenme Aralığı	Parti Boyutu	Rastgele Öğrenme Durumu	Değerlendirme
76	CL.2.13	2185 (%64)	547 (%16)	100	TD(Conv1D)	32	2	Tanh	Adam 0,001	100	Evet	MSE = 0,044 MAE = 0,159 RMSE = 0,209 MAPE = 4,736 R ² = 0,778
					MaxPooling1D	2						
					Flatten							
					LSTM	50		Tanh				
					Dropout	0,01						
	CL2.14	2185 (%64)	547 (%16)	100	Dense	1			Adam 0,001	100	Evet	MSE = 0,053 MAE = 0,185 RMSE = 0,231 MAPE = 5,580 R ² = 0,727
					TD(Conv1D)	32	2	Tanh				
					MaxPooling1D	2						
					Flatten							
					LSTM	50		Tanh				
	CL2.15	2185 (%64)	547 (%16)	100	Dropout	0,03			Adam 0,005	100	Evet	MSE = 0,090 MAE = 0,208 RMSE = 0,299 MAPE = 6,166 R ² = 0,542
					Dense	1						
					TD(Conv1D)	32	2	Tanh				
					MaxPooling1D	2						
					Flatten							

3.4. Model 3

Model 3'teki veri setinde BİST Kimya Petrol Plastik'in (XKMYA), MOEX Oil and Gas (MOEXOG); RTS Oil & Gas (RTSOG) endeks açılış/kapanış fiyatları, Türkiye Petrol Rafinelerileri A.Ş. (TUPRS), Petkim Petrokimya Holding A.Ş (PETKM) ve Lukoil (LKOI) hisse senedi açılış/kapanış fiyatları ve dolar (USD/TRY) fiyatının açılış ve ortalama alış/satış kur değeri girdi, PETKM hisse senedi kapanış fiyatı çıktı olarak kullanılmaktadır.

Model 3'ün RFR algoritmasında, test veri kümesinin boyutu tüm veri setinin %20-%40'ı arasındaki değerlerden belirlenmiştir. RandomForestRegressor metodunun içerisinde kullanılan hiperparametrelerden oluşturulacak ağaç sayısı 10-100 ve kullanılacak özellik miktarı 0,6-Otomatik (auto) değerleri arasında değişmektedir.

Tablo 3.8'de Model 3 için RFR algoritmasında kullanılan hiperparametre değerleri, tahmin değerlerinin gözlem değerleriyle olan hata değerleri (MSE, RMSE, MAE, MAPE) ve tahmin değerlerinin gözlem değerlerine uygunluğu (R^2) verilmiştir.

Model 3 için oluşturulan LSTM algoritması içerisindeki katmanlar sırası ile ilerlediği için Keras kütüphanesindeki models modülünden Sequential (sıralı) ve layers modülünden LSTM, Dropout ve Dense katmanları kullanılmaktadır. Bu model için oluşturulan algorithmda LSTM katmanından birim sayısı 50-100 arasında değişmektedir. LSTM katmanında Tanh, ReLU ve Sigmoid aktivasyon fonksiyonları kullanılmıştır. LSTM katmanına bağlı olan Dropout katmanında eşik değeri 0,02-0,08 aralığında değişmektedir. Dense katmanı tam bağlı çıktı katmanıdır. Algoritmanın tek çıktısı olduğu için bu katmandaki birim sayısı 1'dir. Katman sayısı artırıldığında elde edilen sonuçların hata değerleri arttığı için 2 katmanlı modeller oluşturulmuştur.

Optimizasyon (optimizer) fonksiyonu olarak Adam (Momentum) kullanılmıştır. Optimizasyon fonksiyonunun öğrenme oranı 0,002-0,02 arasında değişmektedir. Kayıp fonksiyonu mean_squared_error olarak belirlenmiştir. Parti Boyutu (batch_size) 30 ve 50 olarak belirlenmiştir. Devir sayısı (Epoch) 50-100 arasında değişmektedir. Tüm testlerde shuffle hiperparametresi Evet olarak seçilmiştir.

Tablo 3.9’da Model 3 için LSTM algoritmasında kullanılan hiperarametre değerleri, tahmin değerlerinin gözlem değerleriyle olan hata değerleri (MSE, RMSE, MAE) ve tahmin değerlerinin gözlem değerlerine uygunluğu (R^2) verilmiştir.

Oluşturulan CNN+LSTM algoritması içerisindeki katmanlar sırası ile ilerlediği için Keras kütüphanesindeki models modülünden Sequential (sıralı) ve layers modülünden TimeDistributed (TD), Conv1D, MaxPooling1D, Flatten, LSTM, Dropout, Dense katmanları kullanılmaktadır. TimeDistributed katmanı Conv1D katmanını LSTM katmanına bağlamak için kullanılmıştır. Conv1D katmanındaki filtre sayısı (filters) 32 ve 64 olarak test edilmiştir. Çekirdek boyutu (kernel_size) 2 ve 3 olarak kullanılmıştır. MaxPooling1D katmanında havuz boyutu (pool_size) 2 ve 3 olarak belirtilmiştir. Flatten katmanı ile veriler düzleştirilmektedir. LSTM katmanından birim sayısı 50 ve 70 olarak değişmektedir. Conv1D, LSTM katmanlarında Tanh, ReLU ve Sigmoid aktivasyon fonksiyonları kullanılmıştır. LSTM katmanına bağlı Dropout katmanındaki eşik değeri 0,001 ve 0,01 olarak verilmiştir. Dense katmanı çıktı katmanıdır. Algoritmanın tek çıktısı olduğu için bu katmandaki birim sayısı 1’dir.

Optimizasyon (optimizer) fonksiyonu olarak Adam (Momentum) kullanılmıştır. Optimizasyon fonksiyonunun öğrenme oranı 0,0002-0,003 arasında değişmektedir. Kayıp fonksiyonu mean_squared_error olarak belirlenmiştir. Parti Boyutu (batch_size) 50 ve 100 olarak belirlenmiştir. Devir sayısı (Epoch) 50-80 arasında değişmektedir. Rastgele Öğrenme Durumu hiperparametresi shuffle, Evet olarak seçilmiştir.

Tablo 3.10’da Model 3 için CNN+LSTM algoritmasında kullanılan hiperarametre değerleri ve modellerin değerlendirme sonuçları verilmiştir.

Tablo 3.8. Model 3 için RFR algoritmasının hiperparametre değerleri ve hata metrik değerleri.

Model3	Eğitim Veri Seti Boyutu	Test Veri Seti Boyutu	Rastgele Veri Bölme	Ağaç Sayısı	Maksimum Özellik Sayısı	Değerlendirme				
						MSE	MAE	RMSE	MAPE	R ²
R3.1	2049 (%60)	1367 (%40)	Hayır	10	Auto	1,635	1,111	1,279	31,833	-2,861
R3.2	2049 (%60)	1367 (%40)	Hayır	50	Auto	1,500	1,059	1,225	30,282	-2,543
R3.3	2049 (%60)	1367 (%40)	Hayır	100	Auto	1,476	1,048	1,215	29,945	-2,487
R3.4	2391 (%70)	1025 (%30)	Hayır	10	Auto	0,233	0,365	0,483	10,499	0,207
R3.5	2391 (%70)	1025 (%30)	Hayır	50	Auto	0,217	0,356	0,466	10,237	0,260
R3.6	2391 (%70)	1025 (%30)	Hayır	100	Auto	0,219	0,354	0,469	10,111	0,283
R3.7	2391 (%70)	1025 (%30)	Hayır	50	0,8	0,220	0,360	0,469	10,316	0,251
R3.8	2391 (%70)	1025 (%30)	Hayır	100	0,8	0,225	0,364	0,475	10,416	0,234
R3.9	2391 (%70)	1025 (%30)	Hayır	100	0,6	0,222	0,363	0,471	10,428	0,245
R3.10	2732 (%80)	684 (%20)	Hayır	10	Auto	0,043	0,156	0,207	4,721	0,781
R3.11	2732 (%80)	684 (%20)	Hayır	50	Auto	0,052	0,177	0,229	5,411	0,733
R3.12	2732 (%80)	684 (%20)	Hayır	100	Auto	0,047	0,167	0,218	5,105	0,757
R3.13	2732 (%80)	684 (%20)	Hayır	50	0,8	0,053	0,182	0,231	5,549	0,726
R3.14	2732 (%80)	684 (%20)	Hayır	100	0,8	0,054	0,189	0,232	5,798	0,723
R3.15	2732 (%80)	684 (%20)	Hayır	100	0,6	0,067	0,205	0,259	6,261	0,657

Tablo 3.9. Model 3 için LSTM algoritmasının hiperparametre değerleri ve hata metrik değerleri.

Model3	Eğitim Veri Seti Boyutu	Doğrulama Veri Seti Boyutu	Test Veri Seti boyutu	Devir Sayısı	Katman	Birim Sayısı/ Eşik Değeri	Parti Boyutu	Aktivasyon Fonksiyonu	Rastgele Öğrenme Durumu	Optimizasyon Fonksiyonu/ Öğrenme Aralığı	Değerlendirme
L3.1	1639 (%48)	410 (%12)	1367 (%40)	50	LSTM Dropout Dense	50 0,02 1	50	Tanh	Evet	Adam 0,002	MSE = 0,221 MAE = 0,412 RMSE = 0,470 MAPE = 12,320 R ² = 0,479
L3.2	1639 (%48)	410 (%12)	1367 (%40)	70	LSTM Dropout Dense	50 0,02 1	50	Tanh	Evet	Adam 0,002	MSE = 0,243 MAE = 0,436 RMSE = 0,493 MAPE = 13,148 R ² = 0,425
L3.3	1639 (%48)	410 (%12)	1367 (%40)	100	LSTM Dropout Dense	50 0,02 1	50	Tanh	Evet	Adam 0,002	MSE = 0,097 MAE = 0,252 RMSE = 0,311 MAPE = 7,794 R ² = 0,772
L3.4	1912 (%56)	479 (%14)	1025 (%30)	50	LSTM Dropout Dense	50 0,02 1	50	Tanh	Evet	Adam 0,002	MSE = 0,044 MAE = 0,168 RMSE = 0,209 MAPE = 4,965 R ² = 0,851
L3.5	1912 (%56)	479 (%14)	1025 (%30)	50	LSTM Dropout Dense	50 0,02 1	50	Tanh	Evet	Adam 0,02	MSE = 0,434 MAE = 0,615 RMSE = 0,659 MAPE = 17,504 R ² = -0,475

Tablo 3.9. (Devamı): Model 3 için LSTM algoritmasının hiperparametre değerleri ve hata metrik değerleri.

Model3	Eğitim Veri Seti Boyutu	Doğrulama Veri Seti Boyutu	Test Veri Seti Boyutu	Devir Sayısı	Katman	Birim Sayısı/ Eşik Değeri	Parti Boyutu	Aktivasyon Fonksiyonu	Rastgele Öğrenme Durumu	Optimizasyon Fonksiyonu/ Öğrenme Aralığı	Değerlendirme
L3.6	1912 (%56)	479 (%14)	1025 (%30)	50	LSTM Dropout Dense	50 0,02 1	50	Sigmoid	Evet	Adam 0,002	MSE = 0,839 MAE = 0,783 RMSE = 0,916 MAPE = 21,567 R ² = -1,850
L3.7	1912 (%56)	479 (%14)	1025 (%30)	50	LSTM Dropout Dense	50 0,02 1	50	ReLU	Evet	Adam 0,002	MSE = 0,143 MAE = 0,326 RMSE = 0,378 MAPE = 9,559 R ² = 0,513
L3.8	1912 (%56)	479 (%14)	1025 (%30)	50	LSTM Dropout Dense	50 0,05 1	50	Tanh	Evet	Adam 0,002	MSE = 0,031 MAE = 0,132 RMSE = 0,177 MAPE = 3,824 R ² = 0,894
L3.9	1912 (%56)	479 (%14)	1025 (%30)	50	LSTM Dropout Dense	50 0,08 1	50	Tanh	Evet	Adam 0,002	MSE = 0,034 MAE = 0,144 RMSE = 0,186 MAPE = 4,317 R ² = 0,883
L3.10	2185 (%64)	547 (%16)	684 (%20)	50	LSTM Dropout Dense	50 0,02 1	50	Tanh	Evet	Adam 0,002	MSE = 0,027 MAE = 0,126 RMSE = 0,165 MAPE = 3,812 R ² = 0,861

Tablo 3.9. (Devamı): Model 3 için LSTM algoritmasının hiperparametre değerleri ve hata metrik değerleri.

Model3	Eğitim Veri Seti Boyutu	Doğrulama Veri Seti Boyutu	Test Veri Seti Boyutu	Devir Sayısı	Katman	Birim Sayısı/ Eşik Değeri	Parti Boyutu	Aktivasyon Fonksyonu	Rastgele Öğrenme Durumu	Optimizasyon Fonksiyonu/ Öğrenme Aralığı	Değerlendirme
L3.11	2185 (%64)	547 (%16)	684 (%20)	50	LSTM Dropout Dense	70 0,05 1	50	Tanh	Evet	Adam 0,002	MSE = 0,027 MAE = 0,118 RMSE = 0,163 MAPE = 3,419 R ² = 0,909
L3.12	2185 (%64)	547 (%16)	684 (%20)	50	LSTM Dropout Dense	100 0,05 1	50	Tanh	Evet	Adam 0,002	MSE = 0,021 MAE = 0,101 RMSE = 0,146 MAPE = 2,926 R ² = 0,928
L3.13	2185 (%64)	547 (%16)	684 (%20)	50	LSTM Dropout Dense	100 0,05 1	30	Tanh	Evet	Adam 0,002	MSE = 0,041 MAE = 0,167 RMSE = 0,203 MAPE = 4,873 R ² = 859
L3.14	2185 (%64)	547 (%16)	684 (%20)	70	LSTM Dropout Dense	100 0,05 1	30	Tanh	Evet	Adam 0,002	MSE = 0,072 MAE = 0,229 RMSE = 0,268 MAPE = 6,459 R ² = 0,757
L3.15	2185 (%64)	547 (%16)	684 (%20)	100	LSTM Dropout Dense	100 0,05 1	30	Tanh	Evet	Adam 0,002	MSE = 0,068 MAE = 0,220 RMSE = 0,260 MAPE = 6,234 R ² = 0,770

Model3	Eğitim Veri Seti Boyutu	Doğrulama Veri Seti Boyutu	Test Veri Seti Boyutu	Devir Sayısı	Katman	Birim Sayısı/Eşik Değeri/Havuz Boyutu	Çekirdek Boyutu	Aktivasyon Fonksiyonu	Optimizasyon Fonksiyonu/ Öğrenme Aralığı	Parti Boyutu	Rastgele Öğrenme Durumu	Değerlendirme
CL3.1	1639 (%48)	410 (%12)	1367 (%40)	50	TD(Conv1D)	32	2	Tanh	Adam 0,002	50	Evet	MSE = 0,297 MAE = 0,399 RMSE = 0,545 MAPE = 11,660 R² = 0,299
					MaxPooling1D	2						
					Flatten		Tanh					
					LSTM	50						
					Dropout	0,01						
					Dense	1						
CL3.2	1639 (%48)	410 (%12)	1367 (%40)	80	TD(Conv1D)	32	2	Tanh	Adam 0,002	50	Evet	MSE = 0,324 MAE = 0,418 RMSE = 0,569 MAPE = 12,224 R² = 0,235
					MaxPooling1D	2						
					Flatten		Tanh					
					LSTM	50						
					Dropout	0,01						
					Dense	1						
CL3.3	1639 (%48)	410 (%12)	1367 (%40)	80	TD(Conv1D)	32	3	Tanh	Adam 0,002	50	Evet	MSE = 0,373 MAE = 0,481 RMSE = 0,611 MAPE = 13,968 R² = 0,119
					MaxPooling1D	2						
					Flatten		Tanh					
					LSTM	50						
					Dropout	0,01						
					Dense	1						

Tablo 3.10. (Devamı): Model 3 için CNN+LSTM algoritmasının hiperparametre değerleri ve hata metrik değerleri.

Model3	Eğitim Veri Seti Boyutu	Doğrulama Veri Seti Boyutu	Test Veri Seti Boyutu	Devir Sayısı	Katman	Birim Sayısı/Eşik Değeri/ Havuz Boyutu	Çekirdek Boyutu	Aktivasyon Fonksiyonu	Optimizasyon Fonksiyonu/ Öğrenme Aralığı	Parti Boyutu	Rastgele Öğrenme Durumu	Değerlendirme
CL3.4	1912 (%56)	479 (%14)	1025 (%30)	80	TD(Conv1D)	32	3	Tanh	Adam 0,002	50	Evet	MSE = 0,420 MAE = 0,512 RMSE = 0,648 MAPE = 13,776 R² = -0,427
					MaxPooling1D	2						
					Flatten							
					LSTM	50						
					Dropout	0,01						
Dense	1											
CL3.5	1912 (%56)	479 (%14)	1025 (%30)	80	TD(Conv1D)	32	2	Tanh	Adam 0,002	50	Evet	MSE = 0,664 MAE = 0,666 RMSE = 0,815 MAPE = 18,097 R² = -1,257
					MaxPooling1D	3						
					Flatten							
					LSTM	50						
					Dropout	0,01						
Dense	1											
CL3.6	1912 (%56)	479 (%14)	1025 (%30)	50	TD(Conv1D)	64	2	Tanh	Adam 0,002	50	Evet	MSE = 0,546 MAE = 0,618 RMSE = 0,739 MAPE = 16,896 R² = -0,854
					MaxPooling1D	2						
					Flatten							
					LSTM	50						
					Dropout	0,001						
Dense	1											

Tablo 3.10. (Devamı): Model 3 için CNN+LSTM algoritmasının hiperparametre değerleri ve hata metrik değerleri.

Model3	Eğitim Veri Seti Boyutu	Doğrulama Veri Seti Boyutu	Test Veri Seti Boyutu	Devir Sayısı	Katman	Birim Sayısı/Eşik Değeri/ Havuz Boyutu	Aktivasyon Fonksiyonu	Çekirdek Boyutu	Optimizasyon Fonksiyonu/ Öğrenme Aralığı	Parti Boyutu	Rastgele Öğrenme Durumu	Değerlendirme
CL3.7	1912 (%56)	479 (%14)	1025 (%30)	50	TD(Conv1D)	32	2	Tanh	Adam 0,002	50	Evet	MSE = 0,442 MAE = 0,510 RMSE = 0,665 MAPE = 13,690 R² = -0,502
					MaxPooling1D	2						
					Flatten							
					LSTM	70		Tanh				
					Dropout	0,01						
				Dense	1							
CL3.8	1912 (%56)	479 (%14)	1025 (%30)	50	TD(Conv1D)	32	2	Tanh	Adam 0,002	100	Evet	MSE = 0,385 MAE = 0,456 RMSE = 0,621 MAPE = 12,130 R² = -0,309
					MaxPooling1D	2						
					Flatten							
					LSTM	70		Tanh				
					Dropout	0,01						
				Dense	1							
CL3.9	1912 (%56)	479 (%14)	1025 (%30)	50	TD(Conv1D)	32	2	Tanh	Adam 0,0002	100	Evet	MSE = 0,198 MAE = 0,307 RMSE = 0,445 MAPE = 8,214 R² = 0,326
					MaxPooling1D	2						
					Flatten							
					LSTM	70		Tanh				
					Dropout	0,01						
				Dense	1							

Tablo 3.10. (Devamı): Model 3 için CNN+LSTM algoritmasının hiperparametre değerleri ve hata metrik değerleri.

Model3	Eğitim Veri Seti Boyutu	Doğrulama Veri Seti Boyutu	Test Veri Seti Boyutu	Devir Sayısı	Katman	Birim Sayısı/Eşik Değeri/Havuz Boyutu	Çekirdek Boyutu	Aktivasyon Fonksiyonu	Optimizasyon Fonksiyonu/ Öğrenme Aralığı	Parti Boyutu	Rastgele Öğrenme Durumu	Değerlendirme
CL3.10	2185 (%64)	547 (%16)	684 (%20)	50	TD(Conv1D)	32	2	Tanh	Adam 0,002	50	Evet	MSE = 0,022 MAE = 0,105 RMSE = 0,150 MAPE = 3,180 R² = 0,885
					MaxPooling1D	2						
					Flatten							
					LSTM	50		Tanh				
					Dropout	0,01						
					Dense	1						
CL3.11	2185 (%64)	547 (%16)	684 (%20)	50	TD(Conv1D)	32	2	Tanh	Adam 0,001	50	Evet	MSE = 0,062 MAE = 0,215 RMSE = 0,250 MAPE = 6,572 R² = 0,681
					MaxPooling1D	2						
					Flatten							
					LSTM	70		Tanh				
					Dropout	0,01						
					Dense	1						
CL3.12	2185 (%64)	547 (%16)	684 (%20)	50	TD(Conv1D)	32	2	Tanh	Adam 0,003	50	Evet	MSE = 0,023 MAE = 0,111 RMSE = 0,153 MAPE = 3,345 R² = 0,881
					MaxPooling1D	2						
					Flatten							
					LSTM	70		Tanh				
					Dropout	0,01						
					Dense	1						

Tablo 3.10. (Devamı): Model 3 için CNN+LSTM algoritmasının hiperparametre değerleri ve hata metrik değerleri.

Model3	Eğitim Veri Seti Boyutu	Doğrulama Veri Seti Boyutu	Test Veri Seti Boyutu	Devir Sayısı	Katman	Birim Sayısı/Eşik Değeri/Havuz Boyutu	Çekirdek Boyutu	Aktivasyon Fonksiyonu	Optimizasyon Fonksiyonu/ Öğrenme Aralığı	Parti Boyutu	Rastgele Öğrenme Durumu	Değerlendirme
CL3.13	2185 (%64)	547 (%16)	684 (%20)	70	TD(Conv1D)	32	2	Tanh	Adam 0,003	50	Evet	MSE = 0,018 MAE = 0,088 RMSE = 0,133 MAPE = 2,637 R² = 0,910
					MaxPooling1D	2						
					Flatten							
					LSTM	70		Tanh				
					Dropout	0,01						
					Dense	1						
CL3.14	2185 (%64)	547 (%16)	684 (%20)	70	TD(Conv1D)	64	2	Tanh	Adam 0,003	50	Evet	MSE = 0,021 MAE = 0,104 RMSE = 0,145 MAPE = 3,171 R² = 0,892
					MaxPooling1D	2						
					Flatten							
					LSTM	70		Tanh				
					Dropout	0,01						
					Dense	1						
CL3.15	2185 (%64)	547 (%16)	684 (%20)	70	TD(Conv1D)	64	2	ReLU	Adam 0,003	50	Evet	MSE = 0,169 MAE = 0,362 RMSE = 0,411 MAPE = 11,248 R² = 0,136
					MaxPooling1D	2						
					Flatten							
					LSTM	70		ReLU				
					Dropout	0,01						
					Dense	1						

3.5. Model 4

Model 4'teki veri setinde BİST Kimya Petrol Plastik'in (XKMYA), QE Consumer Goods & Services'in (QECON), MOEX Oil and Gas (MOEXOG) ve RTS Oil & Gas (RTSOG) endeks açılış/kapanış fiyatlarının, Türkiye Petrol Rafinelerileri A.Ş. (TUPRS), Petkim Petrokimya Holding A.Ş (PETKM), Qatar Fuel Co (QFLS) ve Lukoil (LKOIL) hisse senedi açılı ve kapanı fiyatlarının ve dolar (USD/TRY) açılış ve ortalama alı/satı kur değeri girdi, PETKM hisse senedi kapanış fiyatı çıktı olarak kullanılmaktadır.

Model 4'ün RFR algoritmasında, test veri kümesinin boyutu tüm veri setinin %20-%40'ı arasında değerler arasından belirlenmiştir. RandomForestRegressor metodunun içerisinde kullanılan hiperparametrelerden oluşturulacak ağaç sayısı 10-100 ve kullanılacak özellik miktarı 0,6-Otomatik (auto) değerleri arasında değişmektedir.

Tablo 3.11'de Model 4 için RFR algoritmasında kullanılan hiperparametre değerleri, tahmin değerlerinin gözlem değerleriyle olan hata değerleri (MSE, RMSE, MAE) ve tahmin değerlerinin gözlem değerlerine uygunluğu (R^2) verilmiştir.

Model 4 için oluşturulan LSTM algoritması içerisindeki katmanlar sırası ile ilerlediği için Keras kütüphanesindeki models modülünden Sequential (sıralı) ve layers modülünden LSTM, Dropout ve Dense katmanları kullanılmaktadır. Bu model için oluşturulan algorithmada LSTM katmanından birim sayısı 50-100 arasında değişmektedir. LSTM katmanında Tanh, ReLU ve Sigmoid aktivasyon fonksiyonları kullanılmıştır. LSTM katmanına bağlı olan Dropout katmanında eşik değeri 0,02-0,08 değerleri arasında verilmiştir. Dense katmanı tam bağlı çıktı katmanıdır. Algoritmanın tek çıktısı olduğu için bu katmandaki birim sayısı 1'dir. Katman sayısı artırıldığında elde edilen sonuçların hata değerleri arttığı için 2 katmanlı modeller oluşturulmuştur.

Optimizasyon (optimizer) fonksiyonu olarak Adam (Momentum) kullanılmıştır. Optimizasyon fonksiyonunun öğrenme oranı 0,001-0,01 arasında değişmektedir. Kayıp fonksiyonu mean_squared_error olarak belirlenmiştir. Parti Boyutu (batch_size) 50 ve 100 arasında belirlenmiştir. Devir sayısı (Epoch) 50 ve 100 olarak değişmektedir. Shuffle hiperparametresi Evet olarak seçilmiştir.

Tablo 3.12’de Model 4 için LSTM algoritmasında kullanılan hiperarametre değerleri, tahmin değerlerinin gözlem değerleriyle olan hata değerleri (MSE, RMSE, MAE, MAPE) ve tahmin değerlerinin gözlem değerlerine uygunluğu (R^2) verilmiştir.

Oluşturulan CNN+LSTM algoritması içerisindeki katmanlar sırası ile ilerlediği için Keras kütüphanesindeki models modülünden Sequential (sıralı) ve layers modülünden TimeDistributed (TD), Conv1D, MaxPooling1D, Flatten, LSTM, Dropout, Dense katmanları kullanılmaktadır. TimeDistributed katmanı Conv1D katmanını LSTM katmanına bağlamak için kullanılmıştır. Conv1D katmanındaki filtre sayısı (filters) 32, 64 ve 128 olarak belirlenmiştir. Çekirdek boyutu (kernel_size) 2 ve 3 olarak kullanılmıştır. MaxPooling1D katmanında havuz boyutu (pool_size) 2 olarak belirtilmiştir. Flatten katmanı ile veriler düzleştirilmektedir. LSTM katmanından birim sayısı 30 ve 50 olarak değişmektedir. Conv1D, LSTM katmanlarında Tanh, ReLU ve Sigmoid aktivasyon fonksiyonları kullanılmıştır. LSTM katmanına bağlı Dropout katmanındaki eşik değeri 0,01 - 0,05 olarak verilmiştir. Dense katmanı çıktı katmanıdır. Algoritmanın tek çıktısı olduğu için bu katmandaki birim sayısı 1’dir.

Optimizasyon (optimizer) fonksiyonu olarak Adam (Momentum) kullanılmıştır. Optimizasyon fonksiyonunun öğrenme oranı 0,001 ve 0,005 olarak değişmektedir. Kayıp fonksiyonu mean_squared_error olarak belirlenmiştir. Parti Boyutu (batch_size) 100 olarak belirlenmiştir. Devir sayısı (Epoch) 50 ve 70 olarak belirlenmiştir. Rastgele Öğrenme Durumu hiperparametresi shuffle, Evet olarak seçilmiştir.

Tablo 3.13’te Model 4 için CNN+LSTM algoritmasında kullanılan hiperarametre değerleri ve modellerin değerlendirme sonuçları verilmiştir.

Tablo 3.11. Model 4 için RFR algoritmasının hiperparametre değerleri ve hata metrik değerleri.

Model4	Eğitim Veri Seti Boyutu	Test Veri Seti Boyutu	Rastgele Veri Bölme Durumu	Ağaç Sayısı	Maksimum Özellik Sayısı	Değerlendirme				
						MSE	MAE	RMSE	MAPE	R ²
R4.1	2049 (%60)	1367 (%40)	Hayır	10	Auto	1,538	1,076	1,240	30,848	-2,632
R4.2	2049 (%60)	1367 (%40)	Hayır	50	Auto	1,481	1,050	1,217	29,980	-2,498
R4.3	2049 (%60)	1367 (%40)	Hayır	100	Auto	1,479	1,049	1,216	29,978	-2,492
R4.4	2391 (%70)	1025 (%30)	Hayır	10	Auto	0,243	0,335	0,493	8,847	0,172
R4.5	2391 (%70)	1025 (%30)	Hayır	50	Auto	0,266	0,363	0,515	9,687	0,097
R4.6	2391 (%70)	1025 (%30)	Hayır	100	Auto	0,280	0,366	0,529	9,636	0,049
R4.7	2391 (%70)	1025 (%30)	Hayır	50	0,8	0,026	0,365	0,510	9,851	0,115
R4.8	2391 (%70)	1025 (%30)	Hayır	100	0,8	0,265	0,370	0,515	9,946	0,098
R4.9	2391 (%70)	1025 (%30)	Hayır	100	0,6	0,255	0,357	0,505	9,527	0,132
R4.10	2732 (%80)	684 (%20)	Hayır	10	Auto	0,058	0,188	0,242	5,697	0,701
R4.11	2732 (%80)	684 (%20)	Hayır	50	Auto	0,057	0,191	0,238	5,824	0,709
R4.12	2732 (%80)	684 (%20)	Hayır	100	Auto	0,054	0,180	0,233	5,477	0,721
R4.13	2732 (%80)	684 (%20)	Hayır	50	0,8	0,054	0,186	0,233	5,656	0,723
R4.14	2732 (%80)	684 (%20)	Hayır	100	0,8	0,057	0,191	0,240	5,828	0,705
R4.15	2732 (%80)	684 (%20)	Hayır	100	0,6	0,068	0,213	0,260	6,535	0,653

Tablo 3.12. Model 4 için LSTM algoritmasının hiperparametre değerleri ve hata metrik değerleri.

Model4	Eğitim Veri Seti Boyutu	Doğrulama Veri Seti Boyutu	Test Veri Seti Boyutu	Devir Sayısı	Katman	Birim Sayısı/Eşik Değeri	Parti Boyutu	Aktivasyon Fonksiyonu	Rastgele Öğrenme Durumu	Optimizasyon Fonksiyonu/ Öğrenme Aralığı	Değerlendirme
L4.1	1639 (%48)	410 (%12)	1367 (%40)	50	LSTM Dropout Dense	50 0,02 1	50	Tanh	Evet	Adam 0,002	MSE = 0,048 MAE = 0,186 RMSE = 0,220 MAPE = 5,761 R ² = 0,886
L4.2	1639 (%48)	410 (%12)	1367 (%40)	50	LSTM Dropout Dense	50 0,05 1	50	Tanh	Evet	Adam 0,002	MSE = 0,067 MAE = 0,216 RMSE = 0,258 MAPE = 6,435 R ² = 0,842
L4.3	1639 (%48)	410 (%12)	1367 (%40)	50	LSTM Dropout Dense	50 0,08 1	50	Tanh	Evet	Adam 0,002	MSE = 0,026 MAE = 0,123 RMSE = 0,162 MAPE = 3,802 R ² = 0,938
L4.4	1912 (%56)	479 (%14)	1025 (%30)	50	LSTM Dropout Dense	50 0,08 1	50	Tanh	Evet	Adam 0,002	MSE = 0,032 MAE = 0,141 RMSE = 0,179 MAPE = 4,218 R ² = 0,891
L4.5	1912 (%56)	479 (%14)	1025 (%30)	100	LSTM Dropout Dense	50 0,08 1	50	Tanh	Evet	Adam 0,002	MSE = 0,090 MAE = 0,267 RMSE = 0,299 MAPE = 7,658 R ² = 0,696

Tablo 3.12. (Devamı): Model 4 için LSTM algoritmasının hiperparametre değerleri ve hata metrik değerleri.

Model4	Eğitim Veri Seti Boyutu	Doğrulama Veri Seti Boyutu	Test Veri Seti Boyutu	Devir Sayısı	Katman	Birim Sayısı/Eşik Değeri	Parti Boyutu	Aktivasyon Fonksiyonu	Rastgele Öğrenme Durumu	Optimizasyon Fonksiyonu/ Öğrenme Aralığı	Değerlendirme
L4.6	1912 (%56)	479 (%14)	1025 (%30)	50	LSTM Dropout Dense	70 0,08 1	50	Tanh	Evet	Adam 0,002	MSE = 0,025 MAE = 0,116 RMSE = 0,158 MAPE = 3,446 R ² = 0,915
L4.7	1912 (%56)	479 (%14)	1025 (%30)	50	LSTM Dropout Dense	70 0,08 1	50	Sigmoid	Evet	Adam 0,002	MSE = 2,284 MAE = 1,425 RMSE = 1,511 MAPE = 40,515 R ² = -6,762
L4.8	1912 (%56)	479 (%14)	1025 (%30)	50	LSTM Dropout Dense	70 0,08 1	50	ReLU	Evet	Adam 0,002	MSE = 0,073 MAE = 0,238 RMSE = 0,271 MAPE = 7,112 R ² = 0,751
L4.9	1912 (%56)	479 (%14)	1025 (%30)	50	LSTM Dropout Dense	100 0,08 1	50	ReLU	Evet	Adam 0,002	MSE = 0,031 MAE = 0,121 RMSE = 0,176 MAPE = 3,439 R ² = 0,895
L4.10	2185 (%64)	547 (%16)	684 (%20)	50	LSTM Dropout Dense	50 0,08 1	50	Tanh	Evet	Adam 0,002	MSE = 0,038 MAE = 0,146 RMSE = 0,196 MAPE = 4,314 R ² = 0,804

Tablo 3.12. (Devamı): Model 4 için LSTM algoritmasının hiperparametre değerleri ve hata metrik değerleri.

Model4	Eğitim Veri Seti Boyutu	Doğrulama Veri Seti Boyutu	Test Veri Seti Boyutu	Devir Sayısı	Katman	Birim Sayısı/Eşik Değeri	Parti Boyutu	Aktivasyon Fonksyonu	Rastgele Öğrenme Durumu	Optimizasyon Fonksiyonu/ Öğrenme Aralığı	Değerlendirme
L4.11	2185 (%64)	547 (%16)	684 (%20)	50	LSTM Dropout Dense	50 0,08 1	50	Tanh	Evet	Adam 0,001	MSE = 0,123 MAE = 0,305 RMSE = 0,350 MAPE = 9,100 R ² = 0,373
L4.12	2185 (%64)	547 (%16)	684 (%20)	50	LSTM Dropout Dense	50 0,08 1	50	Tanh	Evet	Adam 0,003	MSE = 0,046 MAE = 0,170 RMSE = 0,215 MAPE = 5,077 R ² = 0,763
L4.13	2185 (%64)	547 (%16)	684 (%20)	50	LSTM Dropout Dense	50 0,08 1	50	Tanh	Evet	Adam 0,005	MSE = 0,023 MAE = 0,111 RMSE = 0,150 MAPE = 3,375 R ² = 0,885
L4.14	2185 (%64)	547 (%16)	684 (%20)	50	LSTM Dropout Dense	50 0,08 1	50	Tanh	Evet	Adam 0,01	MSE = 0,019 MAE = 0,094 RMSE = 0,138 MAPE = 2,797 R ² = 0,903
L4.15	2185 (%64)	547 (%16)	684 (%20)	50	LSTM Dropout Dense	50 0,08 1	50	Tanh	Evet	Adam 0,008	MSE = 0,016 MAE = 0,086 RMSE = 0,128 MAPE = 2,596 R ² = 0,916

Tablo 3.13. Model 4 için CNN+LSTM algoritmasının hiperparametre değerleri ve hata metrik değerleri.

Model4	Eğitim Veri Seti Boyutu	Doğrulama Veri Seti Boyutu	Test Veri Seti Boyutu	Devir Sayısı	Katman	Birim Sayısı/Eşik Değeri/Havuz Boyutu	Çekirdek Boyutu	Aktivasyon Fonksiyonu	Aktivasyon Fonksiyonu/ Öğrenme Aralığı	Parti Boyutu	Rastgele Öğrenme Durumu	Değerlendirme
CL4.1	1639 (%48)	410 (%12)	1367 (%40)	50	TD(Conv1D)	32	2	Tanh	Adam 0,002	100	Evet	MSE = 0,359 MAE = 0,496 RMSE = 0,599 MAPE = 14,669 R² = 0,151
					MaxPooling1D	2						
					Flatten							
					LSTM	50	Tanh					
					Dropout	0,01						
					Dense	1						
CL4.2	1639 (%48)	410 (%12)	1367 (%40)	70	TD(Conv1D)	32	2	Tanh	Adam 0,002	100	Evet	MSE = 0,350 MAE = 0,487 RMSE = 0,592 MAPE = 14,312 R² = 0,172
					MaxPooling1D	2						
					Flatten							
					LSTM	50	Tanh					
					Dropout	0,01						
					Dense	1						
CL4.3	1639 (%48)	410 (%12)	1367 (%40)	50	TD(Conv1D)	64	2	Tanh	Adam 0,002	100	Evet	MSE = 0,397 MAE = 0,532 RMSE = 0,630 MAPE = 15,705 R² = 0,062
					MaxPooling1D	2						
					Flatten							
					LSTM	50	Tanh					
					Dropout	0,01						
					Dense	1						

Tablo 3.13. (Devamı): Model 4 için CNN+LSTM algoritmasının hiperparametre değerleri ve hata metrik değerleri.

Model4	Eğitim Veri Seti Boyutu	Doğrulama Veri Seti Boyutu	Test Veri Seti Boyutu	Devir Sayısı	Katman	Birim Sayısı/Eşik Değer/ Havuz Boyutu	Çekirdek Boyutu	Aktivasyon Fonksiyonu	Optimizasyon Fonksiyonu/ Öğrenme Aralığı	Parti Boyutu	Rastgele Öğrenme Durumu	Değerlendirme
CL4.4	1912 (%56)	479 (%14)	1025 (%30)	50	TD(Conv1D)	32	2	Tanh	Adam 0,002	100	Evet	MSE = 0,400 MAE = 0,495 RMSE = 0,633 MAPE = 13,336 R ² = -0,360
					MaxPooling1D	2						
					Flatten	50	Tanh					
					LSTM	50						
					Dropout	0,01						
					Dense	1						
CL4.5	1912 (%56)	479 (%14)	1025 (%30)	50	TD(Conv1D)	32	3	Tanh	Adam 0,002	100	Evet	MSE = 0,305 MAE = 0,372 RMSE = 0,552 MAPE = 9,654 R ² = -0,037
					MaxPooling1D	2						
					Flatten	50	Tanh					
					LSTM	50						
					Dropout	0,01						
					Dense	1						
CL4.6	1912 (%56)	479 (%14)	1025 (%30)	50	TD(Conv1D)	32	2	Tanh	Adam 0,002	100	Evet	MSE = 0,476 MAE = 0,483 RMSE = 0,690 MAPE = 12,605 R ² = -0,618
					MaxPooling1D	3						
					Flatten	50	Tanh					
					LSTM	50						
					Dropout	0,01						
					Dense	1						

Tablo 3.13. (Devamı): Model 4 için CNN+LSTM algoritmasının hiperparametre değerleri ve hata metrik değerleri.

Model4	Eğitim Veri Seti Boyutu	Doğrulama Veri Seti Boyutu	Test Veri Seti Boyutu	Devir Sayısı	Katman	Birim Sayısı/Eşik Değeri/Havuz Boyutu	Çekirdek Boyutu	Aktivasyon Fonksiyonu	Optimizasyon Fonksiyonu/ Öğrenme Aralığı	Parti Boyutu	Rastgele Öğrenme Durumu	Değerlendirme
CL4.7	1912 (%56)	479 (%14)	1025 (%30)	50	TD(Conv1D)	32	2	Tanh	Adam 0,001	100	Evet	MSE = 0,165 MAE = 0,259 RMSE = 0,406 MAPE = 6,731 R ² = 0,440
					MaxPooling1D	2						
					Flatten		Tanh					
					LSTM	50						
					Dropout	0,01						
					Dense	1						
CL4.8	1912 (%56)	479 (%14)	1025 (%30)	50	TD(Conv1D)	32	2	Tanh	Adam 0,001	100	Evet	MSE = 0,225 MAE = 0,307 RMSE = 0,475 MAPE = 7,979 R ² = 0,234
					MaxPooling1D	2						
					Flatten		Tanh					
					LSTM	30						
					Dropout	0,01						
					Dense	1						
CL4.9	1912 (%56)	479 (%14)	1025 (%30)	50	TD(Conv1D)	128	2	Tanh	Adam 0,001	100	Evet	MSE = 0,269 MAE = 0,362 RMSE = 0,519 MAPE = 9,497 R ² = 0,084
					MaxPooling1D	2						
					Flatten		Tanh					
					LSTM	30						
					Dropout	0,01						
					Dense	1						

Tablo 3.13. (Devamı): Model 4 için CNN+LSTM algoritmasının hiperparametre değerleri ve hata metrik değerleri.

Model4	Eğitim Veri Seti Boyutu	Doğrulama Veri Seti Boyutu	Test Veri Seti Boyutu	Devir Sayısı	Katman	Birim Sayısı/Eşik Değeri/ Havuz Boyutu	Çekirdek Boyutu	Aktivasyon Fonksiyonu	Optimizasyon Fonksiyonu/ Öğrenme Aralığı	Parti Boyutu	Rastgele Öğrenme Durumu	Değerlendirme
CL4.10	2185 (%64)	547 (%16)	684 (%20)	50	TD(Conv1D)	32	2	Tanh	Adam 0,002	100	Evet	MSE = 0,046
					MaxPooling1D	2						MAE = 0,168
					Flatten							RMSE = 0,215
					LSTM	50		Tanh				MAPE = 5,021
					Dropout	0,01						R ² = 0,763
CL4.11	2185 (%64)	547 (%16)	684 (%20)	50	Dense	1			Adam 0,001	100	Evet	
					TD(Conv1D)	128	2	Tanh				MSE = 0,149
					MaxPooling1D	2						MAE = 0,352
					Flatten							RMSE = 0,386
					LSTM	30		Tanh				MAPE = 10,671
CL4.12	2185 (%64)	547 (%16)	684 (%20)	70	Dropout	0,01			Adam 0,001	100	Evet	R ² = 0,237
					Dense	1						
					TD(Conv1D)	128	2	Tanh				MSE = 0,101
					MaxPooling1D	2						MAE = 0,269
					Flatten							RMSE = 0,318
					LSTM	30		Tanh				MAPE = 8,064
					Dropout	0,01						R ² = 8,064
					Dense	1						
					TD(Conv1D)	128	2	Tanh				
					MaxPooling1D	2						

Tablo 3.13. (Devamı): Model 4 için CNN+LSTM algoritmasının hiperparametre değerleri ve data metrik değerleri.

Model4	Eğitim Veri Seti Boyutu	Doğrulama Veri Seti Boyutu	Test Veri Seti Boyutu	Devir Sayısı	Katman	Birim Sayısı/Eşik Değeri/Havuz Boyutu	Çekirdek Boyutu	Aktivasyon Fonksiyonu	Optimizasyon Fonksiyonu/ Öğrenme Aralığı	Parti Boyutu	Rastgele Öğrenme Durumu	Değerlendirme
CL4.13	2185 (%64)	547 (%16)	684 (%20)	70	TD(Conv1D)	32	2	Tanh	Adam 0,005	100	Evet	MSE = 0,020 MAE = 0,098 RMSE = 0,143 MAPE = 2,894 R² = 0,896
					MaxPooling1D	2						
					Flatten		Tanh					
					LSTM	50						
					Dropout	0,05						
					Dense	1						
CL4.14	2185 (%64)	547 (%16)	684 (%20)	70	TD(Conv1D)	32	2	ReLU	Adam 0,005	100	Evet	MSE = 0,031 MAE = 0,130 RMSE = 0,177 MAPE = 3,961 R² = 0,840
					MaxPooling1D	2						
					Flatten		ReLU					
					LSTM	50						
					Dropout	0,05						
					Dense	1						
CL4.15	2185 (%64)	547 (%16)	684 (%20)	70	TD(Conv1D)	32	2	Sigmoid	Adam 0,005	100	Evet	MSE = 0,309 MAE = 0,395 RMSE = 0,556 MAPE = 10,938 R² = -0,578
					MaxPooling1D	2						
					Flatten		Sigmoid					
					LSTM	50						
					Dropout	0,05						
					Dense	1						

4. BULGULAR VE TARTIŞMA

Bu bölümde uygulama bölümünde verilen örnek test değerleri karşılaştırılmış ve tahmin sonuçlarına etkisi tartışılmıştır.

Uygulamada kullanılan verisetindeki özellikler farklı ülkelere ait olduğundan hisse senedi, endeks ve fiyat değerlerinin eksik olduğu tarihler bulunmaktadır. Bu problemi gidermek için, eksik veri bulunan tarihin veri setinden çıkartılması, eksik verilerin yerine 0 yazılması, eksik verinin ortalama değer ile doldurulması ve eksik verinin doğrusal enterpolasyon ile bulunması yöntemleri Model1 veri setine ayrı ayrı uygulanmıştır. En kötü sonuçlar eksik değerlerin yerine 0 yazıldığında alınmıştır. Eksik verinin bulunduğu tarihlerin veri setinden çıkarılması veri kayıplarına neden olduğundan ve diğer yöntemlerden daha kötü sonuç verdiğinden tercih edilmemiştir. Eksik verinin doldurulmasında kullanılan diğer iki yöntem olan ortalama ve enterpolasyon yöntemleri ile elde edilen tahmin sonuçlarına bakıldığında, genel olarak en iyi sonuçların eksik verinin doğrusal enterpolasyon yöntemi ile doldurulduğu durumda olduğundan çalışmada eksik verinin doldurulması için doğrusal enterpolasyon yöntemi kullanılmıştır.

Algoritmaların hiperparametrelerinin karşılaştırılması, bir hiperparametrenin değişken olarak kullanılıp diğer hiperparametrelerin sabit tutulduğu örnek test verileriyle yapılmıştır.

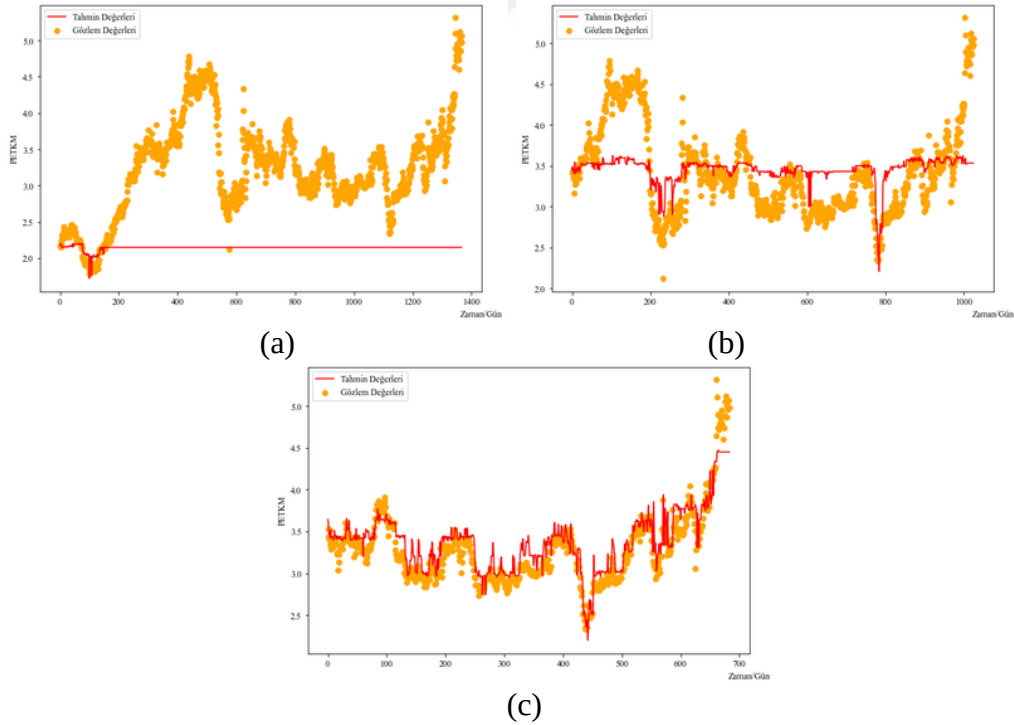
Tüm modellerde RFR algoritmasının eğitim ve test veri seti boyutu hiperparametreleri kontrol edildiğinde tahmin sonucuna büyük oranda etki ettiği görülmektedir. Tablo 4.1’de modellere uygulanan RFR algoritmasının eğitim ve test veri setleri boyutunun tahmin sonuçlarına etkisi gösterilmiştir. Tablodaki metrik hata değerlerine ve R^2 değerine bakıldığında eğitim veri setindeki veriler artırıldığında tahmin başarısı da arttığı görülmektedir.

Aynı zamanda bu oluşturulan 4 farklı veri seti açısından ayrı ayrı kontrol edildiğinde sırasıyla en iyi sonucu veren modellerin Model1, Model3, Model2 ve Model4 olduğu görülmüştür. Bu sonuçlar arasındaki fark RFR algoritmasının verilerden elde edilen

bilgileri ayırt etmeksizin kullanmasından kaynaklanmaktadır. Algoritma testlerinde R^2 değerinin negatif çıkması veri seti değişkenlerinin ve/veya oluşturulan algoritma hiperparametrelerinin PETKM hisse senedi kapanış fiyatı tahminlerine uyum sağlamadığının bir göstergesidir. Şekil 4.1’de Model 1’in test örneklerinin grafikleri verilmiştir.

Tablo 4.1. Tüm modeller için RFR algoritmasındaki eğitim ve test veri seti boyutları hiperparametrelerinin tahmin sonucuna etkisi.

Model1	Eğitim Veri Seti Boyutu	Test Veri Seti Boyutu	Değerlendirme				
			MSE	RMSE	MAE	MAPE	R^2
R1.2	2049 (%60)	1367 (%40)	1,573	1,254	1,088	31,182	-2,714
R1.5	2391 (%70)	1025 (%30)	0,225	0,475	0,368	10,621	0,233
R1.11	2732 (%80)	684 (%20)	0,042	0,206	0,162	4,934	0,782
R2.2	2049 (%60)	1367 (%40)	1,541	1,076	1,241	30,798	-2,640
R2.5	2391 (%70)	1025 (%30)	0,253	0,358	0,503	9,625	0,141
R2.11	2732 (%80)	684 (%20)	0,046	0,167	0,215	5,083	0,763
R3.1	2049 (%60)	1367 (%40)	1,635	1,111	1,279	31,833	-2,861
R3.4	2391 (%70)	1025 (%30)	0,233	0,365	0,483	10,499	0,207
R3.10	2732 (%80)	684 (%20)	0,043	0,156	0,207	4,721	0,781
R4.3	2049 (%60)	1367 (%40)	1,479	1,049	1,216	29,978	-2,492
R4.6	2391 (%70)	1025 (%30)	0,280	0,366	0,529	9,636	0,049
R4.12	2732 (%80)	684 (%20)	0,054	0,180	0,233	5,477	0,721



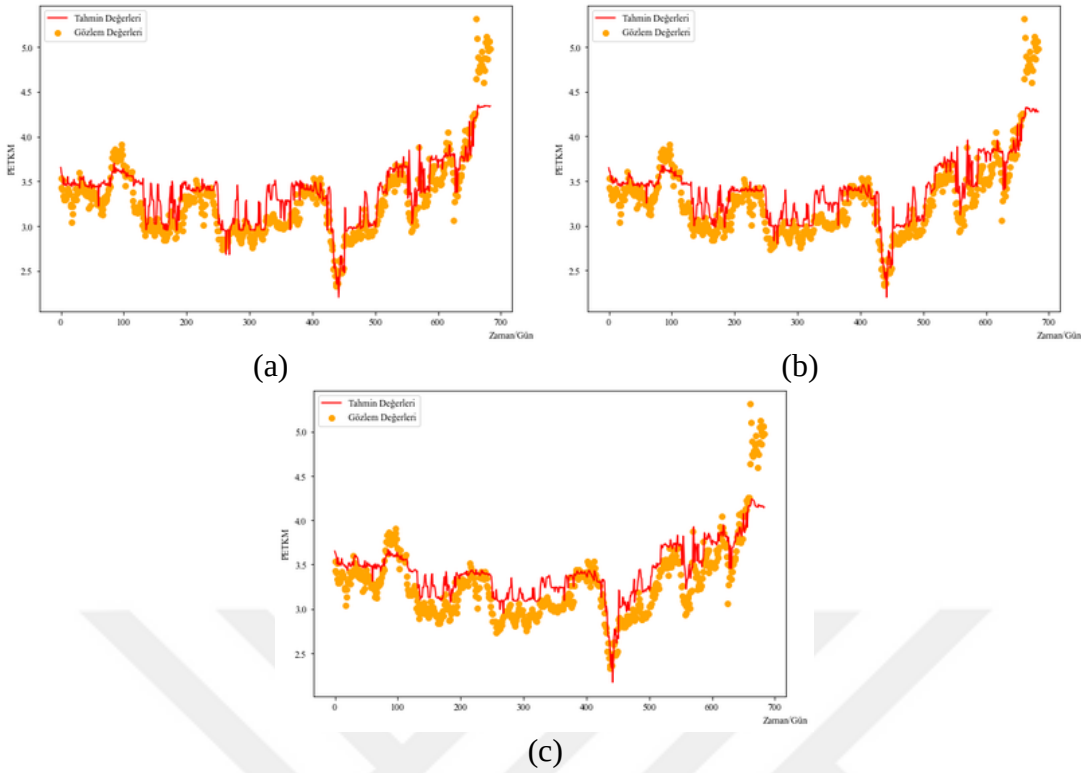
Şekil 4.1. (a) R1.2 (b) R1.5 (c) R.1.11 örneklerinin tahmin grafikleri.

RFR algoritmasında kullanılan özellik sayısı arttığında genellikle algoritmanın başarısı da artmıştır. Ancak bu değişimin eğitim ve test veri kümelerinin sağladığı değişim kadar büyük bir fark yaratmadığı görülmektedir. Maksimum özellik sayısı hiperparametresinin değeri “Otomatik” ve “0,8” olduğunda yakın değerler elde edilirken, bu parameter “0,6” olduğunda tüm denemelerde daha düşük sonuçlar vermiştir. Tablo 4.2’de RFR algoritmasının uygulandığı tüm modellerden maksimum özellik sayısı hiperparametresinin algoritma değerlendirme sonuçlarına etkisi verilmiştir.

Tablo 4.2. Tüm modeller için RFR algoritmasındaki maksimum özellik sayısı hiperparametresinin tahmin sonucuna etkisi.

Model1	Maksimum Özellik Sayısı	Değerlendirme				
		MSE	RMSE	MAE	MAPE	R ²
R1.12	Auto	0,048	0,220	0,174	5,339	0,752
R1.14	0,8	0,050	0,225	0,185	5,717	0,741
R1.15	0,6	0,069	0,263	0,225	6,954	0,644
R2.12	Auto	0,049	0,175	0,222	5,309	0,748
R2.14	0,8	0,052	0,185	0,228	5,635	0,733
R2.15	0,6	0,078	0,232	0,280	7,132	0,598
R3.12	Auto	0,047	0,167	0,218	5,105	0,757
R3.14	0,8	0,054	0,189	0,232	5,798	0,723
R3.15	0,6	0,067	0,205	0,259	6,261	0,657
R4.12	Auto	0,054	0,180	0,233	5,477	0,721
R4.14	0,8	0,057	0,191	0,240	5,828	0,705
R4.15	0,6	0,068	0,213	0,260	6,535	0,653

Maksimum özellik sayıları “Otomatik” olarak belirlendiğinde algoritma kullanılacak verileri seçmektedir. “0,8” veya “0,6” olarak belirlendiğinde veri setindeki özelliklerin 0,8 veya 0,6 katı kullanılmaktadır. Tablo 4.2’deki test verileri incelendiğinde Model1’in değerlendirme sonuçları arasındaki farkın diğer modellerden daha az olduğu göstermektedir. Bu hiperparametre 0,8 veya 0,6 olarak belirlendiğinde verilerin bir kısmının kullanılmaması modelin başarısını artıran özelliklerin kullanılmamasına neden olabilmektedir. Şekil 4.2’de Model4’ün Tablo 4.2’deki değerlerinin grafikleri verilmiştir.

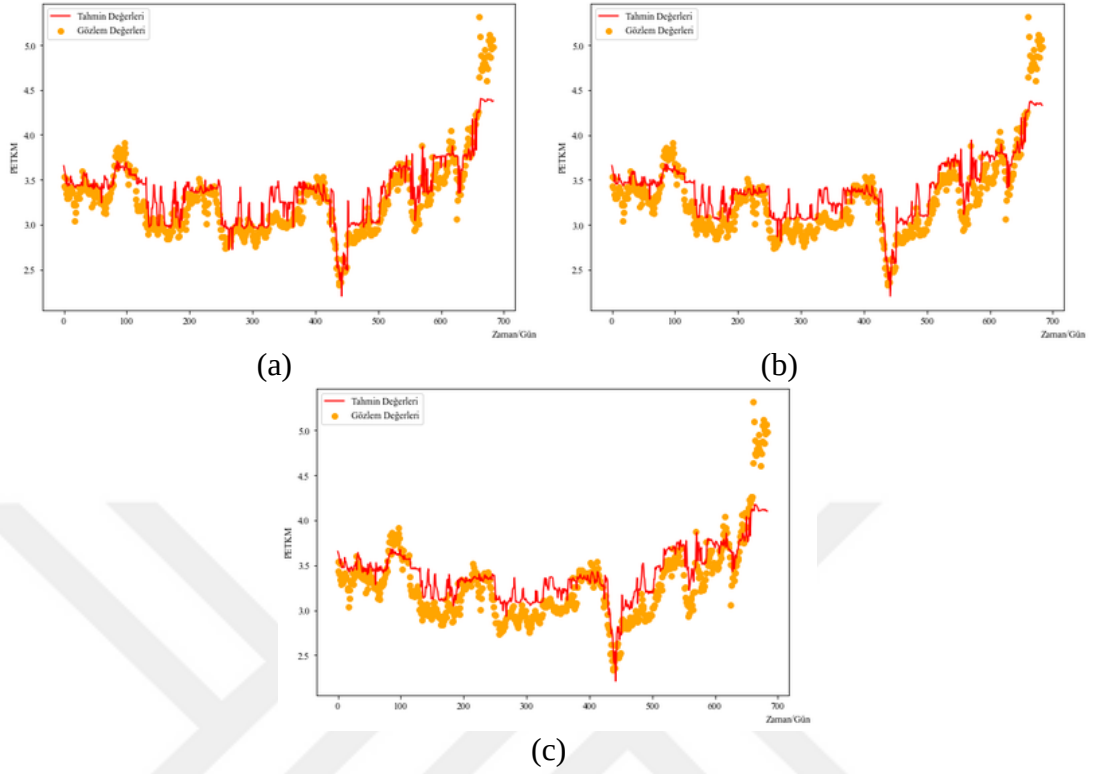


Şekil 4.2. (a) R4.12, (b)R4.14 (c) R4.15 örneklerinin tahmin grafikleri.

Modellerde Ağaç Sayısı hiperparametresi 10, 50 ve 100 olarak verilmiştir. 4 model için de kontrol edildiğinde genellikle 50 değerinin daha iyi sonuç verdiği görülmektedir. Ancak bu hiperparametrenin de tahmin sonuçları üzerinde bu veriler için çok büyük bir etkisinin olmadığı görülmektedir. Tablo 4.3'te Ağaç Sayısı hiperparametresinin tahmin sonuçlarına etkisi verilmiştir. Şekil 4.3'te Tablo 4.3'teki Model 3 test örneklerinin grafikleri verilmiştir.

Tablo 4.3. Tüm modeller için RFR algoritmasındaki ağaç sayısı hiperparametresinin tahmin sonucuna etkisi.

Model1	Maksimum Özellik Sayısı	Değerlendirme				
		MSE	RMSE	MAE	MAPE	R ²
R1.12	Auto	0,048	0,220	0,174	5,339	0,752
R1.14	0,8	0,050	0,225	0,185	5,717	0,741
R1.15	0,6	0,069	0,263	0,225	6,954	0,644
R2.12	Auto	0,049	0,175	0,222	5,309	0,748
R2.14	0,8	0,052	0,185	0,228	5,635	0,733
R2.15	0,6	0,078	0,232	0,280	7,132	0,598
R3.12	Auto	0,047	0,167	0,218	5,105	0,757
R3.14	0,8	0,054	0,189	0,232	5,798	0,723
R3.15	0,6	0,067	0,205	0,259	6,261	0,657
R4.12	Auto	0,054	0,180	0,233	5,477	0,721
R4.14	0,8	0,057	0,191	0,240	5,828	0,705
R4.15	0,6	0,068	0,213	0,260	6,535	0,653



Şekil 4.3. (a)R3.12 (a) R3.14 (c) R3,15 örneklerinin tahmin grafikleri.

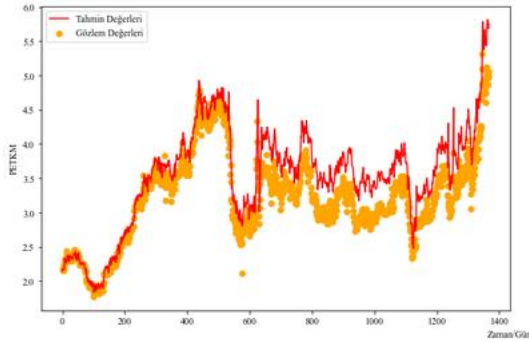
RFR algoritması için tüm bu hiperparametre sonuçları karşılaştırıldığında en büyük etkinin eğitim ve test kümelerinin boyutu olduğu görülmektedir.

Eğitim-doğrulama-test veri seti boyutu LSTM algoritmasında kullanılan önemli hiperparametrelerdendir. Çalışmalarda bu hiperparametrenin eğitim-test ve eğitim doğrulama-test olmak üzere 2 türü bulunmaktadır. Doğrulama kümesi bulunmayan durumda test verileri aynı zamanda doğrulama veri seti olarak kullanılmaktadır. Çalışmadaki verilerde ayrı bir doğrulama kümesinin kullanılması tahmin sonuçlarına olumlu etki ettiği için çalışmada veri eğitim-doğrulama ve test olmak üzere 3'e bölünmüştür. Tablo 4.4'te bu hiperparametrenin değerlendirme sonuçları için örnekler verilmiştir. Verilen örneklerde genellikle eğitim veri seti boyutu arttığında tahmin sonuçlarının da başarısının arttığı gözlemlenmiştir.

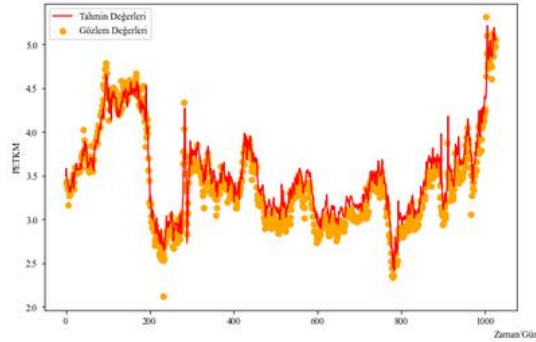
Tablo 4.4. Tüm modeller için LSTM algoritmasındaki eğitim, doğrulama ve test veri seti boyutu hiperparametrelerinin tahmin sonucuna etkisi.

Model	Eğitim Veri Seti Boyutu	Doğrulama Veri Seti Boyutu	Test Veri seti Boyutu	MSE	MAE	RMSE	MAPE	R ²
L1.3	1639 (%48)	410 (%12)	1367 (%40)	0,140	0,313	0,374	9,625	0,670
L1.4	1912 (%56)	479 (%14)	1025 (%30)	0,032	0,145	0,180	4,416	0,890
L2.1	1639 (%48)	410 (%12)	1367 (%40)	0,368	0,497	0,607	14,260	0,130
L2.4	1912 (%56)	479 (%14)	1025 (%30)	0,117	0,239	0,342	6,393	0,603
L2.10	2185 (%64)	547 (%16)	684 (%20)	0,027	0,127	0,164	3,915	0,863
L3.1	1639 (%48)	410 (%12)	1367 (%40)	0,221	0,412	0,470	12,320	0,479
L3.4	1912 (%56)	479 (%14)	1025 (%30)	0,044	0,168	0,209	4,965	0,851
L3.10	2185 (%64)	547 (%16)	684 (%20)	0,027	0,126	0,165	3,812	0,861
L4.3	1639 (%48)	410 (%12)	1367 (%40)	0,026	0,123	0,162	3,802	0,938
L4.4	1912 (%56)	479 (%14)	1025 (%30)	0,032	0,141	0,179	4,218	0,891
L4.10	2185 (%64)	547 (%16)	684 (%20)	0,038	0,146	0,196	4,314	0,804

Tablo 4.4'teki Model 3 ve Model 4'e ait test örneklerinin tahmin değerleri arasındaki fark az iken Model 1 ve Model 2'deki test örneklerinde eğitim-doğrulama ve test veri seti boyutu hiperparametresinin sonuçlara etkisinin büyük olduğu görülmektedir. Şekil 4.4'te Tablo 4.4'te bulunan Model 1'in örnek test değerlerinin tahmin grafiği ve Şekil 4.5'te eğitim ve doğrulama kümelerinin kayıp fonksiyonu grafikleri verilmiştir. Kayıp fonksiyon grafikleri eğitim ve doğrulama veri kümelerinin öğrenme değerlerinin verildiği grafiklerdir. Eğitim ve doğrulama hata değerlerinin benzer bir şekilde değişime uğraması oluşturulan modelin veri seti ile uyumlu olduğunu göstermektedir.

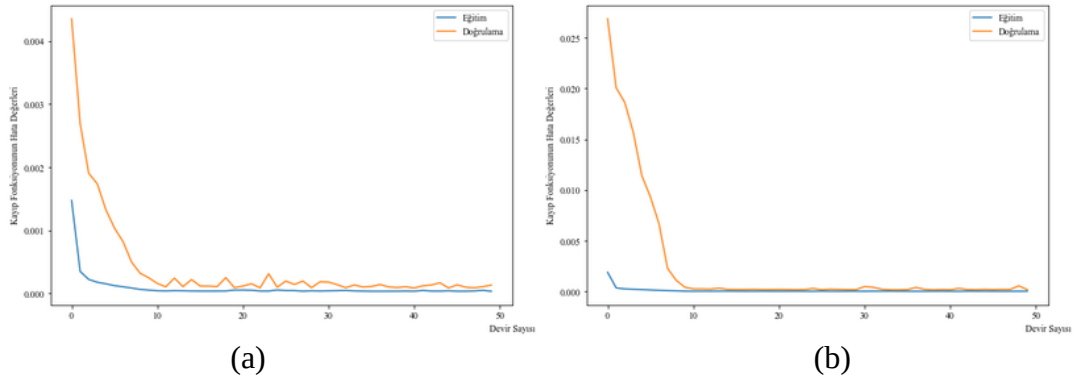


(a)



(b)

Şekil 4.4. (a) L1.3 (b) L1.4 örneklerinin tahmin grafikleri.

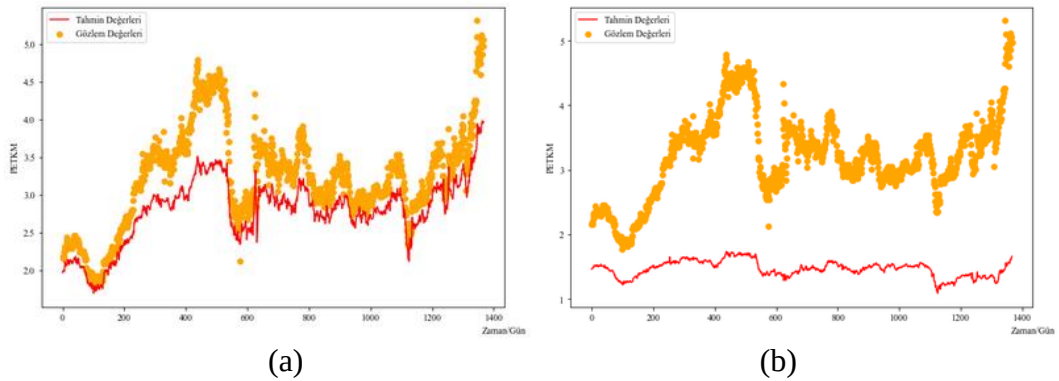


Şekil 4.5. (a) L1.3 (b) L1.4 örneklerinin kayıp fonksiyon grafikleri.

LSTM algoritmasında Model 2 üzerinde algoritmanın rastgele öğrenme durumu (shuffle) hiperparametresi kontrol edilmiştir. Testler sonucunda bu hiperparametrenin modelin başarısında önemli bir rolü olduğu görülmektedir. Rastgele öğrenme durumu hiperparametresi “Evet” olarak seçilen testlerde “Hayır” seçilenlere oranla çok daha başarılı sonuçlar vermiştir. Bu nedenle diğer modellerin testlerinde bu hiperparametre “Evet” olarak seçilmiştir. Tablo 4.5’te Rastgele öğrenme durumu hiperparametresinin tahmin sonuçlarına etkisi verilmiştir. Şekil 4.6’da Tablo 4.5’te verilen değerlerin tahmin grafikleri verilmiştir.

Tablo 4.5. Model2 için LSTM algoritmasındaki rastgele öğrenme durumu hiperparametresinin tahmin sonucuna etkisi.

Model	Rastgele Öğrenme Durumu	MSE	MAE	RMSE	MAPE	R ²
L2.2	Evet	0,263	0,436	0,513	12,625	0,378
L2.3	Hayır	3,435	1,757	1,853	53,001	-7,111



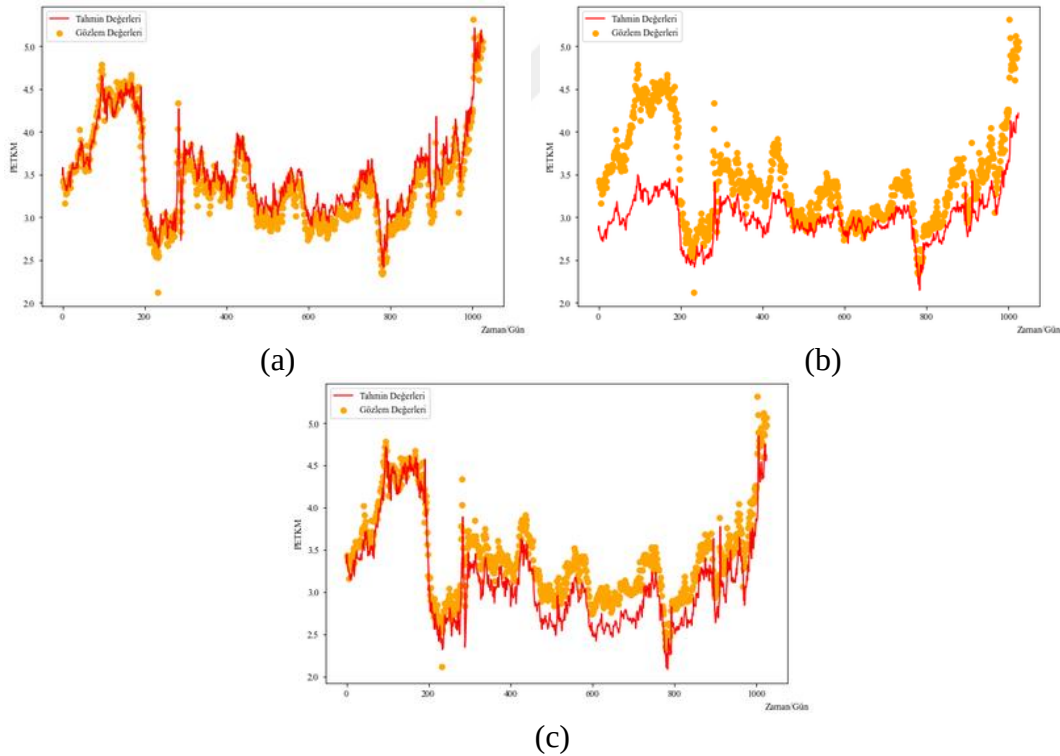
Şekil 4.6. (a) L2.2 (b) L2.3 örneklerinin tahmin grafikleri.

Tüm modeller için LSTM algoritmasının aktivasyon fonksiyonu hiperparametresi kontrol edilmiştir. Aktivasyon fonksiyonu olarak Tanh, ReLU ve Sigmoid

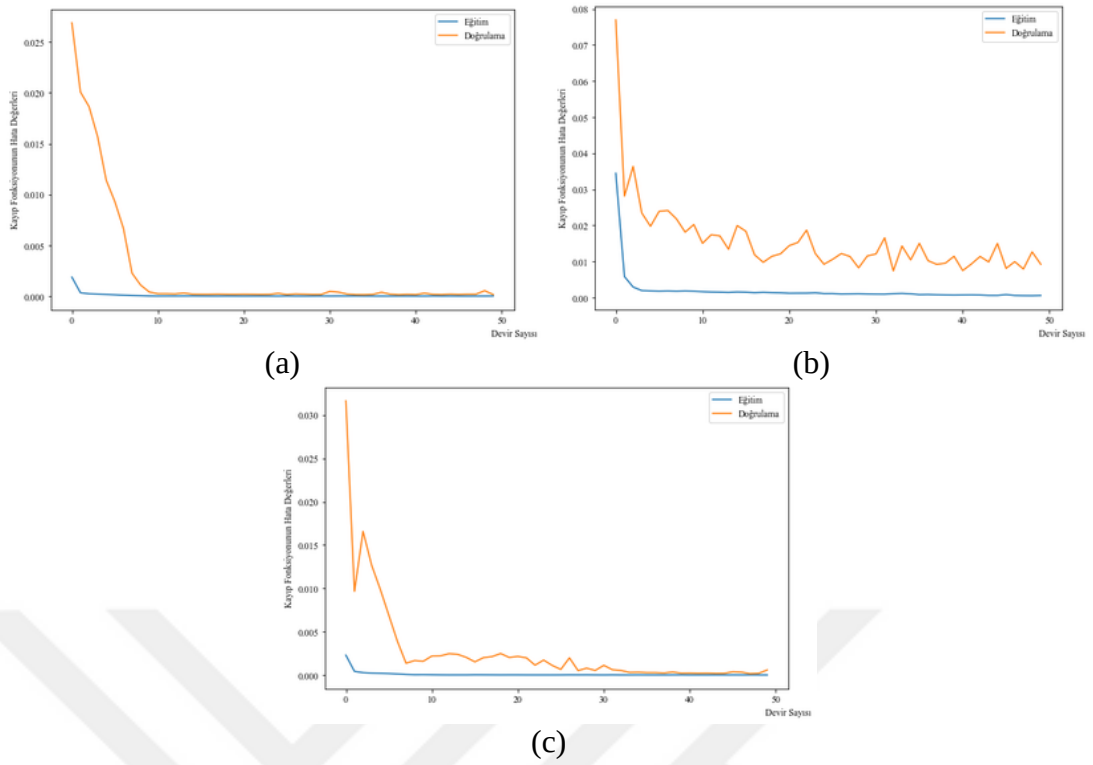
fonksiyonları kullanılmıştır. Model 3 dışındaki tüm modellerde en iyi sonucu Tanh fonksiyonunun kullanıldığı testler vermiştir. Model 3'te en iyi sonuç ReLU fonksiyonu ile alınmıştır. Tüm modellerde en kötü sonucu ise Sigmoid fonksiyonun kullanıldığı testler vermiştir. Tablo 4.6'da Aktivasyon Fonksiyonlarının tahmin sonuçlarına etkisi paylaşılmıştır. Model 1 veri seti için yapılan aktivasyon fonksiyonu testlerinin tahmin grafikleri Şekil 4.7'de ve kayıp fonksiyonları Şekil 4.8'de verilmiştir.

Tablo 4.6. Modeller için LSTM algoritmasındaki aktivasyon fonksiyonu hiperparametresinin tahmin sonucuna etkisi.

Model	Aktivasyon Fonksiyonu	MSE	MAE	RMSE	MAPE	R ²
L1.4	Tanh	0,032	0,145	0,180	4,416	0,890
L1.5	Sigmoid	0,286	0,423	0,535	11,418	0,028
L1.6	ReLU	0,101	0,282	0,318	8,458	0,656
L2.13	Tanh	0,076	0,235	0,276	7,038	0,610
L2.14	Sigmoid	0,116	0,306	0,340	9,637	0,409
L2.15	ReLU	0,469	0,605	0,685	17,955	-1,396
L3.5	Tanh	0,434	0,615	0,659	17,504	-0,475
L3.6	Sigmoid	0,839	0,783	0,916	21,567	-1,850
L3.7	ReLU	0,143	0,326	0,378	9,559	0,513
L4.6	Tanh	0,025	0,116	0,158	3,446	0,915
L4.7	Sigmoid	2,284	1,425	1,511	-40,515	-6,762
L4.8	ReLU	0,073	0,238	0,271	7,112	0,751



Şekil 4.7. (a) L1.4 (b) L1.5 (c) L1.6 örneklerinin tahmin grafikleri.



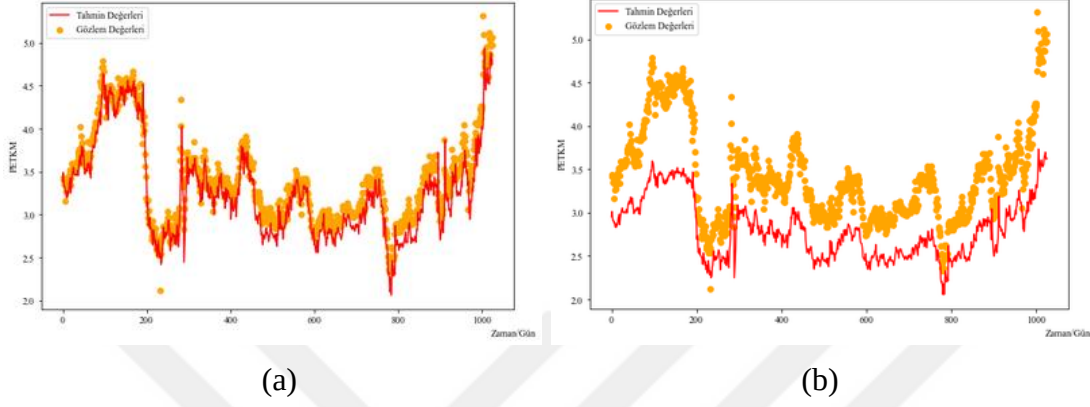
Şekil 4.8. (a) L1.4 (b) L1.5 (c) L1.6 örneklerinin kayıp fonksiyon grafikleri.

LSTM algoritmasının öğrenme hızını etkileyen öğrenme aralığı hiperparametresine yüksek bir değer verildiğinde algoritma gizli bilgileri yeterli düzeyde öğrenememektedir, çok düşük bir değer verildiğinde ise öğrenme süresi çok uzamaktadır. Tablo 4.7’de tüm modeller için yapın testlerden öğrenme aralığı hiperparametresinin değerlendirme sonuçları verilmiştir.

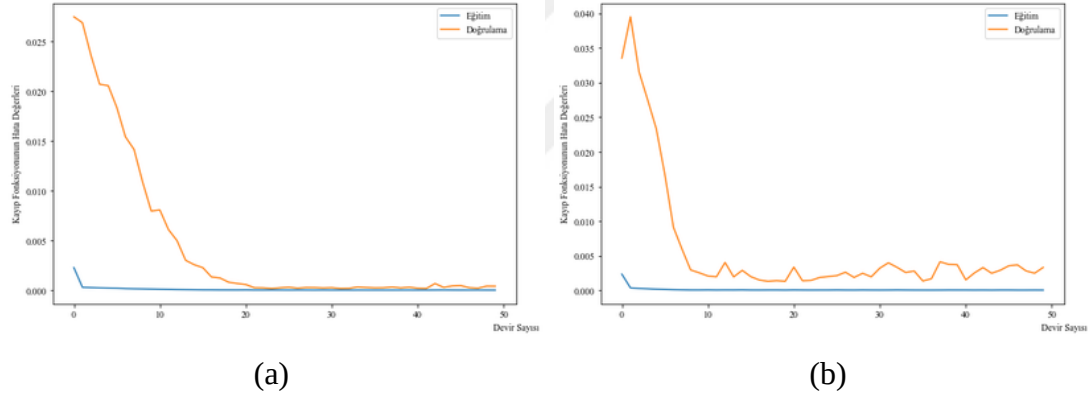
Tablo 4.7. Tüm modeller için LSTM algoritmasındaki öğrenme aralığı hiperparametresinin tahmin sonucuna etkisi.

Model	Öğrenme Aralığı	MSE	MAE	RMSE	MAPE	R ²
L1.4	0,01	0,032	0,145	0,180	4,416	0,890
L1.7	0,03	0,218	0,382	0,467	10,720	0,260
L2.4	0,01	0,117	0,239	0,342	6,393	0,603
L2.5	0,001	0,020	0,099	0,142	2,893	0,931
L3.4	0,002	0,044	0,168	0,209	4,965	0,851
L3.5	0,02	0,434	0,615	0,659	17,504	-0,475
L4.11	0,001	0,123	0,305	0,350	9,100	0,373
L4.12	0,003	0,046	0,170	0,215	5,077	0,763
L4.13	0,005	0,023	0,111	0,150	3,375	0,885
L4.14	0,01	0,019	0,094	0,138	2,797	0,903
L4.15	0,008	0,016	0,086	0,128	2,596	0,916

Tablo 4.7’deki değerlendirme sonuçlarına bakıldığında Model 1, Model 2 ve Model 3’te öğrenme aralığı değeri azaldıktan tahmin başarısı artmakta, Model 4’te ise öğrenme aralığı değeri arttıkça tahmin başarısı artmaktadır. Tablo 4.7’deki Model 3 için yapılan testlerin tahmin grafikleri Şekil 9’da kayıp fonksiyon grafikleri Şekil 10’da verilmiştir.



Şekil 4.9. (a) L3.4 (b) L3.5 örneklerinin tahmin grafikleri.

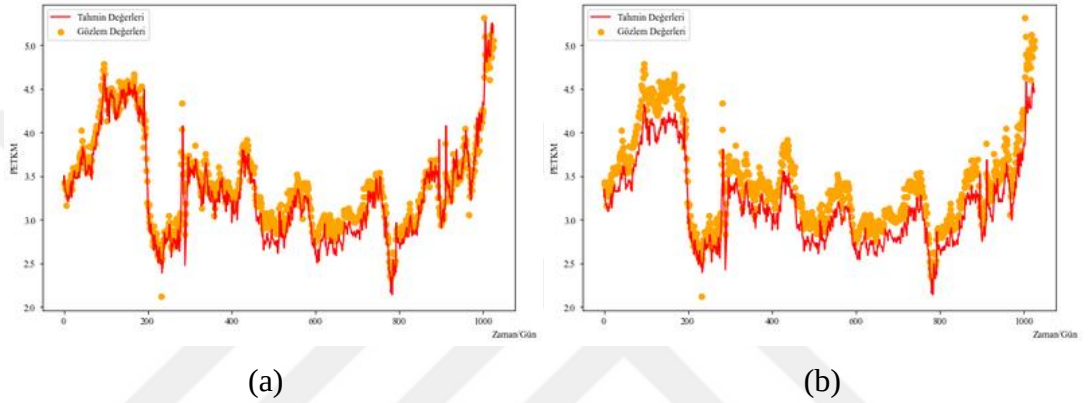


Şekil 4.10. (a) L3.4 (b) L3.5 örneklerinin kayıp fonksiyon grafikleri.

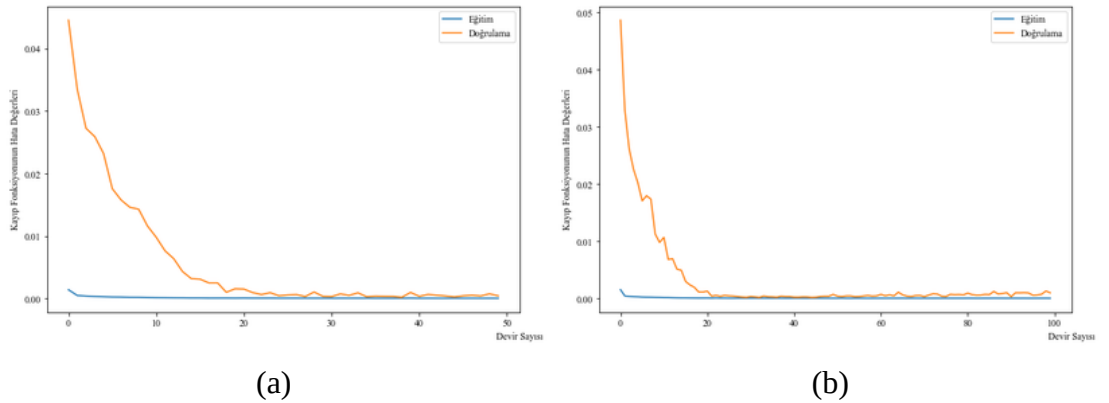
Devir sayısı hiperparametresi verilerin ağa öğretilmesinde önemli bir hiperparametredir. Tablo 4.8’da devir sayısı hiperparametresinin tahmin sonuçlarına etkisi gösterilmiştir. Genellikle devir sayısı arttığında tahmin sonuçlarının başarısının da arttığı görülmektedir. Ancak algoritmaya etki eden diğer önemli hiperparametrelerin ve veri seti farklılığının etkisiyle devir sayısı azaldığında da daha iyi sonuçlar elde edilebilmekte veya tahmin sonuçlarında önemli bir değişikliğe neden olmadığı durumlar da oluşmaktadır. Şekil 4.11’de Tablo 4.8’deki Model 4’ün test örneklerinin tahmin grafikleri, Şekil 4.12’ kayıp fonksiyon grafikleri verilmiştir.

Tablo 4.8. Modeller İçin LSTM algoritmasındaki devir sayısı hiperparametresinin tahmin sonucuna etkisi.

Model	Devir Sayısı	MSE	MAE	RMSE	MAPE	R ²
L1.10	50	0,034	0,147	0,184	4,447	0,827
L1.11	70	0,023	0,102	0,150	3,065	0,885
L1.12	100	0,019	0,101	0,139	3,057	0,902
L3.1	50	0,221	0,412	0,470	12,320	0,479
L3.2	70	0,243	0,436	0,493	13,148	0,425
L3.3	100	0,097	0,252	0,2311	7,794	0,772
L3.13	50	0,041	0,167	0,203	4,873	0,859
L3.14	70	0,072	0,229	0,268	6,459	0,757
L3.15	100	0,068	0,220	0,260	6,234	0,770
L4.4	50	0,032	0,141	0,179	4,218	0,891
L4.5	100	0,090	0,267	0,299	7,658	0,696



Şekil 4.11. (a) L4.4 (b) L4.5 örneklerinin kayıp fonksiyon grafikleri.



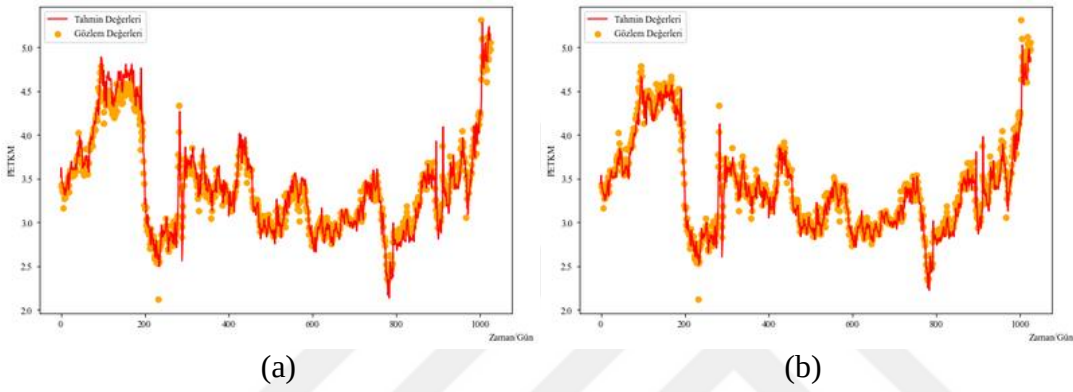
Şekil 4.12. (a) L4.4 (b) L4.5 örneklerinin kayıp fonksiyon grafikleri.

LSTM algoritmalarında bulunan LSTM katmanında bulunan birimler veriler içerisindeki gizli bilgileri öğrenen bileşenlerdir. Bu nedenle önemli hiperparametrelerdendir. Tablo 4.9’da LSTM algoritmalarındaki birim sayısının tahmin sonuçlarına etkisi verilmiştir. Tablodaki değerlere bakıldığında birim sayısı azaldıkça tahmin sonuçlarının başarı oranı artmaktadır. Tablo 4.9’daki L3.11 ve

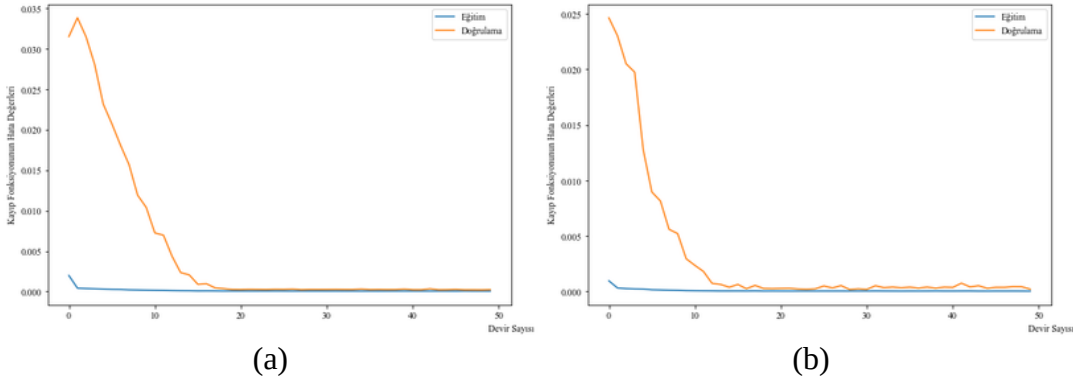
L3.12'nin tahmin grafikleri Şekil 4.13'te, kayıp fonksiyon grafikleri Şekil 4.14'te verilmektedir.

Tablo 4.9. Modeller İçin LSTM algoritmasındaki birim sayısı hiperparametresinin tahmin sonucuna etkisi.

Model	Birim Sayısı	MSE	MAE	RMSE	MAPE	R ²
L1.1	50	0,096	0,255	0,310	7,777	0,774
L1.2	100	0,082	0,234	0,287	7,167	0,805
L3.11	70	0,027	0,118	0,163	3,419	0,909
L3.12	100	0,021	0,101	0,146	2,926	0,928
L4.8	70	0,073	0,238	0,271	7,112	0,751
L4.9	100	0,031	0,121	0,176	3,439	0,895



Şekil 4.13. (a) L3.11 (b) L3.12 örneklerinin tahmin grafikleri.



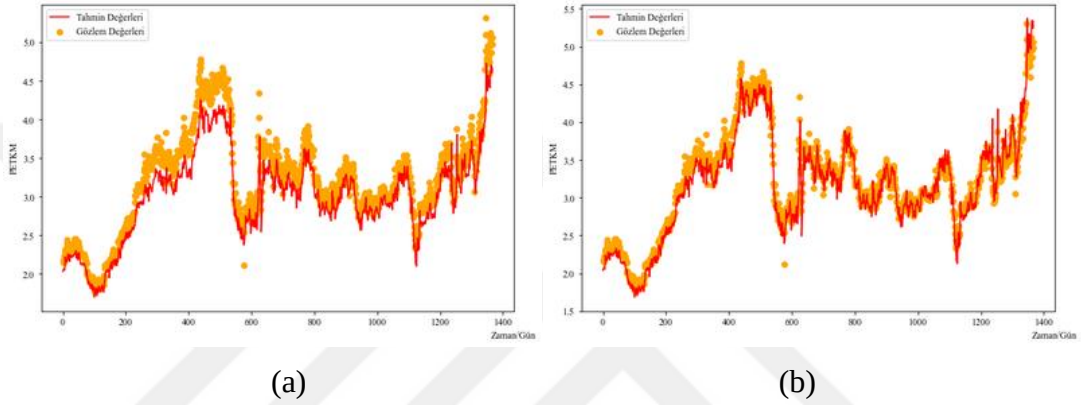
Şekil 4.14. (a) L3.11, (b) L3.12 örneklerinin kayıp fonksiyon grafikleri.

LSTM algoritmalarında Dropout katmanında bulunan eşik değerler, oluşturulan algoritmanın aşırı öğrenme probleminin önüne geçilmek için kullanılmaktadır. Oluşturulan algoritmanın verileri ezberlemesini engellemekte ve tahmin edebilmesini sağlamaktadır. Tablo 4.10'da Model 3 ve Model 4'ten örnekler verilmiştir. Verilen test örneklerinde eşik değeri arttıkça algoritmaların başarısının da arttığı görülmektedir.

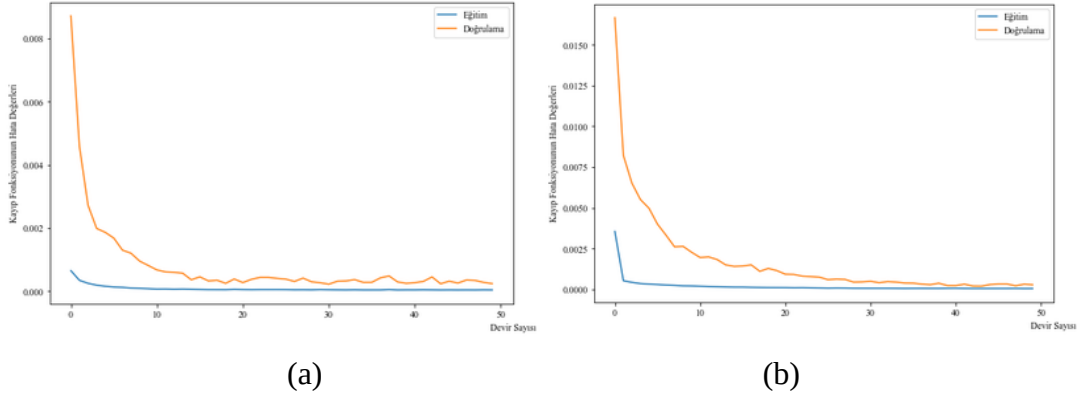
Tablo 4.10'daki Model 4 test örneklerinden L4.2 ve L4.3'ün tahmin grafikleri Şekil 4.15'te ve kayıp fonksiyon grafikleri Şekil 4.16'da verilmiştir.

Tablo 4.10. Modeller için LSTM algoritmasındaki eşik değeri hiperparametresinin tahmin sonucuna etkisi.

Model	Eşik Değeri	MSE	MAE	RMSE	MAPE	R ²
L3.5	0,02	0434	0,615	0,659	17,504	-0,475
L3.8	0,05	0,031	0,132	0,177	3,824	0,894
L3.9	0,08	0,034	0,144	0,186	4,317	0,883
L4.1	0,02	0,048	0,186	0,220	5,761	0,886
L4.2	0,05	0,067	0,216	0,258	6,435	0,842
L4.3	0,08	0,026	0,123	0,162	3,802	0,938



Şekil 4.15. (a) L4.2 (b) L4.3 örneklerinin tahmin grafikleri.



Şekil 4.16. (a) L4.2 (b) L4.3 örneklerinin kayıp fonksiyon grafikleri.

Parti boyutu hiperparametresi, algoritmanın aynı anda kaç adet veriyi işleyeceğini gösteren hiperparametredir. Tablo 4.11'de parti boyutu hiperparametresinin tahmin sonuçlarına etkisini göstermektedir. Verilen örneklerde parti boyutu hiperparametresinin azalması veya artmasının tahmin sonucuna belirli yönde bir etkisinin olmadığı görülmüştür.

Tablo 4.11. Modeller için LSTM algoritmasındaki parti boyutu hiperparametresinin tahmin sonucuna etkisi.

Model	Parti Boyutu	MSE	MAE	RMSE	MAPE	R ²
L1.12	100	0,019	0,101	0,139	3,057	0,902
L1.13	80	0,018	0,094	0,133	2,827	0,909
L1.14	50	0,015	0,082	0,121	2,456	0,926
L1.15	20	0,019	0,096	0,137	2,874	0,903
L2.6	50	0,126	0,319	0,355	9,415	0,573
L2.7	100	0,074	0,211	0,271	6,036	0,750
L2.8	20	0,032	0,140	0,891	4,190	0,891

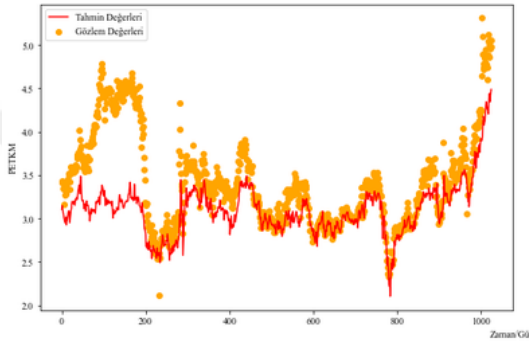
LSTM algoritmaları için verilen örneklere bakıldığında tahmin sonuçlarını en çok etkileyen hiperparametrenin rastgele öğrenme durumu olduğu görülmektedir. Rastgele öğrenme durumu değeri Hayır olarak seçildiğinde eğitim veri seti boyutunun büyümesi tahmin sonuçlarını olumlu etkilerken, Rastgele öğrenme durumu değeri Evet olarak seçildiğinde eğitim veri seti boyutu azaldığında da başarılı tahmin sonuçlarına ulaşılabilmektedir. Modeller için uygulanan aktivasyon fonksiyonları da göz önüne alındığında kullanılan veri setleri için genellikle en uygun aktivasyon fonksiyonunun Tanh olduğu görülmüştür.

CNN+LSTM modeli evrişimli sinir ağı (CNN) ve LSTM algoritmalarından oluşmuş hibrit bir algortmadır. Tablo 4.12’de Evrişim katmanındaki (Conv1D) filtre sayısı değerleri verilmiştir. Verilen örneklerde filtre sayısı hiperparametre değeri azaldıkça tahmin başarısının genellikle arttığı görülmüştür. CL2.6 ve CL2.7 test örneklerinin tahmin grafikleri Şekil 4.17’de ve kayıp fonksiyon grafikleri Şekil 4.18’de verilmiştir.

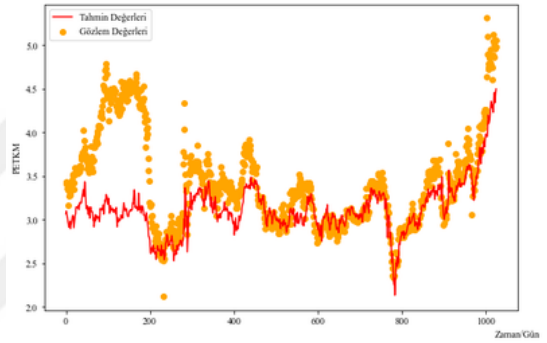
Tablo 4.13’te CNN+LSTM algoritmasının CNN kısmında bulunan havuzlama katmanı (MaxPooling1D) havuz boyutu hiperparametresinin tahmin sonuçlarına etkisi verilmiştir. Havuzlama boyutu 2 olarak verildiğinde daha iyi tahmin sonuçları elde edildiği görülmüştür. CL1.11 ve CL1.12 test örneklerinin tahmin grafikleri Şekil 4.19’da ve kayıp fonksiyon grafikleri Şekil 4.20’de verilmiştir.

Tablo 4.12. Modeller için CNN+LSTM algoritmasındaki filtre sayısı hiperparametresinin tahmin sonucuna etkisi.

Model	Filtre Sayısı	MSE	MAE	RMSE	MAPE	R ²
CL2.1	64	0,413	0,544	0,642	13,116	0,025
CL2.2	32	0,458	0,555	0,677	16,550	-0,082
CL2.6	32	0,241	0,329	0,491	8,601	0,181
CL2.7	16	0,267	0,337	0,517	8,781	0,092
CL3.13	32	0,018	0,088	0,155	2,637	0,910
CL3.14	64	0,021	0,104	0,145	3,171	0,892
CL4.1	32	0,359	0,496	0,599	14,669	0,151
CL4.3	64	0,397	0,532	0,630	15,705	0,062
CL4.8	32	0,225	0,307	0,475	7,979	0,234
CL4.9	128	0,269	0,362	0,519	9,497	0,084

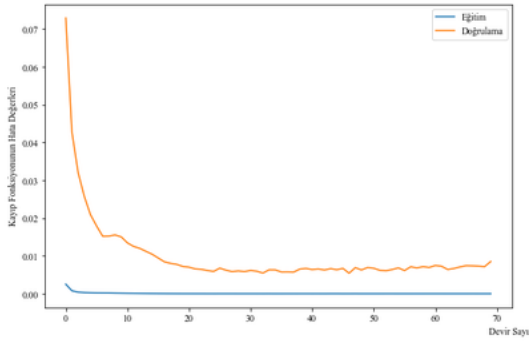


(a)

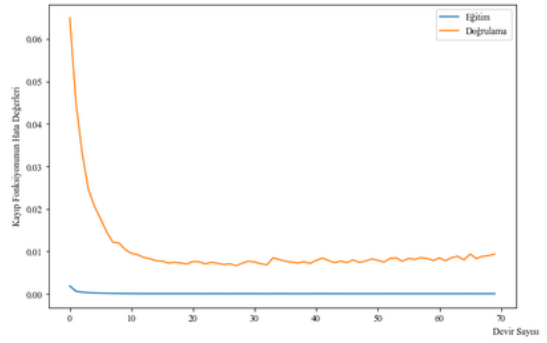


(b)

Şekil 4.17. (a) CL2.6 (b) CL2.7 örneklerinin tahmin grafikleri.



(a)

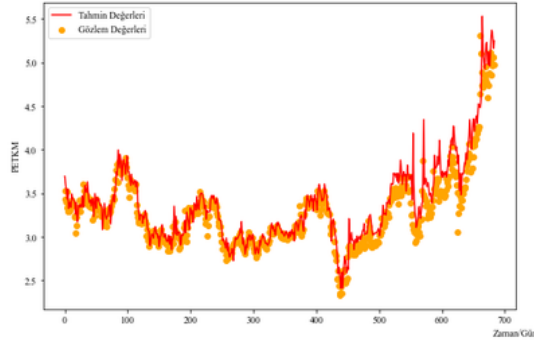


(b)

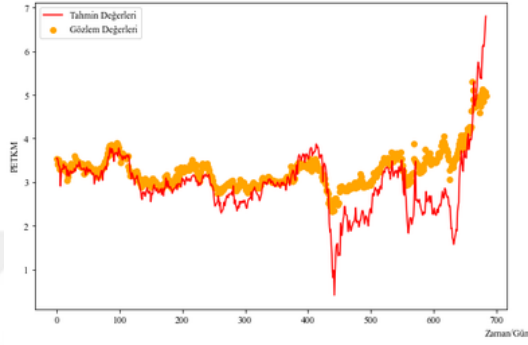
Şekil 4.18. (a) CL2.6 (b) CL2.7 örneklerinin kayıp fonksiyon grafiği.

Tablo 4.13. Modeller için CNN+LSTM algoritmasındaki havuz boyutu hiperparametresinin tahmin sonucuna etkisi.

Model	Havuzlama Boyutu	MSE	MAE	RMSE	MAPE	R ²
CL1.11	2	0,032	0,130	0,180	3,894	0,835
CL1.12	3	0,325	0,404	0,570	012,365	-0,662
CL3.3	2	0,373	0,481	0,611	13,968	0,119
CL3.5	3	0,064	0,666	0,815	18,097	-1,257
CL4.4	2	0,400	0,495	0,633	13,336	-0,360
CL4.6	3	0,476	0,483	0,690	12,605	-0,618

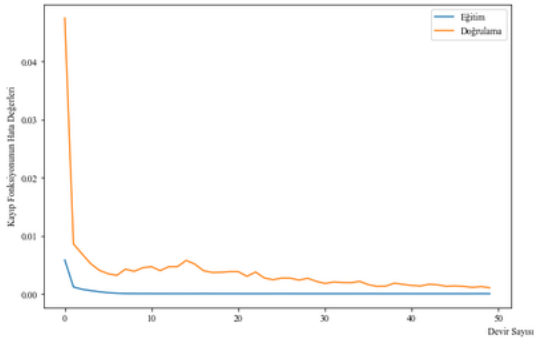


(a)

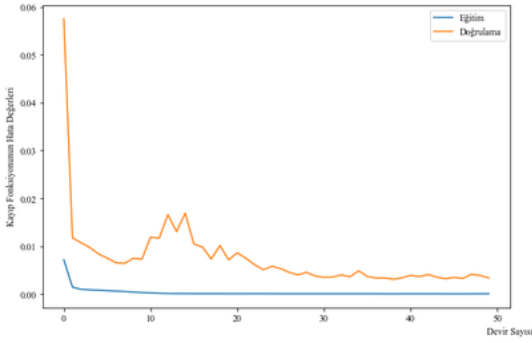


(b)

Şekil 4.19. (a) CL1.11 (b) CL1.12 örneklerinin tahmin grafikleri.



(a)



(b)

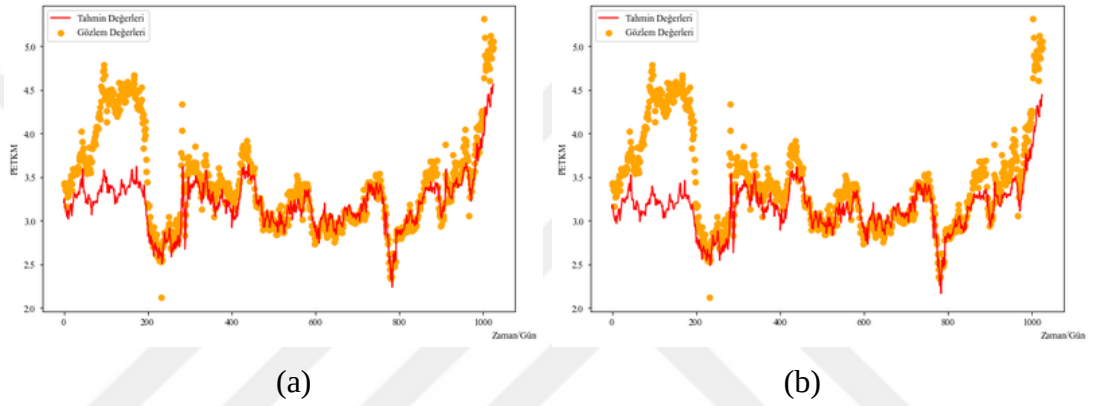
Şekil 4.20. (a) CL1.11 (b) CL2.12 örneklerinin kayıp fonksiyon grafikleri.

CNN+LSTM algoritmasının LSTM kısmında ise LSTM katmanlarındaki Birim Sayısı ve Dropout katmanlarındaki Eşik Değeri hiperparametreleri bulunmaktadır. LSTM katmanlarındaki birim sayısı hiperparametresi için Tablo 4.14'te test örnekleri verilmiştir. Verilen örneklere göre genellikle birim sayısının artması ya da azalması tahmin sonuçlarında çok büyük bir değişikliğe neden olmamıştır. CL4.7 ve CL4.8 test örneklerinin tahmin grafikleri Şekil 4.21'de ve kayıp fonksiyon grafikleri Şekil 4.22'de verilmiştir.

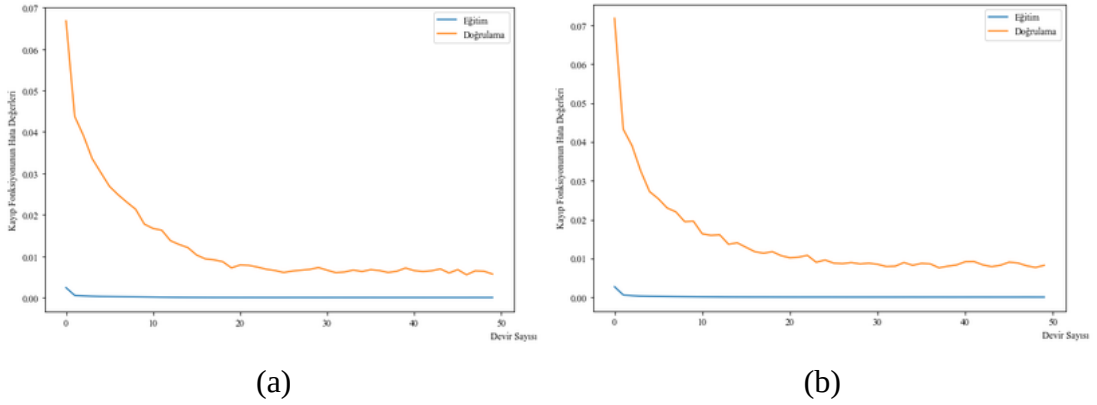
Dropout katmanındaki eşik değeri hiperparametresinin tahmin sonuçlarına etkisi Tablo 4.15'te verilmiştir. Verilen test örneklerine bakıldığında eşik değeri azaldıkça tahmin başarısının arttığı görülmüştür. CL2.13 ve CL2.14 test örneklerinin tahmin grafikleri Şekil 4.23'te ve kayıp fonksiyon grafikleri Şekil 4.24'te verilmiştir.

Tablo 4.14. Modeller için CNN+LSTM algoritmasındaki birim sayısı hiperparametresinin tahmin sonucuna etkisi.

Model	Birim Sayısı	MSE	MAE	RMSE	MAPE	R ²
CL1.7	50	0,198	0,281	0,445	7,432	0,372
CL1.8	100	0,208	0,283	0,457	7,368	0,236
CL4.7	50	0,165	0,259	0,406	6,731	0,440
CL4.8	30	0,225	0,307	0,475	7,979	0,234



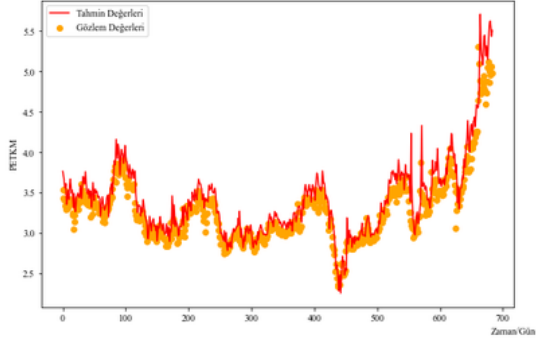
Şekil 4.21. (a) CL4.7 (b) CL4.8 örneklerinin tahmin grafikleri.



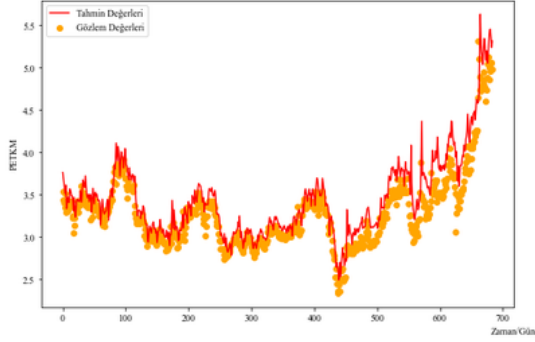
Şekil 4.22. (a) CL4.7 (b) CL4.8 örneklerinin kayıp fonksiyon grafikleri.

Tablo 4.15. Modeller için CNN+LSTM algoritmasındaki eşik değeri hiperparametresinin tahmin sonucuna etkisi.

Model	Eşik Değeri	MSE	MAE	RMSE	MAPE	R ²
CL2.13	0,01	0,044	0,159	0,209	4,736	0,778
CL2.14	0,03	0,053	0,185	0,231	5,580	0,727

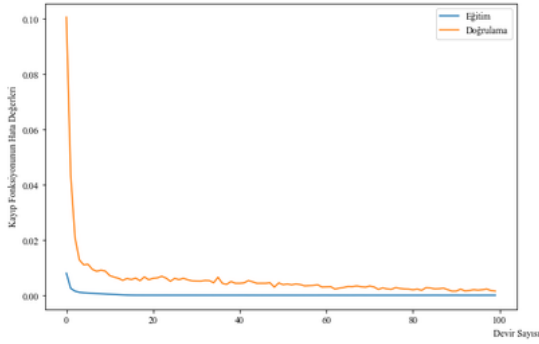


(a)

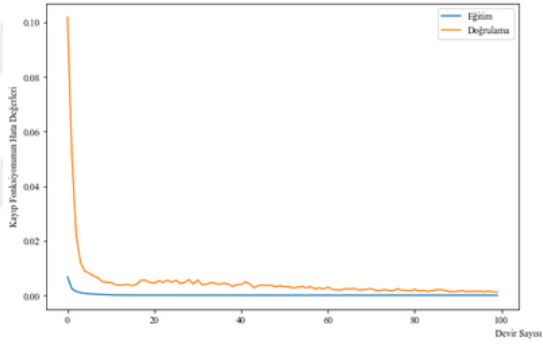


(b)

Şekil 4.23. (a) CL2.13 (b) CL2.14 örneklerinin tahmin grafikleri.



(a)



(b)

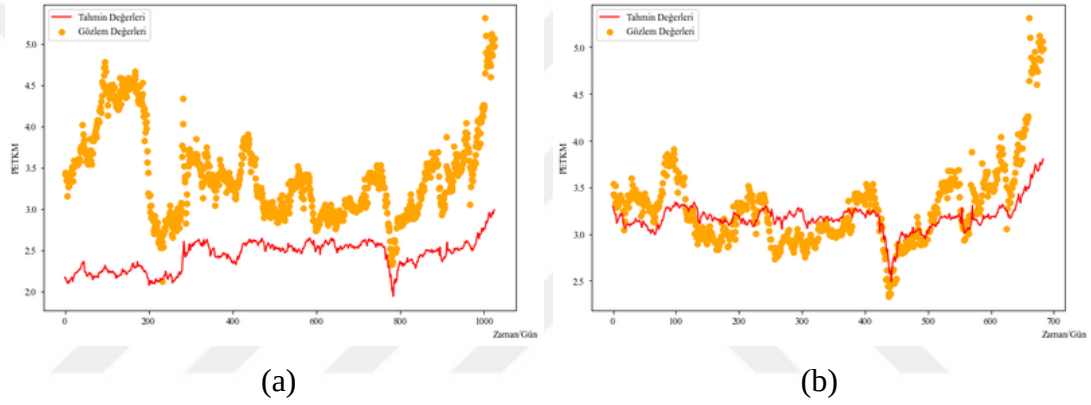
Şekil 4.24. (a) CL2.13 (b) CL2.14 örneklerinin kayıp fonksiyon grafikleri.

Rastgele öğrenme durumu hiperparametresinin LSTM algoritmasındaki başarısı yüksek bir düzeyde olduğu için CNN+LSTM algoritmasındaki tüm testlerde rastgele öğrenme durumu hiperparametresi Evet olarak seçilmiştir. LSTM algoritmasında olduğu gibi bu hiperparametre Evet olarak seçildiğinde CNN+LSTM algoritmasında da eğitim ve test veri seti boyutlarındaki değişiklik genellikle tolere edebilmiş ve tüm eğitim ve test veri seti boyutlarında başarılı sonuçlar elde edilebilmiştir. Tablo 4.16'da CNN+LSTM algoritmasındaki eğitim ve test veri seti boyutu hiperparametrelerindeki değişimin tahmin sonucuna etkisini göstermek için örnek test değerleri verilmiştir. Değerler kontrol edildiğinde eğitim-doğrulama-test veri seti boyutu %56-%14-%30 olduğunda tahmin sonuçlarının genel olarak başarısız olduğu görülmektedir. CL2.9 ve

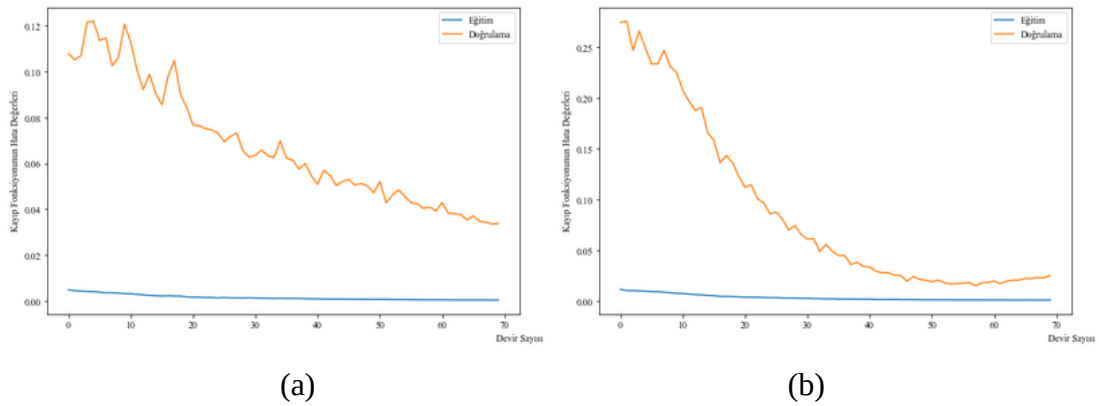
CL2.10 test örneklerinin tahmin grafikleri Şekil 4.25'te ve kayıp fonksiyon grafikleri Şekil 4.26'da verilmiştir.

Tablo 4.16. Modeller için CNN+LSTM algoritmasındaki eğitim, doğrulama ve test veri seti boyutu hiperparametresinin tahmin sonucuna etkisi.

Model	Eğitim Veri Seti Boyutu	Doğrulama Veri Seti Boyutu	Test Veri seti Boyutu	MSE	MAE	RMSE	MAPE	R ²
CL1.3	1639 (%48)	410 (%12)	1367 (%40)	0,141	0,298	0,376	9,200	0,666
CL1.4	1912 (%56)	479 (%14)	1025 (%30)	0,173	0,262	0,416	6,858	0,412
CL1.11	2185 (%64)	547 (%16)	684 (%20)	0,051	0,160	0,225	4,719	0,742
CL2.9	1912 (%56)	479 (%14)	1367 (%40)	1,275	0,984	1,129	27,220	-3,333
CL2.10	2185 (%64)	547 (%16)	684 (%20)	0,125	0,262	0,353	7,503	0,364
CL3.3	1639 (%48)	410 (%12)	1367 (%40)	0,373	0,481	0,611	13,968	0,119
CL3.4	1912 (%56)	479 (%14)	1025 (%30)	0,420	0,512	0,648	13,776	-0,427
CL4.1	1639 (%48)	410 (%12)	1367 (%40)	0,359	0,496	0,599	14,669	0,151
CL4.4	1912 (%56)	479 (%14)	1025 (%30)	0,400	0,495	0,633	13,336	-0,360
CL4.10	2185 (%64)	547 (%16)	684 (%20)	0,046	0,168	0,215	5,021	0,763



Şekil 4.25. (a) CL2.9 (b) CL2.10 örneklerinin tahmin grafikleri.



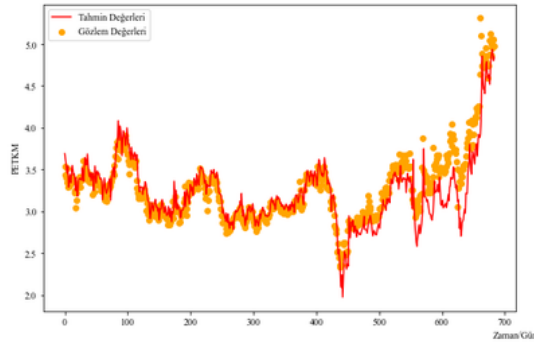
Şekil 4.26. (a) CL2.9 (b) CL2.10 örneklerinin kayıp fonksiyon grafikleri.

CNN+LSTM algoritmasının CNN kısmında bulunan Çekirdek Boyutu hiperparametresinin tahmin sonuçlarına etkisi Tablo 4.17'de verilmiştir. Bu hiperparametreye verilen değerler 2 veya 3'tür. Model 1 ve Model 3'te bu

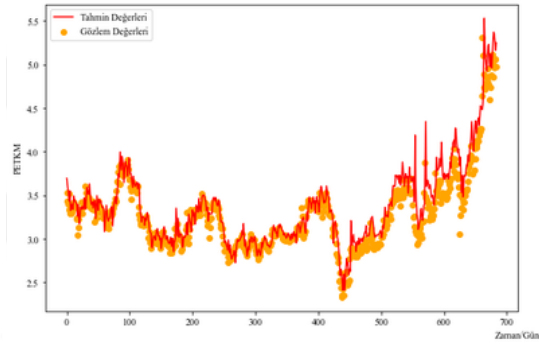
hiperparametre 2 değerinde daha iyi sonuç verirken Model 4'te bu hiperparametrenin 3 değerini almasının tahmin başarısını arttırdığı görülmüştür. CL1.10 ve CL1.11 test örneklerinin tahmin grafikleri Şekil 4.27'de ve kayıp fonksiyon grafikleri Şekil 4.28'de verilmiştir.

Tablo 4.17. Modeller için CNN+LSTM algoritmasındaki çekirdek boyutu hiperparametresinin tahmin sonucuna etkisi.

Model	Çekirdek Boyutu	MSE	MAE	RMSE	MAPE	R ²
CL1.4	2	0,173	0,262	0,416	6,858	0,412
CL1.9	3	0,225	0,315	0,474	8,278	0,236
CL1.10	3	0,051	0,160	0,225	4,719	0,742
CL1.11	2	0,032	0,130	0,180	3,894	0,835
CL3.2	2	0,324	0,418	0,569	12,224	0,235
CL3.3	3	0,373	0,481	0,611	13,968	0,119
CL4.4	2	0,400	0,495	0,633	13,336	-0,360
CL4.5	3	0,305	0,372	0,552	9,654	-0,037

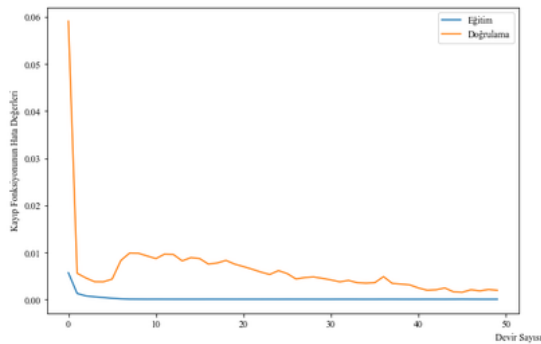


(a)

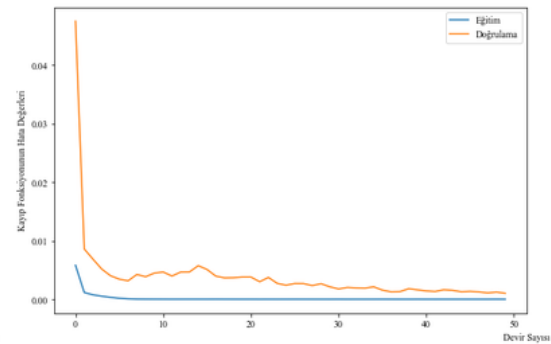


(b)

Şekil 4.27. (a) CL1.10 (b) CL1.11 örneklerinin tahmin grafikleri.



(a)



(b)

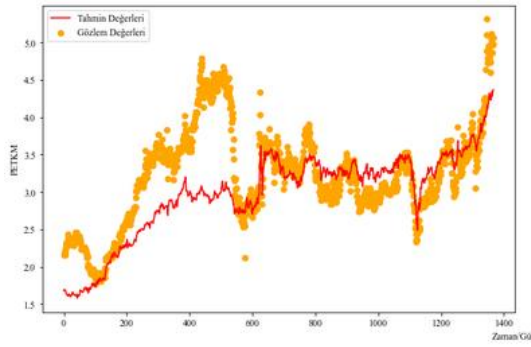
Şekil 4.28. (a) CL1.10 (b) CL1.11 örneklerinin kayıp fonksiyon grafikleri.

Tablo 4.18'de CNN+LSTM algoritması için yapılan testlerden öğrenme aralığı hiperparametresinin değerlendirme sonuçları verilmiştir. Değerlendirme sonuçlarına

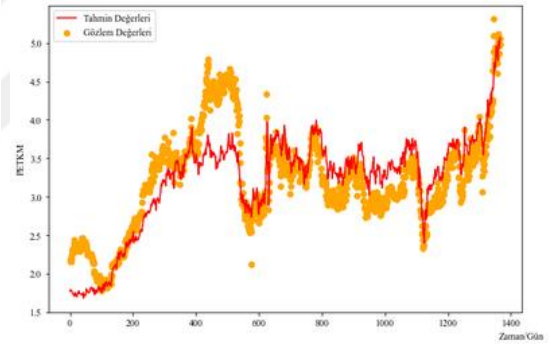
bakıldığında öğrenme aralığı değeri azaldığında tahminin başarısı da genellikle artmaktadır. Model 3'te ise öğrenme aralığı değeri 0,002 iken tahmin sonucu başarısız, 0,0002 iken tahmin başarısının orta bir düzeyde olduğu ve 0,003 olduğunda ise tahmin başarısının arttığı görülmektedir. Verilen örnek değerlerden CL1.1 ve CL1.3'ün tahmin grafiği Şekil 4.29'da, kayıp fonksiyon grafikleri ise Şekil 4.30'da verilmiştir.

Tablo 4.18. Modeller için CNN+LSTM algoritmasındaki öğrenme aralığı hiperparametresinin tahmin sonucuna etkisi.

Model	Öğrenme Aralığı	MSE	MAE	RMSE	MAPE	R ²
CL1.1	0,02	0,318	0,403	0,564	11,929	0,249
CL1.3	0,002	0,141	0,298	0,376	9,200	0,666
CL2.14	0,001	0,053	0,185	0,231	5,580	0,727
CL2.15	0,005	0,090	0,208	0,299	6,166	0,542
CL3.8	0,002	0,385	0,456	0,621	12,130	-0,309
CL3.9	0,0002	0,198	0,307	0,445	8,214	0,326
CL3.11	0,001	0,062	0,215	0,250	6,572	0,681
CL3.12	0,003	0,023	0,111	0,153	3,345	0,881
CL4.4	0,002	0,400	0,495	0,633	13,336	-0,360
CL4.7	0,001	0,225	0,307	0,475	7,979	0,234

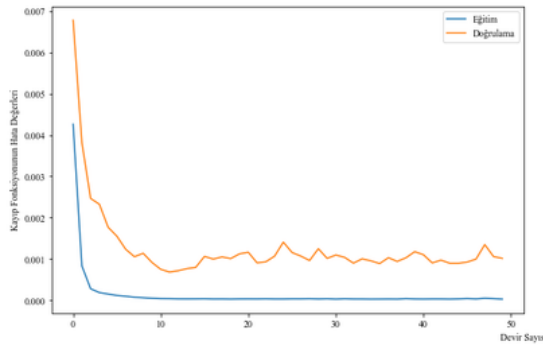


(a)

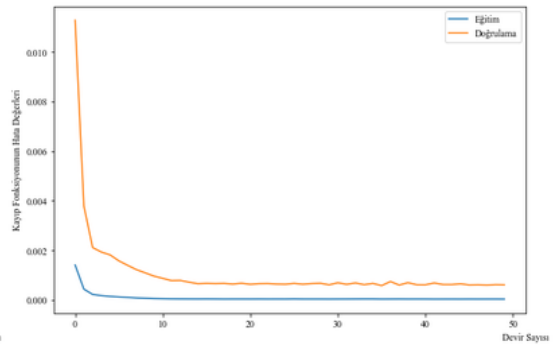


(b)

Şekil 4.29. (a) CL1.1 (b) CL1.3 örneklerinin tahmin grafikleri.



(a)



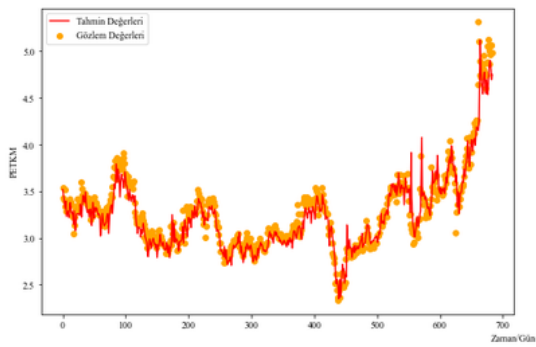
(b)

Şekil 4.30. (a) CL1.1 (b) CL1.3 örneklerinin kayıp fonksiyon grafikleri.

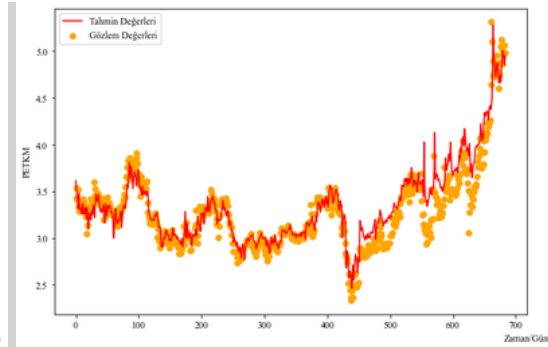
Tablo 4.19’da CNN+LSTM algoritması için yapılan testlerden aktivasyon fonksiyonu hiperparametresinin değerlendirme sonuçları verilmiştir. Değerlendirme sonuçlarına göre, Model 3 ve Model 4 için Tanh fonksiyonunun genel olarak daha iyi sonuçlar verdiği görülmektedir. Model 1 ve Model 2’de Sigmoid fonksiyonunun da iyi değerler verdiği görülmektedir. Tabloda verilen örnek değerlerden CL4.13 ve CL4.14’ün tahmin grafiği Şekil 4.31’de, kayıp fonksiyon grafiği Şekil 4.32’de verilmiştir.

Tablo 4.19. Modeller için CNN+LSTM algoritmasındaki aktivasyon fonksiyonu hiperparametresinin tahmin sonucuna etkisi.

Model	Katman	Aktivasyon Fonksiyonu	MSE	MAE	RMSE	MAPE	R ²
CL1.12	Conv1D LSTM	Tanh Tanh	0,325	0,404	0,570	12,365	-0,662
CL1.13	Conv1D LSTM	ReLU ReLU	0,257	0,435	0,507	12,924	-0,315
CL1.14	Conv1D LSTM	Sigmoid Sigmoid	0,170	0,255	0,328	7,702	0,451
CL1.15	Conv1D LSTM	Sigmoid Tanh	0,034	0,145	0,183	4,474	0,828
CL2.7	Conv1D LSTM	Tanh Tanh	0,267	0,337	0,517	8,781	0,092
CL2.8	Conv1D LSTM	ReLU ReLU	0,387	0,445	0,622	12,263	-3,13
CL2.9	Conv1D LSTM	Sigmoid Sigmoid	1,275	0,984	1,129	27,220	-3,333
CL2.10	Conv1D LSTM	Sigmoid Sigmoid	0,125	0,262	0,353	7,503	0,364
CL2.11	Conv1D LSTM	Tanh Tanh	0,129	0,296	0,359	8,849	0,340
CL3.14	Conv1D LSTM	Tanh Tanh	0,021	0,104	0,145	3,171	0,892
CL3.15	Conv1D LSTM	ReLU ReLU	0,169	0,362	0,411	11,248	0,136
CL4.13	Conv1D LSTM	Tanh Tanh	0,020	0,098	0,143	2,894	0,896
CL4.14	Conv1D LSTM	ReLU ReLU	0,031	0,130	0,177	3,961	0,840
CL4.15	Conv1D LSTM	Sigmoid Sigmoid	0,309	0,395	0,556	10,938	-0,578

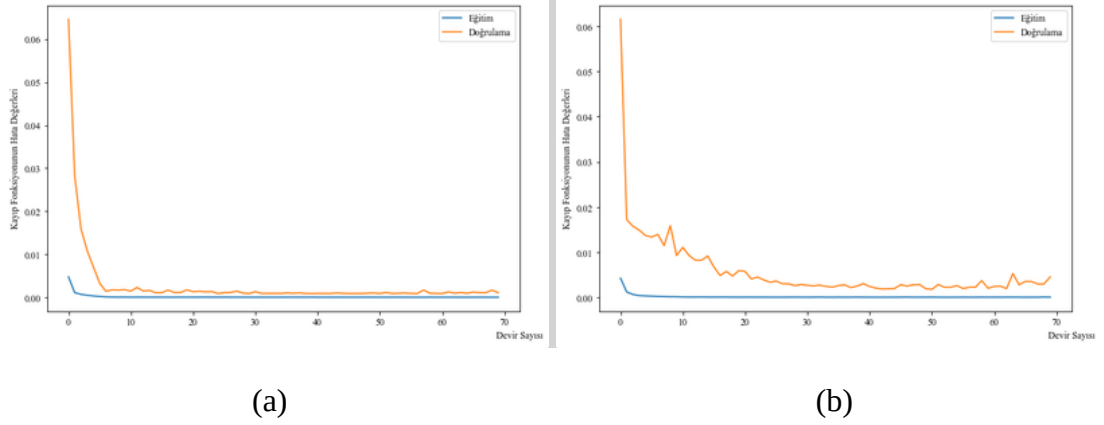


(a)



(b)

Şekil 4.31. (a) CL4.13 (b) CL4.14 örneklerinin tahmin grafikleri.



Şekil 4.32. (a) CL4.13 (b) CL4.14 örneklerinin kayıp fonksiyon grafikleri.

Tablo 4.20’de parti boyutu hiperparametresinin tahmin sonuçlarına etkisi verilmektedir. Model 1 ve Model 3’te parti boyutu hiperparametresinin değeri arttıkça tahmin sonuçlarının daha iyi olduğu görülmektedir.

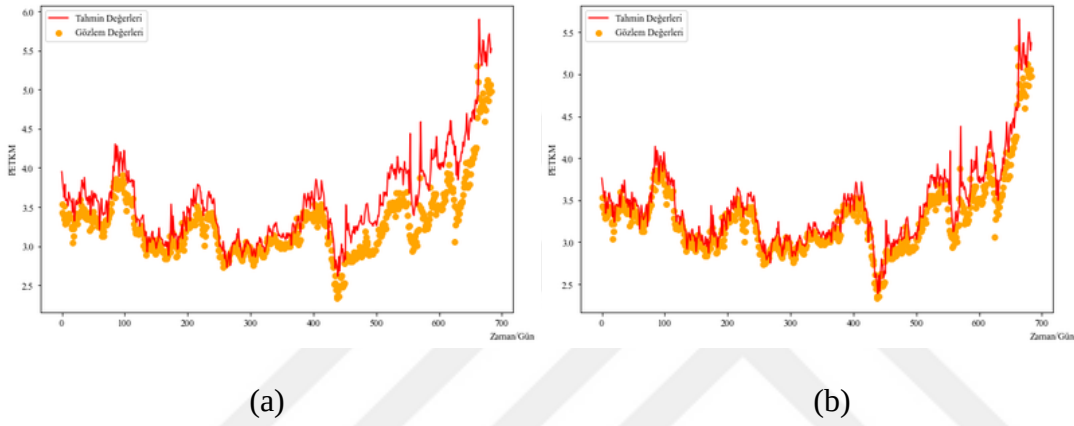
Tablo 4.20. Modeller için CNN+LSTM algoritmasındaki parti boyutu hiperparametresinin tahmin sonucuna etkisi.

Model	Parti Boyutu	MSE	MAE	RMSE	MAPE	R ²
CL1.1	100	0,318	0,403	0,564	11,929	0,249
CL1.2	50	0,467	0,492	0,683	14,311	-0,102
CL3.7	50	0,442	0,510	0,665	13,690	-0,502
CL3.8	100	0,385	0,456	0,621	12,130	-0,309

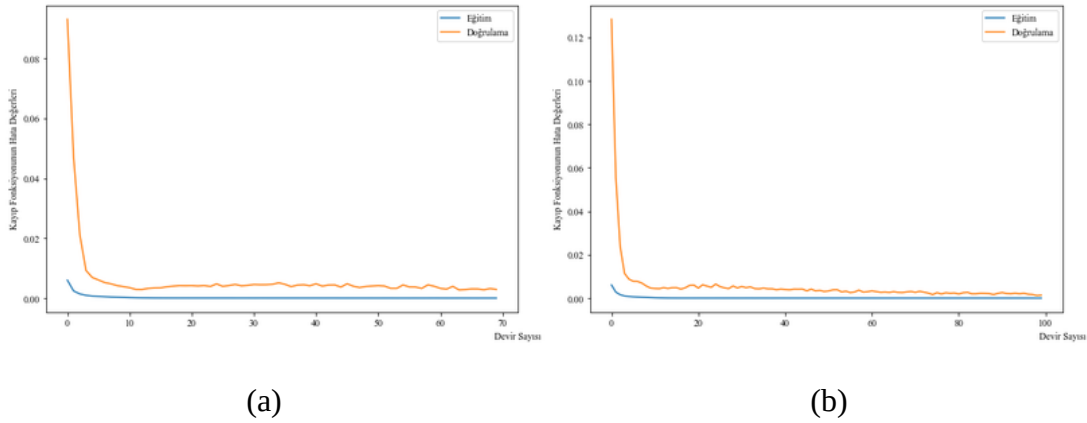
Tablo 4.21’de devir sayısı hiperparametresinin tahmin sonuçlarına etkisi gösterilmektedir. Model 2 ve Model 4’te devir sayısı arttıkça tahmin başarısının arttığı görülmektedir. Model 1 için ise devir sayısı arttıkça algoritmanın tahmin başarısı düşmektedir. Model 3 için ise devir sayısı hiperparametresinin belli bir artış veya düşüş yönü olduğu bilgisine ulaşılamamıştır. CL2.11 ve CL2.12 test örneklerini tahmin grafiği Şekil 4.33’te kayıp fonksiyon grafiği ise Şekil 4.34’te verilmiştir.

Tablo 4.21. Modeller için CNN+LSTM algoritmasındaki devir sayısı hiperparametresinin tahmin sonucuna etkisi.

Model	Devir Sayısı	MSE	MAE	RMSE	MAPE	R ²
CL1.4	50	0,173	0,262	0,416	6,858	0,412
CL1.5	80	0,199	0,279	0,446	7,289	0,325
CL1.6	100	0,231	0,308	0,481	8,120	0,213
CL2.11	70	0,129	0,296	0,359	8,849	0,340
CL2.12	100	0,047	0,170	0,216	5,074	0,762
CL3.1	50	0,297	0,399	0,545	11,660	0,299
CL3.2	80	0,324	0,418	0,569	12,224	0,235
CL3.12	50	0,023	0,111	0,153	3,345	0,881
CL3.13	70	0,018	0,088	0,133	2,637	0,910
CL4.1	50	0,359	0,496	0,599	14,669	0,151
CL4.2	70	0,350	0,487	0,592	14,312	0,172



Şekil 4.33. (a) CL2.11 (b) CL2.12 örneklerinin tahmin grafikleri.



Şekil 4.34. (a) CL2.11 (b) CL2.12 örneklerinin kayıp fonksiyon grafikleri.

CNN+LSTM algoritmaları için yapılan testlerin örnekleri kontrol edildiğinde, çalışma içerisindeki modellere en çok etki eden hiperparametrelerin havuzlama boyutu, öğrenme aralığı ve çekirdek boyutu olduğu görülmüştür.

5. SONUÇ VE ÖNERİLER

Günümüzde büyük miktarlarda gelir elde edilmesini sağladığı için borsalar önemli bir kazanç kapısı haline gelmiştir. Yatırımcılar tecrübe ve sezgilerini kullanarak doğru hisse senetlerine yatırım yapmaya ve doğru zamanda al-sat işlemi yapmaya çalışmaktadırlar. Ancak hisse senetlerinin iç ve dış pek çok faktörden etkinlenmesi, onların sezgisel olarak tahmininde hatalı sonuçlara ulaşma olasılığını artırmaktadır. Bu nedenle tahmin işleminin bilimsel yöntemlere dayalı olarak yapılması, tahmin sonucunun doğruluğunu ve güvenilirliğini artırmaktadır. Bunun için pek çok farklı yöntem ile tahmin işlemi yapılmaya çalışılmıştır. Günümüzde en iyi sonuçları elde eden yöntemlerin makine öğrenmesi ve derin öğrenme gibi yapay zeka yöntemleri olduğu birçok çalışmada ortaya konulmuştur.

Petrol, Türkiye ve diğer ülkeler için büyük önem taşıyan bir hammadDEDİR ve ülkemizde de ham ve işlenmiş türlerinin üretiminde, işlenmesinde, dağıtımında ve satışında pek çok büyük şirket bulunmaktadır. Bu şirketlerin elde ettiği gelişmeler de göz önüne alındığında, petrol şirketlerinin borsadaki hisse senetlerine yatırımcıların ilgisi her geçen gün artmaktadır. Bu çalışmanın amaçlarından biri Türkiye'deki önemli petrol şirketlerinden olan Petkim Petrokimya Holding A.Ş.'nin hisse senedi (PETKM) fiyat tahminlerinde ülke içindeki ve dışındaki petrol ile ilgili endeks ve petrol üretimi, dağıtımı vb. işindeki şirketlerin hisse senetlerinin bir etkisi olup olmadığını tespit etmektir. Bir diğeri ise çalışma içerisindeki algoritmaların geliştirilmesinde kullanılan hiperparametrelerinin etkinliğini gözlemlemek ve veri setine göre en iyi tahminin elde edildiği algoritmayı bulmaktır.

Bu çalışma içerisinde PETKM hisse senedinin kapanış fiyat tahmini yapılmıştır. Girdi olarak dolar alış/satış kur fiyatı (USD/TRY), Türkiye'deki ve Türkiye'nin petrol temin ettiği ülkelerden Katar ve Rusya'daki petrol endeksleri, öncü hisse senetlerinin açılış ve kapanış fiyatları kullanılmıştır. Tahmin işlemi için bu verilerle 4 farklı girdi modeli oluşturulmuştur. Bu modeller:

BİST Kimyasal, Petrol & Plastik (XKMYA) endeksi, Türkiye Petrol Rafinelerileri A.Ş. (TUPRS); PETKM hisse senedi fiyatlarının açılış ve kapanış değerleri, USD/TRY (dolar) açılış ve ortalama alıř/satıř kur fiyatları Model 1'i,

QE Consumer Goods & Services (-Katar'ın tüketiciler malları endeksi-QECON); XKMYA endeksi , Qatar Fuel (QFLS); PETKM hisse senedi fiyatlarının açılıř-kapanıř deęerleri ve dolar açılıř ve ortalama alıř/satıř kur fiyatları Model 2'yi,

XKMYA; MOEX Oil and Gas (-Rusya'nın petrol endeksi- MOEXOG); RTS Oil & Gas (-Rusya'nın petrol endeksi- RTSOG), LUKOIL (LKOIL); PETKM hisse senedi fiyatları; dolar fiyatının açılıř ve ortalama alıř/satıř kur deęerleri Model 3'ü,

XKMYA; QECON; MOEXOG; RTSOG endeksleri, TUPRS; QFLS; LUKOIL; PETKM hisse senedi fiyatları, dolar fiyatının açılıř ve ortalama alıř/satıř kur deęerleri Model 4'ü oluřturmuřtur.

PETKM hisse senedinin oluřturulan bu girdi modelleriyle tahmini için Rastal Orman Regresyonu (RFR), Uzun-Kısa Süreli Bellek (LSTM) ve Evriřimli Sinir Aęı + Uzun Kısa Süreli Bellek (CNN+LSTM) algoritmaları kullanılmıřtır. Tahmin sonularının optimizasyonu için algoritmalarındaki hiperparametre deęerleri kullanılmıřtır. Sonuların deęerlendirilmesi için Ortalama Kareler Hata (MSE), Kök Ortalama Kareler Hata (RMSE), Ortalama Mutlak Hata (MAE), Ortalama Mutlak Yüzde Hata (MAPE) gibi hata metrik yöntemleri ve Determinasyon Katsayısı (R^2) kullanılmıřtır.

RFR, LSTM ve CNN+LSTM algoritmalarının alıřmadaki performansları karřılařtırıldıęında, RFR algoritmasının veri ierisindeki aykırı deęerlere toleransının dięer yöntemlere göre daha düşük olduęu görölmüřtür. Eęitim veri seti boyutunun en yüksek (%80) ve test veri seti boyutunun en düşük (%20) olduęu senaryolarda iyi sonular vermiřtir. Ancak dięer senaryolar için dięer algoritmalarından daha düşük bařarıda sonular vermiřtir. LSTM algoritmasındaki tahmin konusunda en önemli hiperparametresinin rastgele öęrenme durumu (shuffle) olduęu görölmüřtür. Eęitim ařamasında öęrenilen veri gruplarının rastgele bir sıraya göre seilmesi tahmin konusunda bařarı sonucunu büyük oranda artırmakta ve zaman serisi modelini de bozmamaktadır. Bu řekilde eęitim-doęrulama ve test kümesi için yapılan bölünmelerin tümünde bařarılı sonular elde edilebilmektedir. CNN+LSTM algoritmalarının bařarısını en çok etkileyen hiperparametrelerin havuzlama boyutu, öęrenme aralıęı ve ekirdek boyutu olduęu görölmektedir.

Çalışma için oluşturulmuş 4 model kendi içlerinde karşılaştırıldığında genel olarak Model 1'in Model 2, Model 3 ve Model4'ten daha iyi sonuçlar verdiği görülmüştür. Model 2, Model 3ve Model 4'teki Rusya ve Katar endeks değerleri ve hisse senedi fiyatlarının PETKM hisse senedi fiyat tahmin başarısına önemli bir katkı sağladığı görülmemektedir.

İleride yapılacak çalışmalarda, Türkiye'nin petrol temin ettiği diğer ülkelerin hisse senedi ve endeks değerleri veri setine eklenebilir. Literatürde son yıllarda kullanılmaya başlanmış olan GRU-LSTM, ANN-LSTM gibi diğer hibit derin öğrenme yöntemleri tahminlerde kullanılabilir, karşılaştırmaları yapılabilir.





KAYNAKLAR

- [1] Meija, J., Avelar-Sosa, L., Mederos, B., Ramírez, E. Santiago, Roman, J. David Diaz., Prediction of time series using an analysis filter bank of LSTM units. *Computers & Industrial Engineering*, 157, 2021.
- [2] Jackson, C., PrassannaQuadir, J., Sivakumar, M. V., Stock Market Analysis and Prediction Using Time Series Analysis. *Materials Today: Proceedings*, 2021.
- [3] Liu, Z., Loo, C. Kiong, Pasupa, K., A novel error-output recurrent two-layer extreme learning machine for multi-step time series prediction. *Sustainable Cities and Society*, 66, 2021.
- [4] Qitou, G., Qun, D., Huihui, Y., Rui, Y., Integrating Multi-Source Transfer Learning Active Learning and Metric Learning Paradigms for Time Series Prediction, *Applied Soft Computing*, 109, 2021.
- [5] Mateńczuk, K., Kozina, A., Markowska, A., Czerniachowski, K., Kaczmarczyk, K., Golec, P., Hernes, M., Lutosławski, K., Kozierkiewicz, A., Pietranik, M., Rot, A., Dyvak, M., Financial Time Series Forecasting: Comparison of Traditional and Spiking Neural Networks, *Procedia Computer Science*, 192, 5023-5029, 2021.
- [6] Mohanty, D. K., Prida, A. Kumar, Khuntia, S. Suman, Financial market prediction under deep learning framework using auto encoder and kernel extreme learning machine, *Applied Soft Computing*, 99, 2021.
- [7] Cheng, D., Yang, F., Xiang, S., Liu, J., Financial time series forecasting with multi-modality graph neural network, *Pattern Recognition*, 121, 2022.
- [8] Liu, H., Long, Z., An improved deep learning model for predicting stock market price time series, *Digital Signal Processing*, 102, 2020.
- [9] Urolagin, S., Sharma, N., Datta, T. Kumar, A combined architecture of multivariate LSTM with Mahalanobis and Z-Score Transformations for Oil Price Forecasting, *Energy*, 231, 2021.
- [10] Yan, X., Weiha, W., Chang, M, Research on financial assets transaction prediction model based on LSTM neural network. *Neural Computing and Applications*, 33, 257-270, 2021.
- [11] Karakoyun, E. Şeyma, Derin öğrenme ile zaman serilerinin gerçek zamanlı tahmini, Necmettin Erbakan Üniversitesi, Fen Bilimleri Enstitüsü, Endüstri Mühendisliği Bölümü, Yüksek Lisans Tezi, 2018.
- [12] Manjula, K. A., Karthikeyan, P., Gold Price Prediction using Ensemble based Machine Learning Techniques, 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), 1360-1364, 2019.

- [13] Raşo, H., Demirci, M., Predicting the Turkish Stock Market BIST 30 Index using Deep Learning, *International Journal of Engineering Research and Development*, 11(1), 253-265, 2019.
- [14] Yadav, A., Jha, C. K., Sharan, A., Optimizing LSTM for time series prediction in Indian stock market, *Procedia Computer Science*, 167, 2091-2100, 2020.
- [15] Vidal, A., Kristjanpoller, W., Gold volatility prediction using a CNN-LSTM approach, *Expert Systems with Applications*, 157, 2020.
- [16] Jing, N., Wu, Z., Wang, H., A hybrid model integrating deep learning with investor sentiment analysis for stock price prediction, *Expert Systems with Applications*, 178, 2021.
- [17] Hsu, TY., Machine learning applied to stock index performance enhancement, *Journal of Banking and Financial Technology*, 5, 21-33, 2021.
- [18] Demirel, U., Çam, H., Ünlü, R., Predicting Stock Prices Using Machine Learning Methods and Deep Learning Algorithms: The Sample of the Istanbul Stock Exchange, *Gazi University Journal of Science*, 34(1), 63-82, 2021.
- [19] <https://tr.investing.com/> ., Erişim Tarihi: 01.06.2021.
- [20] <https://www.petkim.com.tr/>., Erişim Tarihi: 01.02.2022.
- [21] <https://www.tupras.com.tr/>., Erişim Tarihi: 01.02.2022.
- [22] <https://www.woqod.com/>., Erişim Tarihi: 01.02.2022.
- [23] <https://www.lukoil.com/>., Erişim Tarihi: 01.02.2022
- [24] Garcia, S., Luengo, J., Herrera, F., *Data Preprocessing in Data Mining*, Intelligent Systems Reference Library, 72, 2015.
- [25] Awad, M., Khanna, R., *Efficient Learning Machines*, Apress Berkeley, CA, 2015.
- [26] Kotsiantis, S., Kanellopoulos, D., Pintelas, P. E., *Data Preprocessing for Supervised Learning*, *International Journal of Computer Science*, 1, 111-117, 2006.
- [27] <https://www.kaggle.com/code/abdulkerimnee/veri-n-leme-02/notebook>, Erişim Tarihi: 10.05.2022.
- [28] Gilik, A., Ogrenci, A. S., Ozmen, A., Air quality prediction using CNN+LSTM-based hybrid deep learning architecture, *Environmental Science and Pollution Research* ,29, 11920-11938, 2022
- [29] Subasi, A., *Machine learning techniques*, İçinde: *Practical Machine Learning for Data Analysis Using Python*, 91-202, 2020.
- [30] Kushwah, J. S., Kumar, A., Patel, S., Soni, R., Gawande, A., Gupta, S., Comparative study of regressor and classifier with decision tree using modern tools, *Materials Today: Proceedings*, 56(6), 3571-3576, 2022.
- [31] Breiman, L., *Random Forests*, *Machine Learning*, 45, 5-32, 2001.
- [32] Wang, P., Jiang, T., Fan, G., Dan, C., Prediction of Torpedo Initial Velocity Based on Random Forests Regression, *2015 7th International Conference on Intelligent Human-Machine Systems and Cybernetics*, 337-339, 2015.

- [33] Yakut, B. E. T., Yakut, Elmas, D., Yavuz, Y., Yapay Sinir Ağları ve Destek Vektör Makineleri Yöntemleriyle Borsa Endeksi Tahmini, Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 19(1), 139-157, 2014.
- [34] Çaparoğlu, Ö. F., Yapay Sinir Ağları İle Pandemi Sürecinde Kısıtlama Stratejilerinin Etkinliğinin Belirlenmesi İçin Bir Model Önerisi, Atatürk Üniversitesi, Fen Bilimleri Enstitüsü, Endüstri Mühendisliği Bölümü, 2021.
- [35] Şimşek Can, G., Havacılık Endüstrisinde Yapay Sinir Ağları İle Ürün Ağaçlarındaki Değişimlere Bağlı Maliyet Tahmini, Karabük Üniversitesi, Lisansüstü Eğitim Enstitüsü, Endüstri Mühendisliği Bölümü, 2021.
- [36] Kilicçarslan, S., Celik, M., RSigELu: A nonlinear activation function for deep neural networks, Expert Systems with Applications, 174, 2021.
- [37] Singh, J, Banerjee, R, A study on single and multi-layer perceptron neural network, 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), 35-40, 2019.
- [38] Zupan, J., Introduction to Artificial Neural Network (ANN) Methods: What They Are and How to Use Them, Acta Chimica Slovenica, 41(3), 327-352, 1994.
- [39] Öztürk, K., Şahin, M. E., Yapay Sinir Ağları ve Yapay Zeka'ya Genel Bir Bakış, Takvim-i Vekayi, 6(2), 25-36, 2018.
- [40] LeCun, Y., Bengio, Y., Hinton, G., Deep learning, Nature, 521, 2015.
- [41] O'Shea, K. T., Nash, R., An Introduction to Convolutional Neural Networks, 2015.
- [42] Albawi, S., Mohammed T. A., Al-Zawi, S., Understand of a convolutional neural network, 2017 International Conference on Engineering and Technology (ICET), 1-6, 2017.
- [43] Mehtab, S., Sen, J., Dasgupta, S., Robust Analysis of Stock Price Time Series Using CNN and LSTM-Based Deep Learning Models, 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), 1481-1486, 2020.
- [44] Lasheras, F.S., Juez, F.J.D.C., Sánchez, A. S., Krzemień, A., Fernández, P. R., Forecasting the COMEX copper spot price by means of neural networks and ARIMA models, Resources Policy, 45, 37-43, 2015.
- [45] Kumar, S., Ningombam, D., Short-Term Forecasting of Stock Prices Using Long Short Term Memory, 2018 International Conference on Information Technology (ICIT), 182-186, 2018.
- [46] Staudemeyer, R. C., Morris, E. R., Understanding LSTM-a tutorial into Long Short-Term Memory Recurrent Neural Networks, 2019.
- [47] Jin, Z., Yang, Y., Liu, Y., Stock closing price prediction based on sentiment analysis and LSTM, Neural Computing and Applications, 32, 9713-9729, 2020.
- [48] Yu, P., Yan, X., Stock price prediction based on deep neural networks, Neural Computing and Applications, 32, 1602-1628, 2020.

- [49] Li, G., Xiao, M., Guo, Y., Application of Deep Learning in Stock Market Valuation Index Forecasting, 2019 IEEE 10th International Conference on Software Engineering and Service Science (ICSESS), 551-554, 2019.
- [50] Charniak, E., Introduction to Deep Learning, MIT Press, 2019.
- [51] Sakarya, Ş., Yılmaz, Ü., Derin Öğrenme Mimarisi Kullanarak BİST30 İndeksinin Tahmini, European Journal of Educational and Social Sciences, 4(2), 106-121, 2019.
- [52] Wu, J.M.T., Li, Z., Herencsar, N., Vo, B., Lin, J.C.W., A graph- based CNN-LSTM stock price prediction algorithm with leading indicators, Multimedia Systems, 2021.
- [53] Brownlee, J., Deep Learning for Time Series Forecasting, 2018.
- [54] <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html>, Erişim Tarihi: 05.06.2022
- [55] <https://keras.io/api/>, Erişim Tarihi: 30.06.2022
- [56] Ciner, C., Do industry returns predict the stock market? A reprise using the random forest, The Quarterly Review of Economics and Finance, 72, 152-158, 2019.
- [57] Sezer, O.B, Gudelek, M.U., Ozbayoglu, A.M, Financial time series forecasting with deep learning: A systematic literature review: 2005-2019, Applied Soft Computing, 90, 2020.
- [58] <https://machinelearningmastery.com/multivariate-time-series-forecasting-lstm-keras/>, Erişim Tarihi: 15.07.2022

ÖZGEÇMİŞ

Ad-Soyad : Şevval Toprak

ÖĞRENİM DURUMU:

- **Lisans** : 2019, Balıkesir Üniversitesi, Mühendislik Mimarlık Fakültesi, Endüstri Mühendisliği
- **Yüksek Lisans** : Devam Ediyor, Sakarya Üniversitesi, Endüstri Mühendisliği Anabilim Dalı, Endüstri Mühendisliği Program

MESLEKİ DENEYİM VE ÖDÜLLER:

- 2021 yılında Experteam Consulting'te Tedarik Zincirleri Modüllerinde Foksiyonel Danışman olarak çalışmaya başladı, halen devam ediyor.

TEZDEN TÜRETİLEN ESERLER:

- Toprak, Ş., Çağıl, G. ve Kökçam, A. H., (Yayın Sürecinde), Makine Öğrenmesi Algoritmalarıyla Hisse Senedi Kapanış Fiyat Tahmini: BIST'te Yer Alan PETKM Hisse Senedi Örneği, Düzce Üniversitesi Bilim ve Teknoloji Dergisi

DİĞER ESERLER:

- Arslankaya, S. ve Toprak, Ş. 2021. Makine Öğrenmesi ve Derin Öğrenme Algoritmalarını Kullanarak Hisse Senedi Fiyat Tahmini, Uluslararası Mühendislik Araştırma ve Geliştirme Dergisi, 13 (1) , 178-192