



**AKIŞ VERİSİNDEKİ DAĞILIM TAHMİNİNE DAYALI TAVSİYE
SİSTEMİ VE BEKLENMEDİK ÖNERİ SUNUMU**

Anıl UTKU

**DOKTORA TEZİ
BİLGİSAYAR MÜHENDİSLİĞİ ANA BİLİM DALI**

**GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

HAZİRAN 2020

Anıl UTKU tarafından hazırlanan “AKIŞ VERİSİNDEKİ DAĞILIM TAHMİNİNE DAYALI TAVSİYE SİSTEMİ VE BEKLENMEDİK ÖNERİ SUNUMU” adlı tez çalışması aşağıdaki jüri tarafından OY BİRLİĞİ ile Gazi Üniversitesi Bilgisayar Mühendisliği Ana Bilim Dalında DOKTORA TEZİ olarak kabul edilmiştir.

Danışman: Prof. Dr. M. Ali AKCAYOL

Bilgisayar Mühendisliği Ana Bilim Dalı, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Doktora Tezi olduğunu onaylıyorum.

.....

Başkan: Prof. Dr. O. Ayhan ERDEM

Bilgisayar Mühendisliği Ana Bilim Dalı, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Doktora Tezi olduğunu onaylıyorum.

.....

Üye: Prof. Dr. Ebru AKÇAPINAR SEZER

Bilgisayar Mühendisliği Ana Bilim Dalı, Hacettepe Üniversitesi

Bu tezin, kapsam ve kalite olarak Doktora Tezi olduğunu onaylıyorum.

.....

Üye: Prof. Dr. Hasan OĞUL

Bilgisayar Mühendisliği Ana Bilim Dalı, Başkent Üniversitesi

Bu tezin, kapsam ve kalite olarak Doktora Tezi olduğunu onaylıyorum.

.....

Üye: Doç. Dr. Oktay YILDIZ

Bilgisayar Mühendisliği Ana Bilim Dalı, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Doktora Tezi olduğunu onaylıyorum.

.....

Tez Savunma Tarihi: 29/06/2020

Jüri tarafından kabul edilen bu tezin Doktora Tezi olması için gerekli şartları yerine getirdiğini onaylıyorum.

.....

Prof. Dr. Sena YAŞYERLİ

Fen Bilimleri Enstitüsü Müdürü

ETİK BEYAN

Gazi Üniversitesi Fen Bilimleri Enstitüsü Tez Yazım Kurallarına uygun olarak hazırladığım bu tez çalışmada;

- Tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi,
- Tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu,
- Tez çalışmada yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi,
- Kullanılan verilerde herhangi bir değişiklik yapmadığımı,
- Bu tezde sunduğum çalışmanın özgün olduğunu,

bildirir, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim.

Anıl UTKU

29/06/2020

AKIŞ VERİSİNDEKİ DAĞILIM TAHMİNİNE DAYALI TAVSİYE SİSTEMİ VE BEKLENMEDİK ÖNERİ SUNUMU

(Doktora Tezi)

Anıl UTKU

GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Haziran 2020

ÖZET

Olayların, sonraki meydana gelme zamanlarının tahmin edilmesine yönelik çalışmalar uzun yıllardır devam etmektedir. Bu sayede olaylar meydana geldiğinde ortaya çıkacak zararı minimuma indirmek ya da elde edilecek faydayı maksimum yapmak amaçlanır. Bir müşterinin sonraki davranışının tahmini, doğal afetlerin sonraki olma zamanının tahmini, belirli bir zaman aralığında gelecek talep sayısının tahmini gibi çok farklı alanlarda çalışmalar yapılmaktadır. Ancak, literatürde olayların meydana gelmeden önce bir sonraki olma zamanının ve olay türünün tahminine yönelik beklenen düzeyde başarılı sonuç veren bir çalışma bulunmamaktadır. Bu tez çalışmasında, olay zamanı tahminine yönelik hibrit bir derin öğrenme modeli olan HDLM geliştirilmiştir. HDLM ile elde edilen sonuçlar, RF, SVM, ARIMA, MLP, CNN, GRU, LSTM ve CNN ile GRU modelleri kullanılarak oluşturulan CNN+GRU hibrit modeli ile kapsamlı bir şekilde karşılaştırılmıştır. Geliştirilen model e-ticaret, DDoS, suç ve acil çağrı veri kümeleri üzerinde uygulanmıştır. Deneysel sonuçlar, geliştirilen hibrit derin öğrenme modelinin ileriye dönük bir sonraki olayın olma zamanının tahmininde diğer modellerden daha başarılı olduğunu göstermiştir.

Bilim Kodu : 92431
Anahtar Kelimeler : Öneri sistemleri, Derin öğrenme, ARIMA, Olay zamanı tahmini
Sayfa Adedi : 223
Danışman : Prof. Dr. M. Ali AKCAYOL

RECOMMENDATION SYSTEM BASED ON DISTRIBUTION PREDICTION IN
STREAM DATA AND UNEXPECTED RECOMMENDATION

(Ph. D. Thesis)

Anıl UTKU

GAZİ UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

June 2020

ABSTRACT

Studies for predicting the occurrence time of events have been going on for many years. In this way, it is aimed to minimize the damage when events occur or to maximize the benefit to be obtained. Studies are carried out in many different areas such as the prediction of a client's next behavior, the prediction of the time of natural disasters, and the prediction of the number of requests that will come in a certain period of time. However, there are no studies in the literature that give the expected level of successful results for the prediction of the next time and event type before the events occur. In this thesis, HDLM, a hybrid deep learning model has been developed for event time prediction. The results obtained with HDLM have been extensively compared with RF, SVM, ARIMA, MLP, CNN, GRU, LSTM and CNN+GRU hybrid model created using CNN and GRU models. The developed model has been applied on e-commerce, DDoS, crime and emergency call datasets. Experimental results showed that the developed hybrid deep learning model is more successful than other models in predicting the time of the next event.

Science Code : 92431

Key Words : Recommendation systems, Deep learning, ARIMA, Event time prediction

Page Number : 223

Supervisor : Prof. Dr. M. Ali AKCAYOL

TEŞEKKÜR

Tez çalışmamda büyük emeği olan değerli Hocam Prof. Dr. M. Ali AKCAYOL'a, tez izleme komitemde bulunan hocalarım Prof. Dr. O. Ayhan ERDEM ve Prof. Dr. Ebru AKÇAPINAR SEZER'e, hayatımın her anında benden desteğini esirgemeyen eşim Esen Damla BALO UTKU'ya, mutluluk kaynağım olan kızım Elis Mila UTKU'ya ve aileme çok teşekkür ederim.

Bu tez çalışmasını destekleyen Gazi Üniversitesi Bilimsel Araştırma Projeleri Birimine teşekkür ederim (Proje No: BAP 06/2019-14).

İÇİNDEKİLER

	Sayfa
ÖZET	iv
ABSTRACT.....	v
TEŞEKKÜR.....	vi
İÇİNDEKİLER	vii
ÇİZELGELERİN LİSTESİ.....	x
ŞEKİLLERİN LİSTESİ.....	xiv
SİMGELER VE KISALTMALAR.....	xvii
1. GİRİŞ.....	1
2. ÖNERİ SİSTEMLERİ	9
2.1. Öneri Sistemlerinin İşlevleri	10
2.2. Öneri Sistemlerinde Kullanılan Yöntemler	13
2.2.1. İşbirlikçi filtreleme yöntemi	13
2.2.2. İçerik tabanlı filtreleme yöntemi.....	18
2.3. Öneri Sistemlerinde Kullanılan Veri Toplama Yöntemleri	21
3. AKIŞ VERİSİNDE GELENEKSEL ÖNERİ SİSTEMLERİ	23
3.1. Akış Verisi Analizinde Kullanılan Yöntemler.....	32
3.1.1. Örneklemeye.....	32
3.1.2. Kayan pencereler	34
3.1.3. Histogramlar	34
3.1.4. Eskizler	34
3.2. Doğrusal Yöntemler.....	35
3.2.1. ARMA model	37

	Sayfa
3.2.2. ARIMA model	38
3.3. Makine Öğrenmesi Tabanlı Yöntemler	40
3.3.1. RF.....	40
3.3.2. SVM.....	40
3.3.3. MLP	41
4. AKIŞ VERİSİNDE DERİN ÖĞRENME TABANLI ÖNERİ SİSTEMLERİ.....	43
4.1. CNN	43
4.2. RNN	44
4.3. GRU	45
4.4. LSTM.....	46
5. GELİŞTİRİLEN HİBRİT DERİN ÖĞRENME MODELİ.....	49
5.1. Kullanılan Veri kümeleri	55
5.2. Veri Önileme	59
5.3. Hiperparametre Seçimi	61
6. DENEYSEL SONUÇLAR.....	87
6.1. Bir Olayın Bir Sonraki Olma Zamanının Saniye Cinsinden Belirlenmesi	88
6.1.1. E-ticaret veri kümesi.....	88
6.1.2. DDoS veri kümesi.....	104
6.1.3. Suç veri kümesi.....	119
6.1.4. Acil çağrı veri kümesi.....	134
6.2. Bir Olayın Bir Sonraki Olma Zamanının Dakika Cinsinden Belirlenmesi.....	150
6.2.1. E-ticaret veri kümesi.....	150
6.2.2. DDoS veri kümesi.....	164

	Sayfa
6.2.3. Suç veri kümesi.....	178
6.2.4. Acil çağrı veri kümesi.....	193
7. SONUÇLAR	209
KAYNAKLAR	213
ÖZGEÇMİŞ	221

ÇİZELGELERİN LİSTESİ

Çizelge	Sayfa
Çizelge 2.1. Kullanıcı derecelendirme matrisi.....	17
Çizelge 5.1. E-ticaret veri kümesi.....	58
Çizelge 5.2. DDoS saldırıları arasındaki zaman farkı.....	58
Çizelge 5.3. Baltimor şehrindeki suç verileri.....	59
Çizelge 5.4. Acil çağrı verileri.....	59
Çizelge 5.5. RF parametre seçimi	62
Çizelge 5.6. SVM parametre seçimi	63
Çizelge 5.7. MLP parametre seçimi.....	64
Çizelge 5.8. CNN parametre seçimi	64
Çizelge 5.9. GRU parametre seçimi	65
Çizelge 5.10. LSTM parametre seçimi	66
Çizelge 5.11. CNN+GRU parametre seçimi.....	66
Çizelge 5.12. HDLM parametre seçimi	67
Çizelge 5.13. RF parametre seçimi	68
Çizelge 5.14. SVM parametre seçimi	69
Çizelge 5.15. MLP parametre seçimi.....	70
Çizelge 5.16. CNN parametre seçimi	70
Çizelge 5.17. GRU parametre seçimi	71
Çizelge 5.18. LSTM parametre seçimi	72
Çizelge 5.19. CNN+GRU parametre seçimi.....	72
Çizelge 5.20. HDLM parametre seçimi	73
Çizelge 5.21. RF parametre seçimi	74
Çizelge 5.22. SVM parametre seçimi	75

Çizelge	Sayfa
Çizelge 5.23. MLP parametre seçimi.....	76
Çizelge 5.24. CNN parametre seçimi	76
Çizelge 5.25. GRU parametre seçimi	77
Çizelge 5.26. LSTM parametre seçimi	78
Çizelge 5.27. CNN+GRU parametre seçimi.....	78
Çizelge 5.28. HDLM parametre seçimi	79
Çizelge 5.29. RF parametre seçimi	80
Çizelge 5.30. SVM parametre.....	80
Çizelge 5.31. MLP parametre seçimi.....	82
Çizelge 5.32. CNN parametre seçimi	82
Çizelge 5.33. GRU parametre seçimi	83
Çizelge 5.34. LSTM parametre seçimi	83
Çizelge 5.35. CNN+GRU parametre seçimi.....	84
Çizelge 5.36. HDLM parametre seçimi	85
Çizelge 6.1. Satın alma olaylarının meydana gelme zamanları arasındaki fark	88
Çizelge 6.2. MSE metriğine göre deneysel sonuçlar	91
Çizelge 6.3. MSE metriğine göre HDLM'nin iyileşme oranları	93
Çizelge 6.4. RMSE metriğine göre deneysel sonuçlar	95
Çizelge 6.5. RMSE metriğine göre HDLM'nin iyileşme oranları.....	97
Çizelge 6.6. MAE metriğine göre deneysel sonuçlar	100
Çizelge 6.7. MAE metriğine göre HDLM'nin iyileşme oranları.....	101
Çizelge 6.8. DDoS saldırıları arasındaki zaman farkı.....	104
Çizelge 6.9. MSE metriğine göre deneysel sonuçlar	106
Çizelge 6.10. MSE metriğine göre HDLM'nin iyileşme oranları	108

Çizelge	Sayfa
Çizelge 6.11. RMSE metriğine göre deneysel sonuçlar	110
Çizelge 6.12. RMSE metriğine göre HDLM'nin iyileşme oranları.....	112
Çizelge 6.13. MAE metriğine göre deneysel sonuçlar	114
Çizelge 6.14. MAE metriğine göre HDLM'nin iyileşme oranları.....	116
Çizelge 6.15. Çete saldırısı olaylarının meydana gelme zamanları arasındaki fark	119
Çizelge 6.16. MSE metriğine göre deneysel sonuçlar	122
Çizelge 6.17. MSE metriğine göre HDLM'nin iyileşme oranları	124
Çizelge 6.18. RMSE metriğine göre deneysel sonuçlar	126
Çizelge 6.19. RMSE metriğine göre HDLM'nin iyileşme oranları.....	128
Çizelge 6.20. MAE metriğine göre deneysel sonuçlar	130
Çizelge 6.21. MAE metriğine göre HDLM'nin iyileşme oranları.....	132
Çizelge 6.22. Trafik kazası olaylarının meydana gelme zamanları arasındaki fark	135
Çizelge 6.23. MSE metriğine göre deneysel sonuçlar	137
Çizelge 6.24. MSE metriğine göre HDLM'nin iyileşme oranları	139
Çizelge 6.25. RMSE metriğine göre deneysel sonuçlar	141
Çizelge 6.26. RMSE metriğine göre HDLM'nin iyileşme oranları.....	143
Çizelge 6.27. MAE metriğine göre deneysel sonuçlar	145
Çizelge 6.28. MAE metriğine göre HDLM'nin iyileşme oranları.....	147
Çizelge 6.29. Bir önceki satın almadan sonra geçen zaman	150
Çizelge 6.30. MSE metriğine göre deneysel sonuçlar	152
Çizelge 6.31. MSE metriğine göre HDLM'nin iyileşme oranları	154
Çizelge 6.32. RMSE metriğine göre deneysel sonuçlar	156
Çizelge 6.33. RMSE metriğine göre HDLM'nin iyileşme oranları.....	158
Çizelge 6.34. MAE metriğine göre deneysel sonuçlar	160

Çizelge	Sayfa
Çizelge 6.35. MAE metriğine göre HDLM'nin iyileşme oranları.....	162
Çizelge 6.36. Bir önceki saldırıdan sonra geçen zaman	165
Çizelge 6.37. MSE metriğine göre deneysel sonuçlar	166
Çizelge 6.38. MSE metriğine göre HDLM'nin iyileşme oranları	168
Çizelge 6.39. RMSE metriğine göre deneysel sonuçlar	170
Çizelge 6.40. RMSE metriğine göre HDLM'nin iyileşme oranları.....	172
Çizelge 6.41. MAE metriğine göre deneysel sonuçlar	174
Çizelge 6.42. MAE metriğine göre HDLM'nin iyileşme oranları.....	176
Çizelge 6.43. Bir önceki çete saldırısından sonra geçen zaman	179
Çizelge 6.44. MSE metriğine göre deneysel sonuçlar	181
Çizelge 6.45. MSE metriğine göre HDLM'nin iyileşme oranları	182
Çizelge 6.46. RMSE metriğine göre deneysel sonuçlar	185
Çizelge 6.47. RMSE metriğine göre HDLM'nin iyileşme oranları.....	186
Çizelge 6.48. MAE metriğine göre deneysel sonuçlar	189
Çizelge 6.49. MAE metriğine göre HDLM'nin iyileşme oranları.....	190
Çizelge 6.50. Bir önceki trafik kazasından sonra geçen zaman (dk).....	193
Çizelge 6.51. MSE metriğine göre deneysel sonuçlar	195
Çizelge 6.52. MSE metriğine göre HDLM'nin iyileşme oranları	197
Çizelge 6.53. RMSE metriğine göre deneysel sonuçlar	199
Çizelge 6.54. RMSE metriğine göre HDLM'nin iyileşme oranları.....	201
Çizelge 6.55. MAE metriğine göre deneysel sonuçlar	203
Çizelge 6.56. MAE metriğine göre HDLM'nin iyileşme oranları.....	205

ŞEKİLLERİN LİSTESİ

Şekil	Sayfa
Şekil 4.1. Tek boyutlu evrişim.....	44
Şekil 4.2. Temel RNN yapısı	44
Şekil 4.3. Temel LSTM yapısı.....	47
Şekil 5.1. Geliştirilen sistemin akış diyagramı	50
Şekil 5.2. HDLM mimarisi	51
Şekil 5.3. Geliştirilen modelleme yaklaşımı	52
Şekil 5.4. Satın alma olayları arasında geçen zamanın ortalama değerleri.....	53
Şekil 5.5. HDLM tarafından yapılan tahmin değerleri	54
Şekil 5.6. Tahmin edilen değerlerin ortalaması ve test verisinin ortalaması	55
Şekil 5.7. İleri yürüyüş doğrulaması	56
Şekil 5.8. Zaman serisi verilerinin denetimli öğrenme verilerine dönüştürülmesi	61
Şekil 5.9. ARIMA parametre seçimi.....	63
Şekil 5.10. ARIMA parametre seçimi.....	69
Şekil 5.11. ARIMA parametre seçimi.....	75
Şekil 5.12. ARIMA parametre seçimi.....	81
Şekil 6.1. Satın alma olayları arasındaki saniye cinsinden zaman farkları	89
Şekil 6.2. Test verisindeki ilk 100 veri örneği için deneysel sonuçlar	90
Şekil 6.3. MSE metriği için karşılaştırmalı deneysel sonuçlar	92
Şekil 6.4. RMSE metriği için karşılaştırmalı deneysel sonuçlar	96
Şekil 6.5. MAE metriği için karşılaştırmalı deneysel sonuçlar	101
Şekil 6.6. DDoS saldırıları arasındaki zaman farkı.....	105
Şekil 6.7. Test verisindeki ilk 100 veri örneği için deneysel sonuçlar	106
Şekil 6.8. MSE metriği için karşılaştırmalı deneysel sonuçlar	107
Şekil 6.9. RMSE metriği için karşılaştırmalı deneysel sonuçlar	111

Şekil	Sayfa
Şekil 6.10. MAE metriği için karşılaştırmalı deneysel sonuçlar	115
Şekil 6.11. Çete saldırısı olayları arasındaki zaman farkı.....	120
Şekil 6.12. Test verisindeki ilk 100 veri örneği için deneysel sonuçlar	121
Şekil 6.13. MSE metriği için karşılaştırmalı deneysel sonuçlar	123
Şekil 6.14. RMSE metriği için karşılaştırmalı deneysel sonuçlar	127
Şekil 6.15. MAE metriği için karşılaştırmalı deneysel sonuçlar	131
Şekil 6.16. Trafik kazası olayları arasındaki zaman farkı.....	135
Şekil 6.17. Test verisinin ilk 100 satırı için deneysel sonuçlar.....	136
Şekil 6.18. MSE metriği için karşılaştırmalı deneysel sonuçlar	138
Şekil 6.19. RMSE metriğine göre karşılaştırmalı deneysel sonuçlar.....	142
Şekil 6.20. MAE metriğine göre karşılaştırmalı deneysel sonuçlar.....	146
Şekil 6.21. Satın alma olayları arasındaki dakika cinsinden zaman farkları	151
Şekil 6.22. Test verisindeki ilk 100 veri örneği için deneysel sonuçlar	152
Şekil 6.23. MSE metriğine göre karşılaştırmalı deneysel sonuçlar	153
Şekil 6.24. RMSE metriğine göre karşılaştırmalı deneysel sonuçlar.....	157
Şekil 6.25. MAE metriğine göre karşılaştırmalı deneysel sonuçlar.....	161
Şekil 6.26. Saldırı olayları arasındaki dakika cinsinden zaman farkları.....	165
Şekil 6.27. Test verisindeki ilk 100 veri örneği için deneysel sonuçlar	166
Şekil 6.28. MSE metriğine göre karşılaştırmalı deneysel sonuçlar	163
Şekil 6.29. RMSE metriğine göre karşılaştırmalı deneysel sonuçlar.....	171
Şekil 6.30. MAE metriğine göre karşılaştırmalı deneysel sonuçlar.....	175
Şekil 6.31. Çete saldırısı olayları arasındaki dakika cinsinden zaman farkları	179
Şekil 6.32. Test verisindeki ilk 100 veri örneği için deneysel sonuçlar	180
Şekil 6.33. MSE metriğine göre karşılaştırmalı deneysel sonuçlar	181

Şekil	Sayfa
Şekil 6.34. RMSE metriğine göre karşılaştırmalı deneysel sonuçlar.....	185
Şekil 6.35. MAE metriğine göre karşılaştırmalı deneysel sonuçlar.....	189
Şekil 6.36. Trafik kazası olayları arasındaki dakika cinsinden zaman farkları	194
Şekil 6.37. Test verisindeki ilk 100 veri örneği için deneysel sonuçlar	195
Şekil 6.38. MSE metriğine göre karşılaştırmalı deneysel sonuçlar	196
Şekil 6.39. RMSE metriğine göre karşılaştırmalı deneysel sonuçlar.....	200
Şekil 6.40. MAE metriğine göre karşılaştırmalı deneysel sonuçlar.....	204

SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış simgeler ve kısaltmalar, açıklamaları ile birlikte aşağıda sunulmuştur.

Kısaltmalar	Açıklamalar
RF	Random Forest
SVM	Support Vector Machine
ARIMA	Autoregressive Integrated Moving Average
MLP	Multilayer Perceptron
CNN	Convolutional Neural Network
GRU	Gated Recurrent Unit
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MSE	Mean Squared Error
RMSE	Root Mean Squared Error

1. GİRİŞ

İnternet teknolojilerindeki gelişmeler ve Web tabanlı uygulamaların yaygınlaşması, kullanıcıların sayısız ürün, film veya etkinlik arasından seçim yapmalarını zorlaştırmıştır. Bu nedenle kişiselleştirme, daha iyi bir kullanıcı deneyimi sağlamak için temel bir strateji olarak ön plana çıkmaktadır [1]. Öneri sistemleri, kullanıcıların aradıkları ürün veya servislere erişme aşamasında yaşadıkları bilişsel maliyeti hafifletmeyi ve karar alma sürecini kolaylaştırmayı amaçlamaktadır [2]. Genel olarak, öneri listeleri kullanıcı tercihlerine, öge özelliklerine ve geçmişteki kullanıcı-öge etkileşimlerine dayanarak oluşturulur. Öneri modelleri temel olarak girdi verisinin türüne göre işbirlikçi filtreleme ve içerik tabanlı filtreleme olarak sınıflandırılmaktadır [3].

Derin öğrenme modelleri, bilgisayarla görme ve konuşma tanıma gibi birçok uygulama alanında başarılı bir şekilde kullanılmaktadır. Son yıllarda, derin öğrenme modelleri öneri performansının artırılması amacıyla kullanılmaktadır [4]. Derin öğrenme tabanlı öneri sistemlerindeki gelişmeler, geleneksel modellerin sınırlamalarını aşarak daha yüksek öneri kalitesi elde edilmesini sağlamaktadır. Derin öğrenme, kullanıcı-öge ilişkilerini etkin bir şekilde çıkarabilmekte ve daha karmaşık soyutlamaların, daha yüksek katmanlardaki veri gösterimleri olarak işlenmesini sağlayabilmektedir [5]. Derin öğrenme modellerinin avantajları içerik tabanlı öneriler hesaplarken ön plana çıkmaktadır [6]. Web üzerindeki öğelerin ve kullanıcıların modellenmesinde geleneksel yöntemlere göre daha başarılı sonuçlar elde edilebilmektedir. Örneğin, tweetler ve blog gibi metin verileri ya da sosyal medya paylaşımları ve ürün görselleri gibi görüntü verileri Evrişimsel Sinir Ağları (Convolutional Neural Network-CNN) ve Tekrarlı Sinir Ağları (Recurrent Neural Network-RNN) gibi derin öğrenme modelleri tarafından başarıyla işlenmektedir.

Zamana bağlı veriler, belirli zaman aralıklarında yapılmış olan ardışık gözlemler dizisini ifade etmektedir [7]. Farklı zaman adımlarındaki gözlemler birbirleriyle temel olarak ilişkilidir. Ardışık gözlemler birbirlerine bağımlı olduğu için gelecekteki değerler geçmiş gözlemler kullanılarak tahmin edilebilmektedir [8]. Gerçek zamanlı veri analiz modeli, zaman serisinin içyapısını tanımlayarak geçmiş değerlerin analiz edilmesini sağlamaktadır [9]. Oluşturulan model, gelecekteki değerleri öngörmek için kullanılmaktadır. Gerçek

zamanlı veri analizi, işletme, ekonomi, finans, enerji yönetimi, telekomünikasyon, biyoinformatik, fen ve mühendislik gibi uygulama alanlarında yaygın bir şekilde kullanılmaktadır [7].

Gerçek zamanlı veri analizinde tahmin doğruluğu, verimli ve etkin modellerin geliştirilmesine bağlıdır. Zamana bağımlı akan veri analizinde, en popüler ve yaygın kullanılan olasılıksal modellerden biri Otoregresif Bütünleşik Hareketli Ortalama modelidir (AutoRegressive Integrated Moving Average-ARIMA) [10]. Bu modeldeki temel varsayım, analiz edilecek zaman serilerinin doğrusal olduğu ve normal dağılım gibi belirli bir istatistiksel dağılıma sahip olduğudur [11]. ARIMA modeli, otoregresif (AutoRegressive-AR), hareketli ortalamalar (Moving Average-MA) ve otoregresif hareketli ortalamalar (Autoregressive Moving Average-ARMA) modelleri gibi alt sınıflara sahiptir.

Box ve ark. tarafından mevsimsel bütünleşik otoregresif hareketli ortalamalar (Seasonal AutoRegressive Integrated Moving Average-SARIMA) modeli, mevsimsel zaman serisi analizi için önerilmiştir [12]. ARIMA modeli, temel olarak uygulamasının basit oluşu ve zaman serisinin birçok çeşidini temsil etme esnekliğinden dolayı yaygın olarak kullanılmaktadır. Ancak, ARIMA modelinin dezavantajı serinin belirlenmiş bir doğrusal modele uyacağı varsayımıdır [13]. Bu dezavantajın üstesinden gelmek için, yapay sinir ağları gibi makine öğrenmesi tabanlı çeşitli doğrusal olmayan modeller önerilmiştir [14].

Chujai ve ark. tarafından 2013 yılında yapılan çalışmada, bir hanedeki elektrik tüketimini günlük, haftalık, aylık ve üç aylık dönemlerde tahmine yönelik bir sistem geliştirilmiştir [15]. Çalışma kapsamında, Aralık 2006 ile Kasım 2010 arasındaki elektrik tüketimi verileri kullanılmıştır. Veri analizleri, ARMA ve ARIMA modelleri kullanılarak gerçekleştirilmiştir. Uygun tahmin yöntemleri ve en uygun tahmin süresi, en düşük Akaike ölçütü (Akaike Information Criterion-AIC) ve kök ortalama kare hata (Root Mean Square Error-RMSE) değerine göre belirlenmiştir. Deneysel çalışmaların sonucunda, ARIMA modelinin aylık ve üç aylık dönemlerde en uygun tahmin süresini bulmak için en iyi model olduğu görülmüştür. Günlük ve haftalık en uygun tahmin dönemini bulmak için ARMA modelinin daha yüksek başarı oranı sergilediği ifade edilmiştir.

Radziukynas ve Klementavičius tarafından 2014 yılında yapılan çalışmada, zaman serileri kullanarak Litvanya’da bulunan bir rüzgâr çiftliğindeki rüzgâr hızının tahmin edilmesine yönelik bir sistem geliştirilmiştir [16]. ARIMA modeli kullanılarak geliştirilen sistemde 4 aylık rüzgâr hızı verileri kullanılmıştır. Tahminlerin doğruluğu, RMSE ve ortalama mutlak hata (Mean Absolute Error-MAE) ölçütleri kullanılarak değerlendirilmiştir. Kullanılan veri kümesi 39 parçaya bölünerek 6–48 saat aralığındaki değerler için tahminler yapılarak sonuçları incelenmiştir.

Kane ve ark. tarafından 2014 yılında yapılan çalışmada, Mısır’daki EMPRES-I sistemi üzerinden elde edilen kuş gribi hastalığı verileri kullanılarak ARIMA ve Rastgele Orman (Random Forest-RF) modelleri kullanılarak hastalığın gelecek zamanlardaki durumu hakkında tahminler yapmak amaçlanmıştır [17]. Deneysel çalışmalar, RF algoritmasının ARIMA modelinden daha yüksek tahmin başarısına sahip olduğunu göstermiştir.

Fu ve ark. tarafından 2016 yılında yapılan çalışmada, ARMA ve ARIMA gibi doğrusal modellerin, araç trafik akışının rastsal ve doğrusal olmayan yapısına uyum sağlayamaması nedeniyle derin öğrenme tabanlı bir sistem geliştirilmiştir [18]. Geliştirilen sistemde, kısa süreli trafik akışını öngörmek için Uzun Kısa Süreli Bellek (Long Short-Term Memory-LSTM) ve Kapılı Tekrarlı Birimler (Gated Recurrent Unit-GRU) kullanılmıştır. Çalışma kapsamında, Kaliforniya’da konuşlandırılmış 15.000’den fazla algılayıcının 30 saniyede bir gönderdiği veriden oluşan PeMS veri kümesi kullanılmıştır. Deneysel çalışmalar, LSTM ve GRU yöntemlerinin ARIMA modelinden daha iyi performans gösterdiğini ortaya koymuştur.

Amini ve ark. tarafından 2016 yılında yapılan çalışmada, ARIMA modeli kullanılarak elektrikli araçların şarj talebini öngörmeye yönelik bir sistem geliştirilmiştir [19]. Şarj talebi tahmin modeli, şarj profillerini belirlemek için günlük sürüş modellerini ve mesafeleri bir girdi olarak almaktadır. Geliştirilen sistemde, ARIMA modelinin parametreleri düzenlenerek karesel hataların ortalamasının (Mean Square Error-MSE) en aza indirilmesi hedeflenmiştir. Deneysel çalışmalar geliştirilen ARIMA tabanlı modelin başarılı sonuçlar verdiğini göstermiştir.

Cesario ve ark. tarafından 2016 yılında yapılan çalışmada, kentsel alanlardaki suç eğiliminin tahmin edilmesine yönelik ARIMA modeli tabanlı bir yöntem geliştirilmiştir [20]. Çalışmada veri kümesi olarak Chicago’da meydana gelmiş suç olaylarının tür, konum ve zaman gibi verilerini içeren, 2001 yılından itibaren haftalık güncellenen veriler kullanılmıştır. Model parametrelerinin belirlenmesi için Akaike ölçütü kullanılarak yapılan çalışmalar sonucunda ARIMA(1,1,1) modeli seçilmiştir. Deneysel çalışmalar, geliştirilen modelin suç tahmininde literatürdeki çalışmalara göre daha başarılı sonuçlar verdiğini göstermiştir.

Kang ve Kang tarafından 2017 yılında yapılan çalışmada, Derin Sinir Ağı (Deep Neural Network- DNN) tabanlı bir model geliştirilmiştir [21]. Çalışma kapsamında kullanılan veri kümesi, Chicago’daki suç istatistikleri, demografik ve meteorolojik bilgiler ile görüntü veritabanlarından toplanan verilerden oluşmaktadır. Çalışmada eğitim verileri oluşturmadan önce istatistiksel analizler yapılarak suçla ilgili veriler seçilmiştir. DNN modeli mekân, zaman, çevresel faktörler ve ortak özellikler kullanılarak eğitilmiştir. Çevresel bağlam bilgisinin çıkarılması için görüntü verileri kullanılmıştır. Deneysel sonuçlar, DNN modelinin suç oluşumunu tahmin etmede diğer modellerden daha doğru sonuçlar verdiğini göstermiştir.

Kouziokas tarafından 2017 yılında yapılan çalışmada, yüksek suç riskine sahip olan toplu ulaşım alanlarını tahmin etmek için mekânsal kümeleme yöntemleri ve yapay sinir ağı modelleri kullanılmıştır [22]. Mekânsal analizler için coğrafi bilgi sistemleri kullanılmıştır. Artan sayıda suç olayına sahip alanları bulmak için mekânsal bir kümeleme tekniği olarak komşuluk analizi kullanılmıştır. Deneysel sonuçlar geliştirilen sistemin başarılı sonuçlar verdiğini göstermiştir.

Lin ve ark. Tarafından 2017 yılında yapılan çalışmada, hibrit bir sinir ağı tabanlı öğrenme yaklaşımı olan TreNet modeli sunulmuştur [23]. TreNet modeli, CNN ve LSTM modellerinin hibrit bir uygulamasıdır. TreNet özellik çıkarma aşamasında CNN’i, sıralı bağımlılıkları belirlemek için ise LSTM’i kullanmaktadır. TreNet modeli CNN, LSTM ve Gizli Markov Modeli yaklaşımları ile gerçek veri kümeleri üzerinde karşılaştırmalı olarak

test edilmiştir. Test sonuçları geliştirilen hibrit yöntemin daha başarılı sonuçlar verdiğini göstermiştir.

Yang ve ark. tarafından 2017 yılında yapılan çalışmada, uçuş güvenliğini sağlamak için uçak motor yağındaki metal içeriğinin değişiminin belirlenmesine yönelik ARIMA tabanlı bir model geliştirilmiştir [24]. ARIMA modeli, hareketli ortalamalar ve üstel düzeltme yöntemleri ile karşılaştırılmıştır. Deneysel çalışmalar, ARIMA modelinin tahmin sonucunun, hareketli ortalama ve üstel düzeltme yöntemlerinden daha başarılı sonuçlar verdiğini göstermiştir.

Zheng ve ark. tarafından 2017 yılında yapılan çalışmada, akıllı şebekelerdeki elektrik yükünün tahmin edilmesi için LSTM tabanlı bir sistem geliştirilmiştir [25]. Yapılan çalışmada Destek Vektör Makinesi (Support Vector Machine-SVM), Otoresif Sinir Ağı (Neural NETwork AutoRegression-NNETAR), LSTM, Doğrusal Olmayan Otoresif Eksojen (Nonlinear Autoregressive Exogenous-NARX) ve ARIMA modeli kullanılmıştır. Yapılan deneysel çalışmalar LSTM'in daha başarılı sonuçlar verdiğini göstermiştir.

Sobreiro ve ark. tarafından 2018 yılında yapılan çalışmada, şirketlerin sağladıkları faaliyetlerin planlanmasına yönelik bir sistem geliştirilmiştir [26]. Geliştirilen sistemde üstel düzeltme, ARIMA ve yapay sinir ağları yöntemleri kullanılmıştır. Veri kümesi olarak bir şirketin 192 aylık etkinlik verileri kullanılmıştır. Deneysel çalışmalar, yapay sinir ağları tabanlı modelin daha yüksek tahmin doğruluğuna sahip olduğunu göstermiştir.

McNally ve ark. tarafından 2018 yılında yapılan çalışmada, Amerikan Doları cinsinden bitcoin fiyatının tahmin edilmesine yönelik ARIMA, RNN ve LSTM yöntemlerinin karşılaştırmalı bir uygulaması sunulmuştur [27]. Geliştirilen sistemde veri kümesi olarak bitcoin fiyat endeksleri kullanılmıştır. Deneysel çalışmalar, grafik işlemci üzerindeki eğitim süresi ve yapılan tahminlerin doğruluğu açılarından karşılaştırmalı olarak yapılmıştır. Deneysel sonuçlar, başta LSTM olmak üzere doğrusal olmayan derin öğrenme yöntemlerinin ARIMA modelinden daha yüksek başarı oranına sahip olduğunu göstermiştir.

Wang ve ark. tarafından 2018 yılında yapılan çalışmada, akıllı ev kullanıcılarının davranışlarını tahmin etmeye yönelik ARIMA modelinin bir uygulaması sunulmuştur [28]. Geliştirilen sistemde, 12 haftalık bir süre boyunca kaydedilen su ısıtıcısının açılma süresi verileri kullanılarak ısıtıcının bir daha hangi zaman aralığında açılacağı belirlenmesi hedeflenmiştir. Çalışma kapsamında fonksiyonel katsayılı otoregresif bir model ile ARIMA modelinin geliştirilmesi ve bu sayede doğrusal olmayan özelliklerin ortaya çıkarılması hedeflenmiştir. Deneysel çalışmalar, geliştirilmiş ARIMA modelinin klasik ARIMA modeline göre daha iyi tahmin performansına sahip olduğunu göstermiştir.

Tian ve ark. tarafından 2018 yılında yapılan çalışmada, tahmin doğruluğunu iyileştirmek için CNN ve LSTM modellerini birleştiren bir model geliştirilmiştir [29]. Veri kümesi olarak Transsoency Platformunun sağladığı İtalya-Kuzey Bölgesi'ndeki elektrik yükü verileri kullanılmıştır. Geliştirilen model, Karar Ağacı, RF, Derin Enerji modeli, LSTM ve CNN ile karşılaştırılmalı olarak incelenmiştir. Deneysel sonuçlar MAE, Ortalama Mutlak Yüzde Hata (Mean Absolute Percentage Error-MAPE) ve RMSE ölçütleri açısından geliştirilen modelin daha başarılı sonuçlar verdiğini göstermiştir.

Stec ve Klabjan tarafından 2018 yılında yapılan çalışmada, suç tahminine yönelik derin öğrenme tabanlı bir sistem geliştirilmiştir [30]. Veri kümesi olarak hava durumu, nüfus verileri ve toplu taşıma verilerinin bulunduğu Chicago ve Portland suç verileri kullanılmıştır. Suç verileri 10 parçaya ayrılarak her bir bölge için günlük seviyede tahminler gerçekleştirilmiştir. Veri kümesi, suç tahmin probleminin mekânsal ve zamansal yönleri ele alınarak İleri Beslemeli Ağlar (Feed Forward Network-FFN), CNN ve RNN kullanılarak karşılaştırmalı olarak analiz edilmiştir. Deneysel çalışmalar RNN+CNN hibrit modelinin Chicago ve Portland veri kümelerinin her biri için CNN ve RNN'e göre daha başarılı sonuçlar verdiğini göstermiştir.

Sülo ve ark. tarafından 2019 yılında yapılan çalışmada, akıllı binalardaki enerji verimliliğini artırmaya yönelik LSTM tabanlı bir derin öğrenme modeli sunulmuştur [31]. Çalışmada New York Şehir Üniversitesinin 25 kampüsünde bulunan binaların elektrik tüketimleri analiz edilerek bu binaların gelecekteki elektrik tüketim değerlerinin tahmin edilmesi amaçlanmıştır. Kullanılan veri kümesi, kampüs ve kampüsteki binalara göre

sınıflandırılmıştır. Veriler sıcaklık, rüzgâr hızı, rüzgâr derecesi, nem, yağış, çiy noktası ve doluluk bilgisi içermektedir. Deneysel çalışmalar, veri kümesinde bulunan özniteliklerin etkilerinin gözlenmesi amacıyla aşamalı bir şekilde yapılmıştır. Deneysel sonuçlar, sıcaklık ve nemin enerji tüketim tahmininde başarı oranını arttırdığını, doluluğun ise tahmin doğruluğunu azalttığını göstermiştir.

Livieris ve ark. tarafından 2020 yılında yapılan çalışmada, altın fiyatının ve artış ya da azalış gibi hareketlerinin tahminine yönelik derin öğrenme tabanlı bir tahmin modeli sunulmuştur [32]. Geliştirilen modelde evrimsel katmanlar özellik çıkarımı aşamasında, LSTM ise öğrenme ile kısa ve uzun vadeli bağımlılıkları tanımlamak için kullanılmıştır. Geliştirilen modelde, geçmiş altın fiyatları analiz edilerek, bir sonraki gün için altın fiyatının tahmin edildiği bir regresyon işlemi yapılmaktadır. Ayrıca ertesi gün için altın fiyatının artacağı veya azalacağı tahmin edilmektedir. Deneysel sonuçlar, geliştirilen modelin tahmin performansını artırmada önemli bir başarı sağladığını göstermektedir.

Ma ve ark. tarafından 2020 yılında yapılan çalışmada, bilgisayar ağlarındaki trafik durumunun tahminine yönelik bir yaklaşım sunulmuştur [33]. Önerilen NN-ARIMA yaklaşımı, ARIMA ve Çok Katmanlı Algılayıcı (Multilayer Perceptron-MLP) modelleri kullanılarak geliştirilmiştir. MLP, trafik akışlarındaki hareket modellerinin belirlenmesi için kullanılmıştır. ARIMA ise konuma özgü trafik özelliklerinin çıkarılması için kullanılmıştır. Deney sonuçlar, önerilen modelin MSE ölçütüne göre trafik durumu tahmin doğruluğunu %13,4 oranında artırdığını göstermiştir.

Çalışmanın literatüre katkısı

Bu tez çalışmasında, CNN ve LSTM modellerinin öne çıkan özelliklerini etkili bir şekilde kullanarak tahmin doğruluğunu ve etkinliğini arttırmak için yeni bir Hibrit Derin Öğrenme Modeli (Hybrid Deep Learning Model-HDLM) geliştirilmiştir. Geliştirilen modelde CNN özellik çıkarımı için, LSTM ise öğrenme ve tahmin için kullanılmıştır. Geliştirilen hibrit model e-ticaret kullanıcı davranış veri kümesi, DDoS saldırı veri kümesi, suç veri kümesi ve acil çağrı veri kümesi olmak üzere 4 farklı veri kümesi üzerinde uygulanmıştır. HDLM,

RF, SVM, ARIMA, MLP, CNN, GRU, LSTM ve CNN ile GRU modelleri kullanılarak, bu çalışmada geliştirilen CNN+GRU hibrit modeli ile kapsamlı bir şekilde karşılaştırılmıştır.

Bu çalışmanın literatüre sunduğu temel katkılar şunlardır:

- CNN ve LSTM modellerinin öne çıkan özelliklerini kullanarak daha yüksek tahmin doğruluğuna sahip hibrit bir derin öğrenme modeli geliştirilmiştir. Geliştirilen modelde, CNN özellik çıkarımı için, LSTM ise öğrenme ve tahmin için kullanılmıştır.
- Bu çalışmada geliştirilen model, bir olayın bir sonraki olma zamanının belirlenmesinde literatürdeki benzerlerine göre belirgin oranda daha yüksek başarı düzeyine sahiptir. Geliştirilen model e-ticaret, DDoS, suç ve acil çağrı veri kümeleri üzerinde uygulanmıştır. Geliştirilen model kullanılarak DDoS, suç ve acil çağrı veri kümesi için olaylar meydana gelmeden önce engelleyici nitelikte önlemler alınması sağlanabilmektedir. E-ticaret veri kümesi için ise satın alma zamanı gibi kullanıcı aktivitelerinin zamanları tahmin edilerek kullanıcılara faydalı ve kişiselleştirilmiş öneriler sunulabilmekte ve tedarik yönetimi yapılabilmektedir.
- Geliştirilen model, literatürdeki RF, SVM, ARIMA, MLP, CNN, GRU, LSTM gibi mevcut yöntemler ve CNN ile GRU modelleri kullanılarak oluşturulan CNN+GRU hibrit modeli ile karşılaştırmalı olarak incelenmiştir. Deneysel sonuçlar, geliştirilen modelin uygulanan farklı veri kümeleri için daha başarılı sonuçlara sahip olduğunu göstermiştir.

2. ÖNERİ SİSTEMLERİ

Web teknolojilerindeki gelişmeler, kişilerin yaşama ve birbirleriyle iletişim kurma biçiminde çok sayıda değişikliği de beraberinde getirmiştir. Ancak bu değişiklikler, Web'de faydalı ve güvenilir bilgileri arayan kullanıcılar için bilgiye erişim maliyetini ve bilişsel yükü de beraberinde getirmiştir. Öneri sistemleri bu gibi sorunların üstesinden gelmek için kullanıcıların faydalı ve ilgili öğeleri bulmalarına yardımcı olmak üzere tercihlerini veya davranışlarını dikkate alarak öneriler sunmaktadır [34]. Bu bağlamda, kullanıcıların beğendikleri veya beğenmedikleri ürünler hakkında geribildirim sağlamaları en önemli unsurdur. Örneğin, Netflix gibi bir içerik sağlayıcısı senaryosu düşünüldüğünde, kullanıcılar basit tıklamalar ile kolayca dolaylı geribildirim sağlayabilir. Geribildirim almak için kullanılan temel yöntem, belirli bir değerlendirme sistemi (beş yıldızlı değerlendirme sistemi, beğendim/beğenmedim) kullanarak kullanıcılardan öğeler hakkındaki düşüncelerini doğrudan değerlendirmelerini beklemektir [35].

Dolaylı geribildirimler, doğrudan geribildirimler gibi açık değildir ancak Web platformlarında toplanması daha kolaydır [36]. Örneğin, bir kullanıcının bir öğeyi satın alması veya incelemesi, o öğenin beğenildiğine yönelik bir onay olarak kabul edilebilir. Bu tür geribildirim formları Amazon.com gibi çevrimiçi satıcılar tarafından yaygın olarak kullanılmaktadır ve bu tür verilerin toplanması kullanıcılar açısından zahmetsizdir [35].

Öneri sistemlerindeki temel fikir, çeşitli veri kaynaklarından kullanıcıların ilgi alanlarını çıkarabilmektir. Bu nedenle, öneri analizi genellikle kullanıcılar ve öğeler arasındaki geçmiş etkileşimlere dayanır [37]. Kullanıcıların geçmişteki etkileşimleri, ilgi alanları ve gelecekteki seçimlerinin muhtemel bir göstergesidir. Örneğin, tarihi bir belgesele ilgi duyan bir kullanıcının bir aksiyon filminden ziyade başka bir tarihi belgesel veya eğitim programına ilgi göstermesi daha olasıdır.

Birçok durumda, çeşitli öğe kategorileri, daha doğru önerilerde bulunmak için kullanılabilecek önemli ilişkiler gösterebilir. Alternatif olarak bağımlılıklar, kategoriler yerine tek tek öğeler bazında ve daha küçük ayrıntı düzeyinde mevcut olabilir. Bu bağımlılıklar değerlendirme matrisinden veriye dayalı bir şekilde öğrenilebilir ve ortaya

çıkan model hedef kullanıcılar için tahminler yapmak amacıyla kullanılır [38]. Bir kullanıcı için mevcut olan değerlendirilmiş öge sayısı ne kadar fazla olursa, kullanıcının gelecekteki davranışı hakkında doğru tahminler yapmak da o kadar kolay olur. Bu görevi yerine getirmek için birçok farklı öğrenme modeli kullanılabilir. Örneğin, çeşitli kullanıcıların toplu satın alma veya değerlendirme davranışları, aynı tür ürünlerle ilgilenen benzer kullanıcı grupları oluşturmak için kullanılabilir. Bu grupların ilgi alanları ve etkinlikleri grupların üyelerine bireysel önerilerde bulunmak için kullanılabilir [39].

İşbirlikçi filtreleme olarak adlandırılan bu öğrenme modeli, eksik kullanıcı değerlendirmelerini tahmin etmek için birden çok kullanıcıdan alınan değerlendirmelerin işbirliğine dayalı bir şekilde kullanılmasını ifade etmektedir [40]. Ayrıca öneri sistemleri çeşitli yardımcı veri türleriyle daha karmaşık ve veri bakımından zengin olabilir. Örneğin içerik tabanlı filtrelemede, öge içerikleri öneri sürecinde birincil rol oynar ve burada kullanıcı değerlendirmeleri ile öğelerin özellik açıklamaları tahminlerde bulunmak için kullanılır. Temel fikir, kullanıcıların ilgi alanlarının geçmişte değerlendirdikleri veya eriştikleri öğelerin özelliklerine göre modellenebilmesidir [41].

2.1. Öneri Sistemlerinin İşlevleri

Öneri sistemleri, kullanıcılara ilgilerini çekebilecek faydalı ve kullanışlı öge önerileri sunan yazılım araçları ve teknikleri olarak tanımlanmaktadır. Ancak öneri sistemlerinin kullanıcılar ve servis sağlayıcıları açısından farklı rolleri vardır. Örneğin, bir seyahat firması öneri sistemi kullanarak daha fazla otel odası satmayı ve cirosunu artırmayı hedeflemektedir. Öte yandan, kullanıcılar sisteme erişimlerinde ziyaret edecekleri bir bölge için konaklama alanları, etkinlikler ve turistik yerler bulmayı hedeflemektedir. İçerik sağlayıcılar ürün satışını arttırmak, satılan ürün çeşitliliğini arttırmak, kullanıcı memnuniyetini arttırmak, kullanıcı bağlılığını arttırmak ve kullanıcı beklentilerini daha iyi anlama açısından öneri sistemlerini kullanmaktadır [1].

Satılan ürün sayısını artırmak özellikle ticari öneri sistemlerinde en önemli işlevdir. Ticari olmayan uygulamalarda da, maliyet kaygısı olmasa dahi benzer hedeflere sahiptir. Örneğin, bir haber sitesi okunan haber öğelerinin sayısını artırmayı amaçlamaktadır. Genel olarak,

servis sağlayıcıları bakış açısından öneri sistemlerinin birincil hedefinin dönüşüm oranını yani öneriyi kabul eden ve kullanan kullanıcı sayısını artırmak olduğu söylenebilir [42].

Öneri sistemlerinin diğer bir önemli işlevi, kullanıcıların bir öneri olmadan bulmalarının zor olabileceği öğeleri seçmesini sağlamaktır [43]. Örneğin, Netflix gibi bir film öneri sisteminde servis sağlayıcı sadece en popüler içeriklerin değil arşivindeki tüm içeriklerin izlenmesiyle ilgilenir. Servis sağlayıcısı, belirli bir kullanıcının beğenisine uygun olmayan filmlerin reklamını yapma riskini öneri sistemi kullanmadan karşılayamayabilir. Kullanıcı memnuniyetini artırma, öneri sistemleri yardımıyla kullanıcıların Web sitesi veya uygulama ile ilgili deneyimlerinin geliştirilmesini ifade etmektedir. Sunulan ilginç ve ilgili öneriler ile etkin tasarlanmış kullanıcı arayüzleri, kullanıcıların sistem kullanımını ve önerilerin kabul edilme olasılığını artıracaktır [44].

Kullanıcı bağlılığını arttırma, bir kullanıcının ziyaret ettiğinde kendisini tanıyan ve değerli bir ziyaretçi olarak kabul eden bir Web sitesine bağlılık göstermesini ifade etmektedir [1]. Bu işlev, öneri sistemlerinin olağan bir özelliğidir. Çoğu öneri sistemi, kullanıcıların geçmiş etkileşimlerinden elde ettiği bilgileri kullanarak öneri oluşturur. Kullanıcıların uygulamayla olan etkileşim süresi uzadıkça kullanıcı modelinin ve sunulan önerilerin kullanıcı tercihlerine göre özelleştirilmesi sağlanır [45]. Kullanıcı beklentilerini daha iyi anlama ise öneri sistemlerinin kullanıcı tercihlerini öngörmesini ifade etmektedir. Bu bilgiler servis sağlayıcılar açısından ürün stoklarının veya üretim yönetimini iyileştirilmesi gibi bir dizi amaç için kullanılabilir [1].

Herlocker ve ark. tarafından yapılan çalışmada, kullanıcıların öneri sistemlerinden olan beklentileri en çok beğenilen öğelerin sunulması, kullanışlı öğelerin sunulması, öğe bağlamlarına dair açıklamaların sunulması, öneri listelerinin sunulması, öneri paketlerinin sunulması, güvenilir bir danışman bulunması, kendini ifade etme, diğer kullanıcılara yardımcı olmak ve diğer kullanıcıları etkilemek olarak gruplandırılmıştır [46]. En çok beğenilen öğelerin sunulması, diğer kullanıcılar tarafından değerlendirme ölçekleri kullanılarak en yüksek puan verilmiş öğelerin bulunmasını ifade etmektedir [42]. Ticari sistemlerin çoğu, kullanıcı tercihlerine göre belirli kategorilerdeki yüksek değerlendirme puanına sahip öğeleri ön plana çıkarmaktadır.

Kullanışlı öğelerin sunulması kullanıcıların ihtiyaç duyabilecekleri tüm öğelerin öneri olarak sunulmasını ifade etmektedir [43]. Bu gibi durumlarda sadece en çok beğenilen öğeleri bulmak yeterli değildir. Özellikle, öğe sayısının az olduğu uygulamalar veya tıp ve finans gibi kritik uygulamalar açısından önemlidir.

Öge bağlamlarına dair açıklamalar, kullanıcıların uzun vadeli tercihlerine bağlı olarak sunulan öneri listesindeki bazı öğeler ile ilgili ek açıklamaların sunulmasını ifade etmektedir. Örneğin bir film öneri sistemi, elektronik program rehberinde görüntülenen filmlerin izlenmeye değer olduğuna dair ek açıklamalar ekleyebilir [47]. Öneri listeleri, öneri sistemlerinin kullanıcılara bir bütün olarak beğenebilecekleri bir dizi öğeyi öneri olarak sunar. Televizyon dizilerinin önerilmesi ya da bir müzik listesinin öneri olarak sunulması örnek verilebilir [35]. Öneri gruplarının sunulması, kullanıcılara farklı türlerdeki öğeler hakkında bir öneri paketinin sunulmasını ifade etmektedir. Örneğin, seyahat planı yapan bir kullanıcıya belirli bir bölgede bulunan eğlence merkezleri, turistik alanlar ve konaklama servislerinden oluşan bir öneri paketi sunulabilir [1]. Kullanıcı açısından, sunulan çeşitli alternatifler tek bir seyahat hedefi olarak düşünülebilir ve seçilebilir.

Güvenilir bir danışman bulunması, bazı kullanıcıların öneri sistemlerine güvenmeyip sunulan önerilerde ne kadar iyi olduklarını görmek için uygulamayı test etmelerini ifade etmektedir. Bu nedenle, bazı sistemler kullanıcıların sadece öneri almaları için gerekenlere ek olarak davranışlarını test etmelerine de izin veren belirli işlevler sunabilir. Kendini ifade etme, kullanıcıların önerilerden ziyade değerlendirmelere katkıda bulunmak ya da görüş ve düşüncelerini açıklamak için uygulamaları kullanmalarını ifade etmektedir [48].

Diğer kullanıcılara yardımcı olmak bazı kullanıcıların, bir öğe hakkında araştırma yapan kişilerin öğe değerlendirmelerinden yararlandığını, bu nedenle öğeler hakkında değerlendirmelerde bulunmanın ya da yorum ve açıklamalar sunmanın diğer kullanıcılara faydalı olacağını düşünmelerini ifade etmektedir [49]. Örneğin, yeni bir araba satın almış olan bir kullanıcı, araba ile ilgili yapacağı değerlendirmenin diğer kullanıcılar için yararlı olacağını farkındadır. Diğer kullanıcıları etkilemenin temel amacının diğer kullanıcıları belirli ürünleri satın almaları konusunda yönlendirmektir.

2.2. Öneri Sistemlerinde Kullanılan Yöntemler

Öneri sistemlerinde kullanılan yöntemler temel olarak değerlendirme veya satın alma gibi kullanıcı-öğe etkileşimleri ile kullanıcılar ve öğeler hakkındaki özellik bilgilerini kullanmaktadır. Kullanıcı-öğe etkileşimlerini kullanan yöntemlere işbirlikçi filtreleme yöntemleri, öğeler hakkındaki özellik bilgilerini kullanan yöntemlere ise içerik tabanlı filtreleme yöntemleri adı verilir [51]. İçerik tabanlı filtreleme yöntemleri çoğu durumda değerlendirme matrislerini kullanır ancak tüm kullanıcılar yerine tek bir kullanıcının değerlendirmesine odaklanır [38]. Bilgi tabanlı öneri sistemlerinde ise öneriler, açıkça belirtilen kullanıcı tercihlerine dayanmaktadır. Geçmiş kullanıcı değerlendirmelerini kullanmak yerine, dış bilgi tabanları ve kısıtlamalar öneri oluşturmak için kullanılır [51]. Hibrit öneri sistemleri ise öneri sistemlerinin bu gibi farklı yönlerini birleştirir. Hibrit sistemler, daha başarılı performans gösterebilecek teknikler oluşturmak için çeşitli türlerdeki öneri sistemlerinin güçlü yönlerini birleştirebilir [35].

2.2.1. İşbirlikçi filtreleme yöntemi

İşbirlikçi filtreleme yöntemi, önerilerde bulunmak için birden fazla kullanıcı tarafından sağlanan değerlendirmeleri kullanır. İşbirlikçi filtreleme yöntemlerinin tasarlanmasındaki temel zorluk temel değerlendirme matrislerinin seyrek olmasıdır [52]. Örneğin, bir film izleme uygulaması senaryosu düşünüldüğünde, çoğu kullanıcı mevcut filmlerin yalnızca küçük bir kısmını değerlendirecektir ve sonuç olarak öğelerin çok büyük bir bölümü için değerlendirme yapılmayacaktır. Kullanıcılar tarafından yapılan değerlendirmeler gözlemlenen değerlendirmeler olarak da adlandırılır. İşbirlikçi filtreleme yöntemlerinin temel fikri, eksik değerlendirmelerin gözlemlenen değerlendirmeler kullanılarak tahmin edilmesidir [37]. Örneğin, A ve B adında benzer zevkleri olan iki kullanıcı düşünüldüğünde, her ikisinin de gözlemlenen değerlendirmeleri benzerse, benzerlikleri bir algoritma ile tanımlanabilir. Bu benzerlik, eksik değerlendirmeler hakkında çıkarımlar yapmak için kullanılabilir.

İşbirlikçi filtreleme yöntemlerinin çoğu tahmin süreci için öğeler arası ilişkilerden ya da kullanıcılar arası ilişkilerden yararlanmaya odaklanmaktadır [53]. Ayrıca, bazı işbirlikçi

filtreleme yöntemleri sınıflandırıcıların etiketli verilerden eğitim modeli oluşturduğu gibi bir eğitim modeli oluşturmak için optimizasyon tekniklerini kullanır. Bu model daha sonra, değerlendirme matrisindeki eksik değerleri belirlemek için kullanılır. İşbirlikçi filtrelemede bellek tabanlı yöntemler ve model tabanlı yöntemler yaygın olarak kullanılmaktadır [3].

Bellek tabanlı yöntemler aynı zamanda komşuluk tabanlı işbirlikçi filtreleme algoritmaları olarak da adlandırılır. Bu yöntemler, kullanıcı-öge kombinasyonlarına ait değerlendirmelerin komşuluklara göre tahmin edilmesi temeline dayanmaktadır [54]. Bu yöntemler kullanıcı tabanlı işbirlikçi filtreleme ve öge tabanlı işbirlikçi filtreleme yöntemleri olarak tanımlanmaktadır.

Kullanıcı tabanlı işbirlikçi filtreleme yönteminde, hedef kullanıcıya benzer seçimlerde bulunan kullanıcılar tarafından yapılan değerlendirmeler, hedef kullanıcıya önerilerde bulunmak için kullanılabilir [55]. Temel fikir, hedef kullanıcıya benzer seçimleri olan kullanıcıları belirlemek ve bu kullanıcılardan oluşan grubun değerlendirmelerinin ağırlıklı ortalamalarını hesaplayarak değerlendirmeleri gözlemlenmeyen öğeleri hedef kullanıcıya öneri olarak sunmaktır. Örneğin, Birol ve Melis geçmişte aynı filmlere benzer değerlendirmede bulundularsa, Birol'un X filmi için gözlemlenmemiş değerlendirmelerini tahmin etmek için Melis'in X filmi için gözlemlenen değerlendirmesi kullanılabilir. Genel olarak, Birol'a en çok benzeyen kullanıcılar Birol için değerlendirme tahminleri yapmak amacıyla kullanılabilir. Benzer kullanıcıları bulmak için değerlendirme matrisinin satırları arasında benzerlik işlevleri hesaplanır [3].

Öge tabanlı işbirlikçi filtreleme yönteminde ise A kullanıcısının B ögesi için değerlendirmesini tahmin etmek için hedef öge B'ye en çok benzeyen bir öge kümesi belirlenmektedir. Öge kümesindeki değerlendirmeler, A kullanıcısının B ögesini beğenip beğenmeyeceğini tahmin etmek için kullanılır [56]. Örneğin, Alien ve Predator gibi benzer bilim kurgu filmleri için yapılan değerlendirme puanları, Terminator filminin değerlendirme puanını tahmin etmek için kullanılabilir. Benzer öğeleri bulmak için değerlendirme matrisinin sütunları arasında benzerlik işlevleri hesaplanır. Bellek tabanlı yöntemlerin avantajı, uygulanmasının basit olması ve ortaya çıkan önerilerin nedenselliğinin açıklanmasının kolay olmasıdır. Ancak, bellek tabanlı algoritmalar seyrek değerlendirme

matrisleri ile çok iyi çalışmaz. Örneğin, Gladiator'u değerlendiren Birol'a yeterince benzer kullanıcılar bulmak zor olabilir. Bu gibi durumlarda, Birol'un Gladyatör değerlendirmesini güçlü bir şekilde tahmin etmek zordur [3].

Model tabanlı yöntemlerde, makine öğrenmesi ve veri madenciliği yöntemleri kullanılmaktadır. Model parametresi kullanıldığı durumlarda, bu modelin parametreleri bir optimizasyon çerçevesi kapsamında öğrenilir. Model tabanlı işbirlikçi filtreleme algoritmaları olasılık tabanlı yaklaşım kullanmaktadır [1]. Bir kullanıcı için değerlendirme tahmini hesaplarken kullanıcının diğer öğeler üzerindeki değerlendirmelerini kullanır. Bu tür model tabanlı yöntemlere karar ağaçları ve kural tabanlı modeller gibi makine öğrenmesi yöntemleri ve derin öğrenme yöntemleri örnek olarak verilebilir.

İşbirlikçi filtreleme yöntemleri, sınıflandırma ve regresyon modellerinin genel bir hali olarak görülebilir. Sınıflandırma ve regresyon problemlerindeki bağımlı değişken, eksik değerlere sahip bir nitelik olarak görülebilir. İşbirlikçi filtreleme herhangi bir sütunun eksik değerlendirme sonuçlarına sahip olması nedeniyle bu çerçevenin genel bir hali olarak görülebilir [35]. Öneri probleminde, her özellik bağımlı ve bağımsız değişken rolü oynadığı için sınıf değişkenleri ve özellik değişkenleri arasında açık bir ayrım yoktur. Bu ayrım, yalnızca eksik girişler özel bir sütunla sınırlı olduğu için sınıflandırma probleminde bulunur. Ayrıca, işbirlikçi filtrelemede herhangi bir satır eksik girişler içerebileceği için eğitim ve test satırları gibi bir ayrım yoktur. Bu nedenle, işbirlikçi filtrelemede eğitim ve test satırlarından ziyade eğitim ve test girişleri kullanılmaktadır. İşbirlikçi filtreleme tahminin girdi olarak yapıldığı sınıflandırma/regresyon modellerinin genelleştirilmiş bir halidir.

Kullanıcı tabanlı işbirlikçi filtreleme

Bu yaklaşımda, değerlendirme tahminlerinin hesaplanacağı hedef kullanıcıya benzer kullanıcıları belirlemek için kullanıcı tabanlı komşuluklar tanımlanır. Hedef kullanıcının komşuluklarını belirlemek için diğer tüm kullanıcılarla olan benzerliği hesaplanır [37]. Bu nedenle kullanıcılar tarafından belirtilen değerlendirmeler arasında bir benzerlik işlevinin tanımlanması gerekir [55]. Farklı kullanıcılar farklı değerlendirme ölçeklerine sahip olabileceğinden böyle bir benzerliğin hesaplanması zor olabilir. Bir kullanıcı çoğu öğeyi

beğenmeye eğilimli olabilirken, başka bir kullanıcı öğelerin çoğunu beğenmemeye eğilimli olabilir. Ayrıca, farklı kullanıcılar farklı öğeleri değerlendirmiş olabilir.

$m \times n$ boyutundaki $R = |r_{u,j}|$ kullanıcı-öge matrisinde m kullanıcıları, n ise öğeleri ifade etmektedir. I_u ise u kullanıcısı tarafından yapılan değerlendirmelerin bulunduğu öge kümesini ifade etmektedir. Örneğin, u kullanıcısı tarafından 1., 3. ve 5. öğeler değerlendirilmiş ve geri kalan öğeler eksikse, $I_u = \{1, 3, 5\}$ olacaktır. Benzer şekilde, hem u hem de v kullanıcıları tarafından değerlendirilen öge kümesi $I_u \cap I_v$ şeklinde elde edilmektedir. Örneğin, v kullanıcısı ilk dört öğeyi değerlendirdiyse, $I_v = \{1, 2, 3, 4\}$ ve $I_u \cap I_v = \{1, 3, 5\} \cap \{1, 2, 3, 4\} = \{1, 3\}$ olacaktır. Değerlendirme matrisleri genellikle seyrek olduğu için $I_u \cap I_v$ 'nin boş küme olması mümkündür. $I_u \cap I_v$ kümesi, u ve v kullanıcıları arasındaki benzerliği hesaplamak için kullanılan karşılıklı gözlenen değerlendirmeleri tanımlar.

u ve v gibi iki kullanıcının değerlendirme puanları arasındaki benzerlik Pearson korelasyon katsayısı ile hesaplanabilmektedir. İlk olarak, her bir u kullanıcısının değerlendirme puanlarını kullanarak ortalama değerlendirme puanı olan μ_u değeri hesaplanmaktadır. u ve v kullanıcıları arasındaki Pearson korelasyon katsayısı Eş.2.1'de görüldüğü gibi hesaplanmaktadır.

$$Pearson(u, v) = \frac{\sum_{k \in I_u \cap I_v} (r_{uk} - \mu_u) \cdot (r_{vk} - \mu_v)}{\sqrt{\sum_{k \in I_u \cap I_v} (r_{uk} - \mu_u)^2} \cdot \sqrt{\sum_{k \in I_u \cap I_v} (r_{vk} - \mu_v)^2}} \quad (2.1)$$

I_u , u kullanıcısı tarafından yapılan değerlendirmelerin bulunduğu öge kümesini, I_v ise v kullanıcısı tarafından yapılan değerlendirmelerin bulunduğu öge kümesini ifade etmektedir.

r_{uk} , u kullanıcısının k ögesi hakkında verdiği değerlendirme puanını, r_{vk} ise v kullanıcısının k ögesi hakkında verdiği değerlendirme puanını ifade etmektedir. μ_u ve μ_v ise u ve v kullanıcılarının ortalama değerlendirme puanını ifade etmektedir [35].

Pearson katsayısı hedef kullanıcı ile diğer tüm kullanıcılar arasında hesaplanır. Hedef kullanıcının komşuluklarını belirlemek için hedef kullanıcı ile en yüksek Pearson katsayısına sahip k kullanıcı kümesi kullanılacaktır. Ancak, hedef kullanıcının komşulukları öğelere göre değişebilmektedir. Bu nedenle en yakın k kullanıcı, hedef kullanıcı için öğelere göre ayrı ayrı bulunur. Bu sayede, sadece mevcut öğe için değerlendirmede bulunan k kullanıcı belirlenmiş olur. Bu değerlendirmelerin ağırlıklı ortalaması mevcut öğe için tahmin değeri olarak hesaplanır [1]. Kullanıcı tabanlı işbirlikçi filtreleme yaklaşımındaki temel sorun, farklı kullanıcıların farklı ölçekler kullanarak değerlendirmeler sunabilmeleridir. Farklı ölçekler, bir kullanıcının tüm öğeleri yüksek oranda değerlendirirken, başka bir kullanıcının ise olumsuz olarak değerlendirmesine neden olabilmektedir [35].

Öge tabanlı işbirlikçi filtreleme

Kullanıcı tabanlı işbirlikçi filtreleme yaklaşımları farklı uygulama alanlarında başarıyla kullanılıyor olsa da, milyonlarca kullanıcı ve ürünün bulunduğu büyük e-ticaret platformları söz konusu olduğunda bazı zorluklar yaşanabilmektedir. Özellikle, çok sayıda potansiyel komşuluğun taranması ihtiyacı, tahminlerin gerçek zamanlı olarak hesaplanmasını zorlaştırmaktadır. Bu nedenle, büyük ölçekli e-ticaret platformları genellikle çevrimdışı ön işleme için daha uygun olan ve böylece çok daha büyük bir değerlendirme matrisi için bile gerçek zamanlı hesaplamaların yapılmasına izin veren öge tabanlı filtreleme yöntemini kullanır [56]. Öge tabanlı algoritmaların ana fikri, kullanıcılar arasındaki benzerliği değil, öğeler arasındaki benzerliği kullanarak tahminleri hesaplamaktır [2]. Çizelge 2.1’de kullanıcı değerlendirme matrisi görülmektedir.

Çizelge 2.1. Kullanıcı derecelendirme matrisi

Kullanıcı	Öge 1	Öge 2	Öge 3	Öge 4	Öge 5
Melis	5	3	4	4	?
Kullanıcı 1	3	1	2	3	3
Kullanıcı 2	4	3	4	3	5
Kullanıcı 3	3	3	1	5	4
Kullanıcı 4	1	5	5	2	1

Çizelge 2.1’de sunulan kullanıcı değerlendirme matrisi kullanılarak, Melis kullanıcısı ve Öğe 5 için derecelendirme puanı tahmini yapmak için öncelikle diğer öğelerin değerlendirme vektörleri karşılaştırılır ve Öğe 5’e benzer değerlendirmelere sahip öğeler belirlenir. Çizelge 2.1’de sunulan değerlendirme matrisinde Öğe 5’in (3, 5, 4, 1), Öğe 1 (3, 4, 3, 1) ve Öğe 4 (3, 3, 5, 2) ile benzer değerlendirme puanına sahip olduğu görülmektedir. Komşu öğeler belirlendikten sonra hedef kullanıcının komşu öğeler hakkındaki değerlendirme puanları belirlenir. Melis kullanıcısı, Öğe 1’e 5 ve Öğe 4’e 4 değerlendirme puanı vermiştir. Öğe tabanlı işbirlikçi filtreleme algoritması diğer değerlendirmelerin ağırlıklı ortalamasını hesaplayarak Öğe 5 için 4 ile 5 arasında bir değerlendirme puanı tahmin edecektir [35].

Öğe tabanlı işbirlikçi filtrelemede, komşuluk grupları kullanıcılar yerine öğelerden oluşturulur. Bu nedenle öğeler arasında benzerliklerin hesaplanması gerekir. Öğe tabanlı işbirlikçi filtreleme yaklaşımlarında kosinüs benzerliği en doğru sonuçları verdiği için standart ölçüt olarak belirlenmiştir. Kosinüs benzerliği, n boyutlu iki vektör arasındaki benzerliği aralarındaki açıya göre ölçmektedir. i ve j gibi iki öğe arasındaki benzerlik, değerlendirme vektörleri kullanılarak Eş. 2.2 kullanılarak hesaplanmaktadır.

$$Cosine(i, j) = \frac{\sum_{u \in U_i \cap U_j} (s_{ui} \cdot s_{uj})}{\sqrt{\sum_{u \in U_i \cap U_j} s_{ui}^2} \cdot \sqrt{\sum_{u \in U_i \cap U_j} s_{uj}^2}} \quad (2.2)$$

U_i , i öğesi için değerlendirmede bulunan kullanıcılar kümesini ve U_j ise j öğesi için değerlendirilmede bulunan kullanıcılar kümesini ifade etmektedir. s_{ui} , u kullanıcısının i öğesi için yaptığı değerlendirme puanını, s_{uj} ise u kullanıcısının j öğesi için yaptığı değerlendirme puanını ifade etmektedir [3].

2.2.2. İçerik tabanlı filtreleme yöntemi

İşbirlikçi filtreleme yöntemleri öneri hesaplamak için kullanıcılar arasındaki değerlendirme

ilişkilerini kullanır. Ancak bu yöntemler tahminleri hesaplamak için öge özniteliklerini kullanmaz [42]. İçerik tabanlı filtreleme yöntemleri öğelerin açıklayıcı nitelik kümeleriyle tanımlanabileceği senaryolardan yararlanmak için tasarlanmıştır. Burada diğer kullanıcıların değerlendirmelerinin önemi yoktur. Öneri sunmak için kullanıcının ilgi alanları ve etkileşimleri kullanılmaktadır. Bu yaklaşım, özellikle değerlendirmenin henüz yapılmadığı yeni öğeler için kullanışlıdır [37].

İçerik tabanlı sistemler temel olarak öğeler hakkındaki açıklayıcı bilgilere ve öğeler ile ilgili kullanıcı geribildirimlerinden üretilen kullanıcı profillerine ihtiyaç duymaktadır. Kullanıcı geribildirimleri doğrudan veya dolaylı olarak elde edilebilir. Doğrudan geribildirimler öge değerlendirmelerine karşılık gelirken, dolaylı geribildirimler kullanıcı eylemlerine karşılık gelebilir. Kullanıcı profilleri, öğelerin özelliklerini kullanıcıların ilgi alanlarıyla ilişkilendirir [52]. Öğeler için katalog bilgilerini tanımlamanın en basit yolu, her öge için öge profilleri olarak adlandırılan bir özellik listesi sunmaktır. Örneğin bir kitap öneri sisteminde kitabın türü, yazarı, yayıncısı veya kitabı tanımlayan herhangi bir bilgi kullanılabilir ve bu bilgiler ilişkisel bir veritabanında depolanabilir. Kullanıcı tercihleri, bu özellik kümeleri kullanılarak ilgi alanları bağlamında açıklandığında öneri görevi, öge özellikleri ile kullanıcı tercihlerinin eşleştirilmesinden oluşur [41].

İçerik tabanlı sistemler öneri listeleri oluştururken henüz görülmemiş bir ögenin, hedef kullanıcının geçmişte beğendiği öğelere ne kadar benzer olduğunu değerlendirerek çalışır. Buradaki benzerlik farklı şekillerde ölçülebilir [38]. Örneğin kullanıcının daha önce görmediği bir kitap düşünüldüğünde, sistem kitabın türünü hedef kullanıcının tercih ettiği türler listesinde olup olmadığını kontrol edebilir. Bu durumda benzerlik 0 veya 1'dir. Diğer bir benzerlik hesaplama yöntemi ise ilgili anahtar kelimelerin benzerliğini veya çakışmasını hesaplamaktır.

İçerik tabanlı sistemler, büyük ölçüde özellik bilgisinin mevcut olduğu senaryolarda kullanılır. Çoğu durumda, bu özellikler ürün açıklamalarından çıkarılan anahtar kelimelerdir. Bu nedenle, içerik tabanlı sistemler özellikle metin açısından zengin ve yapılandırılmamış alanlarda öneriler sunmak için uygundur. Bu tür sistemlerin kullanımına klasik bir örnek Web sayfası önerisidir. Örneğin, içerik tabanlı bir öneri sistemi oluşturmak

için bir kullanıcının tarama davranışı kullanılabilir [1].

İçerik tabanlı sistemlerin ana bileşenleri arasında çevrimdışı ön işleme bölümü, çevrimdışı öğrenme bölümü ve çevrimiçi tahmin bölümü yer alır. Çevrimdışı bölümler genellikle bir sınıflandırma veya regresyon modeli oluşturmak için kullanılır. Bu model daha sonra kullanıcılar için çevrimiçi öneriler hesaplamak amacıyla kullanılır [3].

Önişleme ve özellik çıkarma

İçerik tabanlı sistemler, Web sayfaları, ürün açıklamaları, haberler ve müzik özellikleri gibi çeşitli alanlarda kullanılır. Çoğu durumda, bu gibi çeşitli kaynaklardan anahtar kelime tabanlı vektör-uzay gösterimine dönüştürülmek üzere özellikler çıkarılır. Bu işlem, içerik tabanlı bir öneri sisteminin ilk adımıdır ve yüksek düzeyde alana özgüdür. Bilgilendirici özelliklerin doğru bir şekilde çıkarılması içerik tabanlı bir öneri sisteminin etkili bir şekilde çalışması için gereklidir.

Kullanıcı profillerinin öğrenilmesi

İçerik tabanlı filtreleme yöntemleri belirli bir kullanıcıya yöneliktir. Bu nedenle, kullanıcının satın alma geçmişine veya değerlendirme geçmişine dayanarak kullanıcının ilgi alanlarını tahmin etmek için kullanıcıya özgü bir model oluşturulur. Bu amaca ulaşmak için daha önce yapılan değerlendirmeler (doğrudan geribildirim) veya kullanıcı etkinlikleri (dolaylı geribildirim) şeklinde elde edilen geribildirimler kullanılır. Bu tür geribildirimler, eğitim verileri oluşturmak için öğelerin nitelikleri ile birlikte kullanılır. Bu sayede eğitim verisi üzerinde bir öğrenme modeli oluşturulur. Bu aşama, geribildirimlerin kategorik olup olmadığına (örneğin bir öğeyi seçme ya da seçmeme) veya geribildirimlerin sayısal olup olmadığına (örneğin derecelendirme puanları) bağlı olarak genellikle sınıflandırma veya regresyon modellemesine benzerdir. Ortaya çıkan model, kullanıcının ilgi alanlarını kavramsal olarak öğe nitelikleriyle ilişkilendirdiği için kullanıcı profili olarak adlandırılır.

Filtreleme ve öneri

Bu adımda bir önceki adımdan öğrenilen model kullanıcılar için gerçek zamanlı önerilerde

bulunmak amacıyla kullanılır.

2.3. Öneri Sistemlerinde Kullanılan Veri Toplama Yöntemleri

Öneri algoritmalarının tasarımı, kullanıcı değerlendirmelerini izlemek için kullanılan sistemlerden etkilenir. Değerlendirmeler genellikle, mevcut öge için beğenme veya beğenmeme düzeyini gösteren bir ölçekle belirtilir [57]. Beğeni seviyesini ölçmek için sıralı numaralar kümesinin kullanıldığı aralık tabanlı ölçekler yaygın olarak kullanılmaktadır. Örneğin, -2 çok kötü ve 2 ise çok iyi olmak üzere $\{-2, -1, 0, 1, 2\}$ kümesi kullanılarak 5 puanlı bir değerlendirme ölçeği elde edilebilir. Olası değerlendirmelerin sayısı kullanılan sisteme göre değişebilir. 5 puan, 7 puan ve 10 puan değerlendirmelerinin kullanımı özellikle yaygındır. Benzer şekilde $\{\text{Kesinlikle Katılmıyorum, Katılmıyorum, Kararsızım, Katılıyorum, Kesinlikle Katılıyorum}\}$ gibi sıralı kategorik değerler de kullanılabilir. Bu değerlendirmeler sıralı değerlendirmeler olarak adlandırılır ve sıralı nitelikler kavramından türetilir [35].

İkili değerlendirmelerde ise kullanıcılar öğeler için yalnızca beğenme veya beğenmeme durumunu ifade edebilir. Burada değerlendirmeler 0 ve 1 gibi değerler ile ifade edilebilir. Özel bir değerlendirme durumu ise kullanıcının yalnızca beğeni durumunu belirtebildiği, beğenmeme durumunu belirtmesi için bir mekanizmanın olmadığı tekli değerlendirmelerdir [58]. Tekli değerlendirmeler özellikle dolaylı geribildirimlerde kullanılmaktadır. Bu durumlarda, kullanıcı tercihleri doğrudan kullanıcıya sormak yerine kullanıcı faaliyetlerinden elde edilir. Örneğin, bir müşterinin satın alma davranışı tekli değerlendirmelere dönüştürülebilir. Bir müşteri bir ürün satın aldığı anda bu ürün için bir beğeni olarak görülebilir. Ancak, geniş bir ürün veritabanından bir ürünün satın alınmaması eylemi her zaman memnuniyetsizliği göstermez.

Doğrudan geribildirim kullanılan sistemlerde bir kullanıcı arayüzü aracılığıyla kullanıcılardan sistemde bulunan öğeler hakkında değerlendirmelerde bulunmaları beklenmektedir. Bu sistemlerde sunulan önerilerin doğruluğu kullanıcı tarafından yapılan değerlendirmelerin sayısına bağlıdır. Bu yöntemin eksikliği ise kullanıcı çabası gerektirmesi ve kullanıcıların daima yeterli bilgi vermeye hazır olmamasıdır. Doğrudan geribildirimlerin

daha fazla kullanıcı çabası gerektirmesi gerçeğine rağmen kullanıcı davranışlarından çıkarımlar yapılması gerekli olmadığından elde edilen veriler daha güvenilir olarak görülmekte ve öneri sürecinde şeffaflık sağlanmaktadır [59].

Dolaylı geribildirimler kullanıcıların satın alma geçmişi, gezinme geçmişi, Web sayfalarında kalınan süreler, kullanıcı tarafından tıklanan bağlantılar ve kullanıcı arayüzü üzerindeki tıklamalar gibi farklı kullanıcı eylemlerinin gözlemlenmesi yoluyla elde edilmektedir. Dolaylı geribildirimler kullanıcı tercihlerini kullanıcıların sistemle olan etkileşimlerinden çıkararak kullanıcı iş yükünü azaltmaktadır. Dolaylı geribildirimler, herhangi bir kullanıcı çabası gerektirmemesi ve doğrudan kullanıcı davranışlarının analiz edilmesi yoluyla elde edildiği için daha objektif bir yaklaşım olarak ön plana çıkmaktadır [60].

3. AKIŞ VERİSİNDE GELENEKSEL ÖNERİ SİSTEMLERİ

Gerçek zamanlı gözetim sistemleri, iletişim ağları, İnternet trafiği, finansal piyasalar ve e-ticaret alanlarındaki çevrimiçi işlemler endüstri üretim süreçleri ve algılayıcılar gibi dinamik ortamlar tarafından potansiyel olarak sonsuz miktarda veri üretilmektedir. Büyük miktardaki bu verilerden faydalı bilgilerin nasıl çıkarılacağı giderek önem kazanan bir çalışma konusudur [61]. Öneri sistemleri, kullanıcılara veri madenciliği tekniklerini kullanarak kişiselleştirilmiş servis sunan, bu sayede aşırı bilgi yükü sorununun çözülmesine yardımcı olan uygulamalardır.

Öneri sistemlerinin, akış verilerinin düzensizlik, süreklilik ve yüksek hız gibi karakteristik özellikleri ile başa çıkacak bir yapıda tasarlanması gereklidir. Gerçek zamanlı öneri sistemlerinde gerçek zamanlı güncelleme, büyük veri boyutları ve kavram kayması konuları ön plana çıkmaktadır. Gerçek zamanlı güncelleme akış verilerinin yüksek hızlarına uyum sağlayarak kullanıcı beklentilerine göre gerçek zamanlı model güncellemesi yapılmasını ifade etmektedir [62]. Büyük veri boyutları sistemlere sürekli yeni kullanıcılar ve yeni öğeler eklenmesini ifade etmektedir. Bu nedenle kullanıcı sayısı ve öneri listelerinin boyutu önceden bilinmemektedir. Kavram kayması ise kullanıcı tercihlerinde ya da ürünlerin tercih edilme oranlarındaki değişikliği ifade etmektedir. Örneğin, bir ürünün yeni bir sürümünün piyasaya sürülmesi önceki sürümlerin popülaritesini azaltır. Akış verilerine yönelik geliştirilecek öneri sistemlerinin, akış verilerinin doğası gereği statik veri kümelerine yönelik tasarlanmış geleneksel algoritmalarından farklı bir şekilde tasarlanması gereklidir.

Chen ve ark. tarafından 2013 yılında yapılan çalışmada, çevrimiçi puanlama yaklaşımları genişletilerek TeRec adı verilen zamana dayalı bir tavsiye sistemi önerilmiştir [63]. TeRec'te, kullanıcılar tweet gönderirken gerçek zamanlı olarak ilgi alanlarına göre hashtag önerileri alabilmekte ve öneriler için geribildirimler sunabilmektedir. TeRec, kullanıcıların gerçek zamanlı konu önerilerine erişmesini sağlayan tarayıcı tabanlı istemci arabirimi sağlamak ve sunucu tarafında gerçek zamanlı akış verilerini işleyip saklamaktadır. TeRec, Weibo üzerinde çalışarak kullanıcılara herhangi bir anda tercihlerine göre gerçek zamanlı tavsiyeler sağlamaktadır. TeRec kullanıcıları ve öğeleri daha doğru sonuçlar elde edebilmek için matris çarpanlarına ayırma kullanarak modellemektedir. TeRec'in temel fikri

hashtag'leri ilgi çekici konuların vekilleri olarak kullanmaktır. Kullanıcı bir tweet göndermek üzereyken sistem kullanıcının mevcut ilgi alanlarını öngörmekte ve bu tweet'te kullanmak isteyebileceği çeşitli konuları önermektedir.

TeRec, kullanıcılara tweet gönderdiklerinde kendilerine uygun hashtag'leri seçmelerine yardımcı olan tarayıcı tabanlı bir servis sunmaktadır. Sistemin sunucu tarafı kullanıcı tercihlerini modelleyip depolamakta, kullanıcılar ve hashtag'ler arasında değerlendirme puanı tahminlerini hesaplamakta ve kullanıcılar tweet attıklarında hashtag öneri listesinin oluşturulmasını sağlamaktadır. Geliştirilen sistem üç katmandan oluşmaktadır. Birinci katman, öneri sonuçlarının gösterilmesi ve kullanıcı geribildirimlerinin alınması gibi kullanıcılar ve veriler arasındaki etkileşimlerin gerçekleştirildiği kullanıcı arayüzüdür. İkinci katman olan depolama katmanında kullanıcı tercihleri ve öğe özelliklerinden oluşan bir matris tutulmaktadır. Üçüncü katman ise, depolama katmanı ve kullanıcı arayüzü arasında kalan önerilerin oluşturulduğu ve modelin güncellendiği katmandır. TeRec'in çalışma adımları şu şekildedir: 1. adımda kullanıcı arayüzü, kullanıcılardan gelen talepleri alarak tavsiyede bulunulmasını istemektedir. 2. adımda tavsiye modeli değerlendirme puanlarını hesaplayarak öneri listesini kullanıcı arayüzüne döndürmektedir. 3. adımda kullanıcı arayüzü, kullanıcı geribildirimini alarak öneri modelini güncellemektedir. 4. adımda tavsiye modeli, geçmiş girdilerin veri deposunu güncellemektedir. 5. adımda tavsiye modeli güncellenmiş rezervuarı kullanarak matrisi güncellemektedir.

Verilerin çok büyük olabileceği ve çoğunun yararsız olduğu düşünüldüğünde, model güncellemelerini hızlandırmak için bilgilendirici girdilerin bir alt kümesinin örneklenmesi gerekmektedir. Bu çalışmada, amaçla rezervuar örnekleme tekniği kullanılmıştır. Veri girdilerinin boyutu c sayısına ulaştığında t . sıradaki veri c/t olasılıkla korunacaktır. Çalışma kapsamında sunulan rezervuar örnekleme mekanizmasında ise t . sıradaki veri $1-c/t$ olasılıkla korunacaktır. Yeni verilerin rezervuara konulmasına karar verilirse, rezervuarda bulunan s_i veri örneğinin yer değiştirme olasılığı $1 - P(s_i \in R_{t-1})$, Eş. 3.1 kullanılarak hesaplanmaktadır.

$$P(s_i \in R_{t-1}) \propto \exp \frac{1}{t-i} \quad (3.1)$$

$1 - P(s_i \in R_{t-1})$, zaman serisi analizinde yaygın olarak kullanılan üstel bozunma fonksiyonunu, $t - i$ mevcut zaman sırası t ile sisteme gelen veri örneğinin zaman sırası i arasındaki farkı ifade etmektedir. Geliştirilen sistemde kullanıcıların son tercihlerinin örneklerini saklayan rezervuar, yeni gelen her veri ile dinamik olarak güncellenmekte ve önerilerini her zaman en son güncellenen modele göre üretmektedir. Analiz sonuçları, TeRec sistemimizin tweet akışları için gerçek zamanlı hashtag önerisi sunma konusunda daha başarılı sonuçlar verdiğini göstermiştir.

Werner ve Lommatzsch tarafından 2015 yılında yapılan çalışmada, gerçek zamanlı haber tavsiyesi sunmak amacıyla sınırlı donanım kaynakları ve zaman kısıtlamaları konularında optimize edilmiş bir yöntem önerilmiştir [64]. PLISTA adı verilen sistem ile öneri taleplerine 100 ms içerisinde yanıt verilerek hızlı ve verimli öneriler sunmak hedeflenmiştir. Geliştirilen sistemde bir kullanıcı bir haber portalını ziyaret ettiğinde portal PLISTA sunucusuna bir öneri talebi göndermektedir. PLISTA sunucusu isteği rastgele seçilen bir tavsiye algoritmasına devretmektedir. İsteğe cevap verilmesi için seçilen tavsiye algoritmasının 100 ms içinde bir öneri listesi sunması gerekmektedir. Vagrant4 adı verilen çevrimdışı test sunucusu ile geçmişte oturum açmış olan kullanıcıların etkileşim verileri kullanılarak önerilen algoritmaların performansı analiz edilmektedir. Geliştirilen sistemin sınırlamalarından biri kullanıcılar haber portallarında kullanıcı girişi yapmadıkları için benzersiz kullanıcıların belirlenmesi sorunudur. Bu durum, kullanıcıya dayalı yöntemlerin kişiselleştirilmiş önerileri hesaplamak için kapsamlı veriler elde etmesini zorlaştırmaktadır.

Geliştirilen sistemde, haber makalelerinin sınırlı kullanım ömürleri hesaba katılarak önceden belirlenmiş zaman aralığında en çok incelenen haber makaleleri kullanıcı-öğe etkileşimi istatistiğine göre belirlenerek öneri olarak sunulmaktadır. Bu yaklaşımdaki temel fikir, en popüler makalelerin henüz bu makaleleri görmeyen kullanıcılar için de ilgi çekici olabileceğidir. $p(a)$, a makalesini okuyan kullanıcı sayısı ve $t(a)$, a makalesinin yayınlanma zamanı olmak üzere bir makalenin popülerlik ölçüsü $r(a)$, Eş. 3.2’de görüldüğü gibi hesaplanmaktadır.

$$r(a)_{T,P} = \log_{10} (\max(\text{abs}(p(a) - T), 1.0)) + \text{sign}(p(a) - T) \cdot \frac{-\text{şimdiki zaman} + T(a)}{P} \quad (3.2)$$

Burada P , makalenin yaşının popülerlik ölçüsü üzerindeki etkisini belirlemektedir. T ise bir makalenin öneri değeri kazanmadan önce kaç kullanıcı tarafından okunmuş olması gerektiğini belirlemektedir. Abs fonksiyonu ifadenin mutlak değerini, $sign$ fonksiyonu ise verilen ifadenin pozitif ya da negatif olacak şekilde işaretini terslemektedir.

Klasik işbirlikçi filtreleme yaklaşımlarına benzer olarak ortak özelliklere göre makaleler arasındaki ilişkiler hesaplanmaktadır. Örneğin, bir kullanıcı iki makaleyi okuduysa bu makaleler ilişkili olabilir denilmektedir. Algoritma, bilinen her bir makale için bir dizi ilgili makale sunmaktadır. Bu yaklaşım ile bir kullanıcıya öneri sunulacağı zaman kullanıcının daha önce okuduğu makale ile ilgili makaleler aranarak elde edilen makaleler kümesi kullanılmaktadır. Geliştirilen sistem, kullanıcı etkileşimlerinin en yoğun olduğu durumda, sistemin kaynak tüketim oranı açısından, paralel istekler, maksimum güncelleme ve gösterim sayısı ile tepki süresi açısından ve bir algoritmanın aynı donanım kaynakları ile kaç tane portala cevap verebildiğine yönelik test senaryoları ile analiz edilmiştir.

Ludmann tarafından 2015 yılında yapılan çalışmada, akış tabanlı tavsiye sistemleri için akış verileri üzerinde sorgular yaparak kişiselleştirilmiş öneriler kümesinin hesaplandığı yeni bir yöntem önerilmiştir [65]. Bu sorgular ilişkişel cebir işlemlerini ve veri madenciliği işlemlerini kapsamaktadır. $U = \{u_1, u_2, \dots, u_n\}$ kullanıcılar kümesi, $I = \{i_1, i_2, \dots, i_m\}$ öğeler kümesi ve T zamanı ifade etmek üzere her bir u kullanıcısının i öğesi ile t zamanda yaptığı etkileşimin sonucu doğrudan ya da dolaylı olarak değerlendirme puanı şeklinde elde edilmektedir. Değerlendirme puanı $R = \{(u, i, r, t) | (u, i, r, t) \in U \times I \times R \times T\}$ şeklinde hesaplanmaktadır. RecSys sorguları, geliştirilen modeli eğitmek için değerlendirme girdi akışının verileri kullanılarak eğitilen model ile her bir istek talebi için öneri listeleri oluşturulmaktadır. Değerlendirme veri akışları EXTRACT TEST DATA modülü tarafından test verileri akışı ve öğrenme verileri akışı olmak üzere iki veri akışına bölünmektedir. Öğrenme verileri, bir modeli eğitmek için TRAIN RECSYS MODEL modülü tarafından kullanılmaktadır. Her yeni öğrenme grubu ile bu operatör modeli güncellenmekte ve sonraki operatörler için bir kopyasını çıkarmaktadır. Bu operatör, giden modellerin geçerlilik aralıklarını şu kurallara göre ayarlamaktadır:

- Zamanın her noktasında, geçerli bir tane model vardır. Bu, yeni bir model geçerli olduğunda önceki modelin geçersiz olması gerektiği anlamına gelmektedir.
- Modeli eğitmek için kullanılan tüm öğrenme dizileri, modelin geçerlilik aralığında olmalı ve modelin geçerlilik aralığında kullanılmayan bir öğrenme grubu olmamalıdır. Ayrıca, hafızada tutulan değerlendirme sayısını sınırlamaktadır.

Her bir öneri isteğinde, RECOMM CANDIDATES modülü ile bir dizi öneri adayı belirlenmektedir. Bu adaylar genellikle kullanıcı tarafından değerlendirilmemiş öğelerdir. PREDICT RATING modülü, her bir öneri adayının değerlendirme puanını tahmin etmek için modelleri kullanmaktadır. RECOMMEND modülü değerlendirme sonuçlarına göre önerilmesi gereken öğeleri seçmektedir. TEST PREDICTION operatörü ise RMSE gibi bir değerlendirme metriği uygulayarak tahmin edilen ve gerçek değerlendirme puanlarını karşılaştırmaktadır.

Lommatzsch ve Albayrak tarafından 2015 yılında yapılan çalışmada, çevrimiçi haber portallarındaki kullanıcı-öğe etkileşimlerinin analiz edilerek kalite, sağlamlık, ölçeklenebilirlik ve sıkı zaman kısıtlamaları gibi gereksinimleri karşılayacak akan veri tabanlı öneriler sunmak için optimize edilmiş yeni bir algoritma geliştirilmiştir [66]. Haber portalları ve tartışma platformlarından günlük haber akışları alınarak analiz edilmektedir. Akışların özelliklerini ve gizli kurallarını belirleyebilmek için PLISTA yarışmasındaki (ACM Recsys News Challenge 2013) veri akışları incelenmiştir. PLISTA, araştırmacılara gerçek dünya senaryoları altında algoritmalarını değerlendirme imkânı sağlayan reklamcılar ve yayıncıların bir araya geldiği bir platformdur. PLISTA'nın amacı kullanıcılara ilgi çekici makaleler önermektir. Bir kullanıcı, haber portalındaki bir web sayfasını her ziyaret edişinde makale önerileri hesaplanmakta ve öneri sayfasına yerleştirilmektedir. Bu yarışmada, bir kullanıcı yarışma programına katılan haber sitesini ziyaret ettiğinde PLISTA sunucusu rastgele bir tavsiye ekibi seçerek talebi bu ekibe iletmektedir. Seçilen ekibin algoritması talebe göre değişmekle birlikte altı adet öneri sunmaktadır. Öneri talepleri iletişim süresi de dâhil olmak üzere 100 ms içerisinde cevaplanmaktadır. Ekiplerin performansı kullanıcılar tarafından tıklanan önerilerin sayısı ile ölçülmektedir. Yaşanan sorunlar ise kullanıcılar giriş yapmadığı için benzersiz kullanıcıları

tanımlamanın zor olması ve yarışmaya katılan takımların sundukları önerilerin kullanıcı davranışlarını etkilemesidir. Diğer bir problem ise sunulan önerilere kullanıcıların geribildirimde bulunma süreleridir. Bazı kullanıcılar önerileri hemen tıklarken diğerleri günlerce bekleyebilmektedir.

Çalışma kapsamında eğitim verisi olarak ACM Recsys 2013 verileri kullanılmıştır. Eğitim verisi 84 milyon kullanıcı tıklama verisi ve yaklaşık olarak 1 milyon öneri tıklama olayından oluşmaktadır. Değerlendirme ölçütü olarak kullanıcıların sunulan önerilere tıklama oranı ve çevrimiçi analizler kullanılmıştır. Çoğu kullanıcı yalnızca Web sitesindeki makalelerle ilgilendiği ve sunulan önerilere dikkat etmediği için tıklama oranı genellikle çok düşüktür. Web sayfasındaki önerilerin yerleştirilmesi kullanıcı gruplarının alışkanlıkları ve zamanın etkisi nedeniyle tıklama oranı büyük oranda portala bağlıdır. Çevrimiçi değerlendirmeler tavsiye algoritmasının kullanıcı tıklamalarını ne ölçüde doğru tahmin ettiğini belirlemekte ve hassasiyet (precision) değerini hesaplamaktadır.

Gerçekleştirilen topluluk tabanlı yaklaşım ile öneri talepleri için en uygun algoritmanın seçilmesi sağlanmaktadır. Bir karar ağacı ile bağlama dayalı olarak gelen isteklerin hangi algoritma ile cevaplanacağına karar verilmektedir. Haber makalelerinin öneri olarak uygun olup olmadığının belirlenmesi için eğilim tabanlı bir yaklaşım kullanılmıştır. Bu yaklaşımda algoritmanın yakın zamanda başarılı olmasının gelecekte de başarılı olacağı varsayımı temel alınarak son 60 dakikalık veriler kullanılmıştır. Öneri listesi oluşturulurken kullanılan doldurma yaklaşımı ile sunulacak 6 elemandan oluşan öneri listesinin her elemanı için talepler birkaç farklı öneri algoritmasına iletilmektedir. Her öneri ajanı için en yüksek sırada yer alan tavsiye sonuç kümesindeki bir öneri için kullanılmaktadır. Topluluk stratejisinin temelindeki fikir, her bir öneri algoritması hesaplama yaparken kendi özel kriterlerini kullandığından farklı öneri algoritmalarından alınan öneriler toplamanın çeşitliliğe sahip olacağıdır.

Analiz sonuçları çevrimdışı değerlendirmelerde (sıkı zaman kısıtlamaları olmadan), topluluk stratejisinin yaklaşık olarak %5 daha iyi öneri hassasiyetine ulaştığını göstermiştir. Topluluk stratejisinin avantajı sistemin kullanıcının davranışlarındaki değişikliklere sürekli uyum

sağlamasıdır. Dezavantajı ise, nihai sonuç kümesinin ancak tüm öneri algoritmaları hesaplamalarını bitirdikten sonra tamamlanmasıdır.

Subbian ve ark. tarafından 2016 yılında yapılan çalışmada, gerçek zamanlı öneriler sunabilmek için olasılıksal komşuluk tabanlı yeni bir yöntem geliştirilmiştir [67]. Akan verilerin kullanıldığı tavsiye sistemi uygulamalarındaki varsayım, belirli bir kullanıcının tüm değerlendirmelerinin veya belirli bir öğenin tüm değerlendirmelerinin aynı anda alınamamasıdır. Tavsiye sistemi uygulamaları aynı kaydın tüm boyutlarının her zaman eşzamanlı olarak alındığı çok boyutlu akış uygulamalarından farklıdır. Tavsiye sistemi uygulamalarında kullanıcı herhangi bir zamanda belirli bir öğe için bir değerlendirmede bulunabilmektedir. Ayrıca, yeni kullanıcılar veya öğeler herhangi bir zamanda sisteme dâhil olabilmektedir. Genellikle değerlendirmeler hiçbir zaman silinmediğinden kullanıcıların ve öğelerin sayısı zamanla artış göstermektedir. Bu nedenle, t zamanındaki kullanıcı sayısı $m(t)$, t zamandaki öğe sayısının $n(t)$ ve t zamandaki değerlendirme matrisinin boyutunun $m(t) \times n(t)$ olduğu kabul edilmiştir. Kullanıcıların öğeler hakkında bulundukları değerlendirmeler (KullanıcıId, ÖğeId, Değerlendirme Puanı) şeklinde alınmaktadır. Kullanıcı değerlendirmeleri, +1 beğenme durumunu, -1 ise beğenmeme durumunu ifade etmek üzere binary olarak alınmıştır. Çevrimiçi uygulama alanlarında kullanıcılara sunulacak öneri listeleri ya da belirli bir öğe ile ilgilenen kullanıcıların listesi belirlenmek istenebilir. Akış verisi üzerinde tek kullanıcı değerlendirme güncellemesinin olması, öğeler arasındaki mesafenin yeniden hesaplanmasını gerektirdiği için klasik komşuluk tabanlı yaklaşımlarda doğru sonuçlar üretememektedir. Ayrıca değerlendirme matrisinin tümünün hafızada tutulması beklenmemektedir. Komşuluk uzaklıkları hesaplanacak i ve j öğeleri için kullanıcılar tarafından t zamanına kadar yapılmış olumlu (+1) değerlendirmeler $P(i, t)$ ve $P(j, t)$ ile olumsuz değerlendirmeler ise $N(i, t)$ ve $N(j, t)$ ile ifade edilmektedir. t zamanda, i ve j öğeleri arasındaki benzerlik Eş. 3.3'te görüldüğü gibi hesaplanmaktadır.

$$S^+(i, j, t) = \frac{P(i, t) \cap P(j, t)}{P(i, t) \cup P(j, t)} \quad (3.3)$$

Olumlu değerlendirmelere dayalı olarak hesaplanan bu ölçüt Jaccard benzerliğidir. $\alpha = |P(i, t)| + |P(j, t)|$ ve $\beta = |N(i, t)| + |N(j, t)|$ olmak üzere olumlu ve olumsuz değerlendirmelerin ağırlıklandırılması ile Eş. 3.4 elde edilmektedir.

$$S(i, j, t) = \frac{\alpha \cdot S^+(i, j, t) + \beta \cdot S^-(i, j, t)}{\alpha + \beta} \quad (3.4)$$

Kullanıcı değerlendirmelerinin tahmin edilmesi için ilk olarak belirli bir ögenin tüm öğeler ile benzerliği hesaplanmaktadır. Belirli bir i ögesine en benzer öğelerin değerlendirmelerinin ağırlıklı ortalaması, o öğe için kullanıcı değerlendirmesi olarak tahmin edilmektedir. $l_i(u)$, u kullanıcısının değerlendirmede bulunduğu i ögesine benzer öğelerin kümesini ve $r_{u,j}$ ise u kullanıcısının j ögesi ile ilgili değerlendirme puanını ifade etmektedir. u kullanıcısı ve i ögesi için t zamanındaki komşuluk tabanlı tahmin puanı Eş. 3.5 kullanılarak hesaplanmaktadır.

$$P_{u,i}(t) = \frac{\sum_{j \in I_{i(u)}} (S(i, j, t) \cdot r_{u,i})}{\sum_{j \in I_{i(u)}} S(i, j, t)} \quad (3.5)$$

Çalışma kapsamında, geliştirilen olasılıksal komşuluk tabanlı yöntemler ile öğeler arasındaki benzerliklerin olasılıksal olarak hesaplandığı min-hash tekniği önerilmiştir. Benzerlikler, min-hash indeksindeki her bir kullanıcı izlenerek yaklaşık olarak hesaplanmaktadır. Temel fikir bir hash fonksiyonu kullanarak kullanıcılara sıralama düzeni uygulamaktır. Bu sıralama düzeninde i ögesi için pozitif değerlendirmede bulunan ilk kullanıcının, j ögesi için de olumlu bir değerlendirmede bulunan ilk kullanıcı olma ihtimali Jaccard ölçütü ile benzerdir. $f_1(\cdot), \dots, f_d(\cdot)$ hash fonksiyonları, j ögesini olumlu olarak ve olumsuz olarak değerlendirmiş kullanıcılara uygulanmaktadır. Hash fonksiyonları uygulandıktan sonra olumlu değerlendirme yapan kullanıcılar M^+ veri yapısında, olumsuz değerlendirme yapan kullanıcılar ise M^- veri yapısında tutulmaktadır. Bu veri yapılarının boyutları kullanıcı-öge değerlendirme matrisinden daha küçük olduğu için ve kolaylıkla güncellenebildiği için bellekte muhafaza edilebilmektedir. i ve j öğeleri arasındaki benzerlik Eş. 3.6 kullanılarak hesaplanmaktadır.

$$S^+(i, j, t) \approx R^+(i, j, t) = \frac{\sum_{s=1}^d \delta(i_s = j_s)}{d} \quad (3.6)$$

$\delta(\cdot)$, i ve j öğeleri benzerse 1 benzer değilse 0 değerini alan bir fonksiyondur.

Chang ve ark. tarafından 2017 yılında yapılan çalışmada, sonsuz ve değişken boyuttaki akış verileri için gerçek zamanlı güncelleme yapabilen sRec sistemi geliştirilmiştir [68]. Geliştirilen sistemde girdi akışı, kullanıcı geri bildirim etkinlikleri, yeni kullanıcılar ve yeni öğeler olarak modellenmiştir. sRec değişen kullanıcı ve ürün sayısı ile kullanıcı tercihlerinde yaşanabilecek içerik kayması durumlarını belirleyerek gerçek zamanlı öneriler sunmaktadır. sRec temel olarak çevrimiçi öneri modülü ve çevrimdışı parametre öğrenme modülünden oluşmaktadır. Sistem, kullanıcı dinamiklerini yakalamak ve gerçek zamanlı öneriler sunabilmek için modelini sürekli olarak güncellemektedir. sRec sisteminde kullanıcı öğe değerlendirmeleri zamana dayalı bir fonksiyon ile modellenmektedir. Çevrimdışı parametre güncellemeleri yalnızca olaylar meydana geldiğinde ve önceki olaylara bağlı olarak gerçekleştirilmektedir. Çevrimiçi öneri modülü ise oluşturulan kullanıcı-öğ-e-zaman modellerini kullanarak kullanıcıların yapmış oldukları tercihlerin son olasılık dağılımına göre öneri sunmaktadır.

Mehmood ve ark. tarafından 2019 yılında yapılan çalışmada, gerçek zamanlı bir seyahat rotası öneri sistemi sunulmuştur [69]. Gerçek zamanlı büyük veri Güney Kore'de bulunan Jeju Adası'ndaki farklı alanlarda konumlandırılan kablosuz ağ yönlendiricilerden toplanmıştır. Kullanılan veri kümesi, 2016-2017 yılları arasında binlerce turistten toplanan mobil hareket örüntülerini içermektedir. Çalışmada kullanıcı tercihleri, demografik bilgiler ve turistlerin hareket modellerinin istatistiksel analizi yapılarak en uygun seyahat rotası önerileri hesaplanmaktadır. Kullanıcı tercihlerinin zaman içinde değişebileceği ya da aynı mobil cihazı kullanan kişiden kişiye farklılık gösterebileceği için öneriler hesaplanırken zaman, mesafe ve konumun popülerliği ile hava ve trafik durumu da göz önüne alınmıştır.

Geliştirilen sistem depolama katmanı, hesaplama katmanı ve uygulama katmanı olmak üzere üç katmandan oluşmaktadır. Depolama katmanında kablosuz ağlardan elde edilen kullanıcıların enlem-boylam bilgileri bulunmaktadır. Uygulama katmanı kullanıcıların

tercihlerini girdiği arayüzdür. Kullanıcı tercihleri, konumun ücretli ya da ücretsiz olması, yaş ve cinsiyet ile ilişkilidir. Optimal güzergâhlar önerilmeden önce yaş ve cinsiyet faktörü hesaba katılmaktadır. Hesaplama katmanı ise buluttaki verilerin işlenmesinden sorumludur. Büyük verilerin işlenmesi ve hesaplanması için Amazon Web servisleri kullanılmıştır. Bu katmanda ön işleme yapılarak ziyaret edilebilecek konumlar belirlenmiş ve veri kümesindeki toplam yol sayısına göre konumların popülerliği hesaplanmıştır.

3.1. Akış Verisi Analizinde Kullanılan Yöntemler

Akış verilerini etkili bir şekilde işlemek için yeni veri yapıları, teknikler ve algoritmalara ihtiyaç duyulmaktadır. Akış verilerini depolayacak sonsuz miktarda alan olmadığı için doğruluk ve depolama alanı arasında ters orantı mevcuttur. Ayırıştırma işlemleri, yapılan sorgulara yaklaşık cevapları döndürmek için kullanılabilen verilerin özetlerini sunmaktadır. Ayırıştırıcılar, esas veri kümesinden akış verisi önemli ölçüde daha küçük bir veri yapısı olan özet veri yapılarını kullanmaktadır. Özetleme işlemleri altta yatan verilerin tarafsız bir tahminini gerçekleştirmektedir. Örneklem yöntemleri kayıtların orijinal gösterimini kullandıkları için herhangi bir veri madenciliği uygulaması veya veritabanı işlemi ile kullanılabilir [70].

3.1.1. Örneklem

Akış verilerinin örneklenmesindeki temel amaç, istatistiksel olarak akış verisini temsil edebilecek bir altküme seçmektir. Verilerin rastgele olarak alınan küçük örnekleme ile genel olarak veri kümesinin özellikleri yakalanmaktadır. Verilerin örnekleminin hafızada tutulması ile bellek tasarrufu sağlanarak sorgulara cevap verilebilmektedir. Akış verisinin örneklenmesinde kullanılan temel yöntem rezervuar örneklemedir [71].

Rastgele örneklem

Rastgele örneklem tekniği genellikle yüksek boyutlu uygulamalar için kullanılmaktadır. Akış verisinin boyutunun bilinmediği durumlarda rezervuar örnekleme kullanılmaktadır. Akış verisinden rezervuar olarak adlandırılan rastgele boyuttaki bir s örneği alınması

temeline dayanmaktadır. Rezervuar örnekleme, özellikle büyük boyuttaki verilerde maliyetli bir işlem olacağı için akış içerisinde belirli bir zamana kadar görülen öğelerin rastgele bir örneğini tutan bir s aday kümesi oluşturmaktadır. Akış verisi gerçekleştikçe, her yeni elemanın rezervuardaki eski bir elemanın yerine geçme ihtimali vardır. Örneğin, akış içerisinde N adet farklı eleman görülürse, yeni bir elemanın eski bir elemanın yerine geçme olasılığı s/N olur [72].

Rezervuar örnekleme

Çevrimiçi olarak ulaşan, n öğeden oluşan bir akış göz önüne alındığında, herhangi bir anda yapılan rezervuar örnekleme ile o zamana kadar gözlemlenen akış bölümünün m boyutunda rastgele bir S örneği elde edilmektedir. Doğal örnekleme, öncelikle akışın ilk m öğesinin S 'ye eklenmesi ile başlamaktadır. İkinci adımda x_t öğesi t zamanında akış içerisinde görüldüğünde, x_t öğesi S örneklemine m/t olasılıkla eklenmektedir. x_t öğesi S örneklemine eklenirse, S den rastgele bir öğe çıkarılmaktadır. Her seferinde $|S| = m$ eşitliğinin sağlanması amaçlanmaktadır [73].

Rezervuar örnekleme, n maddeli bir akış verisinden m örneklerini rastgele seçmek için kullanılan yaklaşımlardır. Burada n değeri, genellikle ana belleğe sığmayacak kadar büyüktür. Örneğin, Google ve Facebook'daki arama sorgularının listesi n değerine örnek verilebilir. Burada $1 \leq m \leq n$ ve giriş dizisi akış verisi olacak şekilde maksimum boyutu m olan bir rezervuar dizisi oluşturmak en basit çözümdür. $[0..n-1]$ akışından rastgele tek tek öğelerin seçilerek rezervuar dizisine eklenmesi, algoritmanın zaman karmaşıklığının $O(m^2)$ olmasına neden olmaktadır. Ancak m değeri büyüdükçe ve girişler akış verisi formunda oldukça bu yöntem verimli olmamaktadır. $O(n)$ zamanda bu işlemi gerçekleştirmek için $[0..m-1]$ boyutunda rezervuar dizisi oluşturularak akış verisi dizisinin ilk m öğesi rezervuar dizisine eklenmektedir. Daha sonra $m+1$. öğeden n . öğeye kadar olan öğeler düşünüldüğünde i mevcut öğenin akış verisi dizisindeki indeksini ifade etmek üzere 0 ile i arasında rastgele bir sayı üretilmektedir. Üretilen sayının j olduğu varsayılırsa ve eğer j , 0 ile $k-1$ aralığında ise rezervuar dizisinin j . elemanı ile akış dizisinin i . elemanı değiştirilmektedir [71].

3.1.2. Kayan pencereler

Akış verilerini rastgele örneklemek yerine akış verilerini analiz etmek için kayan pencere modeli kullanılmaktadır. Temel fikir belirli bir ana kadar görülen veriler üzerinde hesaplamalar yapmadan yalnızca son verilere dayanan kararlar verebilmedir. Her t zamanında yeni bir veri ögesi geldiği ve w pencere uzunluğunu ifade etmek üzere bu elemanın, $t+w$ zamanında sona erdiği temel alınmaktadır. Kayan pencereler modeli son olayların önemli olabileceği hisse senetleri veya algılayıcı ağları için kullanışlıdır. Kayan pencereler modeli sadece küçük bir veri penceresi depolandığı için bellek gereksinimlerini azaltmaktadır [74].

3.1.3. Histogramlar

Histogram tabanlı yöntemler statik veri kümeleri için yaygın olarak kullanılmaktadır. Histogram, bir akış verisindeki öge değerlerinin sıklık dağılımını yaklaşık olarak hesaplayabilmek için kullanılabilen özet bir veri yapısıdır. Histogramlar verileri bitişik kümeler halinde bölümlere ayırmaktadır. Kullanılan bölümlendirme kuralına bağlı olarak, genişlik (kova değeri aralığı) ve derinlik (kova başına öge sayısı) değişebilmektedir. Kümeleme işlemine benzer mantıkla çalışan V-optimal histogramlar verilerin dağılımına göre her kova içindeki frekans değişimini en aza indirgeyen kova boyutları belirlemektedir. Oluşturulan histogramlar ile sorgu ifadelerine cevap döndürülmektedir [75].

3.1.4. Eskizler

Özetleme yöntemleri, depolama hassasiyetini nasıl değiştirdiklerine göre farklılık göstermektedir. Örnekleme teknikleri ve kayan pencere modelleri, verilerin küçük bir bölümüne odaklanırken, diğer özetler genel olarak çok detaylı ayrıntıları özetlemeye çalışmaktadır. Histogramlar ve dalgacıklar gibi yöntemler veriler üzerinde birden çok geçiş gerektirirken, eskizler gibi yöntemler ise tek geçişte çalışabilmektedir. Örnek olarak $U = (1, 2, \dots, v)$ ögelerin evrenini ve $A = (a_1, a_2, \dots, a_N)$ ise akışı ifade ettiği bir histogramda evren büyüdükçe A sırasındaki i frekansının belirlenmesi maliyetli olabilmektedir. Bu durumda moment tahmini ön plana çıkmaktadır ve Eş. 3.7’de görüldüğü gibi hesaplanmaktadır.

$$F_k = \sum_{i=1}^v m_i^k \quad (3.7)$$

Eş. 2.3' te görüldüğü gibi, v evren veya etki alanı boyutu, m_i , i ' nin frekansı ve $k \geq 0$ olmak üzere F_0 , dizideki farklı öğelerin sayısını ifade etmektedir. F_1 , dizinin uzunluğunu (N), F_2 ise tekrar oranını ifade etmektedir. Bir veri kümesinin sıklık anları sorgu cevaplama gibi veritabanı uygulamalarında veriler hakkında bilgiler sağlamaktadır. Ayrıca, uygun bölümlleme algoritmasını belirlemek için paralel veritabanı uygulamalarında kullanışlı olan verilerdeki çarpıklık veya asimetri derecesini göstermektedir. U evreni, v değer sayısını ve N eleman sayısını ifade etmek üzere F_0 , F_1 ve F_2 eskizleri $O(\log v + \log N)$ uzayına yakınsamaktadır. Temel fikir, her elemanı rastgele olarak $z_i \in \{-1, +1\}$ aralığına hash'lemek ve $X = \sum_i m_i z_i$ rastgeleliğini korumaktır. Daha iyi tahmin değerleri elde etmek için, birden fazla rastgele X_i değişkeni tutulabilmektedir. Bu değişkenlerin karesinin medyan değeri seçilerek, tahmin değerinin F_2 'ye yakın olduğu doğrulanabilir. Veritabanı açısından eskiz bölme, akış verisi sorgulama optimizasyonlarında eskiz performansını artırmak için kullanılmaktadır [74].

3.2. Doğrusal Yöntemler

Akış verisi analizi, temel olarak verinin doğasını anlamaya çalışan ve geleceğe dönük tahminler ve simülasyonlar için kullanılabilecek yöntemleri içermektedir. Akış verisi analizinde, geçmiş gözlemler kullanılarak veri oluşturma sürecini belirleyen uygun bir matematik modeli geliştirilmektedir [76]. Akış verileri gerçek zamanlı zaman serisi verileridir.

Zaman serileri, ardışık zaman aralıklarında ölçülen veri noktası dizileridir. Matematiksel olarak, t geçen süreyi ifade etmek üzere y_t , $t = 0, 1, 2, \dots, T$ vektörlerinin kümesi olarak tanımlanmaktadır. Bir zaman serisi, her bir y değişkeninin, t zamanındaki değerini temsil eden y_t değerleri tarafından oluşturulmaktadır. Bu nedenle tahmin işlemi, $y_{t+1} + (y_{t+1})$ değerinin belirlenmesi ve $y_{t+1} - (y_{t+1})$ işlevi olarak temsil edilen hata oranını en aza

indirmeyi hedeflemektedir. y_t değişkeni rastgele bir değişken olarak kabul edilmektedir. Bir zaman serisinde, bir olay sırasında alınan ölçümler belirli bir zamansal düzene uygun olarak sıralanmıştır [77].

Akış verilerinin gerçek zamanlı modellenmesi için yaygın olarak istatistiksel modeller ve makine öğrenmesi yöntemleri kullanılmaktadır. İstatistiksel modeller, ekonomi ve finans verilerinin modellenmesinde yaygın olarak kullanılıyor olsa da, bazı önemli kısıtlamaları vardır [78]. Örneğin, ARIMA modelinde değişkenler arasındaki doğrusal olmayan ilişkileri modellemek zordur. Ayrıca, ARIMA modelinde hataların sabit bir standart sapmaya sahip olduğu varsayılmaktadır. İstatistiksel modellerde yaşanan bu gibi zorlukları aşabilmek için derin öğrenme tabanlı yeni teknikler geliştirilmiştir.

Akış verileri, doğası gereği durağan olmadığı için gelecekte ne olacağı konusunda kesin bir tahminde bulunulamayan gerçek zamanlı zaman serileridir. Genel olarak $\{y_t, t = 0, 1, 2, \dots\}$ şeklindeki bir akış verisinin rastgele bir y_t değişkenin birleşik olasılık dağılımını izlediği varsayılmaktadır [9]. Dolayısıyla serinin gözlem dizisi aslında onu üreten olasılıksal sürecin bir örneklemdir [76]. Zaman serilerindeki genel varsayım, serideki y_t değişkenlerinin normal dağılımı takiben bağımsız ve özdeş olarak dağıtılmış olmasıdır. Ancak zaman serileri uzun vadede daha düzenli bir örüntü izlerler. Örneğin, belirli bir şehrin bugünkü sıcaklığı çok yüksekse yarınki sıcaklığının da muhtemelen yüksek olacağı tahmin edilebilir. Bu durum zaman serilerinde durağanlık kavramı ile ilişkilidir [10].

Olasılıksal bir süreçte durağanlık kavramı istatistiksel bir denge biçimi olarak görülebilir. Durağan bir sürecin ortalaması ve standart sapması gibi istatistiksel özellikleri zamana bağlı değildir. Bu istatistiksel özellikler, geleceğe dair tahminlerde bulunmak için kullanılacak bir zaman serisi modeli oluşturmak ve oluşturulan modelin karmaşıklığını azaltmak için gerekli bir koşuldur [7].

Gerçek zamanlı modellemede, serinin alt yapısını yansıttığı için uygun modelin seçilmesi yapılacak tahminlerin doğruluğunu birinci dereceden etkilemektedir. Genel olarak, kullanılacak modeller farklı istatistiksel süreçleri içerebilir [79]. Literatürde yaygın olarak

AR ve MA modelleri kullanılmaktadır. Ayrıca, bu iki modeli birleştiren ARMA ve ARIMA modelleri kullanılmaktadır [12].

3.2.1. ARMA model

Tek değişkenli serilerin modellenmesi için kullanılan ARMA(p, q) modeli, AR(p) ve MA(q) modellerinin birleşimidir [80]. AR(p) modelinde, bir değişkenin gelecekteki değerinin geçmiş p adet gözlemin doğrusal bir birleşimi olduğu ve sabit bir hata terimine sahip olduğu varsayılmaktadır. Matematiksel olarak AR(p) modeli Eş.3.8 kullanılarak ifade edilmektedir [77].

$$y_t = c + \sum_{i=1}^p \varphi_i y_{t-i} + \varepsilon_t = c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varphi_p y_{t-p} + \varepsilon_t \quad (3.8)$$

Burada, y_t ve ε_t sırasıyla t zamandaki gerçek değeri ve rastgele hatayı ifade etmektedir. $\varphi_i (i=1,2,\dots,p)$ model parametrelerini ve c ise model sabitini ifade etmektedir. Tamsayı sabiti p , modelin sırası olarak tanımlanır ve ihmal edilebilmektedir.

MA(q) modeli ise Eş.3.9'da görüldüğü gibi açıklayıcı değişken olarak geçmiş hataları kullanmaktadır [9].

$$y_t = \mu + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t = \mu + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (3.9)$$

Burada μ serinin ortalamasını, $\theta_j (j=1,2,\dots,q)$ ise model parametrelerini ifade etmektedir. Rastgele hatanın 0 ortalamaya sahip rastgele dağıtılmış ve normal dağılıma sahip değişkenler dizisi olduğu varsayılmaktadır. AR ve MA modelleri, ARMA modeli olarak bilinen genel ve kullanışlı bir zaman serisi modeli oluşturmak için etkin bir şekilde birleştirilebilir. Matematiksel olarak bir ARMA (p, q) modeli, Eş.3.10 kullanılarak ifade edilmektedir [13].

$$y_t = c + \varepsilon_t + \sum_{i=1}^p \varphi_i y_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} \quad (3.10)$$

Modeldeki p , q parametrelerinden p parametresi otoregresifliği, q parametresi ise hareketli ortalamayı ifade etmektedir.

ARMA modelleri yalnızca sabit zaman serisi verileri için kullanılabilen modellerdir. Ancak e-ticaret uygulamaları gibi gerçek dünya uygulamalarındaki zaman serileri durağan olmayan davranış göstermektedir. Bu nedenle, ARMA modelleri uygulamalarda sıklıkla karşılaşılan durağan olmayan zaman serilerini doğru bir şekilde tanımlamakta yetersiz kalmaktadır.

3.2.2. ARIMA model

ARIMA, otoregresif $AR(p)$ ve hareketli ortalama $MA(q)$ süreçlerini birleştirerek zaman serilerinde birleşik süreç oluşturan bir modeldir. ARIMA modeli, ARMA modelinin durağan olmayan durumları da içerecek şekilde güncellenmiş bir halidir [9]. ARIMA modeli, ARMA modelini kullanarak eşit aralıklı tek değişkenli zaman serisi verilerini analiz eden doğrusal bir modeldir. ARIMA modeli, zaman serisindeki geçmiş değerlerin ve geçmiş hataların doğrusal bir birleşimine göre tahmin sürecini işletir [78]. Durağan bir zaman serisi için ARIMA modeli, bağımlı değişkenin gecikme sürelerinden ve tahmin hatalarının gecikmesinden oluşan doğrusal bir modeldir. Modelde tahmin edilecek y değeri, y değişkeninin son değerleri ile hataların ağırlıklı bir toplamıdır [7].

ARIMA modeli p , d ve q parametrelerine sahiptir. ARIMA modelindeki otoregresif $AR(p)$, bir gözlem ile bir dizi gecikmeli gözlem arasındaki bağımlılıkları kullanan bir regresyon modelidir [8]. Entegrasyon, farklı zamanlardaki gözlem farklılıklarını ölçerek zaman serilerini durağan hale getirmeyi ifade etmektedir (d). ARIMA modelindeki hareketli ortalama $MA(q)$ ise gözlemler arasındaki bağımlılığı ve kalan hata terimlerini dikkate alan bir yaklaşımdır (q).

ARIMA modeli, Box ve Jenkins tarafından 1976 yılında tanıtılan Box Jenkins metodolojisini izlemektedir [12]. Box ve Jenkins metodolojisine göre ARIMA modeli modelin

tanımlanması, parametrelerin tahmini, modelin doğrulanması ve tahmin aşamalarından oluşmaktadır.

Modelin tanımlanması, zaman serisinin durağan olup olmadığının yani olasılıksal bir sürecin ortalama ve standart sapmasının zamanla değişip değişmediğinin belirlenmesidir. Zaman serileri bu kriteri karşılamıyorsa dönüşüm uygulanarak serinin durağanlığa ulaşana kadar farklılaştırılması gerekmektedir. Serinin farklılaştırılması gereken zaman sayısı d ile gösterilir ve ARIMA modellerinde belirlenecek parametrelerden biridir [81].

Parametre tahmini, p ve q parametrelerinin en küçük kareler minimizasyonu ya da maksimum olasılık algoritmaları kullanılarak tahmin edilmesini ifade etmektedir. En küçük kareler minimizasyonu eğitim seti tahminindeki karesel hatayı minimuma indirirken maksimum olasılık algoritması ise benzerlik fonksiyonunu en üst düzeye çıkarmayı hedeflemektedir [77].

Model doğrulanması, oluşturulan modelin uygun olup olmadığının belirlenmesini ifade etmektedir. Veri bütünüyle rastsal ise beyaz gürültüye sahip olduğu kabul edilerek model onaylanmaktadır [8]. Farklı modellerin performanslarını değerlendirmek için AIC, MAE veya RMSE gibi ölçütler kullanılabilir. Tahmin aşaması ise parametreleri doğru bir şekilde belirlenmiş ve onaylanmış modelin tahminleri yapmaya hazır olduğunu ifade etmektedir. ARIMA modeli ile tahminler Eş.3.11 kullanılarak hesaplanmaktadır.

$$y_t = \mu + \varphi_1 y_{t-1} + \dots + \varphi_p y_{t-p} - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q} \quad (3.11)$$

Burada y_t tahmin değerini, e_t , t zamanındaki hatayı, φ_i ve θ_j ise katsayıları ifade etmektedir. μ , otoregresif $AR(p)$ ve hareketli ortalama $MA(q)$ polinomlarının sayısı olan beyaz gürültüyü ifade etmektedir. AR-I-MA modelinin AR bölümü, y_t değişkenin önceki değerlerinden elde edildiğini, MA bölümü ise regresyon hatasının önceki hataların doğrusal bir birleşimi olduğunu göstermektedir. I bölümü, normal veri değerlerinin, veriler ve önceki değerler arasındaki farkla değiştirildiğini gösterir [9].

3.3. Makine Öğrenmesi Tabanlı Yöntemler

Verilerin geçmiş değerlerinden gelecekteki değerlerinin tahmin edilmesi, geçmiş ve gelecek arasındaki olasılıksal bağımlılıkları işleyebilen tekniklerin kullanımını gerektirmektedir. Zaman serisi tahmininde ARIMA gibi istatistiksel modeller yaygın olarak kullanılmaktadır. Zaman serisi verileri, pencere yöntemi kullanılarak denetimli öğrenme problemi haline getirilebilir. Bu sayede makine öğrenmesi yöntemleri kullanılarak verilerin geçmiş değerlerinden gelecekteki değerleri hakkında tahminler oluşturulabilmektedir.

3.3.1. RF

RF, torbalama tekniğini kullanan bir makine öğrenmesi algoritmasıdır. Torbalama tekniği, genelleştirilmiş bir sonuç elde etmek için birden fazla modelin sonuçlarının birleştirilmesini ifade etmektedir. RF, birden fazla karar ağacı oluşturarak çalışan öğrenme yöntemini kullanır. Tahminler, ağaç oluştururken kullanılan eğitim verilerinin rastgele örneklenmesi ve düğümleri bölerken göz önüne alınan özelliklerin rastgele alt kümelerinin seçilmesi yoluyla elde edilmektedir. RF'daki temel tahminciler karar ağaçlarıdır. RF, ağaçları büyütürken modele bir rastgelelik ekler. Bir düğümü bölerken en önemli özelliği aramak yerine, rastgele bir özellik alt kümesinde en iyi özelliği arar. Bu durum, genellikle daha iyi bir modelle sonuçlanan geniş bir çeşitlilik sağlar [82].

3.3.2. SVM

SVM, sınıflandırma ve regresyon problemleri için kullanılan en popüler denetimli öğrenme algoritmalarından biridir. SVM'nin amacı, n boyutlu bir alanı sınıflara ayırabilecek en iyi çizgi veya karar sınırını oluşturmaktır. Bu sayede yeni veri noktalarının gelecekte doğru sınıflara atanabilmesini sağlamayı hedeflemektedir. Bu en iyi karar sınırına hiper düzlem adı verilir. SVM temel olarak çok boyutlu uzayda bir hiper düzlemdeki farklı sınıfların temsilidir. Hiper düzlem hatanın en aza indirilebilmesi için SVM tarafından yinelemeli bir şekilde oluşturulmaktadır. SVM'nin amacı, maksimum marjinal hiper düzlemi bulmak için veri kümelerini sınıflara ayırmaktır [83].

3.3.3. MLP

Yapay Sinir Ağları (YSA), biyolojik sinir ağlarının yapı ve işlevlerine dayanan bir hesaplama modelidir [86]. YSA, girdiler ve çıktılar arasındaki karmaşık ilişkilerin modellendiği ya da örüntülerin bulunduğu doğrusal olmayan veri modelleme aracı olarak kabul edilmektedir. YSA, günümüzde robotik, nesne tanıma, konuşma ve el yazısı tanıma gibi çeşitli uygulama alanlarında kullanılmaktadır [87].

YSA, bir giriş katmanı, gizli katmanlar ve bir çıkış katmanı olmak üzere en az üç katmandan oluşmaktadır. Veri kümesindeki özelliklerinin sayısı, giriş katmanının boyutunu veya düğümlerin sayısını belirler. Bu düğümler, gizli katmanlarda oluşturulan düğümlere sinopsis adı verilen yapılar ile bağlanmaktadır [86]. Sinopsis bağlantıları, giriş katmanındaki her düğüm için bazı ağırlıklar taşımaktadır. Ağırlıklar temel olarak, hangi sinyalin veya girdinin geçebileceğine karar veren bir karar verici rolü oynar. Ağırlıklar gizli katmanın gücünü veya kapsamını gösterir [85]. Bir YSA, temel olarak, her bir sinopsis için ağırlığı ayarlayarak öğrenir.

Gizli katmanlarda düğümler, girişleri çıktılara veya tahminlenen değerlere dönüştürmek için girişlerin ağırlıklı toplamı üzerine bir aktivasyon fonksiyonu (sigmoid veya hiperbolik tanjant) uygular. Çıkış katmanı, SoftMax adlı bir fonksiyon kullanarak, maliyet olarak da bilinen, beklenen ve tahmin edilen değerler arasındaki farkı en aza indirmeye çalışır. Ağırlık vektörüne yapılan atamalar ve ağırlık eğitimi elde edilen hatalar beklenen düzeyde olmayabilir [88]. Bu durumda, en uygun hata değerlerini bulmak için hatalar, çıkış katmanından gizli katmanlara doğru ağırlıklı geri yayılır ve sonuç olarak ağırlıklar ayarlanır. Aynı gözlemler tekrardan yapılır ve ağırlıklar tahminlenen değerlerde ve maliyette bir iyileşme sağlanana kadar yeniden ayarlanır. Maliyet fonksiyonu en aza indirildiğinde model eğitilir [13].

MLP, geriye yayılım algoritmasını kullanan ve sinir ağı uygulamalarında yaygın olarak kullanılan bir modeldir. MLP, verilen giriş verilerini beklenen hedef verilere eşleyen bir sinir ağı türüdür [89]. MLP, yönlü bir grafa birden çok düğüm katmanından oluşur ve her katman bir sonrakine bağlıdır. Giriş ve çıkış katmanları arasında bir veya daha fazla gizli katman

vardır. Nöronlar katmanlar halinde düzenlenir ve bağlantılar her zaman alt katmanlardan üst katmanlara yönlendirilir. Aynı katmandaki nöronlar birbirine bağlı değildir [90].

MLP'ler birden fazla algılayıcıdan oluşur. Sinyali almak için bir giriş katmanı, giriş hakkında bir karar veren veya tahmin yapan bir çıkış katmanı ve bu iki katman arasında bulunan hesaplamaların yapıldığı gizli katmanlardan oluşur [91]. Modelin eğitiminde temel olarak üç adım vardır. İleri geçiş, girdi modele verilerek düğümler arasındaki ağırlıklar ile çarpılır. Her bir katmanda bias değeri eklenerek modelin hesaplanan çıktısı bulunur. Hata veya kaybın hesaplanması model tarafından işlendiğinde sonuç olarak tahmin edilen çıktı değeri elde edilir. Gerçek değer ve tahmin edilen değer ile geriye yayılım algoritması kullanılarak kayıp değeri hesaplanır. Geri geçiş, kayıp değerine göre modelin ağırlıklarının güncellenmesini ifade etmektedir. Bu işlem, modelin eğitimindeki temel adımdır [92].

4. AKIŞ VERİSİNDE DERİN ÖĞRENME TABANLI ÖNERİ SİSTEMLERİ

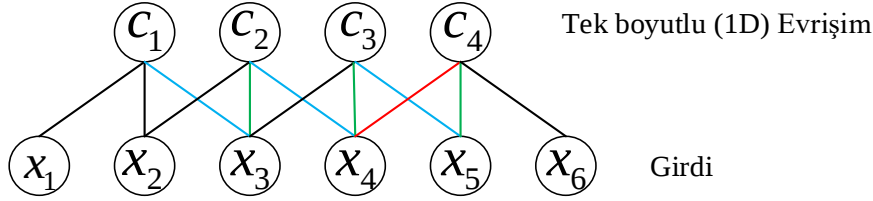
Zaman serileri, borsadaki fiyat değişimleri, hastalıkların yayılma süreçleri ya da ses sinyallerindeki değişimlerin izlenmesi gibi birçok gerçek dünya uygulaması vasıtasıyla elde edilebilmektedir. Farklı uygulama alanları için kullanılan uygulamalar verilerin niteliğine ve amacına bağlı olarak değişmektedir [84]. Örneğin, doktorlar hastanın uyku düzenindeki anormallikleri belirlemekle ilgilenirken, ekonomistler faiz oranının gelecek değerini belirlemekle ilgilenmektedirler. Bu tür problemler, mevcut verilerden anlamlı örüntülerin çıkarılması için istatistiksel yöntemlerin ya da makine öğrenmesi yöntemlerinin kullanımı gerekli kılmaktadır [85]. Ancak, bu yöntemler büyük boyutlu ve karmaşık veri kümelerinde yeterli düzeyde başarılı olamayabilir. Bu tür veriler için derin öğrenme tabanlı modeller başarılı sonuçlar vermektedir [18, 21, 23, 25].

4.1. CNN

CNN, görüntü tanıma ve sınıflandırma gibi uygulama alanlarındaki başarısı sayesinde popülerlik kazanmış bir derin sinir ağıdır. CNN'ler bilgisayarla görü, görüntü sınıflandırma, yüz tanıma, nesne tanımlama ve sınıflandırılma ile otonom araçlarda görüntü işleme gibi uygulama alanları için başarılı bir şekilde kullanılmaktadır. Genel olarak CNN, 2 veya 3 renk kanallı görüntüleri analiz etmek için 2 boyutlu veya 3 boyutlu nöral katmanları kullanırken, 1 boyutlu katmanlara sahip CNN'ler ise algılayıcı verileri, ses sinyalleri ve doğal dil işleme gibi uygulama alanlarında özellik çıkarımı için kullanılmaktadır [93].

Genel itibariyle CNN, 3 adımda işlem gerçekleştirmektedir. Basit bir sınıflandırma örneğinde ilk olarak görüntünün bir anda birkaç pikseli tarayan bir evrişim katmanı ile her bir özelliğin ait olması gereken sınıfa ait olasılıkları ile bir özellik haritası oluşturulur. İkinci olarak her bir özelliğin boyutunun azaltıldığı alt örnekleme yapıldığı pooling işlemi gerçekleştirilir. Pooling aşaması görüntüdeki en önemli özelliklerin bir özetini oluşturur. Örneğin, ilk aşamada görüntü insan, gemi ya da ağaç gibi nesnelere ayrıştırılmaktadır. İkinci aşamada CNN her bir nesnenin içindeki el, yüz, kol, bacak gibi özellikleri tanımlamaktadır. Üçüncü aşamada ise CNN derin bir analiz ile yüzdeki özellikleri analiz edebilir.

CNN, ağırlık paylaşma kavramını kullanmaktadır. Şekil 4.1'de, $x_1, \dots, x_6$ girişleri, $c_1, \dots, c_4$ ise tek boyutlu (1D) evrişimden sonraki özellik haritalarını ifade etmektedir.

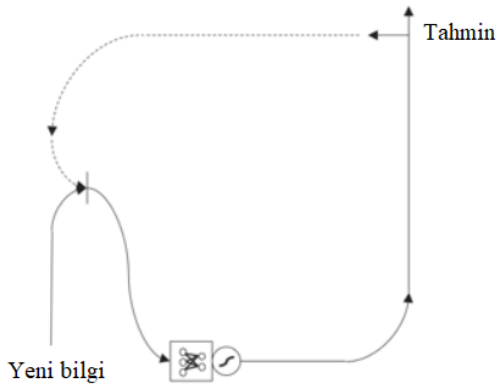


Şekil 4.1. Tek boyutlu evrişim

Giriş katmanı ve evrişim katmanı kırmızı, yeşil ve mavi bağlantılar ile bağlıdır. Her bağlantının kendi ağırlık değeri vardır ve aynı renkteki bağlantılar aynı değere sahiptir. Bu nedenle Şekil 4.1'de, evrişim işlemini gerçekleştirmek için sadece 3 ağırlık değerine ihtiyaç vardır. CNN'nin avantajı, ağırlıklarının sayısının tam bağlı mimarilere göre az olması nedeniyle eğitimin daha kolay olmasıdır. Bu sayede önemli özellikler etkili bir şekilde çıkarılabilir [94].

4.2. RNN

RNN, dizinin gözlenen önceki adımlarına göre bir sonraki adımın tahmin edilmesini amaçlamaktadır. RNN'lerin ardındaki fikir Şekil 4.2'de görüldüğü gibi sıralı gözlemlerden faydalanmak ve gelecekteki eğilimleri tahmin etmek için önceki aşamalardan öğrenmektir [88].



Şekil 4.2. Temel RNN yapısı

RNN'de gizli katmanlar, sıralı veri okumanın önceki aşamalarında elde edilen bilgilerin saklanması için dâhili depolama görevi görmektedir. RNN'ler, dizinin her elemanı için aynı görevi yaptıkları ve daha önce görülmemiş sıralı verileri tahmin etmek için elde edilen eski bilgileri kullandıkları için tekrarlı bir yapıya sahiplerdir [79]. RNN'in en büyük zorluğu, dizideki yakın zamanda işlenen verilerin hatırlanması ve dolayısıyla, daha uzun veri dizilerini hatırlamaya uygun olmamasıdır. Bu sorun LSTM'de bellek hattı kullanılarak çözülmektedir [13].

4.3. GRU

GRU, LSTM'in daha basit bir türü olan tekrarlı sinir ağıdır ve sinir ağındaki hücreler arasındaki bilgi akışını kontrol etmek ve yönetmek için geçit mekanizmaları kullanır. GRU'nun yapısı, dizinin önceki bölümlerinden bilgi atmadan büyük veri dizilerinden bağımlılıkları uyarlamalı olarak öğrenmesini sağlar. Bu durum RNN'deki azalan gradyan (vanishing gradient) problemini çözen LSTM'dekine benzer kapı üniteleri ile elde edilir. Bu kapılar, her zaman adımında saklanacak veya atılacak bilgilerin düzenlenmesinden sorumludur.

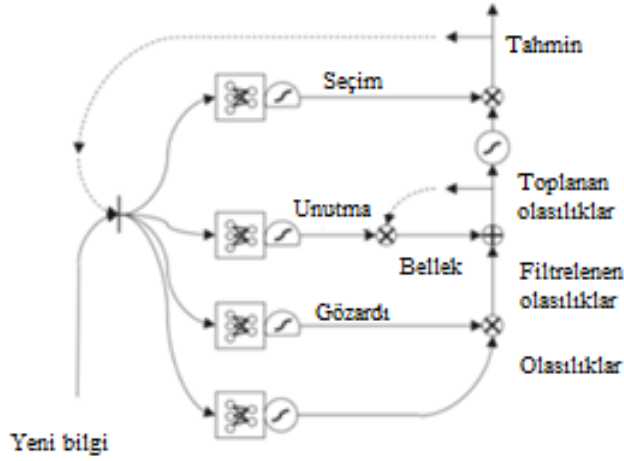
GRU'da unutma ve giriş kapılarını bir güncelleme geçidi olarak birleştirir ve ek olarak bir sıfırlama geçidi bulunmaktadır [95]. Her GRU düğümü güncelleme geçidi ve sıfırlama geçidi olmak üzere iki kapıdan oluşur. LSTM'lerdeki kapılar gibi GRU'daki bu kapılar da yararlı olanları korurken alakasız bilgileri seçici olarak filtrelemek için eğitilmiştir. Bu kapılar, giriş verileri veya gizli durum ile çarpılacak olan 0 ya da 1 değerlerini içeren vektörlerdir. Geçit vektörlerindeki 0 değeri, giriş veya gizli durumdaki verilerin önemsiz olduğunu ve bu nedenle sıfır olarak döneceğini gösterir. 1 değeri ise, karşılık gelen verilerin önemli olduğu ve kullanılacağı anlamına gelir. Güncelleme kapısı, birimin etkinliğini veya içeriğini ne kadar güncelleyeceğine karar verir. Sıfırlama kapısı ise daha önceden hesaplanan durumların unutulmasını sağlar [96]. Tüm ağ geri yayılma ile eğitildiğinde, denklemdeki ağırlıklar güncellenir, böylece vektör sadece yararlı özellikleri korumayı öğrenir.

4.4. LSTM

Yapay sinir ağıları, temel olarak giriş verileri kullanılarak çıkış verilerinin hesaplanmasını sağlamaktadır. Elde edilen çıkış değerlerinin, gerçek veriler ile karşılaştırılması sonucunda geriye yayılım yöntemi ile ağıın ağırlık değerleri üzerinde güncellemeler yapılmaktadır [97]. Ağ yapısının çok karmaşık olması durumunda ağırlık değerleri üzerinde güncellemeler yapılmaz ve ağıın eğitimi sonlanmış olur. Bu problem, ağıdaki önceki katmanların parametrelerini öğrenmeyi ve ayarlamayı zorlaştırmaktadır [84]. LSTM, tekrarlı sinir ağı yapılarındaki zaman içinde geriye bağılılık sorunlarına çözüm olarak geliştirilmiş bir sinir ağı mimarisidir [97].

LSTM, zaman içinde geriye yayılım kullanılarak eğitilen ve azalan gradyan problemini ortadan kaldıran tekrarlı bir sinir ağıdır. Azalan gradyan problemi, geriye yayılım temelli yöntemlerde eğitim aşamasında ilk katmanlardaki ağırlıkların değişiminin azalmasına yol açan problemidir [98]. LSTM, nöronlar yerine katmanlara bağılı bellek bloklarına sahiptir. Bellek bloklarının bileşenleri, blokları klasik bir nörondan daha akıllı hale getiren, bloğun durumunu ve çıkışını yöneten kapılar içermektedir. Bir blok, bir giriş dizisi üzerinde çalışır ve blok içindeki her bir kapı, tetiklenip tetiklenmediklerini kontrol etmek için sigmoid aktivasyon ünitelerini kullanır. Sigmoid aktivasyon üniteleri de durumun değişmesini ve blok koşulundan akan bilginin eklenmesini sağlar [85].

LSTM, veri dizisini öğrenmek için ek özelliklere sahip bir tür RNN'dir. Verilerin daha önceki eğiliminin öğrenilmesi Şekil 4.3'te görüldüğü gibi LSTM'e dâhil edilen bir bellek hattı ile birlikte diğer bileşenlerin kullanılması ile mümkün hale gelmiştir.



Şekil 4.3. Temel LSTM yapısı

LSTM, veri akışlarının elde edildiği ve saklandığı bir dizi hücre veya sistem modülüdür. Hücreler, bir modülden diğerine verileri ileten ve toplayan bir nakil hattına benzemektedir. Her hücrede bazı kapıların kullanılması nedeniyle hücrelerdeki veriler, bir sonraki hücreler için atılabilir, filtrelenebilir veya eklenebilir. Bu nedenle, sigmoidal sinir ağı katmanına dayanan kapılar, hücrelerin isteğine bağlı olarak verilerin geçmesini veya engellenmesini sağlar [97]. Her bir sigmoid katmanı, her hücrede geçirilmesine izin verilen veri bölümünün miktarını gösteren 0 ve 1 sayıları çıkarır. 0 değeri, hiçbir şeyin geçmesine izin verilmemesi, 1 değeri ise her şeyin geçmesine izin verilmesi anlamına gelmektedir [13].

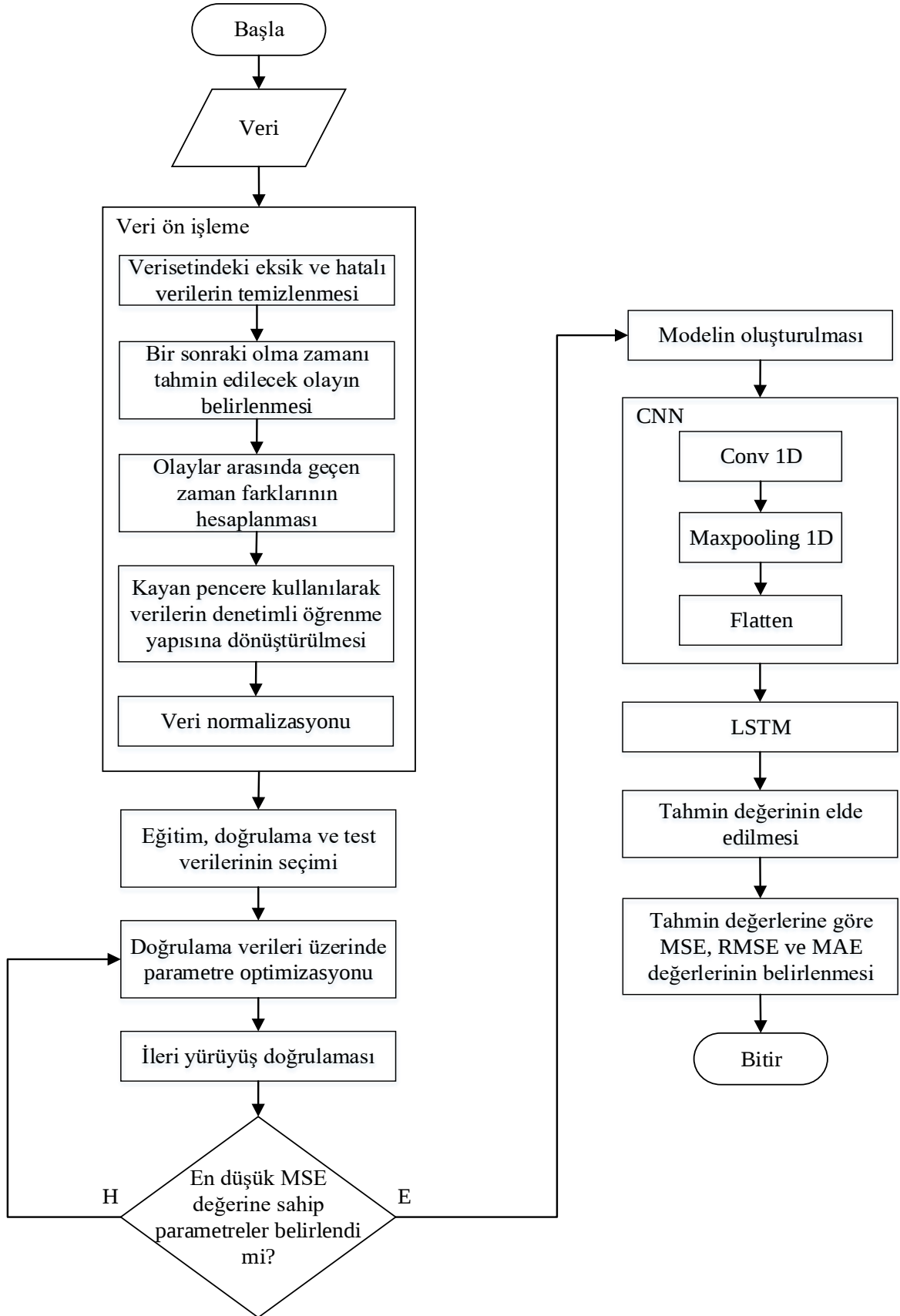
Sigmoid aktivasyon ünitelerinde her bir hücrenin durumunu kontrol etmek amacıyla giriş kapısı, çıkış kapısı ve unutma kapısı olmak üzere üç tip kapı bulunmaktadır. Giriş kapısı, girişteki hangi değerlerin bellek durumunu güncelleyeceğine koşullu olarak karar verir. Çıkış kapısı, girişe ve bloğun belleğine bağlı olarak neyin çıkış yapılacağına koşullu olarak karar verir. Unutma kapısı, şartlı olarak hangi bilgilerin bloktan atılacağına karar verir. Burada her bir ünite, ünite kapılarının eğitim sırasında öğrenilen ağırlıklara sahip olduğu bir durum makinesi gibidir [84].

5. GELİŞTİRİLEN HİBRİT DERİN ÖĞRENME MODELİ

Gerçek zamanlı veri analizi ve dinamik modelleme, işletme, ekonomi, finans ve bilgisayar bilimleri gibi birçok uygulama alanı bulunan bir araştırma konusudur [8]. Gerçek zamanlı veri analizinin amacı, zaman serilerindeki gözlemlenen veri değerlerini incelemek ve tanımlanan veri yapılarıyla zaman serilerinin gelecekteki değerlerini tahmin etmek için bir model oluşturmaktır [12]. Tahmin doğruluğunun iyileştirilmesi temel olarak oluşturulan modelin etkinliğine bağlıdır [78].

Bu tez çalışmasında, tahmin doğruluğunu ve etkinliğini arttırmak için CNN'in özellik çıkarma aşamasındaki başarısını, LSTM'in ise ardışık verilerde öğrenme ve tahmin aşamasındaki başarısını kullanan hibrit bir derin öğrenme modeli geliştirilmiştir. Geliştirilen model ile bir olayın bir sonraki olma zamanının belirlenmesi hedeflenmiştir. HDLM, bir e-ticaret sistemindeki kullanıcı etkileşimlerinden oluşan anonim veriler [99], DDoS saldırı verileri [100], suç verileri [101] ve acil çağrı verileri [102] kullanılarak RF, SVM, ARIMA, MLP, CNN, GRU, LSTM ve CNN+GRU ile karşılaştırmalı olarak analiz edilmiştir. Kullanıcı davranış analizi ile çok sayıda ürünün bulunduğu e-ticaret platformlarında kullanıcıları aşırı bilgi yükünden kurtarmak ve kullanıcı bağlılığını arttırmak hedeflenmiştir. Saldırı analizi ile Web tabanlı sistemlerin bütünlüğünü ve güvenliğini tehlikeye sokabilecek kötücül faaliyetlerin ortaya çıkma zamanından önce tespit edilmesi hedeflenmiştir. Suç analizi ile tehdit altındaki bölgelerin güvenliğini sağlamak, suç işleme oranını azaltmak ve kolluk kuvveti gücünden tasarruf etmek hedeflenmiştir. Benzer şekilde acil çağrı analizi ile çağrılara yanıt süresini kısaltmak ve kaynak kullanımını optimize etmek hedeflenmiştir.

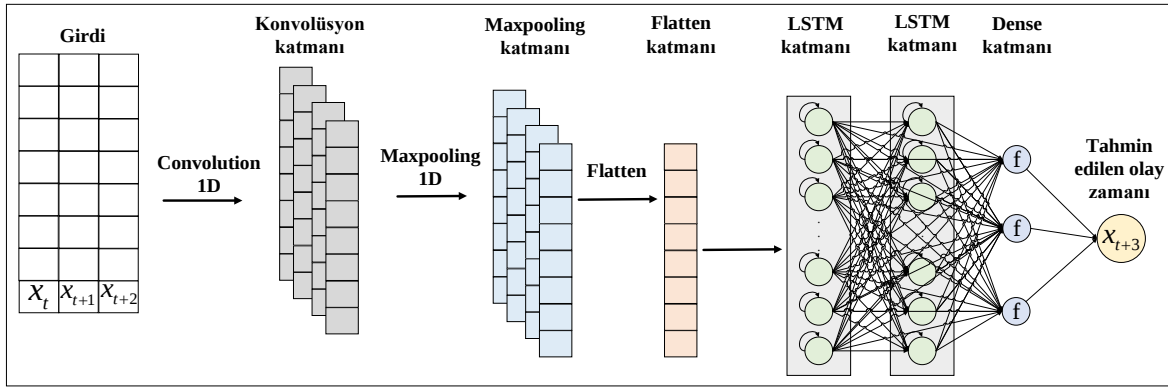
Modelin geliştirilmesi aşamasındaki işlem adımlarına ait akış diyagramı Şekil 5.1'de görülmektedir.



Şekil 5.1. Geliştirilen sistemin akış diyagramı

Bu çalışma kapsamında, geliştirilen derin öğrenme modeli kullanılarak bir olayın, bir sonraki olma zamanının belirlenmesi hedeflenmiştir. Geliştirilen modelde öncelikle, tarih ve saat bilgisi alınan sıralı olaylar arasındaki zaman farkları hesaplanmaktadır. Sistem, olayın tarihini ve saatini girdi olarak almaktadır. Sistemin çıktısı ise belirlenen olayın bir sonraki olma zamanıdır.

CNN, tek değişkenli zaman serisi gibi tek boyutlu dizi verilerinden otomatik olarak özellik çıkarmak ve öğrenmek için kullanılan etkili bir modeldir. Bu çalışmada, CNN, LSTM'in girdisi olan alt dizileri yorumlamak için kullanılarak hibrit bir model elde edilmiştir. Şekil 5.2'de HDLM mimarisi görülmektedir.



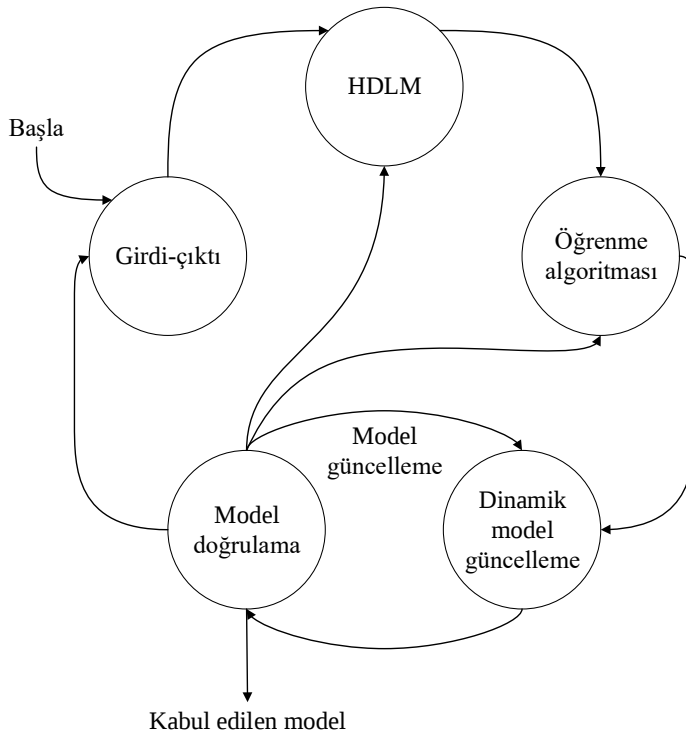
Şekil 5.2. HDLM mimarisi

Tahmin modelini desteklemek için geliştirilen HDLM mimarisi, girdi verilerinden özellik çıkarmak için LSTM ile birleştirilen CNN katmanlarını içermektedir. Ön işleme aşamasında ilk olarak, CNN giriş verilerinden önemli bilgileri çıkarır ve tek değişkenli giriş verilerini evrişim kullanarak çok boyutlu gruplara dönüştürür. İkinci aşamada, çok boyutlu veri grupları, tahmin için LSTM'e iletilir. Girdi verileri LSTM'e iletildikten sonra olayın bir sonraki zamanının belirleneceği tahmin aşamasına geçilmektedir.

Geliştirilen hibrit modelde CNN özellik çıkarımı aşamasında, LSTM ise CNN tarafından çıkarılan özellikleri analiz etmek ve tahmin için kullanılmaktadır. Geliştirilen hibrit modelin, bir olayın bir sonraki olma zamanının belirlenmesi probleminde kullanılabilmesi için ilk adım, giriş dizilerinin CNN modeli tarafından işlenebilecek alt dizilere bölünmesidir. Tek

değişkenli zaman serisi verileri, 3 adet girdi ve bir adet çıktı şeklinde girdi/çıktı örneklerine ayrılabilir.

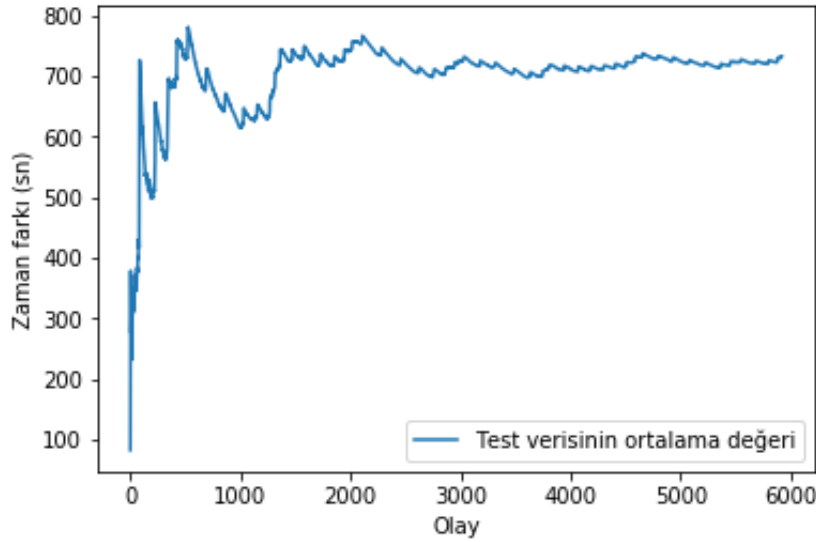
CNN, bu alt dizi örneklerini yorumlayarak girdi olarak işlenmesi için LSTM'e iletmektedir. Burada CNN modeli alt dizileri okumak için kernel boyutu 1 ve filtre sayısı 64 olan tek boyutlu bir evrişimli katmana sahiptir. Filtrelerin sayısı, giriş dizisinin okunma sayısını ifade etmektedir. Evrişimli katmanın ardından, giriş özelliğini yorumlamak için kullanılan bir max pooling katmanı ve modelin evrişim katmanı tarafından çıkarılan özellikleri yorumlayan dense katmanı tanımlanmıştır. Evrişim ve pooling katmanları 3 boyutlu katmanlardır. Bu nedenle özellik haritalarının LSTM'e giriş olarak kullanılacak tek boyutlu bir vektöre indirgenmesi için flatten katmanı kullanılmaktadır. Geliştirilen modelleme yaklaşımı Şekil 5.3'te görülmektedir.



Şekil 5.3. Geliştirilen modelleme yaklaşımı

Geliştirilen hibrit derin öğrenme modelinde, öğrenme algoritması dinamik olarak güncellenen modele dayanmaktadır. Parametreler, modelin doğrulanması sırasında optimize edilmektedir. En düşük hata oranına sahip model doğrulandıktan sonra öğrenme ve tahmin yapılmaktadır.

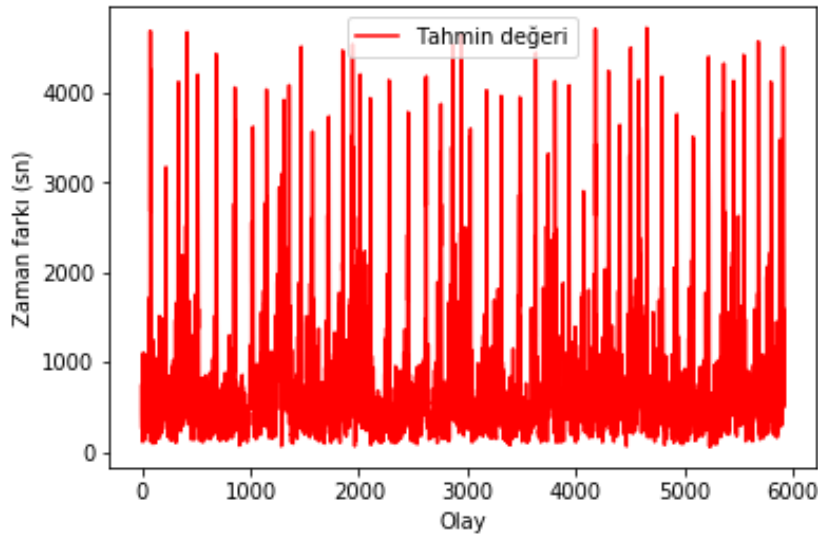
Tanımlama algoritmasının yakınsama noktası zaman serisi verilerinin ortalama değeridir. Algoritma, her döngü için t_n 'e yakınsayarak bir sonraki olay zamanını tahmin eder. Örnek olarak e-ticaret veri kümesindeki test verileri için, satın alma olayları arasında geçen zamanın ortalama değerleri Şekil 5.4'te görülmektedir.



Şekil 5.4. Satın alma olayları arasında geçen zamanın ortalama değerleri

Şekil 5.4'te, e-ticaret veri kümesindeki test verileri için, satın alma olayları arasında geçen zamanın ortalama değerleri görülmektedir. Satın alma olayları arasında geçen zamanların ortalaması belirli bir ortalama değere yakınsamaktadır.

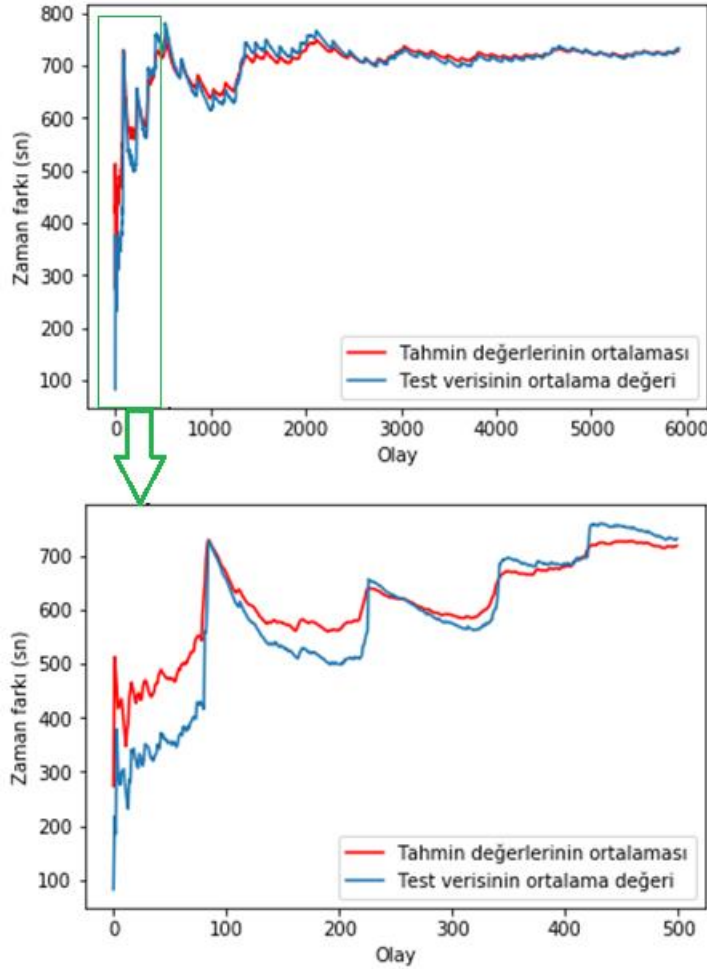
HDLM tarafından tahmin edilen değerler ise Şekil 5.5'te görülmektedir.



Şekil 5.5. HDLM tarafından yapılan tahmin değerleri

Şekil 5.5'te HDLM tarafından yapılan tahmin değerleri görülmektedir. Satın alma olaylarının meydana gelme zamanlarındaki düzensizlik nedeniyle yapılan tahmin değerlerindeki sapmalar şekilde görülmektedir.

Şekil 5.6'da, tahmin edilen değerlerin ortalaması ile test verisinin ortalaması görülmektedir.



Şekil 5.6. Tahmin edilen değerlerin ortalaması ve test verisinin ortalaması

Şekil 5.6’da Tahmin edilen değerlerin ortalaması ve test verisinin ortalaması görülmektedir. Mavi çizgi ile gösterilen değerler ortalama değeri, kırmızı çizgi ile gösterilen değerler ise tahmin edilen değerleri ifade etmektedir. Test verilerinin ortalama değeri, belirli bir ortalama değere yakınsamaktadır. Tahmin edilen değerlerin ortalaması da test verilerinin ortalamasına yakınsamaktadır. Bu durum, geliştirilen modelin tahmin doğruluğunu ve tutarlılığını göstermektedir.

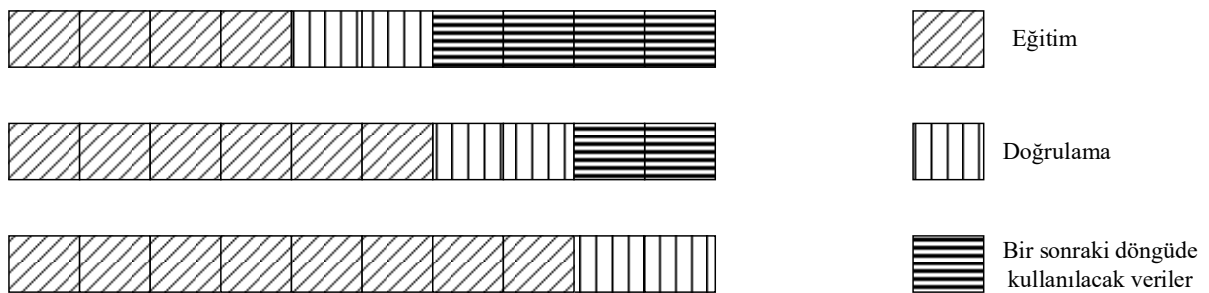
5.1. Kullanılan Veri kümeleri

Bu çalışmada bir olayın, bir sonraki olma zamanının belirlenmesi hedeflenmiştir. Bu nedenle kullanılacak veri kümelerinin zaman ve olay bilgisini içermesi gereklidir. Bu çalışma kapsamında e-ticaret, DDoS saldırı, suç ve acil çağrı veri kümeleri kullanılmıştır. Her bir veri kümesi için öncelikle veri ön işleme işlemleri gerçekleştirilmiştir. Veri kümelerindeki

eksik ve hatalı veriler kontrol edilerek temizlenmiştir. E-ticaret veri kümesinde satın alma davranışının önemi nedeniyle olay olarak satın alma etkinliği seçilmiştir. Suç ve acil çağrı veri kümelerinde ise veri kümelerindeki dağılımlarının büyüklüğü nedeniyle çete saldırısı ve trafik kazası olay olarak seçilmiştir. Olay seçiminden sonra belirlenen olayların meydana gelme zamanları arasındaki saniye ve dakika cinsinden zaman farkları hesaplanmıştır. Zaman serisi verilerinin denetimli öğrenme verilerine dönüştürülebilmesi için boyutu 3 olan kayan pencere kullanılmıştır. Bu sayede t_1 , t_2 ve t_3 zamanındaki veriler girdi, t_4 zamanındaki veri ise çıktı olacak şekilde yapılandırılmıştır.

Daha sonra olaylar arasındaki zaman farkları normalize edilmiştir. Normalleştirme, makine öğrenmesi için genellikle veri hazırlamanın bir parçası olarak uygulanan bir tekniktir. Normalleştirmenin amacı, veri kümesindeki sayısal sütunların değerlerini, değerler aralığındaki farklılıkları bozmadan ortak bir ölçeğe değiştirmektir. Normalizasyon geliştirilen modelin performansını ve eğitimin kararlılığını etkilemektedir. Veriler, 0-1 aralığında yeniden ölçeklendirilmiştir. Normalizasyon için verideki maksimum ve minimum değerler belirlenerek $y = (x - \min) / (\max - \min)$ eşitliği ile x verisi için normalize edilmiş değer olan y elde edilmiştir.

Normalize edilen verilerin %67'si eğitim ve %33'ü test için ayrılmıştır. Eğitim verilerinin %10'u doğrulama için ayrılmıştır. Doğrulama verileri model parametrelerinin optimizasyonu için kullanılmaktadır. Şekil 5.7'de görüldüğü gibi ileri yürüyüş doğrulaması ile doğrulama verileri üzerinde parametre optimizasyonu yapıldıktan sonra en düşük MSE değerine sahip parametreler belirlenerek model oluşturulmuştur.



Şekil 5.7. İleri yürüyüş doğrulaması

İleri yürüyüş doğrulaması, çapraz doğrulama yönteminin zaman serisi için uyarlanmış bir halidir. Çapraz doğrulama, parametreleri ayarlamak ve model performansının ölçümünü yapmak için kullanılan bir tekniktir. Çapraz doğrulamada model eğitim kümesi ile eğitilir ve doğrulama kümesinde hatayı en aza indiren parametreler seçilir. Son olarak, seçilen parametreler kullanılarak model tam eğitim seti üzerinde eğitilir ve test setindeki hata kaydedilir. Çapraz doğrulama tekniği, zaman serisi verilerindeki zamana olan bağımlılık nedeniyle kullanılamamaktadır. İleri yürüyüş doğrulaması yöntemi, Şekil 5.7’de görüldüğü gibi test veri kümesinin elemanlarının her döngü sonunda eğitim kümesine dâhil edilerek model parametrelerinin ayarlanmasını sağlamaktadır.

Eğitim verileri LSTM’in girdisi olan alt dizileri yorumlamak için kullanılacak olan CNN’e iletilmiştir. CNN giriş verilerinden özellik çıkarımı yaparak tek değişkenli giriş verilerini evrişim kullanarak çok boyutlu gruplara dönüştürür. İkinci aşamada, çok boyutlu veri grupları flatten katmanı ile tek boyuta indirgenerek tahmin için LSTM’e iletilir. Girdi verileri LSTM’e iletildikten sonra tahmin değerleri elde edilmekte ve sonuçlar test verisi ile karşılaştırılarak MSE, RMSE ve MAE değerleri hesaplanmaktadır.

E-ticaret veri kümesinde film görüntüleme, sepete ekleme ve satın alma olmak üzere üç farklı kullanıcı davranışı bulunmaktadır. Bu veri kümesi 2756101 satır kullanıcı davranış verisinden oluşmaktadır. Davranış verileri, 4,5 aylık bir süre boyunca toplanmıştır. Toplamda 2664312 adet film görüntüleme, 69332 adet sepete ekleme ve 17939 adet satın alma olmak üzere 1407580 tekil ziyaretçi tarafından üretilen toplam 2756101 etkinlik bulunmaktadır.

Çizelge 5.1’de, e-ticaret veri kümesindeki kullanıcı etkileşimlerinin ilk 5 satırı görülmektedir.

Çizelge 5.1. E-ticaret veri kümesi

Zaman bilgisi	Ziyaretçi ID	Olay	Film ID	Satın alma ID
1433221332117	257597	Görüntüleme	355908	-
1433224214164	992329	Görüntüleme	248676	-
1433221999827	111016	Görüntüleme	318965	-

Çizelge 5.1. (devam) E-ticaret veri kümesi

1433221955914	483717	Görüntüleme	253185	-
1433221337106	951259	Görüntüleme	367447	-

Örnek olarak veri kümesindeki “1433221332117,257597,view,355908,” satırı 257597 numaralı kullanıcının, 355908 numaralı filmi 1433221332117 (02/06/2015, 08:02:12) zamanında görüntülediğini ifade etmektedir.

DDoS saldırı verileri, Hacking and Countermeasure Research Lab tarafından sunulan verileri içermektedir. Bu veri kümesi 10000 saldırı olayı arasındaki saniye cinsinden zaman farklarından oluşmaktadır.

Çizelge 5.2’de, DDoS saldırılarının meydana gelmesi arasındaki saniye cinsinden farklılıklara göre üretilen veri kümesinin ilk 5 satırı görülmektedir.

Çizelge 5.2. DDoS saldırıları arasındaki zaman farkı

Saldırı	Bir önceki saldırıdan sonra geçen zaman (sn)
1	271
2	224
3	241
4	247
5	256

Suç verileri ise 285808 satır ve 12 sütundan oluşmaktadır. Suç işlenme tarihi, saati, kodu, mekânı, açıklaması, kapalı mekân/açık alan, silah türü, posta kodu, semt, yakın komşuluk, konum ve olay sayısı sütunları bulunmaktadır. Bu tez çalışmasında veri kümesindeki suçun işlenme tarihi, saati ve kodu alanları kullanılmıştır.

Çizelge 5.3’te ise suç tarihi, saati ve suç koduna göre düzenlenmiş suç veri kümesinin ilk 5 satırı görülmektedir.

Çizelge 5.3. Baltimor şehrindeki suç verileri

Suç tarihi	Suç saati	Suç kodu
2011-01-01	00:00:00	6E
2011-01-01	00:00:00	6G
2011-01-01	00:01:00	6J
2011-01-01	00:01:00	6G
2011-01-01	00:05:00	4B

Acil çağrı veri kümesi 423909 satır ve 8 sütundan oluşmaktadır. Veri kümesinde acil çağrı yapılan bölgenin enlem ve boylam bilgileri, acil durum hakkındaki açıklama bilgisi, posta kodu, acil çağrının dâhil olduğu kategori, tarih ve saat bilgisi, kasaba ve adres sütunları bulunmaktadır. Bu tez çalışmasında, veri kümesindeki acil çağrının alındığı tarih, saat ve kategori alanları kullanılmıştır.

Çizelge 5.4'te ise acil çağrının alındığı tarih, saat ve kategori alanlarına göre düzenlenmiş acil çağrı veri kümesinin ilk 5 satırını görülmektedir.

Çizelge 5.4. Acil çağrı verileri

Tarih	Saat	Kategori
10-12-2015	14:39:21	Yangın
10-12-2015	16:17:05	Kafa travması
10-12-2015	16:32:10	Trafik kazası
10-12-2015	16:46:48	Bulantı/kusma
10-12-2015	16:47:36	Kalp krizi

5.2. Veri Önışleme

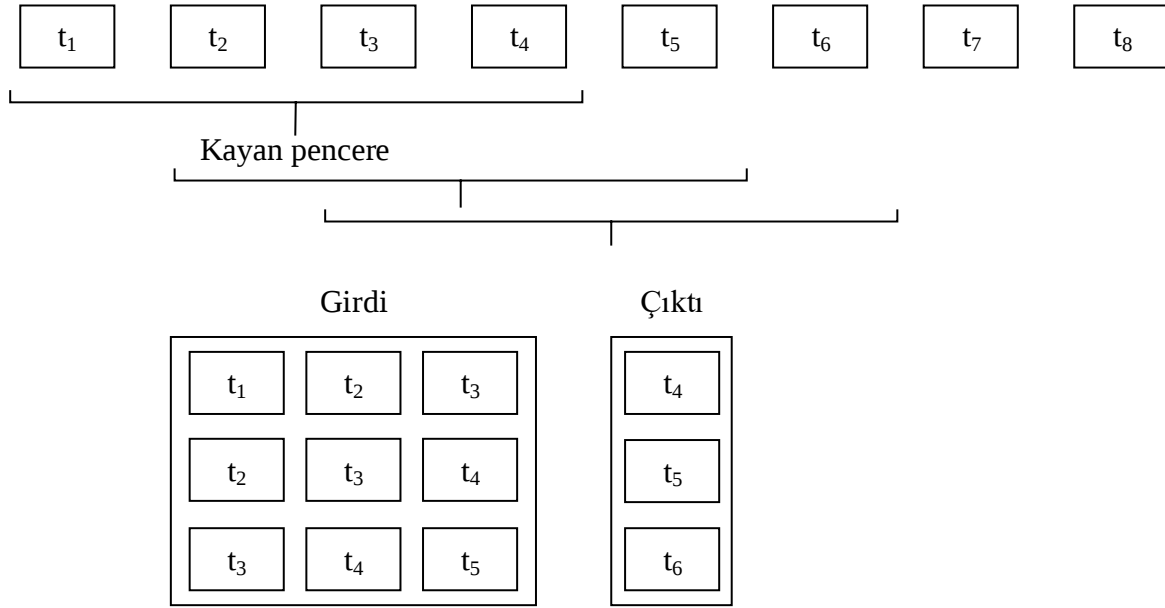
Geliştirilen uygulamada olay zamanı tahmininin yapılabilmesi için özel bir olay durumunun belirlenmesi gereklidir. Bu nedenle, satın alma, sepete ekleme ve film görüntüleme etkinliklerinden, satın alma etkinliği e-ticaret sistemlerindeki önemi nedeniyle olay olarak seçilmiştir. Eksik ve tekrarlı veriler temizlendikten sonra veri kümesinde 17939 adet satın

alma etkinliđi bulunmaktadır. DDoS saldırı verileri ise DDoS saldırıları arasında geen zaman farklarından oluřtuđu iin bu veri kumesi iin zel olarak bir olay belirlenmesine gerek kalmamıřtır.

Su veri kumesinde eksik ve tekrarlı veriler atıldıktan sonra 256688 satır olay verisi elde edilmiřtir. 81 adet su tipinden oluřan veri kumesinde su koduna gre yapılan inceleme sonucunda 41967 4E, 32523 adet 6D, 24328 adet 5A, 22482 adet 7A ve 15324 adet 6G olayı bulunmaktadır. 4E kodu ete saldırısını, 6D kodu ara ii soygunu, 5A kodu ev soygununu, 7A kodu alıntı ara olayını ve 6G kodu ise iřyerinden hırsızlık olayını ifade etmektedir. Geri kalan 120064 olay ise 76 adet su tipinin toplam sayısıdır. 4E kodlu ete saldırısı olayı, veri kumesindeki dađılımının byklđ nedeniyle olay olarak seilmiřtir.

Acil ađrı veri kumesi, eřitli kategorilerdeki 423909 acil durum verisinden oluřmaktadır. 141 adet olay trnden oluřan veri kumesinde olay trne gre yapılan inceleme sonucunda, 98401 adet ile veri kumesindeki diđer acil ađrılara gre daha fazla sayıda gerekleřmiř olan trafik kazası ađrıları olay olarak seilmiřtir.

Zaman serileri, bir zaman dizini tarafından sıralanan sayı dizilerini ifade etmektedir. Denetimli đrenme problemleri ise bir algoritmanın, girdi deđiřkenlerinden (x) ıktı deđiřkenlerinin (y) nasıl tahmin edileceđini đrenebileceđi problemlerdir. $y=f(x)$ şeklinde kullanılan bir fonksiyon ile giriř verilerinden ıkıř verilerinin tahmin edilmesi amalanmaktadır. Zaman serisi verileri, giriř deđiřkeni olarak belirlenen zaman adımlarından bir sonraki zaman adımındaki ıkıř deđiřkeninin belirlenmesi řeklinde denetimli đrenme problemine dnřtrlebilir.



Şekil 5.8. Zaman serisi verilerinin denetimli öğrenme verilerine dönüştürülmesi

Örneğin, Şekil 5.8’de görülmekte olan zaman serisi verileri, bir sonraki zaman adımındaki değeri tahmin etmek için önceki zaman adımındaki değerler kullanılarak denetimli öğrenme problemi şeklinde yapılandırılabilir. Şekil 5.8’de, bir önceki zaman adımının girdisinin (x), denetimli öğrenme problemindeki bir sonraki zaman adımının çıktısı (y) olduğu görülmektedir.

Kayan pencere yöntemi bir sonraki zaman adımını tahmin etmek için önceki zaman adımlarının kullanımına dayalı olan bir yöntemdir. Kayan pencerenin boyutu, önceki zaman adımlarının sayısına göre belirlenmektedir. Bu çalışma kapsamında kayan pencere boyu 3 olarak seçilmiştir.

5.3. Hiperparametre Seçimi

Çalışma kapsamında kullanılan modellerde kullanılacak parametrelerin belirlenmesi için bir analiz çalışması yapılmıştır. Kullanılan veri kümelerinin öncelikle %67’si eğitim, %33’ü ise test için ayrılmıştır. Eğitim verilerinin %10’u ise parametre seçimi aşamasında kullanılmak üzere doğrulama için ayrılmıştır.

E-ticaret veri kümesi için parametre seçimi

Veri kümesi içinde satın alma olayı 17939 defa gerçekleşmiştir. Veri kümesinin %67'si eğitim, %33'ü test olmak üzere ayrılmıştır. Ayırma sonucunda toplam 12019 elemandan oluşan bir eğitim verisi elde edilmiştir. Eğitim verisinin %10'u hiper parametre seçiminde kullanılmak üzere doğrulama verisi olarak ayrılmıştır. Ayırma sonucunda 1201 elemandan oluşan bir doğrulama verisi elde edilmiştir. RF, SVM, ARIMA, MLP, CNN, GRU, LSTM, CNN+GRU ve HDLM'nin alacağı hiper parametrelerin belirlenmesi için doğrulama verileri üzerinde bir analiz çalışması yapılmıştır.

RF modelinin alacağı parametrelerinin belirlenmesi için Çizelge 5.5'te görüldüğü gibi MSE metriği kullanılarak bir analiz çalışması yapılmıştır.

Çizelge 5.5. RF parametre seçimi

Ağaç sayısı	İç düğümü bölmek için gereken örnek sayısı	MSE
100	2	1654738,83
300	2	1601182,39
500	2	*1598352,08
1000	2	1599545,16
2000	2	1598541,92

* Seçilen parametreler

Çizelge 5.5'te görüldüğü gibi RF için yapılan parametre analizinde en düşük MSE değerine sahip parametreler belirlenmiştir. Ağaç sayısının 500 olduğu durumda en düşük MSE değerine ulaşıldığı için bu parametreler deneysel çalışmalarda kullanılmak üzere seçilmiştir.

SVM modelinin alacağı parametrelerinin belirlenmesi için Çizelge 5.6'da görüldüğü gibi MSE metriği kullanılarak bir analiz çalışması yapılmıştır.

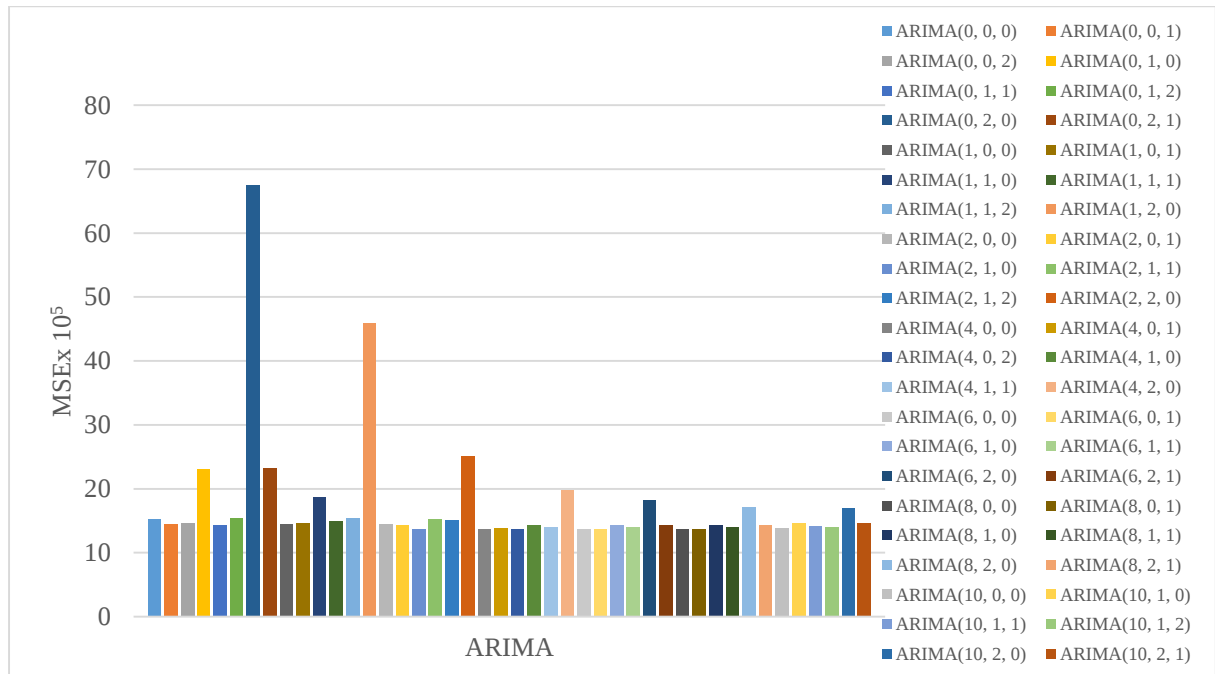
Çizelge 5.6. SVM parametre seçimi

c	e	Kernel	MSE
1,5	0,1	rbf	1494736,21
1,0	0,1	rbf	1491480,54
0,5	0,1	rbf	1493446,28
0,5	0,2	rbf	1478754,48
0,5	0,3	rbf	*1474363,02
0,5	0,5	rbf	1602093,75

* Seçilen parametreler

Çizelge 5.6’da görüldüğü gibi SVM için yapılan parametre analizinde $c = 0,5$ ve $e = 0,3$ değerleri en düşük MSE değerine sahip olduğu için bu parametreler deneysel çalışmalarda kullanılmak üzere seçilmiştir.

Doğrulama verisi üzerinde ARIMA’nın alacağı p , d ve q parametrelerinin belirlenmesi için Şekil 5.9’da görüldüğü gibi MSE metriği kullanılarak bir analiz çalışması yapılmıştır.



Şekil 5.9. ARIMA parametre seçimi

Analiz sonuçları ARIMA(6, 0, 0) modelinin en düşük MSE değerine sahip olduğunu gösterdiği için bu model deneysel çalışmalarda kullanılmak üzere seçilmiştir.

MLP modelinin alacağı parametrelerinin belirlenmesi için Çizelge 5.7’de görüldüğü gibi MSE metriği kullanılarak bir analiz çalışması yapılmıştır.

Çizelge 5.7. MLP parametre seçimi

Yığın boyutu	Katman sayısı	Eğitim sayısı	MSE
2	8	100	*1397314,21
2	16	200	1397638,23
2	32	100	1398812,23
4	8	100	1398931,42
16	8	100	1398226,47
64	8	100	1399076,13
256	8	100	1400119,57

* Seçilen parametreler

Analiz sonuçları MLP’nin en düşük MSE değerine yığın boyutu 2, katman sayısı 8 ve eğitim sayısı 100 olduğunda sahip olduğunu gösterdiği için, bu değerler deneysel çalışmalarda kullanılmak üzere seçilmiştir.

CNN’in yığın boyutu, eğitim sayısı ve gizli katmandaki nöron sayısı gibi parametrelerin belirlenebilmesi için Çizelge 5.8’de görüldüğü gibi MSE metriği kullanılarak bir analiz çalışması yapılmıştır.

Çizelge 5.8. CNN parametre seçimi

Filtre sayısı	Eğitim sayısı	Kernel sayısı	MSE
32	200	1	*1410249,29
64	100	2	1596508,06
64	200	2	1509089,40
64	300	2	1511621,07

Çizelge 5.8. (devam) CNN parametre seçimi

16	200	2	1534501,56
32	200	2	1536087,57
32	200	3	1430266,59
64	200	3	1442160,81

* Seçilen parametreler

Analiz sonuçları CNN'in en düşük MSE değerine filtre sayısı 32, eğitim sayısı 200 ve kernel sayısı 1 olduğunda sahip olduğunu gösterdiği için, bu değerler deneysel çalışmalarda kullanılmak üzere seçilmiştir.

GRU'nun yığın boyutu, eğitim sayısı ve gizli katmandaki nöron sayısı gibi parametrelerin belirlenebilmesi için Çizelge 5.9'da görüldüğü gibi MSE metriği kullanılarak bir analiz çalışması yapılmıştır.

Çizelge 5.9. GRU parametre seçimi

Yığın boyutu	Eğitim sayısı	Nöron sayısı	MSE
32	100	4	1467741,16
64	100	4	1559341,54
256	100	4	1521335,72
512	100	4	1507619,23
1024	100	4	1510972,31
512	100	8	1522436,16
512	100	16	1494887,49
512	100	32	*1456501,02
512	200	32	1492753,27

* Seçilen parametreler

Analiz sonuçları GRU'nun en düşük MSE değerine yığın boyutu 512, eğitim sayısı 100 ve gizli katmandaki nöron sayısı 32 olduğunda sahip olduğunu gösterdiği için, bu değerler deneysel çalışmalarda kullanılmak üzere seçilmiştir.

LSTM'in yığın boyutu, eğitim sayısı ve gizli katmandaki nöron sayısı gibi parametrelerin belirlenebilmesi için Çizelge 5.10'da görüldüğü gibi MSE metriği kullanılarak bir analiz çalışması yapılmıştır.

Çizelge 5.10. LSTM parametre seçimi

Yığın boyutu	Eğitim sayısı	Nöron sayısı	MSE
1	100	4	1426055,16
16	100	4	1442976,84
32	100	4	1437500,39
64	100	4	1414555,30
256	100	4	1392441,75
512	100	4	1398286,37
1024	100	4	1418868,37
512	100	8	1392404,08
512	100	16	*1387064,60
512	100	32	1394094,64
512	200	32	1388986,57

* Seçilen parametreler

Analiz sonuçları LSTM'in en düşük MSE değerine yığın boyutu 512, eğitim sayısı 100 ve gizli katmandaki nöron sayısı 16 olduğunda sahip olduğunu gösterdiği için bu değerler deneysel çalışmalarda kullanılmak üzere seçilmiştir.

CNN+GRU'nun yığın boyutu, eğitim sayısı ve gizli katmandaki nöron sayısı gibi parametrelerin belirlenebilmesi için Çizelge 5.11'de görüldüğü gibi MSE metriği kullanılarak bir analiz çalışması yapılmıştır.

Çizelge 5.11. CNN+GRU parametre seçimi

CNN filtre sayısı	Eğitim sayısı	GRU nöron sayısı	MSE
32	100	32	1031721,06
64	100	16	1036765,56

Çizelge 5.11. (devam) CNN+GRU parametre seçimi

64	50	32	979614,34
64	100	32	*976015,64
64	200	32	992819,07
64	100	50	999126,15

* Seçilen parametreler

Analiz sonuçları CNN+GRU'nun en düşük MSE değerine CNN filtre sayısı 64, eğitim sayısı 100 ve GRU gizli katmandaki nöron sayısı 32 olduğunda sahip olduğunu gösterdiği için, bu değerler deneysel çalışmalarda kullanılmak üzere seçilmiştir.

HDLM'nin yığın boyutu, eğitim sayısı ve gizli katmandaki nöron sayısı gibi parametrelerin belirlenebilmesi için Çizelge 5.12'de görüldüğü gibi MSE metriği kullanılarak bir analiz çalışması yapılmıştır.

Çizelge 5.12. HDLM parametre seçimi

CNN filtre sayısı	Eğitim sayısı	LSTM nöron sayısı	MSE
64	200	4	1269308,66
64	200	16	978489,39
32	200	32	888242,34
64	100	32	879505,45
64	200	32	*859538,92
64	100	50	1058917,26
64	200	50	958280,97
64	300	50	1053988,43

* Seçilen parametreler

Analiz sonuçları HDLM'nin en düşük MSE değerine CNN filtre sayısı 64, CNN kernel sayısı 1, eğitim sayısı 200 ve LSTM gizli katmandaki nöron sayısı 32 olduğunda sahip olduğunu gösterdiği için, bu değerler deneysel çalışmalarda kullanılmak üzere seçilmiştir.

DDoS veri kümesi için parametre seçimi

DDoS veri kümesi 10000 saldırı olayı arasındaki saniye cinsinden zaman farklarından oluşmaktadır. Verilerin %67'si eğitim, %33'ü test olarak ayrılmıştır. Ayırma sonucunda toplam 6700 elemandan oluşan bir eğitim verisi elde edilmiştir. Eğitim verisinin %10'u parametre seçiminde kullanılmak üzere doğrulama verisi olarak ayrılmıştır. Ayırma sonucunda 670 elemandan oluşan bir doğrulama verisi elde edilmiştir.

RF modelinin alacağı parametrelerinin belirlenmesi için Çizelge 5.13'te görüldüğü gibi MSE metriği kullanılarak bir analiz çalışması yapılmıştır.

Çizelge 5.13. RF parametre seçimi

Ağaç sayısı	İç düğümü bölmek için gereken örnek sayısı	MSE
100	2	1012331,26
300	2	1013547,75
500	2	*1012093,27
1000	2	1013356,11
2000	2	1013989,89

* Seçilen parametreler

Çizelge 5.13'te görüldüğü gibi RF için yapılan parametre analizinde en düşük MSE değerine sahip parametreler belirlenmiştir. Ağaç sayısının 500 olduğu durumda en düşük MSE değerine ulaşıldığı için bu parametreler deneysel çalışmalarda kullanılmak üzere seçilmiştir.

SVM modelinin alacağı parametrelerinin belirlenmesi için Çizelge 5.14'te görüldüğü gibi MSE metriği kullanılarak bir analiz çalışması yapılmıştır.

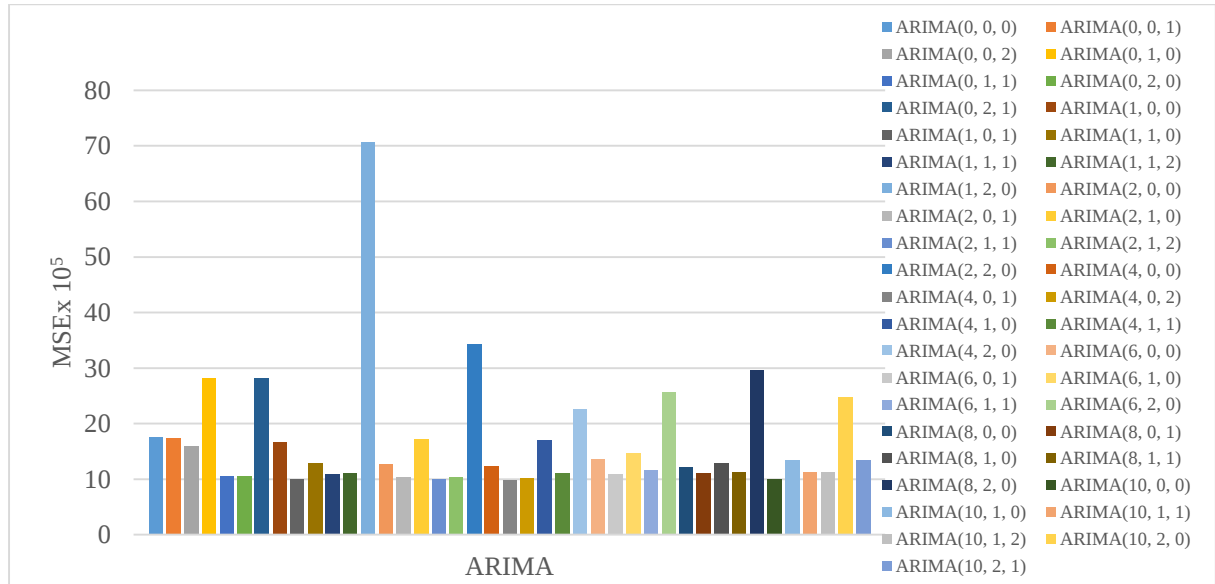
Çizelge 5.14. SVM parametre seçimi

c	e	Kernel	MSE
1,5	0,1	rbf	1066495.56
1,0	0,1	rbf	*1039622.68
0,5	0,1	rbf	1075247.91
0,5	0,2	rbf	1059487.25
0,5	0,3	rbf	1060596.21
0,5	0,5	rbf	1067497.28

* Seçilen parametreler

Çizelge 5.14'te görüldüğü gibi SVM için yapılan parametre analizinde $c = 1,0$ ve $e = 0,1$ değerleri en düşük MSE değerine sahip olduğu için bu parametreler deneysel çalışmalarda kullanılmak üzere seçilmiştir.

Doğrulama verisi üzerinde ARIMA'nın alacağı p , d ve q parametrelerinin belirlenmesi için Şekil 5.10'da görüldüğü gibi MSE metriği kullanılarak bir analiz çalışması yapılmıştır.



Şekil 5.10. ARIMA parametre seçimi

Analiz sonuçları ARIMA(4,0,1) modelinin en düşük MSE değerine sahip olduğunu gösterdiği için bu model deneysel çalışmalarda kullanılmak üzere seçilmiştir.

MLP modelinin alacağı parametrelerinin belirlenmesi için Çizelge 5.15'te görüldüğü gibi MSE metriği kullanılarak bir analiz çalışması yapılmıştır.

Çizelge 5.15. MLP parametre seçimi

Yığın boyutu	Katman sayısı	Eğitim sayısı	MSE
32	32	100	926605,06
16	50	100	932118,55
32	50	50	891776,20
32	50	100	*853153,15
32	50	200	913254,47
64	50	100	889263,78
128	50	100	925412,54

* Seçilen parametreler

Analiz sonuçları MLP'nin en düşük MSE değerine yığın boyutu 32, katman sayısı 50 ve eğitim sayısı 100 olduğunda sahip olduğunu gösterdiği için, bu değerler deneysel çalışmalarda kullanılmak üzere seçilmiştir.

CNN'in yığın boyutu, eğitim sayısı ve gizli katmandaki nöron sayısı gibi parametrelerin belirlenebilmesi için Çizelge 5.16'da görüldüğü gibi MSE metriği kullanılarak bir analiz çalışması yapılmıştır.

Çizelge 5.16. CNN parametre seçimi

Filtre sayısı	Eğitim sayısı	Kernel sayısı	MSE
32	200	1	970694,30
64	200	2	972511,57
16	200	2	966384,37
32	200	2	973083,04
32	200	3	*840750,51
64	200	3	856220,06

* Seçilen parametreler

Analiz sonuçları CNN'in en düşük MSE değerine filtre sayısı 32, eğitim sayısı 200 ve kernel sayısı 3 olduğunda sahip olduğunu gösterdiği için, bu değerler deneysel çalışmalarda kullanılmak üzere seçilmiştir.

GRU'nun yığın boyutu, eğitim sayısı ve gizli katmandaki nöron sayısı gibi parametrelerin belirlenebilmesi için Çizelge 5.17'de görüldüğü gibi MSE metriği kullanılarak bir analiz çalışması yapılmıştır.

Çizelge 5.17. GRU parametre seçimi

Yığın boyutu	Eğitim sayısı	Nöron sayısı	MSE
8	100	4	911245,11
32	100	50	908871,36
64	100	32	902145,63
64	50	50	867968,47
64	100	50	*825952,26
64	200	50	841674,32
64	100	64	889472,31

* Seçilen parametreler

Analiz sonuçları GRU'nun en düşük MSE değerine yığın boyutu 64, eğitim sayısı 100 ve gizli katmandaki nöron sayısı 50 olduğunda sahip olduğunu gösterdiği için, bu değerler deneysel çalışmalarda kullanılmak üzere seçilmiştir.

LSTM'in yığın boyutu, eğitim sayısı ve gizli katmandaki nöron sayısı gibi parametrelerinin belirlenebilmesi için Çizelge 5.18'de görüldüğü gibi MSE metriği kullanılarak bir analiz çalışması yapılmıştır.

Çizelge 5.18. LSTM parametre seçimi

Yığın boyutu	Eğitim sayısı	Nöron sayısı	MSE
1	100	4	874228,60
16	100	4	860170,87
32	100	4	857779,34
64	100	4	829292,80
128	100	4	845326,37
256	100	4	854450,07
512	100	4	862407,37
1024	100	4	861636,69
64	100	8	833872,96
64	100	16	840276,05
64	100	32	*822274,33
64	100	64	850563,65
64	200	32	834293,09

* Seçilen parametreler

Analiz sonuçları LSTM'in en düşük MSE değerine yığın boyutu 64, eğitim sayısı 100 ve gizli katmandaki nöron sayısı 32 olduğunda sahip olduğunu gösterdiği için, bu değerler deneysel çalışmalarda kullanılmak üzere seçilmiştir.

CNN+GRU'nun yığın boyutu, eğitim sayısı ve gizli katmandaki nöron sayısı gibi parametrelerin belirlenebilmesi için Çizelge 5.19'da görüldüğü gibi MSE metriği kullanılarak bir analiz çalışması yapılmıştır.

Çizelge 5.19. CNN+GRU parametre seçimi

CNN filtre sayısı	Eğitim sayısı	GRU nöron sayısı	MSE
32	100	50	873594,72
64	100	16	873561,78
64	100	32	872349,47
64	50	50	871235,69

Çizelge 5.19. (devam) CNN+GRU parametre seçimi

64	100	50	*858937,74
64	200	50	865721,32
64	100	64	863574,12

* Seçilen parametreler

Analiz sonuçları CNN+GRU'nun en düşük MSE değerine CNN filtre sayısı 64, eğitim sayısı 100 ve GRU gizli katmandaki nöron sayısı 50 olduğunda sahip olduğunu gösterdiği için, bu değerler deneysel çalışmalarda kullanılmak üzere seçilmiştir.

HDLM'nin yığın boyutu, eğitim sayısı ve gizli katmandaki nöron sayısı gibi parametrelerin belirlenebilmesi için Çizelge 5.20'de görüldüğü gibi MSE metriği kullanılarak bir analiz çalışması yapılmıştır.

Çizelge 5.20. HDLM parametre seçimi

CNN filtre sayısı	Eğitim sayısı	LSTM nöron sayısı	MSE
64	100	4	904073,76
64	100	16	794290,20
64	100	32	829094,61
64	100	50	*769188,95
64	100	64	831171,90
64	200	16	832549,27
64	200	50	778124,59

* Seçilen parametreler

Analiz sonuçları HDLM'nin en düşük MSE değerine CNN filtre sayısı 64, CNN kernel sayısı 1, eğitim sayısı 100 ve LSTM gizli katmandaki nöron sayısı 50 olduğunda sahip olduğunu gösterdiği için, bu değerler deneysel çalışmalarda kullanılmak üzere seçilmiştir.

Suç veri kümesi için parametre seçimi

Suç veri kümesi 41967 adet 4E kodlu çete saldırısı olayından oluşmaktadır. Verilerin %67'si eğitim, %33'ü test olarak ayrılmıştır. Eğitim verilerinin %10'u doğrulama için ayrılmıştır. Ayırma sonucunda toplam 28117 elemandan oluşan bir eğitim verisi elde edilmiştir. Eğitim verisinin %10'u parametre seçiminde kullanılmak üzere doğrulama verisi olarak ayrılmıştır. Ayırma sonucunda 2811 elemandan oluşan bir doğrulama verisi elde edilmiştir.

RF modelinin alacağı parametrelerinin belirlenmesi için Çizelge 5.21'de görüldüğü gibi MSE metriği kullanılarak bir analiz çalışması yapılmıştır.

Çizelge 5.21. RF parametre seçimi

Ağaç sayısı	İç düğümü bölmek için gereken örnek sayısı	MSE
100	2	32996465,63
200	2	32991396,77
500	2	32986527,77
1000	2	*32953302,88
2000	2	32953752,70
3000	2	32968080,85

* Seçilen parametreler

Çizelge 5.21'de görüldüğü gibi RF için yapılan parametre analizinde en düşük MSE değerine sahip parametreler belirlenmiştir. Ağaç sayısının 1000 olduğu durumda en düşük MSE değerine ulaşıldığı için bu parametreler deneysel çalışmalarda kullanılmak üzere seçilmiştir.

SVM modelinin alacağı parametrelerinin belirlenmesi için Çizelge 5.22'de görüldüğü gibi MSE metriği kullanılarak bir analiz çalışması yapılmıştır.

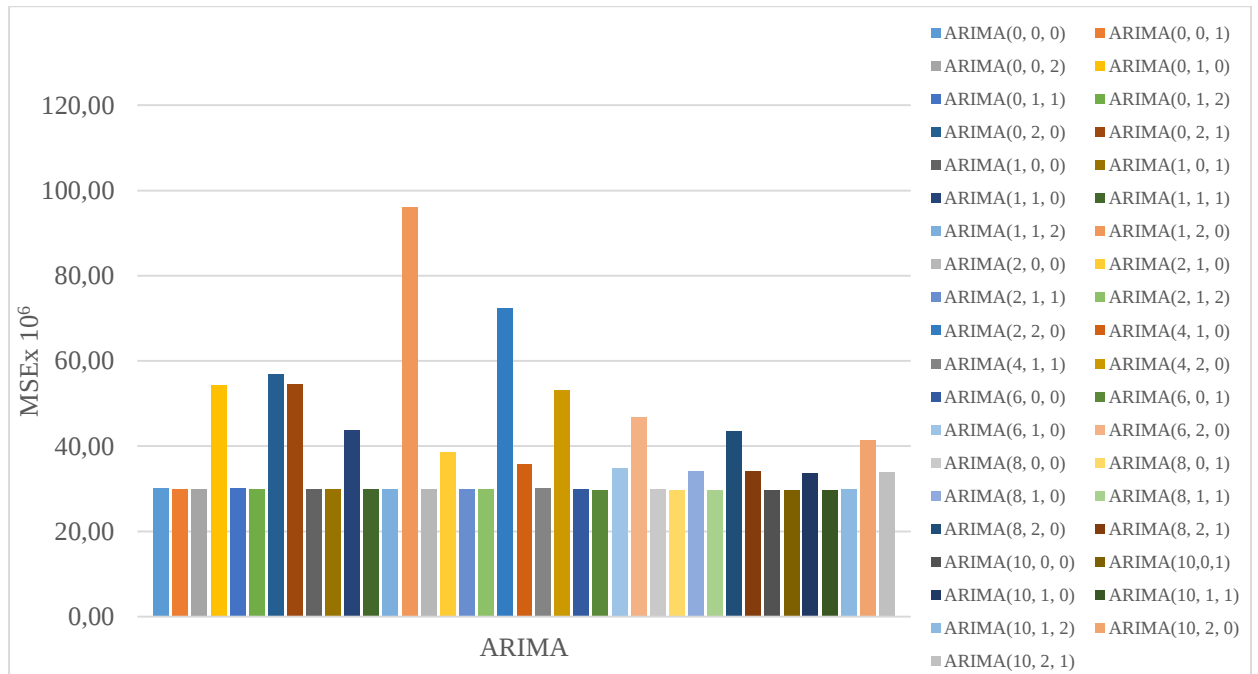
Çizelge 5.22. SVM parametre seçimi

c	e	Kernel	MSE
1,5	0,1	rbf	32252233,93
1,0	0,1	rbf	*32246176,95
0,5	0,1	rbf	32213028,75
0,5	0,2	rbf	31769881,60
0,5	0,3	rbf	31111228,13
0,5	0,5	rbf	29656221,28

* Seçilen parametreler

Çizelge 5.22’de görüldüğü gibi SVM için yapılan parametre analizinde $c = 1,0$ ve $e = 0,1$ değerleri en düşük MSE değerine sahip olduğu için bu parametreler deneysel çalışmalarda kullanılmak üzere seçilmiştir.

Doğrulama verisi üzerinde ARIMA’nın alacağı p , d ve q parametrelerinin belirlenmesi için Şekil 5.11’de görüldüğü gibi MSE metriği kullanılarak bir analiz çalışması yapılmıştır.



Şekil 5.11. ARIMA parametre seçimi

Analiz sonuçları ARIMA(10,0,1) modelinin en düşük MSE değerine sahip olduğunu gösterdiği için bu model deneysel çalışmalarda kullanılmak üzere seçilmiştir.

MLP modelinin alacağı parametrelerinin belirlenmesi için Çizelge 5.23'te görüldüğü gibi MSE metriği kullanılarak bir analiz çalışması yapılmıştır.

Çizelge 5.23. MLP parametre seçimi

Yığın boyutu	Katman sayısı	Eğitim sayısı	MSE
512	16	50	30347415,32
512	8	100	30257412,12
128	16	100	30242112,23
256	16	100	30123872,66
512	16	100	*29566635,47
1024	16	100	31425345,12
512	32	100	31540278,97

* Seçilen parametreler

Analiz sonuçları MLP'nin en düşük MSE değerine yığın boyutu 512, katman sayısı 16 ve eğitim sayısı 100 olduğunda sahip olduğunu gösterdiği için, bu değerler deneysel çalışmalarda kullanılmak üzere seçilmiştir.

CNN'in yığın boyutu, eğitim sayısı ve gizli katmandaki nöron sayısı gibi parametrelerin belirlenebilmesi için Çizelge 5.24'te görüldüğü gibi MSE metriği kullanılarak bir analiz çalışması yapılmıştır.

Çizelge 5.24. CNN parametre seçimi

Filtre sayısı	Eğitim sayısı	Kernel sayısı	MSE
32	100	2	30028427,97
64	100	2	29004971,12
64	200	2	28966621,98
64	300	2	29001551,52

Çizelge 5.24. (devam) CNN parametre seçimi

16	200	2	28920522,18
32	200	2	28894254,24
32	200	3	*28878241,36
64	200	3	28910772,23

* Seçilen parametreler

Analiz sonuçları CNN'in en düşük MSE değerine filtre sayısı 32, eğitim sayısı 200 ve kernel sayısı 3 olduğunda sahip olduğunu gösterdiği için, bu değerler deneysel çalışmalarda kullanılmak üzere seçilmiştir.

GRU'nun yığın boyutu, eğitim sayısı ve gizli katmandaki nöron sayısı gibi parametrelerin belirlenebilmesi için Çizelge 5.25'te görüldüğü gibi MSE metriği kullanılarak bir analiz çalışması yapılmıştır.

Çizelge 5.25. GRU parametre seçimi

Yığın boyutu	Eğitim sayısı	Nöron sayısı	MSE
128	100	16	29322114,33
128	50	32	29332661,65
64	100	32	29136654,22
128	100	32	*28761914,59
256	100	32	29078564,36
128	200	32	28945124,41

* Seçilen parametreler

Analiz sonuçları GRU'nun en düşük MSE değerine yığın boyutu 128, eğitim sayısı 100 ve gizli katmandaki nöron sayısı 32 olduğunda sahip olduğunu gösterdiği için, bu değerler deneysel çalışmalarda kullanılmak üzere seçilmiştir.

LSTM'in yığın boyutu, eğitim sayısı ve gizli katmandaki nöron sayısı gibi parametrelerin belirlenebilmesi için Çizelge 5.26'da görüldüğü gibi MSE metriği kullanılarak bir analiz çalışması yapılmıştır.

Çizelge 5.26. LSTM parametre seçimi

Yığın boyutu	Eğitim sayısı	Nöron sayısı	MSE
256	100	4	28615740,41
512	100	4	28713313,30
256	100	8	28743462,23
64	100	32	28849620,66
256	100	32	*28573509,95
512	100	32	28679560,62
1024	100	32	28740755,11
256	100	64	28793543,56
256	200	32	28737367,60

* Seçilen parametreler

Analiz sonuçları LSTM'in en düşük MSE değerine yığın boyutu 256, eğitim sayısı 100 ve gizli katmandaki nöron sayısı 32 olduğunda sahip olduğunu gösterdiği için, bu değerler deneysel çalışmalarda kullanılmak üzere seçilmiştir.

CNN+GRU'nun yığın boyutu, eğitim sayısı ve gizli katmandaki nöron sayısı gibi parametrelerin belirlenebilmesi için Çizelge 5.27'de görüldüğü gibi MSE metriği kullanılarak bir analiz çalışması yapılmıştır.

Çizelge 5.27. CNN+GRU parametre seçimi

CNN filtre sayısı	Eğitim sayısı	GRU nöron sayısı	MSE
64	50	8	29436121,54
64	100	4	29288945,12
64	100	8	*28540078,57
64	200	8	29984526,45

Çizelge 5.27. (devam) CNN+GRU parametre seçimi

64	100	16	30177214,55
64	100	32	30342514,21

* Seçilen parametreler

Analiz sonuçları CNN+GRU'nun en düşük MSE değerine CNN filtre sayısı 64, eğitim sayısı 100 ve GRU gizli katmandaki nöron sayısı 8 olduğunda sahip olduğunu gösterdiği için, bu değerler deneysel çalışmalarda kullanılmak üzere seçilmiştir.

HDLM'nin yığın boyutu, eğitim sayısı ve gizli katmandaki nöron sayısı gibi parametrelerin belirlenebilmesi için Çizelge 5.28'de görüldüğü gibi MSE metriği kullanılarak bir analiz çalışması yapılmıştır.

Çizelge 5.28. HDLM parametre seçimi

CNN filtre sayısı	Eğitim sayısı	LSTM nöron sayısı	MSE
64	100	4	29043042,85
64	100	16	27956619,87
64	100	32	26458619,83
64	200	32	*25035735,24
64	300	32	25117914,64
64	100	50	27218699,58

* Seçilen parametreler

Analiz sonuçları HDLM'nin en düşük MSE değerine CNN filtre sayısı 64, CNN kernel sayısı 1, eğitim sayısı 200 ve LSTM gizli katmandaki nöron sayısı 32 olduğunda sahip olduğunu gösterdiği için, bu değerler deneysel çalışmalarda kullanılmak üzere seçilmiştir.

Acil çağrı veri kümesi için parametre seçimi

Acil çağrı veri kümesi 98400 adet trafik kazası olayından oluşmaktadır. Verilerin %67'si eğitim, %33'ü test olarak ayrılmıştır. Ayırma sonucunda toplam 65928 elemandan oluşan bir eğitim verisi elde edilmiştir. Eğitim verisinin %10'u parametre seçiminde kullanılmak üzere

doğrulama verisi olarak ayrılmıştır. Ayırma sonucunda 6592 elemandan oluşan bir doğrulama verisi elde edilmiştir.

RF modelinin alacağı parametrelerinin belirlenmesi için Çizelge 5.29’da görüldüğü gibi MSE metriği kullanılarak bir analiz çalışması yapılmıştır.

Çizelge 5.29. RF parametre seçimi

Ağaç sayısı	İç düğümü bölmek için gereken örnek sayısı	MSE
100	2	2977773.65
200	2	2980414.51
500	2	2978531.85
1000	2	2971364.85
3000	2	*2969472.56

* Seçilen parametreler

Çizelge 5.29’da görüldüğü gibi RF için yapılan parametre analizinde en düşük MSE değerine sahip parametreler belirlenmiştir. Ağaç sayısının 3000 olduğu durumda en düşük MSE değerine ulaşıldığı için bu parametreler deneysel çalışmalarda kullanılmak üzere seçilmiştir.

SVM modelinin alacağı parametrelerinin belirlenmesi için Çizelge 5.30’da görüldüğü gibi MSE metriği kullanılarak bir analiz çalışması yapılmıştır.

Çizelge 5.30. SVM parametre

c	e	Kernel	MSE
1,5	0,1	rbf	3028306,95
0,5	0,1	rbf	3029429,69
0,5	0,2	rbf	2987615,28
0,5	0,3	rbf	2960743,12
0,5	0,5	rbf	3047044,76

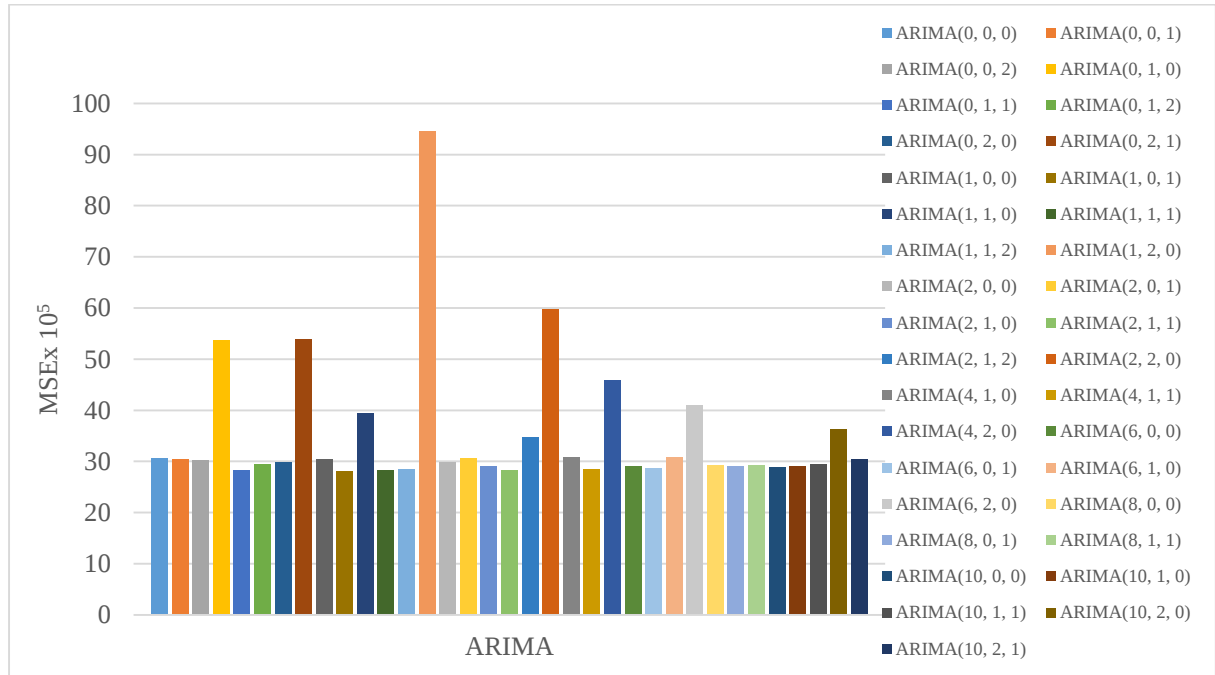
Çizelge 5.30. (devam) SVM parametre

1,0	0,1	rbf	3030485,00
1,0	0,3	rbf	*2957652,45
1,0	0,5	rbf	3045174,12
1,5	0,1	rbf	3028306,95

* Seçilen parametreler

SVM için yapılan parametre analizinde $c=1,0$ ve $e= 0,3$ değerleri en düşük MSE değerine sahip olduğu için bu parametreler deneysel çalışmalarda kullanılmak üzere seçilmiştir.

ARIMA modelinin alacağı p , d ve q parametrelerinin belirlenmesi için Şekil 5.12’de görüldüğü gibi MSE metriği kullanılarak bir analiz çalışması yapılmıştır.



Şekil 5.12. ARIMA parametre seçimi

Analiz sonuçları ARIMA(1,0,1) modelinin en düşük MSE değerine sahip olduğunu gösterdiği için bu model deneysel çalışmalarda kullanılmak üzere seçilmiştir.

MLP modelinin alacağı parametrelerinin belirlenmesi için Çizelge 5.31’de görüldüğü gibi MSE metriği kullanılarak bir analiz çalışması yapılmıştır.

Çizelge 5.31. MLP parametre seçimi

Yığın boyutu	Katman sayısı	Eğitim sayısı	MSE
128	50	100	2823114,55
256	32	100	2811249,61
256	50	50	2737712,25
256	50	100	*2706453,98
256	64	100	2752379,36
256	50	200	2745155,02

* Seçilen parametreler

Analiz sonuçları MLP’nin en düşük MSE değerine yığın boyutu 256, katman sayısı 50 ve eğitim sayısı 100 olduğunda sahip olduğunu gösterdiği için, bu değerler deneysel çalışmalarda kullanılmak üzere seçilmiştir.

CNN’in yığın boyutu, eğitim sayısı ve gizli katmandaki nöron sayısı gibi parametrelerin belirlenebilmesi için Çizelge 5.32’de görüldüğü gibi MSE metriği kullanılarak bir analiz çalışması yapılmıştır.

Çizelge 5.32. CNN parametre seçimi

Filtre sayısı	Eğitim sayısı	Kernel sayısı	MSE
16	100	2	2813651,18
32	100	2	2805624,63
64	100	2	2813443,40
32	200	2	*2800366,68
32	200	3	2803426,02

* Seçilen parametreler

Analiz sonuçları CNN'in en düşük MSE değerine filtre sayısı 32, eğitim sayısı 200 ve kernel sayısı 2 olduğunda sahip olduğunu gösterdiği için, bu değerler deneysel çalışmalarda kullanılmak üzere seçilmiştir.

GRU'nun yığın boyutu, eğitim sayısı ve gizli katmandaki nöron sayısı gibi parametrelerin belirlenebilmesi için Çizelge 5.33'te görüldüğü gibi MSE metriği kullanılarak bir analiz çalışması yapılmıştır.

Çizelge 5.33. GRU parametre seçimi

Yığın boyutu	Eğitim sayısı	Nöron sayısı	MSE
256	100	2	3056781,65
512	100	4	3035605,61
1024	100	4	3014672,23
512	100	16	3013719,49
512	100	32	*2943789,59
512	200	32	2981791,33

* Seçilen parametreler

Analiz sonuçları GRU'nun en düşük MSE değerine yığın boyutu 512, eğitim sayısı 100 ve gizli katmandaki nöron sayısı 32 olduğunda sahip olduğunu gösterdiği için, bu değerler deneysel çalışmalarda kullanılmak üzere seçilmiştir.

LSTM'in yığın boyutu, eğitim sayısı ve gizli katmandaki nöron sayısı gibi parametrelerinin belirlenebilmesi için Çizelge 5.34'te görüldüğü gibi MSE metriği kullanılarak bir analiz çalışması yapılmıştır

Çizelge 5.34. LSTM parametre seçimi

Yığın boyutu	Eğitim sayısı	Nöron sayısı	MSE
64	100	4	2778670,04
128	100	4	2782464,61
256	100	4	*2770119,22

Çizelge 5.34. (devam) LSTM parametre seçimi

512	100	4	2773075,41
512	100	32	2787202,43
1024	100	32	2772160,15

* Seçilen parametreler

Analiz sonuçları LSTM'in en düşük MSE değerine yığın boyutu 256, eğitim sayısı 100 ve gizli katmandaki nöron sayısı 4 olduğunda sahip olduğunu gösterdiği için, bu değerler deneysel çalışmalarda kullanılmak üzere seçilmiştir.

CNN+GRU'nun yığın boyutu, eğitim sayısı ve gizli katmandaki nöron sayısı gibi parametrelerin belirlenebilmesi için Çizelge 5.35'te görüldüğü gibi MSE metriği kullanılarak bir analiz çalışması yapılmıştır.

Çizelge 5.35. CNN+GRU parametre seçimi

CNN filtre sayısı	Eğitim sayısı	GRU nöron sayısı	MSE
4	100	64	2619633,23
8	50	64	2601713,69
8	100	64	*2308616,30
8	200	64	2361370,77
16	100	64	2564279,50
8	100	32	2501646,72
8	100	80	2641534,08

* Seçilen parametreler

Analiz sonuçları CNN+GRU'nun en düşük MSE değerine CNN filtre sayısı 8, eğitim sayısı 100 ve GRU gizli katmandaki nöron sayısı 64 olduğunda sahip olduğunu gösterdiği için, bu değerler deneysel çalışmalarda kullanılmak üzere seçilmiştir.

HDLM'nin yığın boyutu, eğitim sayısı ve gizli katmandaki nöron sayısı gibi parametrelerin belirlenebilmesi için Çizelge 5.36'da görüldüğü gibi MSE metriği kullanılarak bir analiz çalışması yapılmıştır.

Çizelge 5.36. HDLM parametre seçimi

CNN filtre sayısı	Eğitim sayısı	LSTM nöron sayısı	MSE
64	100	4	2376653,32
64	100	16	2478087,99
64	100	32	*2172418,30
64	200	32	2257376,60
64	100	50	2154276,10
64	100	50	2329454,26

* Seçilen parametreler

Analiz sonuçları HDLM'nin en düşük MSE değerine CNN filtre sayısı 64, CNN kernel sayısı 1, eğitim sayısı 100 ve LSTM gizli katmandaki nöron sayısı 32 olduğunda sahip olduğunu gösterdiği için, bu değerler deneysel çalışmalarda kullanılmak üzere seçilmiştir.

6. DENEYSEL SONUÇLAR

Bu tez çalışmasında, olaylar arasındaki zaman farkları saniye ve dakika cinsinden hesaplanmıştır. Her iki senaryo için geliştirilen hibrit derin öğrenme modeli, e-ticaret veri kümesi, DDoS veri kümesi, suç veri kümesi ve acil çağrı veri kümesi kullanılarak RF, SVM, ARIMA, MLP, CNN, GRU, LSTM ve CNN+GRU modelleri ile MSE, RMSE ve MAE hata ölçütlerine göre karşılaştırmalı olarak incelenmiştir.

Doğruluk ve hata ölçütlerinin amacı, hata dağılımının özetini elde etmektir. Kayıp fonksiyonu hesaplandıktan sonra ortalamanın hesaplanarak hata ölçümlerinin belirlenmesi yaygın bir uygulamadır. y_t değeri t zamanda gözlemlenen değeri, $\hat{y}_t$ ise tahmin değerini ifade etmektedir. Hata değeri E_t ise $y_t - \hat{y}_t$ eşitliği ile hesaplanmaktadır [9].

Hata ölçütleri, verilerin ölçeğine bağlı olarak, aynı veri kümesi üzerinde farklı yöntemlerin karşılaştırması yapılırken kullanışlıdır. Yaygın olarak kullanılan ölçütler ortalama mutlak hata ve karesel hata ölçütleridir. Ortalama mutlak hata metriği Eş.6.1’de görüldüğü gibi hataların aritmetik ortalaması alınarak hesaplanmaktadır.

$$MAE = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t| \quad (6.1)$$

Hataların karelerinin ortalamasının alınması ile hesaplanan MSE metriği Eş.6.2’de, hata karelerinin ortalamasının karekökü alınarak hesaplanan RMSE metriği ise Eş.6.3’te görülmektedir.

$$MSE = \frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2 \quad (6.2)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2} \quad (6.3)$$

MAE, kullanılan yöntemlerin tek bir veri kümesi üzerinde karşılaştırıldığı durumlarda, anlaşılması ve hesaplanması kolay olduğu için yaygın olarak kullanılmaktadır. MAE aşırı tahmin hatalarını cezalandırmamakla birlikte, MSE ve RMSE, toplam tahmin hatasının büyük bireysel hatalardan daha fazla etkilendiğini vurgulamaktadır. MSE ve RMSE aykırı değerlere MAE'dan daha duyarlıdır [13].

6.1. Bir Olayın Bir Sonraki Olma Zamanının Saniye Cinsinden Belirlenmesi

6.1.1. E-ticaret veri kümesi

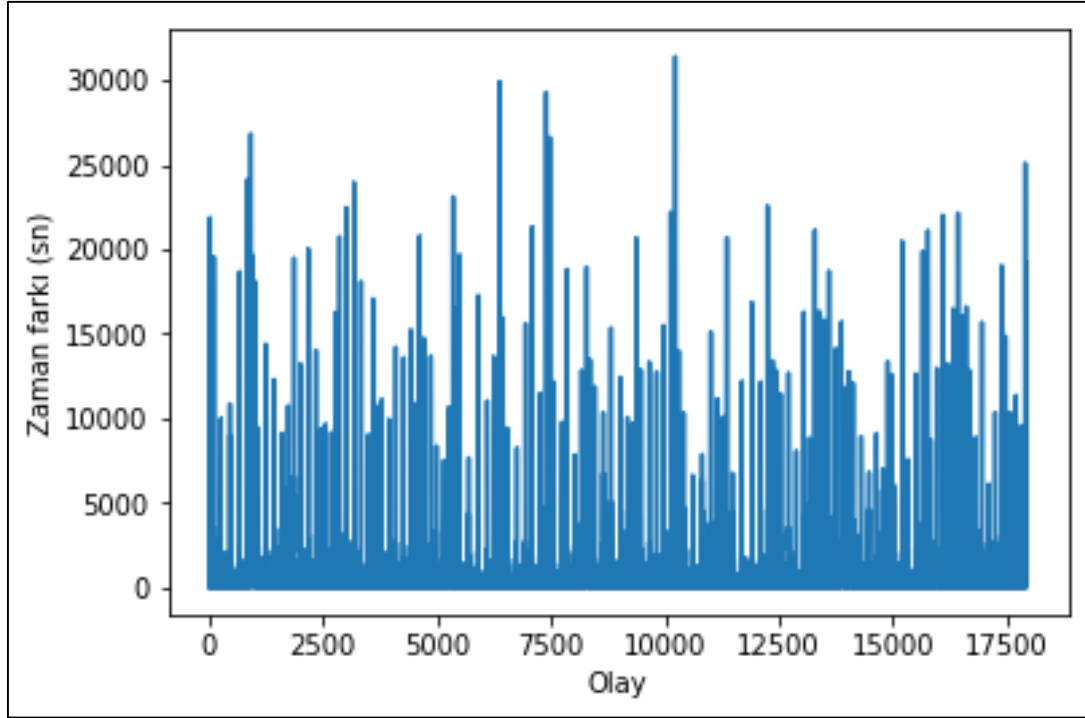
E-ticaret veri kümesi için satın alma olaylarının meydana gelme zamanları arasındaki saniye cinsinden farklara göre oluşturulan verinin ilk 5 satırı Çizelge 6.1’de görülmektedir.

Çizelge 6.1. Satın alma olaylarının meydana gelme zamanları arasındaki fark

Olay	Bir önceki satın almadan sonra geçen zaman (sn)
1	460
2	1606
3	351
4	1416
5	30

Çizelge 6.1’de satın alma olaylarının meydana gelme zamanları arasındaki fark görülmektedir. E-ticaret veri kümesi 17939 satın alma olayından oluşmaktadır.

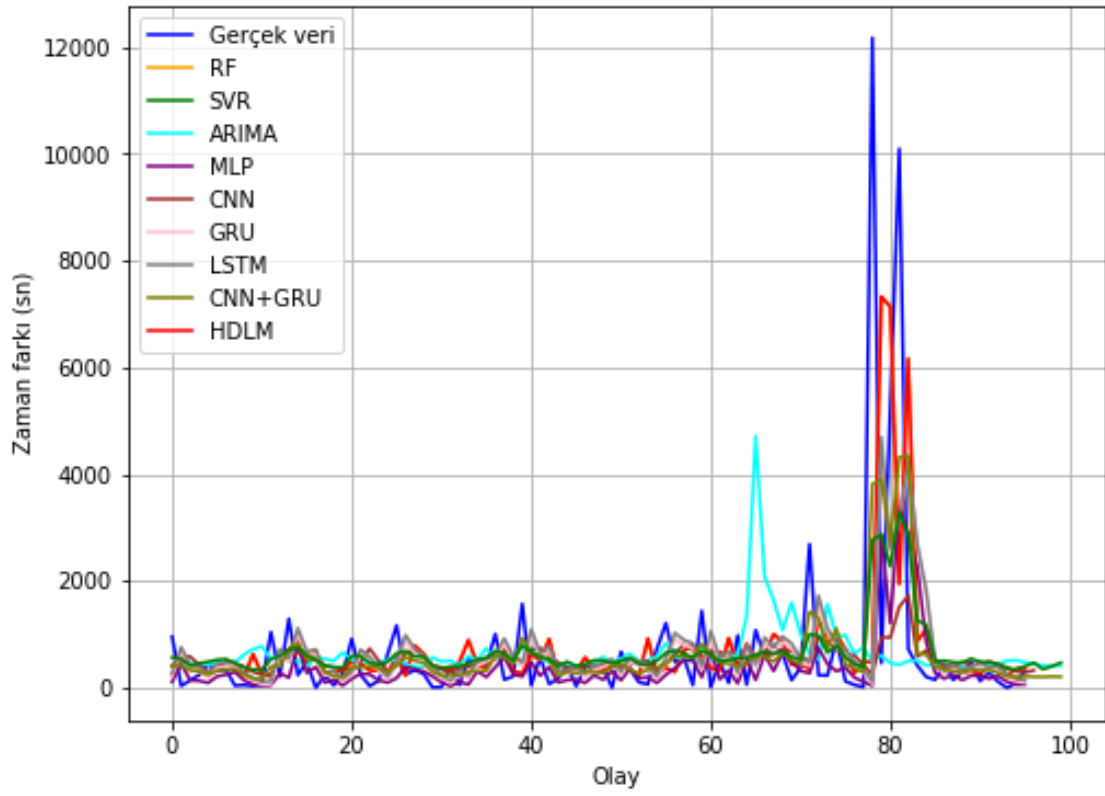
Satın alma olayları arasındaki saniye cinsinden zaman farkları Şekil 6.1’de görülmektedir.



Şekil 6.1. Satın alma olayları arasındaki saniye cinsinden zaman farkları

E-ticaret veri kümesindeki satın alma olayları arasındaki saniye cinsinden zaman farkları Şekil 6.1’de görülmektedir. Olayların dağılımındaki düzensizlikler nedeniyle meydana gelen sapmalar şekilde görülmektedir.

Şekil 6.2’de örnek olarak test verisinin 100 elemanı için deneysel sonuçlar görülmektedir.



Şekil 6.2. Test verisindeki ilk 100 veri örneği için deneysel sonuçlar

Şekil 6.2’de test verisindeki ilk 100 veri örneği için deneysel sonuçlar görülmektedir. HDLM, diğer modellere kıyasla veri kümesindeki ani sapmaları daha başarılı bir şekilde belirleyerek daha yüksek tahmin doğruluğuna sahip olmuştur.

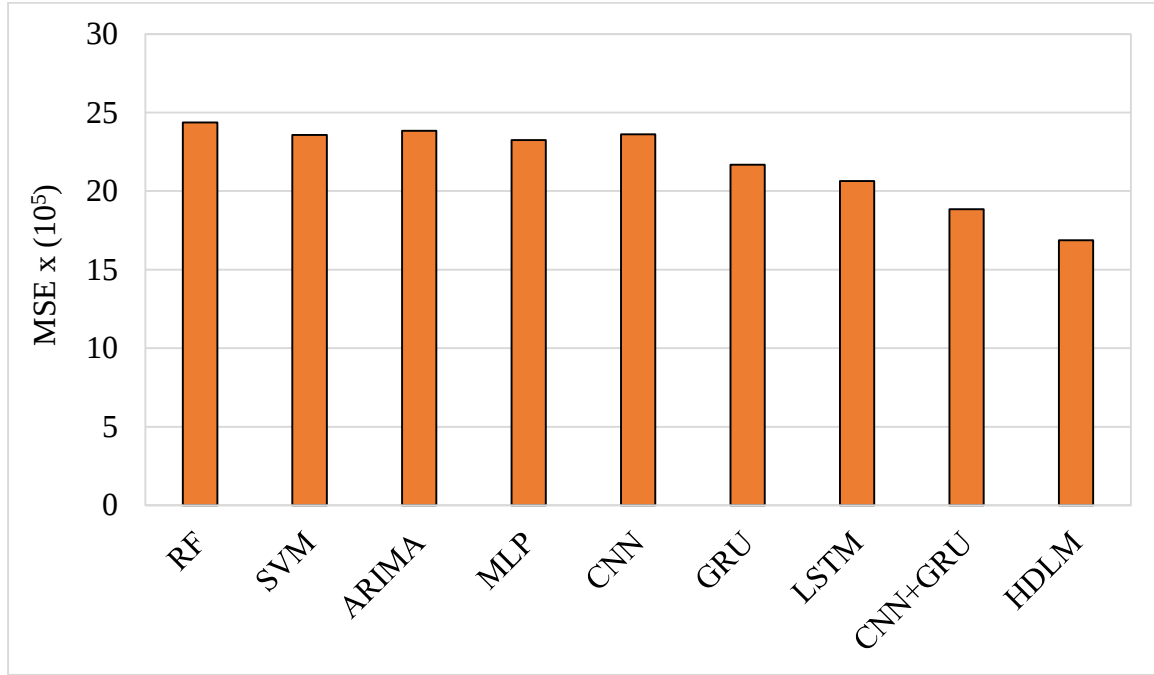
MSE metriğine göre deneysel sonuçlar Çizelge 6.2’de görülmektedir.

Çizelge 6.2. MSE metriğine göre deneysel sonuçlar

Test	RF	SVM	ARIMA	MLP	CNN	GRU	LSTM	CNN+GRU	HDLN
1	2437427,19	2358518,46	2383163,05	2317023,33	2338472,25	2167520,58	2064187,40	1844031,49	1623966,63
2	2437427,19	2358518,46	2383163,05	2322314,46	2342609,34	2167521,17	2064346,14	1849769,64	1658696,76
3	2437427,19	2358518,46	2383163,05	2323738,82	2355577,03	2167532,80	2064349,11	1857424,70	1661081,24
4	2437427,19	2358518,46	2383163,05	2324489,55	2355610,90	2167532,32	2064360,08	1865010,80	1677921,90
5	2437427,19	2358518,46	2383163,05	2325388,83	2360438,57	2167534,04	2064369,32	1877193,40	1677982,80
6	2437427,19	2358518,46	2383163,05	2325485,06	2363212,32	2167539,48	2064449,69	1894230,85	1686848,16
7	2437427,19	2358518,46	2383163,05	2325693,99	2363929,76	2167539,65	2064472,17	1896960,79	1687874,96
8	2437427,19	2358518,46	2383163,05	2325759,97	2368117,63	2167541,38	2064479,26	1909612,23	1697667,30
9	2437427,19	2358518,46	2383163,05	2327537,06	2381283,45	2167542,76	2064528,98	1922249,57	1730683,07
10	2437427,19	2358518,46	2383163,05	2328349,80	2400102,93	2167566,68	2064610,41	1932424,21	1771379,72
Ortalama	2437427,19	2358518,46	2383163,05	2324578,08	2362935,41	2167537,08	2064415,25	1884890,76	1687410,25

MSE metriğine göre yapılan deneysel çalışmalar sonucunda, RF'ın 2437427,19, SVM'in 2358518,46, ARIMA'nın 2383163,05, MLP'nin 2324578,08, CNN'in 2362935,41, GRU'nun 2167537,08, LSTM'in 2064415,25, CNN+GRU'nun 1884890,76 ve HDLM'nin 1687410,25 ortalama MSE değerine sahip olduğu görülmüştür.

Çizelge 6.2'de ve Şekil 6.3'te HDLM'nin RF, SVM, ARIMA, MLP, CNN, GRU, LSTM ve CNN+GRU'ya göre MSE metriği için karşılaştırmalı deneysel sonuçları görülmektedir.



Şekil 6.3. MSE metriği için karşılaştırmalı deneysel sonuçlar

$$\text{HDLM'nin } \frac{\text{RF-HDLM}}{\text{RF}} \times 100, \frac{\text{SVR-HDLM}}{\text{SVR}} \times 100, \frac{\text{ARIMA-HDLM}}{\text{ARIMA}} \times 100, \frac{\text{MLP-HDLM}}{\text{MLP}} \times 100, \\ \frac{\text{CNN-HDLM}}{\text{CNN}} \times 100, \frac{\text{GRU-HDLM}}{\text{GRU}} \times 100, \frac{\text{LSTM-HDLM}}{\text{LSTM}} \times 100 \text{ ve } \frac{(\text{CNN+GRU})-\text{HDLM}}{(\text{CNN+GRU})} \times 100$$

eşitlikleri kullanılarak MSE metriğine göre hesaplanan iyileşme oranları Çizelge 6.3'te görülmektedir.

Çizelge 6.3. MSE metriğine göre HDLM'nin iyileşme oranları

	HDLM'nin en kötü değeri	HDLM'nin en iyi değeri	HDLM'nin ortalama değeri
RF'in değerine göre iyileşme oranı (%)	27,32	33,37	30,77
SVM'nin değerine göre iyileşme oranı (%)	24,89	31,14	28,45
ARIMA'nın değerine göre iyileşme oranı (%)	25,67	31,85	29,19
MLP'nin en kötü değerine göre iyileşme oranı (%)	23,92	30,25	27,52
MLP'nin en iyi değerine göre iyileşme oranı (%)	23,54	29,91	27,17
MLP'nin ortalama değerine göre iyileşme oranı (%)	23,87	30,13	27,41
CNN'in en kötü değerine göre iyileşme oranı (%)	26,19	32,33	29,69
CNN'in en iyi değerine göre iyileşme oranı (%)	24,25	30,55	27,84
CNN'in ortalama değerine göre iyileşme oranı (%)	25,29	31,27	28,58
GRU'nun en kötü değerine göre iyileşme oranı (%)	18,27	25,07	22,15
GRU'nun en iyi değerine göre iyileşme oranı (%)	18,27	25,07	22,15
GRU'nun ortalama değerine göre iyileşme oranı (%)	18,27	25,07	22,15
LSTM'in en kötü değerine göre iyileşme oranı (%)	14,20	21,34	18,26
LSTM'in en iyi değerine göre iyileşme oranı (%)	14,18	21,32	18,25

Çizelge 6.3. (devam) MSE metriğine göre HDLM'nin iyileşme oranları

LSTM'in ortalama değerine göre iyileşme oranı (%)	14,19	21,33	18,26
CNN+GRU'nun en kötü değerine göre iyileşme oranı (%)	8,33	15,96	12,67
CNN+GRU'nun en iyi değerine göre iyileşme oranı (%)	3,93	11,93	8,49
CNN+GRU'nun ortalama değerine göre iyileşme oranı (%)	6,15	13,84	10,47

Yapılan 10 farklı deney sonucunda elde edilen MSE değerlerine göre HDLM'nin RF, SVM, ARIMA, MLP, CNN, GRU, LSTM ve CNN+GRU'ya göre karşılaştırmalı bir analizi görülmektedir. HDLM'nin en kötü değeri RF'a göre %27,32, en iyi değeri %33,37 ve ortalama değeri ise %30,77 iyileşme sağlamıştır. HDLM'nin en kötü değeri SVM'e göre %24,89, en iyi değeri %31,14 ve ortalama değeri ise %28,45 iyileşme sağlamıştır. HDLM'nin en kötü değeri ARIMA'ya göre %25,67, en iyi değeri %31,85 ve ortalama değeri ise %29,19 iyileşme sağlamıştır.

HDLM'nin en kötü değeri MLP'nin en kötü değerine göre %23,92, MLP'nin en iyi değerine göre %23,54 ve MLP'nin ortalama değerine göre %23,87 iyileşme sağlamıştır. HDLM'nin en iyi değeri MLP'nin en kötü değerine göre %30,25, MLP'nin en iyi değerine göre %29,91 ve MLP'nin ortalama değerine göre %30,13 iyileşme sağlamıştır. HDLM'nin ortalama değeri MLP'nin en kötü değerine göre %27,52, MLP'nin en iyi değerine göre %27,17 ve MLP'nin ortalama değerine göre %27,41 iyileşme sağlamıştır.

HDLM'nin en kötü değeri CNN'in en kötü değerine göre %26,19, CNN'in en iyi değerine göre %24,25 ve CNN'in ortalama değerine göre %25,29 iyileşme sağlamıştır. HDLM'nin en iyi değeri CNN'in en kötü değerine göre %32,33, CNN'in en iyi değerine göre %30,55 ve CNN'in ortalama değerine göre %31,27 iyileşme sağlamıştır. HDLM'nin ortalama değeri CNN'in en kötü değerine göre %29,69, CNN'in en iyi değerine göre %27,84 ve CNN'in ortalama değerine göre %28,58 iyileşme sağlamıştır.

HDLM'nin en kötü değeri GRU'nun en kötü değerine göre %18,27, GRU'nun en iyi değerine göre %18,27 ve GRU'nun ortalama değerine göre %18,27 iyileşme sağlamıştır. HDLM'nin en iyi değeri GRU'nun en kötü değerine göre %25,07, GRU'nun en iyi değerine göre %25,07 ve GRU'nun ortalama değerine göre %25,07 iyileşme sağlamıştır. HDLM'nin ortalama değeri GRU'nun en kötü değerine göre %22,15, GRU'nun en iyi değerine göre %22,15 ve GRU'nun ortalama değerine göre %22,15 iyileşme sağlamıştır.

HDLM'nin en kötü değeri LSTM'in en kötü değerine göre %14,20, LSTM'in en iyi değerine göre %14,18 ve LSTM'in ortalama değerine göre %14,19 iyileşme sağlamıştır. HDLM'nin en iyi değeri LSTM'in en kötü değerine göre %21,34, LSTM'in en iyi değerine göre %21,32 ve LSTM'in ortalama değerine göre %21,33 iyileşme sağlamıştır. HDLM'nin ortalama değeri LSTM'in en kötü değerine göre %18,26, LSTM'in en iyi değerine göre %18,25 ve LSTM'in ortalama değerine göre %18,26 iyileşme sağlamıştır.

HDLM'nin en kötü değeri CNN+GRU'nun en kötü değerine göre %8,33, CNN+GRU'nun en iyi değerine göre %3,93 ve CNN+GRU'nun ortalama değerine göre %6,15 iyileşme sağlamıştır. HDLM'nin en iyi değeri CNN+GRU'nun en kötü değerine göre %15,96, CNN+GRU'nun en iyi değerine göre %11,93 ve CNN+GRU'nun ortalama değerine göre %13,84 iyileşme sağlamıştır. HDLM'nin ortalama değeri CNN+GRU'nun en kötü değerine göre %12,67, CNN+GRU'nun en iyi değerine göre %8,49 ve CNN+GRU'nun ortalama değerine göre %10,47 iyileşme sağlamıştır.

RMSE metriğine göre deneysel sonuçlar Çizelge 6.4'te görülmektedir.

Çizelge 6.4. RMSE metriğine göre deneysel sonuçlar

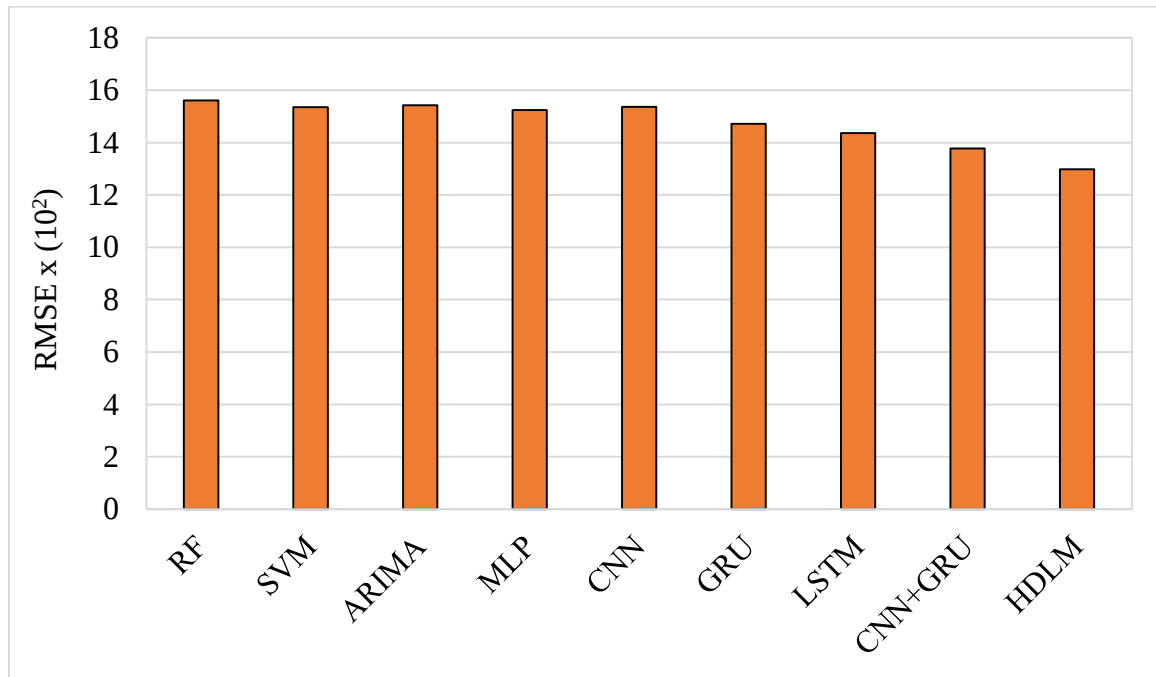
Test	RF	SVM	ARIMA	MLP	CNN	GRU	LSTM	CNN+GRU	HDLM
1	1561,22	1535,74	1543,74	1522,18	1537,50	1472,25	1436,72	1357,95	1274,34
2	1561,22	1535,74	1543,74	1523,91	1529,20	1472,25	1436,78	1360,06	1287,90
3	1561,22	1535,74	1543,74	1524,38	1534,79	1472,25	1436,78	1362,87	1288,82
4	1561,22	1535,74	1543,74	1524,63	1538,86	1472,25	1436,79	1365,65	1295,34
5	1561,22	1535,74	1543,74	1524,92	1549,22	1472,25	1436,79	1370,10	1295,36
6	1561,22	1535,74	1543,74	1524,95	1534,78	1472,26	1436,82	1376,31	1298,78
7	1561,22	1535,74	1543,74	1525,02	1530,55	1472,26	1436,83	1377,30	1299,18
8	1561,22	1535,74	1543,74	1525,04	1537,27	1472,26	1436,83	1381,88	1302,94

Çizelge 6.4. (devam) RMSE metriğine göre deneysel sonuçlar

9	1561,22	1535,74	1543,74	1525,63	1536,37	1472,26	1436,85	1386,45	1315,55
10	1561,22	1535,74	1543,74	1525,89	1543,14	1472,27	1436,88	1390,11	1330,93
Ort.	1561,22	1535,74	1543,74	1524,65	1537,16	1472,25	1436,80	1372,86	1298,91

RMSE metriğine göre yapılan deneysel çalışmalar sonucunda, RF'in 1561,22, SVM'in 1535,74, ARIMA'nın 1543,74, MLP'nin 1524,65, CNN'in 1537,16, GRU'nun 1472,25, LSTM'in 1436,80, CNN+GRU'nun 1372,86 ve HDLM'nin 1298,91 ortalama RMSE değerine sahip olduğu görülmüştür.

Çizelge 6.4'te ve Şekil 6.4'te HDLM'nin RF, SVM, ARIMA, MLP, CNN, GRU, LSTM ve CNN+GRU'ya göre RMSE metriği için karşılaştırmalı deneysel sonuçları görülmektedir.



Şekil 6.4. RMSE metriği için karşılaştırmalı deneysel sonuçlar

$$\begin{aligned}
 &\text{HDLM'nin } \frac{\text{RF-HDLM}}{\text{RF}} \times 100, \frac{\text{SVR-HDLM}}{\text{SVR}} \times 100, \frac{\text{ARIMA-HDLM}}{\text{ARIMA}} \times 100, \frac{\text{MLP-HDLM}}{\text{MLP}} \times 100, \\
 &\frac{\text{CNN-HDLM}}{\text{CNN}} \times 100, \frac{\text{GRU-HDLM}}{\text{GRU}} \times 100, \frac{\text{LSTM-HDLM}}{\text{LSTM}} \times 100 \text{ ve } \frac{(\text{CNN+GRU})\text{-HDLM}}{(\text{CNN+GRU})} \times 100
 \end{aligned}$$

eşitlikleri kullanılarak RMSE metriğine göre hesaplanan iyileşme oranları Çizelge 6.5'te görülmektedir.

Çizelge 6.5. RMSE metriğine göre HDLM'nin iyileşme oranları

	HDLM'nin en kötü değeri	HDLM'nin en iyi değeri	HDLM'nin ortalama değeri
RF'ın değerine göre iyileşme oranı (%)	14,75	18,37	16,80
SVM'nin değerine göre iyileşme oranı (%)	13,33	17,02	15,42
ARIMA'nın değerine göre iyileşme oranı (%)	13,78	17,45	15,85
MLP'nin en kötü değerine göre iyileşme oranı (%)	12,77	16,48	14,87
MLP'nin en iyi değerine göre iyileşme oranı (%)	12,56	16,28	14,66
MLP'nin ortalama değerine göre iyileşme oranı (%)	12,72	16,41	14,80
CNN'in en kötü değerine göre iyileşme oranı (%)	13,75	17,41	15,82
CNN'in en iyi değerine göre iyileşme oranı (%)	13,43	17,11	15,51
CNN'in ortalama değerine göre iyileşme oranı (%)	13,41	17,09	15,49
GRU'nun en kötü değerine göre iyileşme oranı (%)	9,60	13,44	11,77
GRU'nun en iyi değerine göre iyileşme oranı (%)	9,59	13,44	11,77
GRU'nun ortalama değerine göre iyileşme oranı (%)	9,59	13,44	11,77

Çizelge 6.5. (devam) RMSE metriğine göre HDLM'nin iyileşme oranları

LSTM'in en kötü değerine göre iyileşme oranı (%)	7,37	11,31	9,60
LSTM'in en iyi değerine göre iyileşme oranı (%)	7,36	11,30	9,59
LSTM'in ortalama değerine göre iyileşme oranı (%)	7,36	11,30	9,59
CNN+GRU'nun en kötü değerine göre iyileşme oranı (%)	4,25	8,32	6,56
CNN+GRU'nun en iyi değerine göre iyileşme oranı (%)	1,98	6,15	4,34
CNN+GRU'nun ortalama değerine göre iyileşme oranı (%)	3,08	7,17	5,38

Yapılan 10 farklı deney sonucunda elde edilen RMSE değerlerine göre HDLM'nin RF, SVM, ARIMA, MLP, CNN, GRU, LSTM ve CNN+GRU'ya göre karşılaştırmalı bir analizi görülmektedir. HDLM'nin en kötü değeri RF'a göre %14,75, en iyi değeri %18,37 ve ortalama değeri ise %16,80 iyileşme sağlamıştır. HDLM'nin en kötü değeri SVM'e göre %13,33, en iyi değeri %17,02 ve ortalama değeri ise %15,42 iyileşme sağlamıştır. HDLM'nin en kötü değeri ARIMA'ya göre %13,78, en iyi değeri %17,45 ve ortalama değeri ise %15,85 iyileşme sağlamıştır.

HDLM'nin en kötü değeri MLP'nin en kötü değerine göre %12,77, MLP'nin en iyi değerine göre %12,56 ve MLP'nin ortalama değerine göre %12,72 iyileşme sağlamıştır. HDLM'nin en iyi değeri MLP'nin en kötü değerine göre %16,48, MLP'nin en iyi değerine göre %16,28 ve MLP'nin ortalama değerine göre %16,41 iyileşme sağlamıştır. HDLM'nin ortalama değeri MLP'nin en kötü değerine göre %14,87, MLP'nin en iyi değerine göre %14,66 ve MLP'nin ortalama değerine göre %14,80 iyileşme sağlamıştır.

HDLM'nin en kötü değeri CNN'in en kötü değerine göre %13,75, CNN'in en iyi değerine göre %13,43 ve CNN'in ortalama değerine göre %13,41 iyileşme sağlamıştır. HDLM'nin en iyi değeri CNN'in en kötü değerine göre %17,41, CNN'in en iyi değerine göre %17,11

ve CNN'in ortalama deęerine g re %17,09 iyileşme saęlamıştır. HDLM'nin ortalama deęeri CNN'in en k t  deęerine g re %15,82, CNN'in en iyi deęerine g re %15,51 ve CNN'in ortalama deęerine g re %15,49 iyileşme saęlamıştır.

HDLM'nin en k t  deęeri GRU'nun en k t  deęerine g re %9,60, GRU'nun en iyi deęerine g re %9,59 ve GRU'nun ortalama deęerine g re %9,59 iyileşme saęlamıştır. HDLM'nin en iyi deęeri GRU'nun en k t  deęerine g re %13,44, GRU'nun en iyi deęerine g re %13,44 ve GRU'nun ortalama deęerine g re %13,44 iyileşme saęlamıştır. HDLM'nin ortalama deęeri GRU'nun en k t  deęerine g re %11,77, GRU'nun en iyi deęerine g re %11,77 ve GRU'nun ortalama deęerine g re %11,77 iyileşme saęlamıştır.

HDLM'nin en k t  deęeri, LSTM'in en k t  deęerine g re %7,37, LSTM'in en iyi deęerine g re %7,36 ve LSTM'in ortalama deęerine g re %7,36 iyileşme saęlamıştır. HDLM'nin en iyi deęeri LSTM'in en k t  deęerine g re %11,31, LSTM'in en iyi deęerine g re %11,30 ve LSTM'in ortalama deęerine g re %11,30 iyileşme saęlamıştır. HDLM'nin ortalama deęeri LSTM'in en k t  deęerine g re %9,60, LSTM'in en iyi deęerine g re %9,59 ve LSTM'in ortalama deęerine g re %9,59 iyileşme saęlamıştır.

HDLM'nin en k t  deęeri CNN+GRU'nun en k t  deęerine g re %4,25, CNN+GRU'nun en iyi deęerine g re %1,98 ve CNN+GRU'nun ortalama deęerine g re %3,08 iyileşme saęlamıştır. HDLM'nin en iyi deęeri CNN+GRU'nun en k t  deęerine g re %8,32, CNN+GRU'nun en iyi deęerine g re %6,15 ve CNN+GRU'nun ortalama deęerine g re %7,17 iyileşme saęlamıştır. HDLM'nin ortalama deęeri CNN+GRU'nun en k t  deęerine g re %6,56, CNN+GRU'nun en iyi deęerine g re %4,34 ve CNN+GRU'nun ortalama deęerine g re %5,38 iyileşme saęlamıştır.

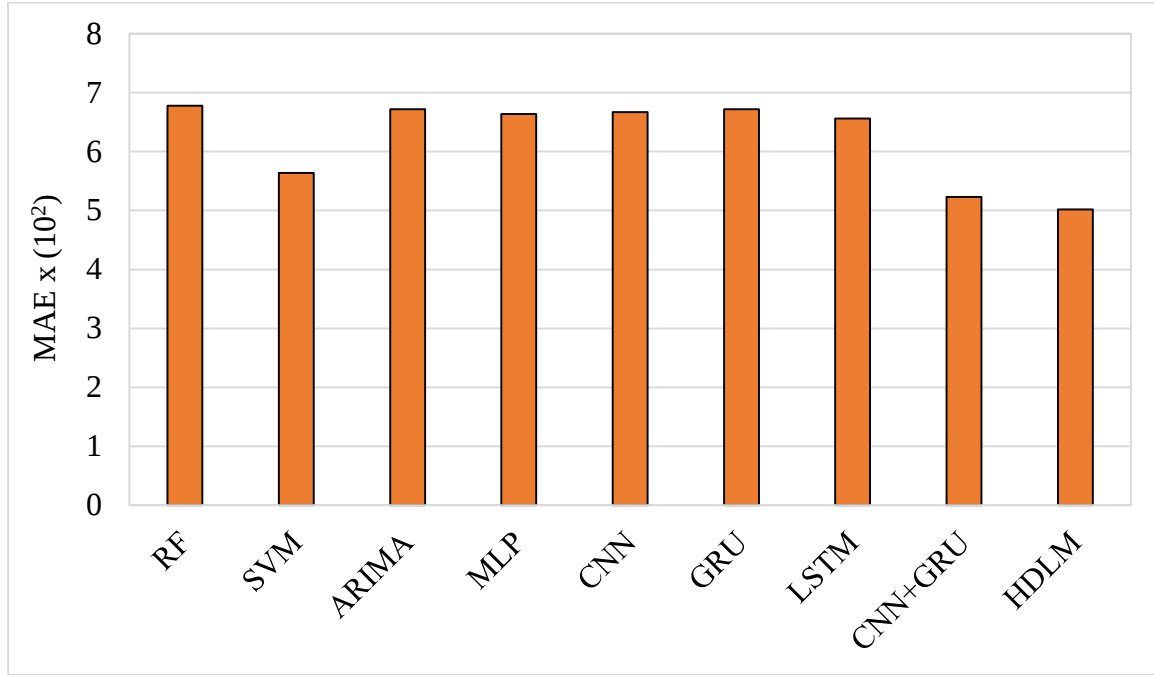
MAE metrięine g re deneysel sonular izelge 6.6'da g r lmektedir.

Çizelge 6.6. MAE metriğine göre deneysel sonuçlar

Test	RF	SVM	ARIMA	MLP	CNN	GRU	LSTM	CNN+GRU	HDLM
1	678,55	564,46	672,19	662,01	644,32	672,50	629,05	518,60	478,30
2	678,55	564,46	672,19	662,46	663,29	672,50	659,95	519,80	482,26
3	678,55	564,46	672,19	662,57	666,23	672,51	659,95	521,10	497,01
4	678,55	564,46	672,19	663,07	666,53	672,51	659,96	521,17	504,00
5	678,55	564,46	672,19	664,02	668,90	672,51	659,97	522,85	505,31
6	678,55	564,46	672,19	665,55	669,09	672,51	659,97	525,79	505,83
7	678,55	564,46	672,19	666,10	671,14	672,52	659,99	525,94	508,43
8	678,55	564,46	672,19	666,20	674,94	672,52	659,99	526,69	509,67
9	678,55	564,46	672,19	667,35	675,75	672,52	659,99	526,84	515,89
10	678,55	564,46	672,19	668,46	676,69	672,53	659,99	530,79	516,27
Ort.	678,55	564,46	672,19	664,77	667,68	672,51	656,88	523,95	502,29

MAE metriğine göre yapılan deneysel çalışmalar sonucunda, RF'in 678,55, SVM'in 564,46, ARIMA'nın 672,19, MLP'nin 664,77, CNN'in 667,68, GRU'nun 672,51, LSTM'in 656,88, CNN+GRU'nun 523,95 ve HDLM'nin 502,29 ortalama MAE değerine sahip olduğu görülmüştür.

Çizelge 6.6'da ve Şekil 6.5'te HDLM'nin RF, SVM, ARIMA, MLP, CNN, GRU, LSTM ve CNN+GRU'ya göre MAE metriği için karşılaştırmalı deneysel sonuçları görülmektedir.



Şekil 6.5. MAE metriği için karşılaştırmalı deneysel sonuçlar

$$\text{HDLM'nin } \frac{\text{RF-HDLM}}{\text{RF}} \times 100, \frac{\text{SVR-HDLM}}{\text{SVR}} \times 100, \frac{\text{ARIMA-HDLM}}{\text{ARIMA}} \times 100, \frac{\text{MLP-HDLM}}{\text{MLP}} \times 100, \\ \frac{\text{CNN-HDLM}}{\text{CNN}} \times 100, \frac{\text{GRU-HDLM}}{\text{GRU}} \times 100, \frac{\text{LSTM-HDLM}}{\text{LSTM}} \times 100 \text{ ve } \frac{(\text{CNN+GRU})\text{-HDLM}}{(\text{CNN+GRU})} \times 100$$

eşitlikleri kullanılarak MAE metriğine göre hesaplanan iyileşme oranları Çizelge 6.7’de görülmektedir.

Çizelge 6.7. MAE metriğine göre HDLM’nin iyileşme oranları

	HDLM’nin en kötü değeri	HDLM’nin en iyi değeri	HDLM’nin ortalama değeri
RF’ın değerine göre iyileşme oranı (%)	23,91	29,51	25,97
SVM’nin değerine göre iyileşme oranı (%)	8,53	15,26	11,01
ARIMA’nın değerine göre iyileşme oranı (%)	23,19	28,84	25,27

Çizelge 6.7. (devam) MAE metriğine göre HDLM'nin iyileşme oranları

MLP'nin en kötü değerine göre iyileşme oranı (%)	22,76	28,44	24,85
MLP'nin en iyi değerine göre iyileşme oranı (%)	22,01	27,75	24,12
MLP'nin ortalama değerine göre iyileşme oranı (%)	22,43	28,05	24,44
CNN'in en kötü değerine göre iyileşme oranı (%)	23,70	29,31	25,77
CNN'in en iyi değerine göre iyileşme oranı (%)	19,87	25,76	22,04
CNN'in ortalama değerine göre iyileşme oranı (%)	23,50	28,36	24,77
GRU'nun en kötü değerine göre iyileşme oranı (%)	23,23	28,88	25,31
GRU'nun en iyi değerine göre iyileşme oranı (%)	23,23	28,87	25,30
GRU'nun ortalama değerine göre iyileşme oranı (%)	23,23	28,87	25,31
LSTM'in en kötü değerine göre iyileşme oranı (%)	21,77	27,52	23,89
LSTM'in en iyi değerine göre iyileşme oranı (%)	17,92	23,96	20,14
LSTM'in ortalama değerine göre iyileşme oranı (%)	22,35	27,18	23,53
CNN+GRU'nun en kötü değerine göre iyileşme oranı (%)	2,73	9,88	5,36
CNN+GRU'nun en iyi değerine göre iyileşme oranı (%)	0,44	7,77	3,14
CNN+GRU'nun ortalama değerine göre iyileşme oranı (%)	1,48	8,71	4,13

Yapılan 10 farklı deney sonucunda elde edilen MAE değerlerine göre HDLM'nin RF, SVM, ARIMA, MLP, CNN, GRU, LSTM ve CNN+GRU'ya göre karşılaştırmalı bir analizi görülmektedir. HDLM'nin en kötü değeri RF'a göre %23,91, en iyi değeri %29,51 ve ortalama değeri ise %25,97 iyileşme sağlamıştır. HDLM'nin en kötü değeri SVM'e göre %8,53, en iyi değeri %15,26 ve ortalama değeri ise %11,01 iyileşme sağlamıştır. HDLM'nin en kötü değeri ARIMA'ya göre %23,19, en iyi değeri %28,84 ve ortalama değeri ise %25,27 iyileşme sağlamıştır.

HDLM'nin en kötü değeri MLP'nin en kötü değerine göre %22,76, MLP'nin en iyi değerine göre %22,01 ve MLP'nin ortalama değerine göre %22,43 iyileşme sağlamıştır. HDLM'nin en iyi değeri MLP'nin en kötü değerine göre %28,44, MLP'nin en iyi değerine göre %27,75 ve MLP'nin ortalama değerine göre %28,05 iyileşme sağlamıştır. HDLM'nin ortalama değeri MLP'nin en kötü değerine göre %24,85, MLP'nin en iyi değerine göre %24,12 ve MLP'nin ortalama değerine göre %24,44 iyileşme sağlamıştır.

HDLM'nin en kötü değeri CNN'in en kötü değerine göre %23,70, CNN'in en iyi değerine göre %19,87 ve CNN'in ortalama değerine göre %23,50 iyileşme sağlamıştır. HDLM'nin en iyi değeri CNN'in en kötü değerine göre %29,31, CNN'in en iyi değerine göre %25,76 ve CNN'in ortalama değerine göre %28,36 iyileşme sağlamıştır. HDLM'nin ortalama değeri CNN'in en kötü değerine göre %25,77, CNN'in en iyi değerine göre %22,04 ve CNN'in ortalama değerine göre %24,77 iyileşme sağlamıştır.

HDLM'nin en kötü değeri GRU'nun en kötü değerine göre %23,23, GRU'nun en iyi değerine göre %23,23 ve GRU'nun ortalama değerine göre %23,23 iyileşme sağlamıştır. HDLM'nin en iyi değeri GRU'nun en kötü değerine göre %28,88, GRU'nun en iyi değerine göre %28,87 ve GRU'nun ortalama değerine göre %28,87 iyileşme sağlamıştır. HDLM'nin ortalama değeri GRU'nun en kötü değerine göre %25,31, GRU'nun en iyi değerine göre %25,30 ve GRU'nun ortalama değerine göre %25,31 iyileşme sağlamıştır.

HDLM'nin en kötü değeri LSTM'in en kötü değerine göre %21,77, LSTM'in en iyi değerine göre %17,92 ve LSTM'in ortalama değerine göre %22,35 iyileşme sağlamıştır. HDLM'nin en iyi değeri LSTM'in en kötü değerine göre %27,52, LSTM'in en iyi değerine göre %23,96

ve LSTM'in ortalama deęerine g re %27,18 iyileşme saęlamıştır. HDLM'nin ortalama deęeri LSTM'in en k t  deęerine g re %23,89, LSTM'in en iyi deęerine g re %20,14 ve LSTM'in ortalama deęerine g re %23,53 iyileşme saęlamıştır.

HDLM'nin en k t  deęeri CNN+GRU'nun en k t  deęerine g re %2,73, CNN+GRU'nun en iyi deęerine g re %0,44 ve CNN+GRU'nun ortalama deęerine g re %1,48 iyileşme saęlamıştır. HDLM'nin en iyi deęeri CNN+GRU'nun en k t  deęerine g re %9,88, CNN+GRU'nun en iyi deęerine g re %7,77 ve CNN+GRU'nun ortalama deęerine g re %8,71 iyileşme saęlamıştır. HDLM'nin ortalama deęeri CNN+GRU'nun en k t  deęerine g re %5,36, CNN+GRU'nun en iyi deęerine g re %3,14 ve CNN+GRU'nun ortalama deęerine g re %4,13 iyileşme saęlamıştır.

Deneyisel sonular, HDLM'nin, RF, SVM, ARIMA, MLP, CNN, GRU, LSTM ve CNN+GRU'ya g re daha d ş k MSE, RMSE ve MAE sonularına sahip olduęunu g stermektedir. Bir sonraki satın alma zamanının belirlenmesinde HDLM'nin dięer modellere g re daha bařarılı ve etkin sonular verdięi g r lm şt r.

6.1.2. DDoS veri k mesi

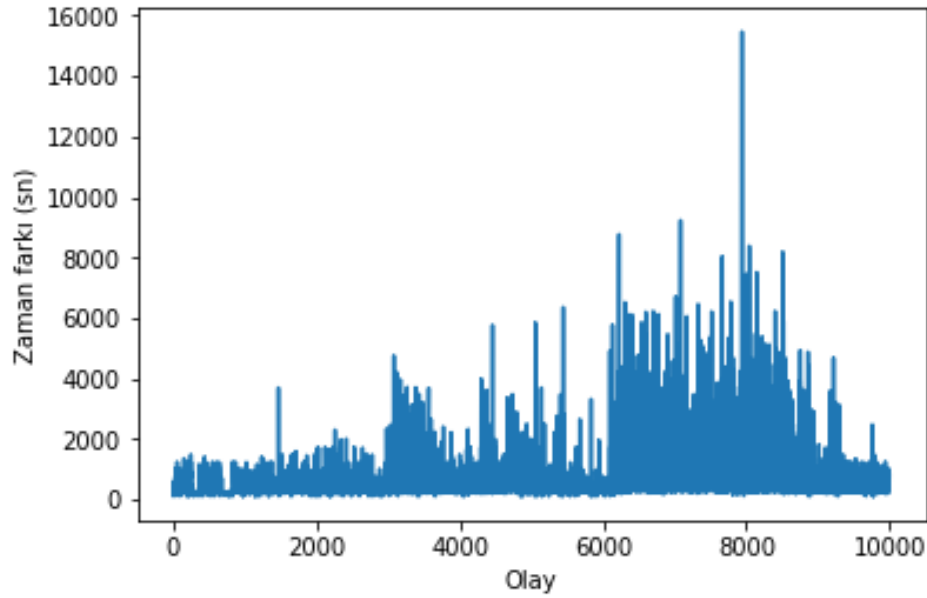
DDoS veri k mesi iin saldırı olaylarının meydana gelme zamanları arasındaki saniye cinsinden farklara g re oluřturulan verinin ilk 5 satırı izelge 6.8'de g r lmektedir.

izelge 6.8. DDoS saldırıları arasındaki zaman farkı

Saldırı	Bir �nceki saldırıdan sonra geen zaman (sn)
1	271
2	224
3	241
4	247
5	256

Çizelge 6.8’de DDoS saldırıları arasındaki zaman farkı görülmektedir. DDoS veri kümesi, e-ticaret veri kümesine göre nispeten daha düzenli bir örüntü izlemektedir.

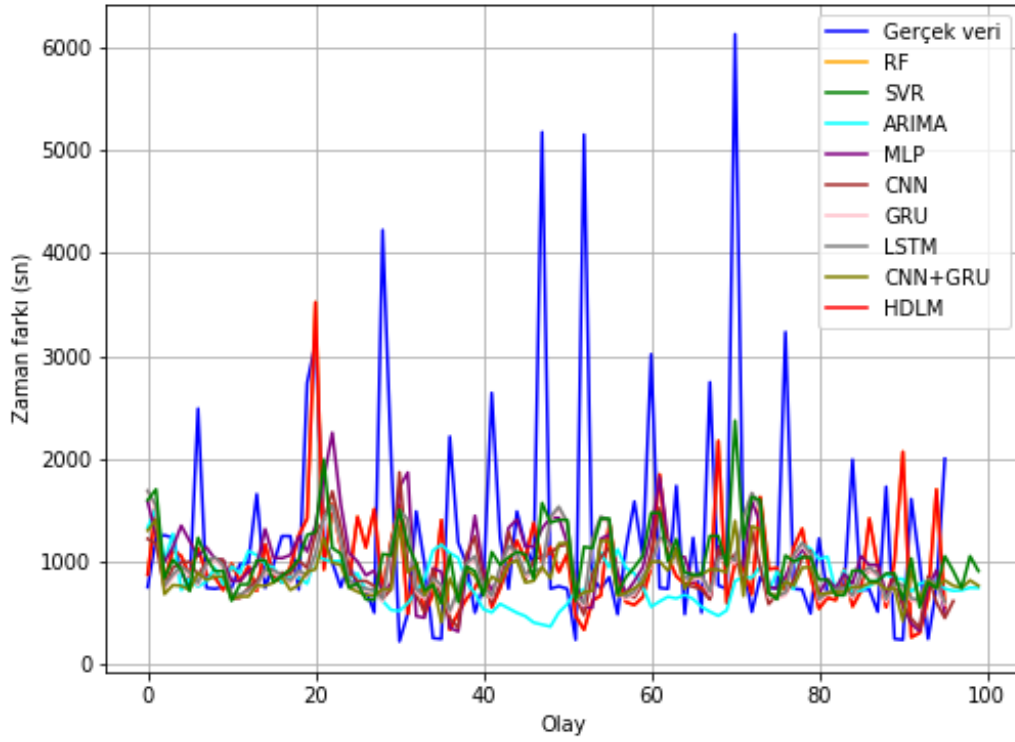
DDoS saldırıları arasındaki zaman farkı, Şekil 6.6’da görülmektedir.



Şekil 6.6. DDoS saldırıları arasındaki zaman farkı

Şekil 6.6’da DDoS saldırıları arasındaki zaman farkı görülmektedir. DDoS veri kümesi 10000 DDoS saldırı olayından oluşmaktadır. Veri kümesinde nadiren de olsa görülen sapmalar mevcuttur.

Şekil 6.7’de örnek olarak test verisinin 100 elemanı için deneysel sonuçlar görülmektedir.



Şekil 6.7. Test verisindeki ilk 100 veri örneği için deneysel sonuçlar

Şekil 6.7’de test verisindeki ilk 100 veri örneği için deneysel sonuçlar görülmektedir. HDLM, diğer modellere kıyasla veri kümesindeki ani sapmaları daha başarılı bir şekilde belirleyerek daha yüksek tahmin doğruluğuna sahip olmuştur.

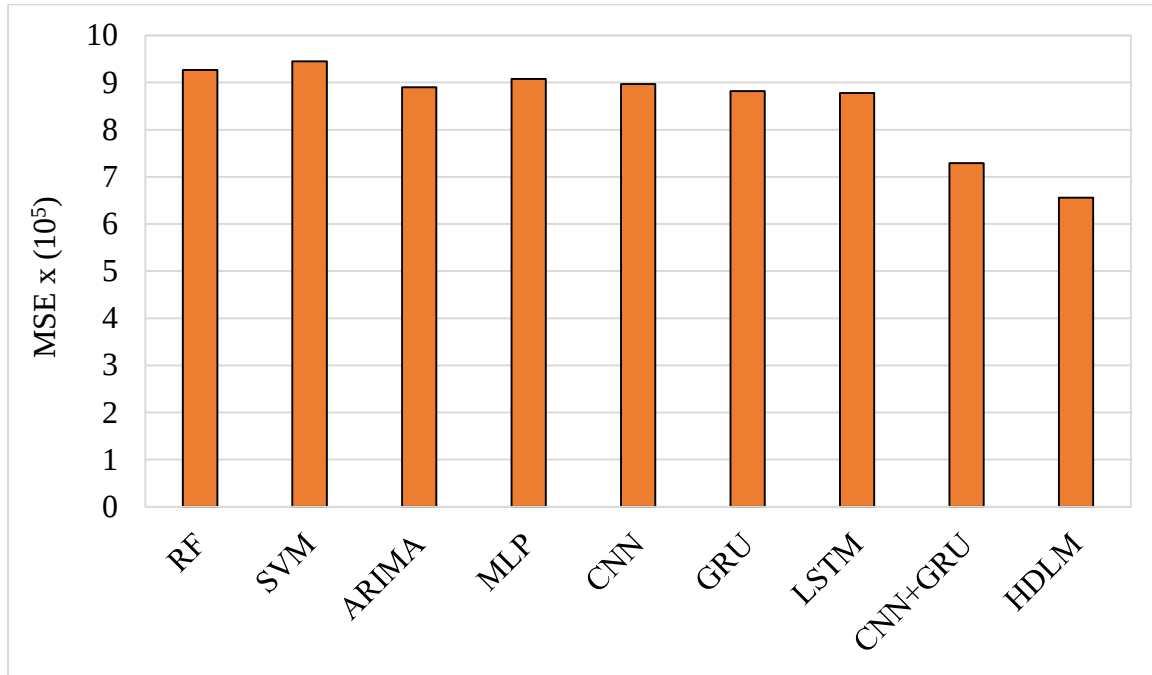
MSE metriğine göre deneysel sonuçlar Çizelge 6.9’da görülmektedir.

Çizelge 6.9. MSE metriğine göre deneysel sonuçlar

Test	RF	SVM	ARIMA	MLP	CNN	GRU	LSTM	CNN+GRU	HDLM
1	927563,75	945917,68	890578,26	907823,38	894625,97	882008,27	878080,72	714511,45	639853,49
2	927563,75	945917,68	890578,26	908035,98	895069,48	882008,27	878081,51	717404,14	650916,98
3	927563,75	945917,68	890578,26	908137,22	895099,43	882008,27	878081,57	718377,84	651453,84
4	927563,75	945917,68	890578,26	908145,54	896101,36	882008,27	878081,81	727102,98	653790,47
5	927563,75	945917,68	890578,26	908150,91	896782,50	882008,27	878083,08	728297,43	654491,66
6	927563,75	945917,68	890578,26	908159,40	897181,65	882008,27	878083,09	732678,30	655599,16
7	927563,75	945917,68	890578,26	908166,77	898928,98	882008,27	878083,18	733330,93	658685,87
8	927563,75	945917,68	890578,26	908179,27	899255,81	882008,27	878083,51	734569,90	664250,79
9	927563,75	945917,68	890578,26	908297,34	901592,20	882008,27	878083,64	743200,72	666075,34
10	927563,75	945917,68	890578,26	908428,87	903493,07	882008,27	878084,64	746606,85	666542,55
Ort.	927563,75	945917,68	890578,26	908152,46	897813,04	882008,27	878082,67	729608,05	656166,01

MSE metriğine göre yapılan deneysel çalışmalar sonucunda, RF'ın 927563,75, SVM'in 945917,68, ARIMA'nın 890578,26, MLP'nin 908152,46, CNN'in 897813,04, GRU'nun 882008,27, LSTM'in 878082,67, CNN+GRU'nun 729608,05 ve HDLM'nin 656166,01 ortalama MSE değerine sahip olduğu görülmüştür.

Çizelge 6.9'da ve Şekil 6.8'de HDLM'nin RF, SVM, ARIMA, MLP, CNN, GRU, LSTM ve CNN+GRU'ya göre MSE metriği için karşılaştırmalı deneysel sonuçları görülmektedir.



Şekil 6.8. MSE metriği için karşılaştırmalı deneysel sonuçlar

$$\text{HDLM'nin } \frac{\text{RF-HDLM}}{\text{RF}} \times 100, \frac{\text{SVR-HDLM}}{\text{SVR}} \times 100, \frac{\text{ARIMA-HDLM}}{\text{ARIMA}} \times 100, \frac{\text{MLP-HDLM}}{\text{MLP}} \times 100, \\ \frac{\text{CNN-HDLM}}{\text{CNN}} \times 100, \frac{\text{GRU-HDLM}}{\text{GRU}} \times 100, \frac{\text{LSTM-HDLM}}{\text{LSTM}} \times 100 \text{ ve } \frac{(\text{CNN+GRU})-\text{HDLM}}{(\text{CNN+GRU})} \times 100$$

eşitlikleri kullanılarak MSE metriğine göre hesaplanan iyileşme oranları Çizelge 6.10'da görülmektedir.

Çizelge 6.10. MSE metriğine göre HDLM'nin iyileşme oranları

	HDLM'nin en kötü değeri	HDLM'nin en iyi değeri	HDLM'nin ortalama değeri
RF'in değerine göre iyileşme oranı (%)	28,14	31,01	29,25
SVM'nin değerine göre iyileşme oranı (%)	29,53	32,35	30,63
ARIMA'nın değerine göre iyileşme oranı (%)	25,15	28,15	26,32
MLP'nin en kötü değerine göre iyileşme oranı (%)	26,62	29,56	27,76
MLP'nin en iyi değerine göre iyileşme oranı (%)	26,57	29,51	27,72
MLP'nin ortalama değerine göre iyileşme oranı (%)	26,61	29,54	27,74
CNN'in en kötü değerine göre iyileşme oranı (%)	26,22	29,18	27,37
CNN'in en iyi değerine göre iyileşme oranı (%)	25,49	28,47	26,65
CNN'in ortalama değerine göre iyileşme oranı (%)	25,85	28,73	26,91
GRU'nun en kötü değerine göre iyileşme oranı (%)	24,42	27,45	25,60
GRU'nun en iyi değerine göre iyileşme oranı (%)	24,42	27,45	25,60
GRU'nun ortalama değerine göre iyileşme oranı (%)	24,42	27,45	25,60
LSTM'in en kötü değerine göre iyileşme oranı (%)	24,09	27,13	25,27
LSTM'in en iyi değerine göre iyileşme oranı (%)	24,09	27,13	25,27

Çizelge 6.10. (devam) MSE metriğine göre HDLM'nin iyileşme oranları

LSTM'in ortalama değerine göre iyileşme oranı (%)	24,09	27,13	25,27
CNN+GRU'nun en kötü değerine göre iyileşme oranı (%)	10,72	14,29	12,11
CNN+GRU'nun en iyi değerine göre iyileşme oranı (%)	6,71	10,44	8,16
CNN+GRU'nun ortalama değerine göre iyileşme oranı (%)	8,82	12,30	10,06

Yapılan 10 farklı deney sonucunda elde edilen MSE değerlerine göre HDLM'nin RF, SVM, ARIMA, MLP, CNN, GRU, LSTM ve CNN+GRU'ya göre karşılaştırmalı bir analizi görülmektedir. HDLM'nin en kötü değeri RF'a göre %28,14, en iyi değeri %31,01 ve ortalama değeri ise %29,25 iyileşme sağlamıştır. HDLM'nin en kötü değeri SVM'e göre %29,53, en iyi değeri %32,35 ve ortalama değeri ise %30,63 iyileşme sağlamıştır. HDLM'nin en kötü değeri ARIMA'ya göre %25,15, en iyi değeri %28,15 ve ortalama değeri ise %26,32 iyileşme sağlamıştır.

HDLM'nin en kötü değeri MLP'nin en kötü değerine göre %26,62, MLP'nin en iyi değerine göre %26,57 ve MLP'nin ortalama değerine göre %26,61 iyileşme sağlamıştır. HDLM'nin en iyi değeri MLP'nin en kötü değerine göre %29,56, MLP'nin en iyi değerine göre %29,51 ve MLP'nin ortalama değerine göre %29,54 iyileşme sağlamıştır. HDLM'nin ortalama değeri MLP'nin en kötü değerine göre %27,76, MLP'nin en iyi değerine göre %27,72 ve MLP'nin ortalama değerine göre %27,74 iyileşme sağlamıştır.

HDLM'nin en kötü değeri CNN'in en kötü değerine göre %26,22, CNN'in en iyi değerine göre %25,49 ve CNN'in ortalama değerine göre %25,85 iyileşme sağlamıştır. HDLM'nin en iyi değeri CNN'in en kötü değerine göre %29,18, CNN'in en iyi değerine göre %28,47 ve CNN'in ortalama değerine göre %28,73 iyileşme sağlamıştır. HDLM'nin ortalama değeri CNN'in en kötü değerine göre %27,37, CNN'in en iyi değerine göre %26,65 ve CNN'in ortalama değerine göre %26,91 iyileşme sağlamıştır.

HDLM'nin en kötü değeri GRU'nun en kötü değerine göre %24,42, GRU'nun en iyi değerine göre %24,42 ve GRU'nun ortalama değerine göre %24,42 iyileşme sağlamıştır. HDLM'nin en iyi değeri GRU'nun en kötü değerine göre %27,45, GRU'nun en iyi değerine göre %27,45 ve GRU'nun ortalama değerine göre %27,45 iyileşme sağlamıştır. HDLM'nin ortalama değeri GRU'nun en kötü değerine göre %25,60, GRU'nun en iyi değerine göre %25,60 ve GRU'nun ortalama değerine göre %25,60 iyileşme sağlamıştır.

HDLM'nin en kötü değeri LSTM'in en kötü değerine göre %24,09, LSTM'in en iyi değerine göre %24,09 ve LSTM'in ortalama değerine göre %24,09 iyileşme sağlamıştır. HDLM'nin en iyi değeri LSTM'in en kötü değerine göre %27,13, LSTM'in en iyi değerine göre %27,13 ve LSTM'in ortalama değerine göre %27,13 iyileşme sağlamıştır. HDLM'nin ortalama değeri LSTM'in en kötü değerine göre %25,27, LSTM'in en iyi değerine göre %25,27 ve LSTM'in ortalama değerine göre %25,27 iyileşme sağlamıştır.

HDLM'nin en kötü değeri CNN+GRU'nun en kötü değerine göre %10,72, CNN+GRU'nun en iyi değerine göre %6,71 ve CNN+GRU'nun ortalama değerine göre %8,82 iyileşme sağlamıştır. HDLM'nin en iyi değeri CNN+GRU'nun en kötü değerine göre %14,29, CNN+GRU'nun en iyi değerine göre %10,44 ve CNN+GRU'nun ortalama değerine göre %12,30 iyileşme sağlamıştır. HDLM'nin ortalama değeri CNN+GRU'nun en kötü değerine göre %12,11, CNN+GRU'nun en iyi değerine göre %8,16 ve CNN+GRU'nun ortalama değerine göre %10,06 iyileşme sağlamıştır.

RMSE metriğine göre deneysel sonuçlar Çizelge 6.11'de görülmektedir.

Çizelge 6.11. RMSE metriğine göre deneysel sonuçlar

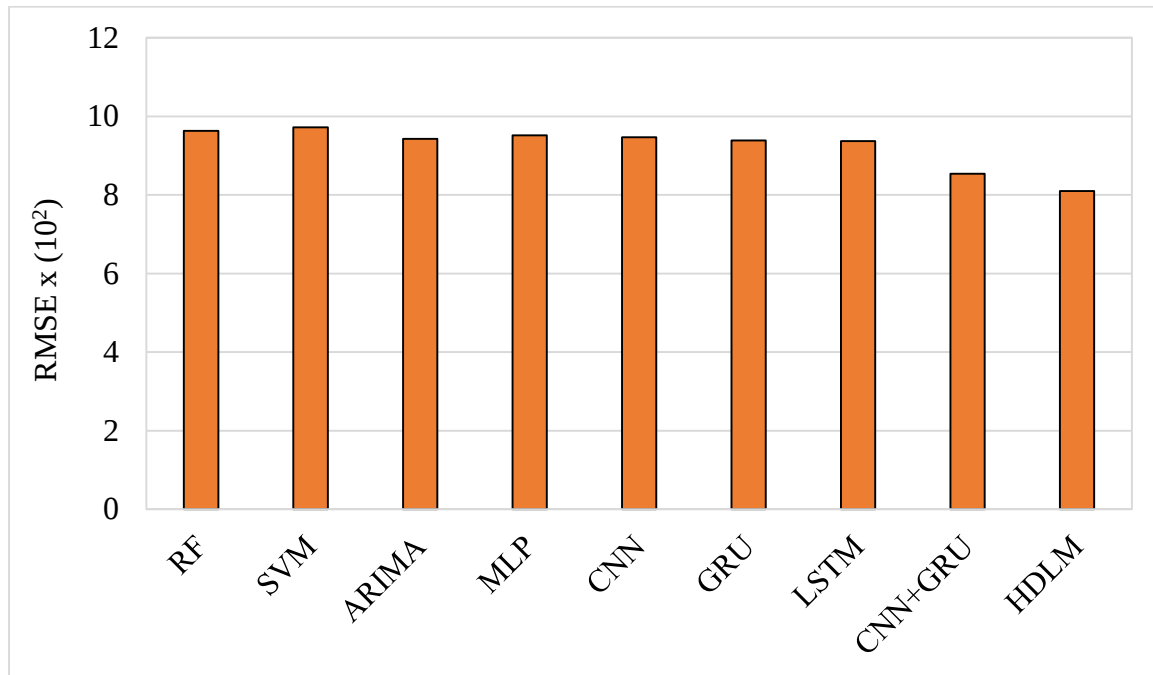
Test	RF	SVM	ARIMA	MLP	CNN	GRU	LSTM	CNN+GRU	HDLM
1	963,10	972,58	943,70	952,80	945,84	939,15	937,06	845,28	799,90
2	963,10	972,58	943,70	952,91	946,08	939,15	937,06	846,99	806,79
3	963,10	972,58	943,70	952,96	946,09	939,15	937,06	847,57	807,12
4	963,10	972,58	943,70	952,97	946,62	939,15	937,06	852,70	808,57
5	963,10	972,58	943,70	952,97	946,98	939,15	937,06	853,40	809,00
6	963,10	972,58	943,70	952,97	947,19	939,15	937,06	855,96	809,69
7	963,10	972,58	943,70	952,98	948,11	939,15	937,06	856,34	811,59
8	963,10	972,58	943,70	952,98	948,29	939,15	937,06	857,07	815,01

Çizelge 6.11. (devam) RMSE metriğine göre deneysel sonuçlar

9	963,10	972,58	943,70	953,05	949,52	939,15	937,06	862,09	816,13
10	963,10	972,58	943,70	953,12	950,52	939,15	937,06	864,06	816,42
Ort.	963,10	972,58	943,70	952,97	947,52	939,15	937,06	854,14	810,02

RMSE metriğine göre yapılan deneysel çalışmalar sonucunda, RF'ın 963,10, SVM'in 972,58, ARIMA'nın 943,70, MLP'nin 952,97, CNN'in 947,52, GRU'nun 939,15, LSTM'in 937,06, CNN+GRU'nun 854,14 ve HDLM'nin 810,02 ortalama RMSE değerine sahip olduğu görülmüştür.

Çizelge 6.11'de ve Şekil 6.9'da HDLM'nin RF, SVM, ARIMA, MLP, CNN, GRU, LSTM ve CNN+GRU'ya göre RMSE metriği için karşılaştırmalı deneysel sonuçları görülmektedir.



Şekil 6.9. RMSE metriği için karşılaştırmalı deneysel sonuçlar

$$\begin{aligned}
 &\text{HDLM'nin } \frac{\text{RF-HDLM}}{\text{RF}} \times 100, \frac{\text{SVR-HDLM}}{\text{SVR}} \times 100, \frac{\text{ARIMA-HDLM}}{\text{ARIMA}} \times 100, \frac{\text{MLP-HDLM}}{\text{MLP}} \times 100, \\
 &\frac{\text{CNN-HDLM}}{\text{CNN}} \times 100, \frac{\text{GRU-HDLM}}{\text{GRU}} \times 100, \frac{\text{LSTM-HDLM}}{\text{LSTM}} \times 100 \text{ ve } \frac{(\text{CNN+GRU})\text{-HDLM}}{(\text{CNN+GRU})} \times 100
 \end{aligned}$$

eşitlikleri kullanılarak RMSE metriğine göre hesaplanan iyileşme oranları Çizelge 6.12’de görülmektedir.

Çizelge 6.12. RMSE metriğine göre HDLM’nin iyileşme oranları

	HDLM’nin en kötü değeri	HDLM’nin en iyi değeri	HDLM’nin ortalama değeri
RF’in değerine göre iyileşme oranı (%)	15,22	16,94	15,89
SVM’nin değerine göre iyileşme oranı (%)	16,05	17,75	16,71
ARIMA’nın değerine göre iyileşme oranı (%)	13,48	15,23	14,16
MLP’nin en kötü değerine göre iyileşme oranı (%)	14,34	16,07	15,01
MLP’nin en iyi değerine göre iyileşme oranı (%)	14,31	16,04	14,98
MLP’nin ortalama değerine göre iyileşme oranı (%)	14,33	16,06	15,00
CNN’in en kötü değerine göre iyileşme oranı (%)	14,10	15,84	14,78
CNN’in en iyi değerine göre iyileşme oranı (%)	13,68	15,42	14,35
CNN’in ortalama değerine göre iyileşme oranı (%)	13,86	15,57	14,51
GRU’nun en kötü değerine göre iyileşme oranı (%)	13,06	14,82	13,74
GRU’nun en iyi değerine göre iyileşme oranı (%)	13,06	14,82	13,74
GRU’nun ortalama değerine göre iyileşme oranı (%)	13,06	14,82	13,74

Çizelge 6.12. (devam) RMSE metriğine göre HDLM'nin iyileşme oranları

LSTM'in en kötü değerine göre iyileşme oranı (%)	12,87	14,63	13,55
LSTM'in en iyi değerine göre iyileşme oranı (%)	12,87	14,63	13,55
LSTM'in ortalama değerine göre iyileşme oranı (%)	12,87	14,63	13,55
CNN+GRU'nun en kötü değerine göre iyileşme oranı (%)	5,51	7,42	6,25
CNN+GRU'nun en iyi değerine göre iyileşme oranı (%)	3,41	5,36	4,17
CNN+GRU'nun ortalama değerine göre iyileşme oranı (%)	4,46	6,35	5,16

Yapılan 10 farklı deney sonucunda elde edilen RMSE değerlerine göre HDLM'nin RF, SVM, ARIMA, MLP, CNN, GRU, LSTM ve CNN+GRU'ya göre karşılaştırmalı bir analizi görülmektedir. HDLM'nin en kötü değeri RF'a göre %15,22, en iyi değeri %16,94 ve ortalama değeri ise %15,89 iyileşme sağlamıştır. HDLM'nin en kötü değeri SVM'e göre %16,05, en iyi değeri %17,75 ve ortalama değeri ise %16,71 iyileşme sağlamıştır. HDLM'nin en kötü değeri ARIMA'ya göre %13,48, en iyi değeri %15,23 ve ortalama değeri ise %14,16 iyileşme sağlamıştır.

HDLM'nin en kötü değeri MLP'nin en kötü değerine göre %14,34, MLP'nin en iyi değerine göre %14,31 ve MLP'nin ortalama değerine göre %14,33 iyileşme sağlamıştır. HDLM'nin en iyi değeri MLP'nin en kötü değerine göre %16,07, MLP'nin en iyi değerine göre %16,04 ve MLP'nin ortalama değerine göre %16,06 iyileşme sağlamıştır. HDLM'nin ortalama değeri MLP'nin en kötü değerine göre %15,01, MLP'nin en iyi değerine göre %14,98 ve MLP'nin ortalama değerine göre %15,00 iyileşme sağlamıştır.

HDLM'nin en kötü değeri CNN'in en kötü değerine göre %14,10, CNN'in en iyi değerine göre %13,68 ve CNN'in ortalama değerine göre %13,86 iyileşme sağlamıştır. HDLM'nin en iyi değeri CNN'in en kötü değerine göre %15,84, CNN'in en iyi değerine göre %15,42

ve CNN'in ortalama deęerine g re %15,57 iyileşme saęlamıştır. HDLM'nin ortalama deęeri CNN'in en k t  deęerine g re %14,78, CNN'in en iyi deęerine g re %14,35 ve CNN'in ortalama deęerine g re %14,51 iyileşme saęlamıştır.

HDLM'nin en k t  deęeri GRU'nun en k t  deęerine g re %13,06, GRU'nun en iyi deęerine g re %13,06 ve GRU'nun ortalama deęerine g re %13,06 iyileşme saęlamıştır. HDLM'nin en iyi deęeri GRU'nun en k t  deęerine g re %14,82, GRU'nun en iyi deęerine g re %14,82 ve GRU'nun ortalama deęerine g re %14,82 iyileşme saęlamıştır. HDLM'nin ortalama deęeri GRU'nun en k t  deęerine g re %13,74, GRU'nun en iyi deęerine g re %13,74 ve GRU'nun ortalama deęerine g re %13,74 iyileşme saęlamıştır.

HDLM'nin en k t  deęeri LSTM'in en k t  deęerine g re %12,87, LSTM'in en iyi deęerine g re %12,87 ve LSTM'in ortalama deęerine g re %12,87 iyileşme saęlamıştır. HDLM'nin en iyi deęeri LSTM'in en k t  deęerine g re %14,63, LSTM'in en iyi deęerine g re %14,63 ve LSTM'in ortalama deęerine g re %14,63 iyileşme saęlamıştır. HDLM'nin ortalama deęeri LSTM'in en k t  deęerine g re %13,55, LSTM'in en iyi deęerine g re %13,55 ve LSTM'in ortalama deęerine g re %13,55 iyileşme saęlamıştır.

HDLM'nin en k t  deęeri CNN+GRU'nun en k t  deęerine g re %5,51, CNN+GRU'nun en iyi deęerine g re %3,41 ve CNN+GRU'nun ortalama deęerine g re %4,46 iyileşme saęlamıştır. HDLM'nin en iyi deęeri CNN+GRU'nun en k t  deęerine g re %7,42, CNN+GRU'nun en iyi deęerine g re %5,36 ve CNN+GRU'nun ortalama deęerine g re %6,35 iyileşme saęlamıştır. HDLM'nin ortalama deęeri CNN+GRU'nun en k t  deęerine g re %6,25, CNN+GRU'nun en iyi deęerine g re %4,17 ve CNN+GRU'nun ortalama deęerine g re %5,16 iyileşme saęlamıştır.

MAE metrięine g re deneysel sonular izelge 6.13'te g r lmektedir.

izelge 6.13. MAE metrięine g re deneysel sonular

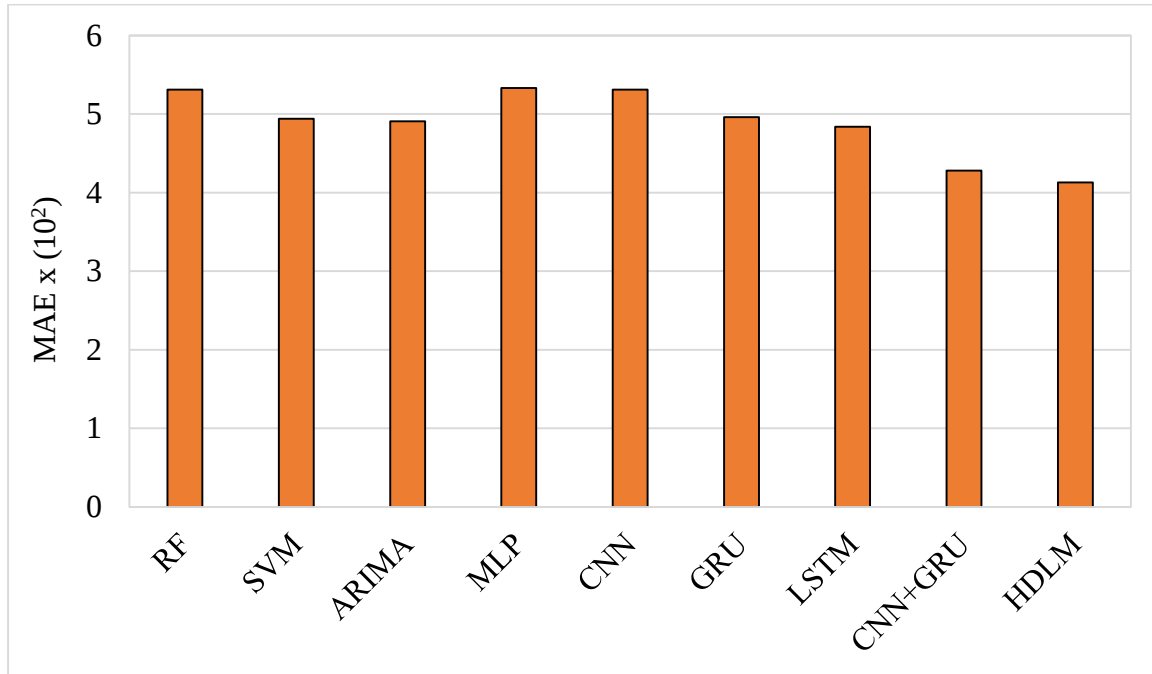
Test	RF	SVM	ARIMA	MLP	CNN	GRU	LSTM	CNN+GRU	HDLM
1	531,13	494,21	491,66	530,78	511,21	496,30	484,76	420,66	399,07
2	531,13	494,21	491,66	530,96	511,85	496,30	484,77	422,26	400,87
3	531,13	494,21	491,66	531,03	512,00	496,30	484,77	424,43	412,55

Çizelge 6.13. (devam) MAE metriğine göre deneysel sonuçlar

4	531,13	494,21	491,66	531,08	512,07	496,30	484,77	427,59	416,92
5	531,13	494,21	491,66	531,08	512,20	496,30	484,77	429,17	417,03
6	531,13	494,21	491,66	531,09	512,37	496,30	484,77	430,50	417,18
7	531,13	494,21	491,66	531,10	512,50	496,30	484,77	431,43	417,98
8	531,13	494,21	491,66	531,17	512,73	496,30	484,77	431,56	418,66
9	531,13	494,21	491,66	531,31	513,10	496,30	484,77	433,26	418,68
10	531,13	494,21	491,66	531,38	513,33	496,30	484,77	437,30	420,50
Ortalama	531,13	494,21	491,66	531,09	512,33	496,30	484,76	428,81	413,94

MAE metriğine göre yapılan deneysel çalışmalar sonucunda, RF'in 531,13, SVM'in 494,21, ARIMA'nın 491,66, MLP'nin 531,09, CNN'in 512,33, GRU'nun 496,30, LSTM'in 484,76, CNN+GRU'nun 428,81 ve HDLM'nin 413,94 ortalama MAE değerine sahip olduğu görülmüştür.

Çizelge 6.13'te ve Şekil 6.10'da HDLM'nin RF, SVM, ARIMA, MLP, CNN, GRU, LSTM ve CNN+GRU'ya göre MAE metriği için karşılaştırmalı deneysel sonuçları görülmektedir.



Şekil 6.10. MAE metriği için karşılaştırmalı deneysel sonuçlar

$$\text{HDLM'nin } \frac{\text{RF-HDLM}}{\text{RF}} \times 100, \frac{\text{SVR-HDLM}}{\text{SVR}} \times 100, \frac{\text{ARIMA-HDLM}}{\text{ARIMA}} \times 100, \frac{\text{MLP-HDLM}}{\text{MLP}} \times 100, \\ \frac{\text{CNN-HDLM}}{\text{CNN}} \times 100, \frac{\text{GRU-HDLM}}{\text{GRU}} \times 100, \frac{\text{LSTM-HDLM}}{\text{LSTM}} \times 100 \text{ ve } \frac{(\text{CNN+GRU})\text{-HDLM}}{(\text{CNN+GRU})} \times 100$$

eşitlikleri kullanılarak MAE metriğine göre hesaplanan iyileşme oranları Çizelge 6.14'te görülmektedir.

Çizelge 6.14. MAE metriğine göre HDLM'nin iyileşme oranları

	HDLM'nin en kötü değeri	HDLM'nin en iyi değeri	HDLM'nin ortalama değeri
RF'in değerine göre iyileşme oranı (%)	20,82	24,86	22,06
SVM'nin değerine göre iyileşme oranı (%)	14,91	19,25	16,24
ARIMA'nın değerine göre iyileşme oranı (%)	14,47	18,83	15,80
MLP'nin en kötü değerine göre iyileşme oranı (%)	20,86	24,89	22,10
MLP'nin en iyi değerine göre iyileşme oranı (%)	20,77	24,81	22,01
MLP'nin ortalama değerine göre iyileşme oranı (%)	20,83	24,85	22,05
CNN'in en kötü değerine göre iyileşme oranı (%)	18,08	22,25	19,36
CNN'in en iyi değerine göre iyileşme oranı (%)	17,74	21,93	19,02
CNN'in ortalama değerine göre iyileşme oranı (%)	17,96	22,10	19,20
GRU'nun en kötü değerine göre iyileşme oranı (%)	15,27	19,59	16,59
GRU'nun en iyi değerine göre iyileşme oranı (%)	15,27	19,59	16,59

Çizelge 6.14. (devam) MAE metriğine göre HDLM'nin iyileşme oranları

GRU'nun ortalama değerine göre iyileşme oranı (%)	15,27	19,59	16,59
LSTM'in en kötü değerine göre iyileşme oranı (%)	13,25	17,67	14,61
LSTM'in en iyi değerine göre iyileşme oranı (%)	13,25	17,67	14,60
LSTM'in ortalama değerine göre iyileşme oranı (%)	13,25	17,67	14,61
CNN+GRU'nun en kötü değerine göre iyileşme oranı (%)	3,84	8,74	5,34
CNN+GRU'nun en iyi değerine göre iyileşme oranı (%)	0,03	5,13	1,59
CNN+GRU'nun ortalama değerine göre iyileşme oranı (%)	1,97	6,93	3,46

Yapılan 10 farklı deney sonucunda elde edilen MAE değerlerine göre HDLM'nin RF, SVM, ARIMA, MLP, CNN, GRU, LSTM ve CNN+GRU'ya göre karşılaştırmalı bir analizi görülmektedir. HDLM'nin en kötü değeri RF'a göre %20,82, en iyi değeri %24,86 ve ortalama değeri ise %22,06 iyileşme sağlamıştır. HDLM'nin en kötü değeri SVM'e göre %14,91, en iyi değeri %19,25 ve ortalama değeri ise %16,24 iyileşme sağlamıştır. HDLM'nin en kötü değeri ARIMA'ya göre %14,47, en iyi değeri %18,83 ve ortalama değeri ise %15,80 iyileşme sağlamıştır.

HDLM'nin en kötü değeri MLP'nin en kötü değerine göre %20,86, MLP'nin en iyi değerine göre %20,77 ve MLP'nin ortalama değerine göre %20,83 iyileşme sağlamıştır. HDLM'nin en iyi değeri MLP'nin en kötü değerine göre %24,89, MLP'nin en iyi değerine göre %24,81 ve MLP'nin ortalama değerine göre %24,85 iyileşme sağlamıştır. HDLM'nin ortalama değeri MLP'nin en kötü değerine göre %22,10, MLP'nin en iyi değerine göre %22,01 ve MLP'nin ortalama değerine göre %22,05 iyileşme sağlamıştır.

HDLM'nin en kötü değeri CNN'in en kötü değerine göre %18,08, CNN'in en iyi değerine göre %17,74 ve CNN'in ortalama değerine göre %17,96 iyileşme sağlamıştır. HDLM'nin en iyi değeri CNN'in en kötü değerine göre %22,25, CNN'in en iyi değerine göre %21,93 ve CNN'in ortalama değerine göre %22,10 iyileşme sağlamıştır. HDLM'nin ortalama değeri CNN'in en kötü değerine göre %19,36, CNN'in en iyi değerine göre %19,02 ve CNN'in ortalama değerine göre %19,20 iyileşme sağlamıştır.

HDLM'nin en kötü değeri GRU'nun en kötü değerine göre %15,27, GRU'nun en iyi değerine göre %15,27 ve GRU'nun ortalama değerine göre %15,27 iyileşme sağlamıştır. HDLM'nin en iyi değeri GRU'nun en kötü değerine göre %19,59, GRU'nun en iyi değerine göre %19,59 ve GRU'nun ortalama değerine göre %19,59 iyileşme sağlamıştır. HDLM'nin ortalama değeri GRU'nun en kötü değerine göre %16,59, GRU'nun en iyi değerine göre %16,59 ve GRU'nun ortalama değerine göre %16,59 iyileşme sağlamıştır.

HDLM'nin en kötü değeri LSTM'in en kötü değerine göre %13,25, LSTM'in en iyi değerine göre %13,25 ve LSTM'in ortalama değerine göre %13,25 iyileşme sağlamıştır. HDLM'nin en iyi değeri LSTM'in en kötü değerine göre %17,67, LSTM'in en iyi değerine göre %17,67 ve LSTM'in ortalama değerine göre %17,67 iyileşme sağlamıştır. HDLM'nin ortalama değeri LSTM'in en kötü değerine göre %14,61, LSTM'in en iyi değerine göre %14,60 ve LSTM'in ortalama değerine göre %14,61 iyileşme sağlamıştır.

HDLM'nin en kötü değeri CNN+GRU'nun en kötü değerine göre %3,84, CNN+GRU'nun en iyi değerine göre %0,03 ve CNN+GRU'nun ortalama değerine göre %1,97 iyileşme sağlamıştır. HDLM'nin en iyi değeri CNN+GRU'nun en kötü değerine göre %8,74, CNN+GRU'nun en iyi değerine göre %5,13 ve CNN+GRU'nun ortalama değerine göre %6,93 iyileşme sağlamıştır. HDLM'nin ortalama değeri CNN+GRU'nun en kötü değerine göre %5,34, CNN+GRU'nun en iyi değerine göre %1,59 ve CNN+GRU'nun ortalama değerine göre %3,46 iyileşme sağlamıştır.

Deneyisel sonuçlar, HDLM'nin, RF, SVM, ARIMA, MLP, CNN, GRU, LSTM ve CNN+GRU'ya göre daha düşük MSE, RMSE ve MAE sonuçlarına sahip olduğunu

göstermektedir. Bir sonraki saldırı zamanının belirlenmesinde HDLM'nin diğer modellere göre daha başarılı ve etkin sonuçlar verdiği görülmüştür.

6.1.3. Suç veri kümesi

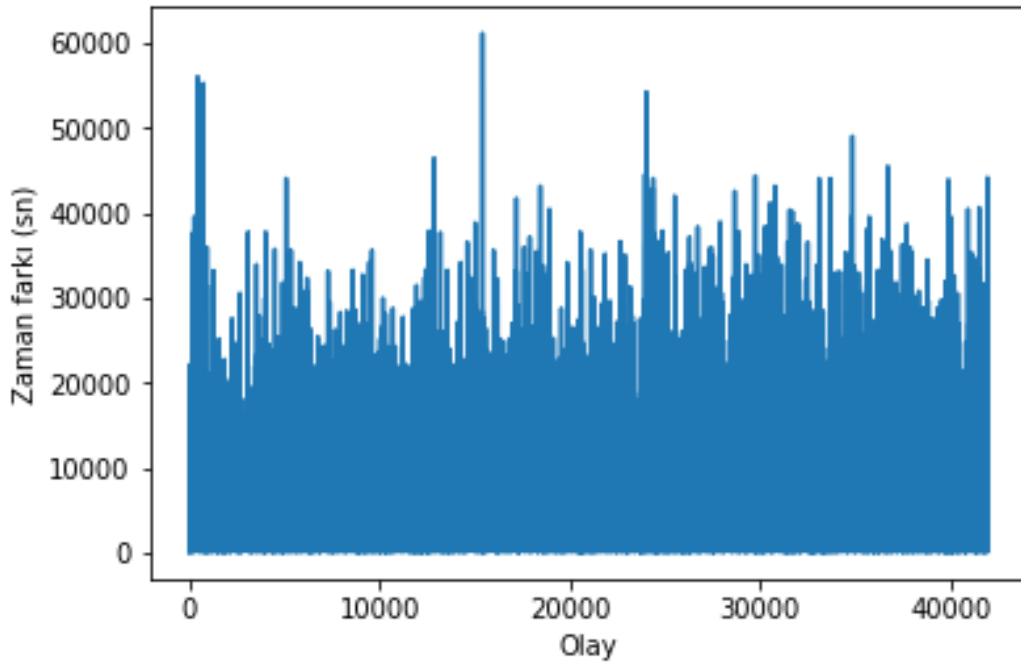
Suç veri kümesindeki 4E kodlu çete saldırısı olaylarının meydana gelme zamanları arasındaki saniye cinsinden farklara göre oluşturulan verinin ilk 5 satırı Çizelge 6.15'te görülmektedir.

Çizelge 6.15. Çete saldırısı olaylarının meydana gelme zamanları arasındaki fark

Olay	Bir önceki çete saldırısından sonra geçen zaman (sn)
1	540
2	900
3	60
4	720
5	120

Çizelge 6.15'te çete saldırısı olayları arasındaki zaman farkı görülmektedir. Suç veri kümesi, DDoS veri kümesine göre daha nispeten daha düzensiz, e-ticaret veri kümesine göre nispeten daha düzenli bir örüntü izlemektedir.

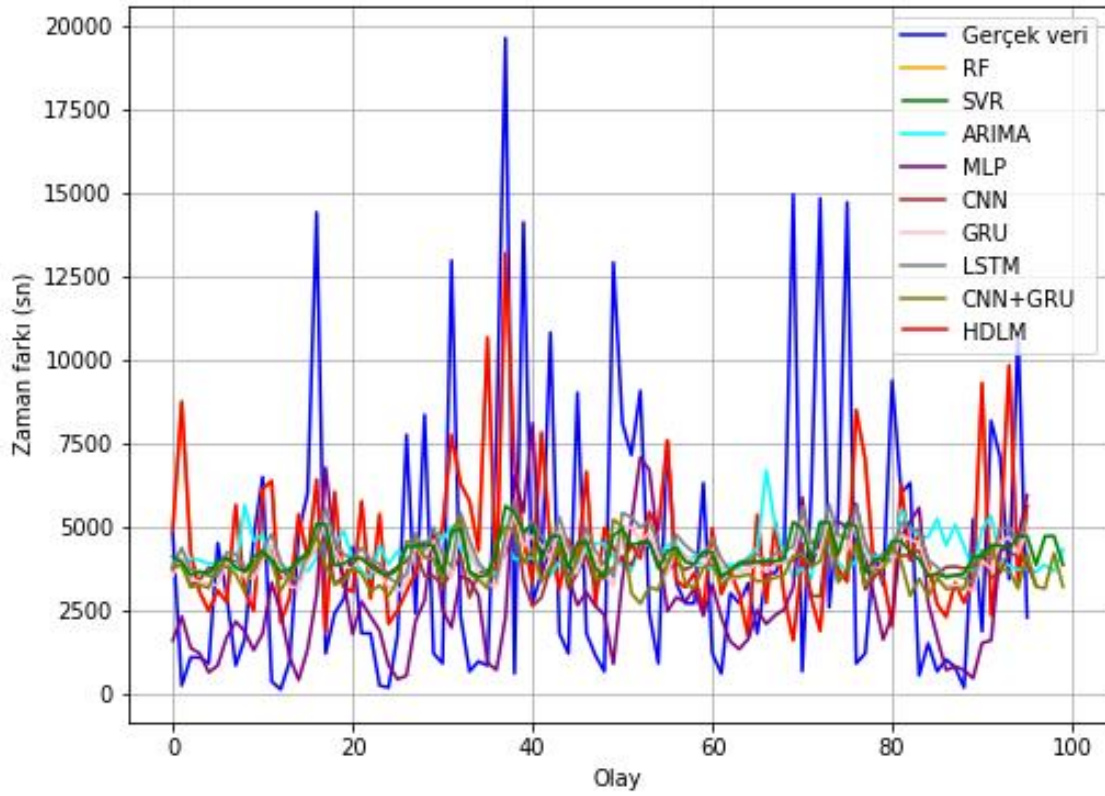
Çete saldırısı olayları arasındaki saniye cinsinden zaman farkları Şekil 6.11'de görülmektedir.



Şekil 6.11. Çete saldırısı olayları arasındaki zaman farkı

Şekil 6.11’de çete saldırısı olayları arasındaki zaman farkı görülmektedir. Suç veri kümesi, 41967 çete saldırısı olayından oluşmaktadır. Veri kümesinde nadiren de olsa görülen sapmalar mevcuttur.

Şekil 6.12’de örnek olarak test verisinin 100 elemanı için deneysel sonuçlar görülmektedir.



Şekil 6.12. Test verisindeki ilk 100 veri örneği için deneysel sonuçlar

Şekil 6.12’de test verisindeki ilk 100 veri örneği için deneysel sonuçlar görülmektedir. HDLM, diğer modellere kıyasla veri kümesindeki ani sapmaları daha başarılı bir şekilde belirleyerek daha yüksek tahmin doğruluğuna sahip olmuştur.

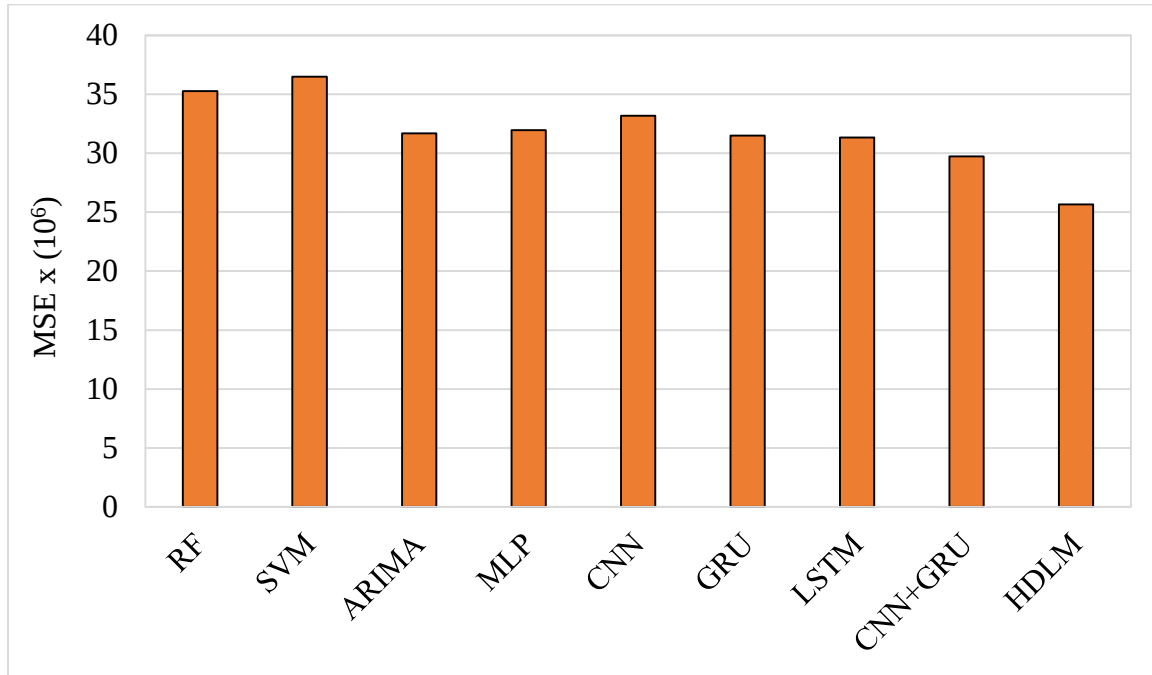
MSE metriğine göre deneysel sonuçlar Çizelge 6.16’da görülmektedir.

Çizelge 6.16. MSE metriğine göre deneysel sonuçlar

Test	RF	SVM	ARIMA	MLP	CNN	GRU	LSTM	CNN+GRU	HDLN
1	35275057,40	36498217,05	31709828,72	31866709,42	32626342,08	31512465,88	31306043,78	28628880,96	25113633,87
2	35275057,40	36498217,05	31709828,72	31948138,18	32837307,54	31512940,23	31308170,63	29051914,23	25311198,31
3	35275057,40	36498217,05	31709828,72	31968051,36	32990589,04	31512999,60	31341493,07	29533753,60	25428661,84
4	35275057,40	36498217,05	31709828,72	31970027,17	33036122,37	31513114,09	31341762,95	29582301,62	25501812,08
5	35275057,40	36498217,05	31709828,72	31973266,39	33245621,50	31513298,17	31342939,61	29661267,01	25582353,56
6	35275057,40	36498217,05	31709828,72	31991166,14	33334171,11	31513437,95	31343726,13	30041565,58	25664099,24
7	35275057,40	36498217,05	31709828,72	31992210,27	33400541,96	31513669,54	31344663,28	30141770,99	25683846,57
8	35275057,40	36498217,05	31709828,72	32003919,30	33435775,26	31514011,75	31345630,36	30161409,16	26005370,49
9	35275057,40	36498217,05	31709828,72	32024498,61	33455645,52	31514262,57	31346412,56	30277096,64	26034210,51
10	35275057,40	36498217,05	31709828,72	32036527,60	33637796,96	31514775,44	31350648,50	30365009,69	26389970,68
Ortalama	35275057,40	36498217,05	31709828,72	31977451,44	33199991,33	31513497,52	31337149,09	29744496,95	25671515,72

MSE metriğine göre yapılan deneysel çalışmalar sonucunda, RF'in 35275057,40, SVM'in 36498217,05, ARIMA'nın 31709828,72, MLP'nin 31977451,44, CNN'in 33199991,33, GRU'nun 31513497,52, LSTM'in 31337149,09, CNN+GRU'nun 29744496,95 ve HDLM'nin 25671515,72 ortalama MSE değerine sahip olduğu görülmüştür.

Çizelge 6.16'da ve Şekil 6.13'te HDLM'nin RF, SVM, ARIMA, MLP, CNN, GRU, LSTM ve CNN+GRU'ya göre MSE metriği için karşılaştırmalı deneysel sonuçları görülmektedir.



Şekil 6.13. MSE metriği için karşılaştırmalı deneysel sonuçlar

$$\text{HDLM'nin } \frac{\text{RF-HDLM}}{\text{RF}} \times 100, \frac{\text{SVR-HDLM}}{\text{SVR}} \times 100, \frac{\text{ARIMA-HDLM}}{\text{ARIMA}} \times 100, \frac{\text{MLP-HDLM}}{\text{MLP}} \times 100, \\ \frac{\text{CNN-HDLM}}{\text{CNN}} \times 100, \frac{\text{GRU-HDLM}}{\text{GRU}} \times 100, \frac{\text{LSTM-HDLM}}{\text{LSTM}} \times 100 \text{ ve } \frac{(\text{CNN+GRU})\text{-HDLM}}{(\text{CNN+GRU})} \times 100$$

eşitlikleri kullanılarak MSE metriğine göre hesaplanan iyileşme oranları Çizelge 6.17'de görülmektedir.

Çizelge 6.17. MSE metriğine göre HDLM'nin iyileşme oranları

	HDLM'nin en kötü değeri	HDLM'nin en iyi değeri	HDLM'nin ortalama değeri
RF'in değerine göre iyileşme oranı (%)	25,18	28,80	27,22
SVM'nin değerine göre iyileşme oranı (%)	27,69	31,19	29,66
ARIMA'nın değerine göre iyileşme oranı (%)	16,77	20,80	19,04
MLP'nin en kötü değerine göre iyileşme oranı (%)	17,62	21,60	19,86
MLP'nin en iyi değerine göre iyileşme oranı (%)	17,18	21,19	19,44
MLP'nin ortalama değerine göre iyileşme oranı (%)	17,53	21,46	19,71
CNN'in en kötü değerine göre iyileşme oranı (%)	21,54	25,34	23,68
CNN'in en iyi değerine göre iyileşme oranı (%)	19,11	23,02	21,31
CNN'in ortalama değerine göre iyileşme oranı (%)	20,87	24,35	22,67
GRU'nun en kötü değerine göre iyileşme oranı (%)	16,26	20,31	18,54
GRU'nun en iyi değerine göre iyileşme oranı (%)	16,25	20,30	18,53
GRU'nun ortalama değerine göre iyileşme oranı (%)	16,25	20,30	18,53
LSTM'in en kötü değerine göre iyileşme oranı (%)	15,82	19,89	18,11
LSTM'in en iyi değerine göre iyileşme oranı (%)	15,70	19,78	17,99

Çizelge 6.17. (devam) MSE metriğine göre HDLM'nin iyileşme oranları

LSTM'in ortalama değerine göre iyileşme oranı (%)	15,80	19,85	18,07
CNN+GRU'nun en kötü değerine göre iyileşme oranı (%)	13,09	17,29	15,45
CNN+GRU'nun en iyi değerine göre iyileşme oranı (%)	7,82	12,27	10,33
CNN+GRU'nun ortalama değerine göre iyileşme oranı (%)	11,71	15,56	13,69

Yapılan 10 farklı deney sonucunda elde edilen MSE değerlerine göre HDLM'nin RF, SVM, ARIMA, MLP, CNN, GRU, LSTM ve CNN+GRU'ya göre karşılaştırmalı bir analizi görülmektedir. HDLM'nin en kötü değeri RF'a göre %25,18, en iyi değeri %28,80 ve ortalama değeri ise %27,22 iyileşme sağlamıştır. HDLM'nin en kötü değeri SVM'e göre %27,69, en iyi değeri %31,19 ve ortalama değeri ise %29,66 iyileşme sağlamıştır. HDLM'nin en kötü değeri ARIMA'ya göre %16,77, en iyi değeri %20,80 ve ortalama değeri ise %19,04 iyileşme sağlamıştır.

HDLM'nin en kötü değeri MLP'nin en kötü değerine göre %17,62, MLP'nin en iyi değerine göre %17,18 ve MLP'nin ortalama değerine göre %17,53 iyileşme sağlamıştır. HDLM'nin en iyi değeri MLP'nin en kötü değerine göre %21,60, MLP'nin en iyi değerine göre %21,19 ve MLP'nin ortalama değerine göre %21,46 iyileşme sağlamıştır. HDLM'nin ortalama değeri MLP'nin en kötü değerine göre %19,86, MLP'nin en iyi değerine göre %19,44 ve MLP'nin ortalama değerine göre %19,71 iyileşme sağlamıştır.

HDLM'nin en kötü değeri CNN'in en kötü değerine göre %21,54, CNN'in en iyi değerine göre %19,11 ve CNN'in ortalama değerine göre %20,87 iyileşme sağlamıştır. HDLM'nin en iyi değeri CNN'in en kötü değerine göre %25,34, CNN'in en iyi değerine göre %23,02 ve CNN'in ortalama değerine göre %24,35 iyileşme sağlamıştır. HDLM'nin ortalama değeri CNN'in en kötü değerine göre %23,68, CNN'in en iyi değerine göre %21,31 ve CNN'in ortalama değerine göre %22,67 iyileşme sağlamıştır.

HDLM'nin en kötü değeri GRU'nun en kötü değerine göre %16,26, GRU'nun en iyi değerine göre %16,25 ve GRU'nun ortalama değerine göre %16,25 iyileşme sağlamıştır. HDLM'nin en iyi değeri GRU'nun en kötü değerine göre %20,31, GRU'nun en iyi değerine göre %20,30 ve GRU'nun ortalama değerine göre %20,30 iyileşme sağlamıştır. HDLM'nin ortalama değeri GRU'nun en kötü değerine göre %18,54, GRU'nun en iyi değerine göre %18,53 ve GRU'nun ortalama değerine göre %18,53 iyileşme sağlamıştır.

HDLM'nin en kötü değeri LSTM'in en kötü değerine göre %15,82, LSTM'in en iyi değerine göre %15,70 ve LSTM'in ortalama değerine göre %15,80 iyileşme sağlamıştır. HDLM'nin en iyi değeri LSTM'in en kötü değerine göre %19,89, LSTM'in en iyi değerine göre %19,78 ve LSTM'in ortalama değerine göre %19,85 iyileşme sağlamıştır. HDLM'nin ortalama değeri LSTM'in en kötü değerine göre %18,11, LSTM'in en iyi değerine göre %17,99 ve LSTM'in ortalama değerine göre %18,07 iyileşme sağlamıştır.

HDLM'nin en kötü değeri CNN+GRU'nun en kötü değerine göre %13,09, CNN+GRU'nun en iyi değerine göre %7,82 ve CNN+GRU'nun ortalama değerine göre %11,71 iyileşme sağlamıştır. HDLM'nin en iyi değeri CNN+GRU'nun en kötü değerine göre %17,29, CNN+GRU'nun en iyi değerine göre %12,27 ve CNN+GRU'nun ortalama değerine göre %15,56 iyileşme sağlamıştır. HDLM'nin ortalama değeri CNN+GRU'nun en kötü değerine göre %15,45, CNN+GRU'nun en iyi değerine göre %10,33 ve CNN+GRU'nun ortalama değerine göre %13,69 iyileşme sağlamıştır.

RMSE metriğine göre deneysel sonuçlar Çizelge 6.18'de görülmektedir.

Çizelge 6.18. RMSE metriğine göre deneysel sonuçlar

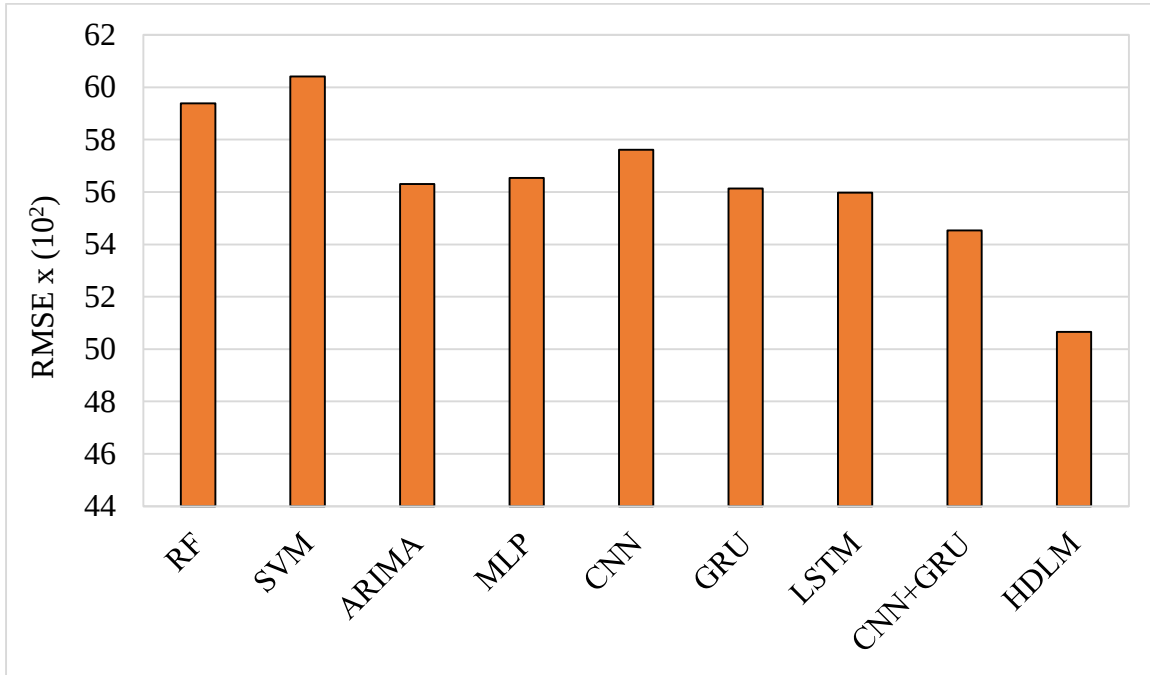
Test	RF	SVM	ARIMA	MLP	CNN	GRU	LSTM	CNN+GRU	HDLM
1	5939,28	6041,37	5631,14	5645,06	5711,94	5613,60	5595,18	5350,59	5011,35
2	5939,28	6041,37	5631,14	5652,27	5730,38	5613,64	5595,37	5389,98	5031,02
3	5939,28	6041,37	5631,14	5654,03	5743,74	5613,64	5598,35	5434,49	5042,68
4	5939,28	6041,37	5631,14	5654,20	5747,70	5613,65	5598,37	5438,96	5049,93
5	5939,28	6041,37	5631,14	5654,49	5765,90	5613,67	5598,48	5446,21	5057,90
6	5939,28	6041,37	5631,14	5656,07	5773,57	5613,68	5598,55	5481,01	5065,97
7	5939,28	6041,37	5631,14	5656,17	5779,32	5613,70	5598,63	5490,15	5067,92
8	5939,28	6041,37	5631,14	5657,20	5782,36	5613,73	5598,72	5491,94	5099,54
9	5939,28	6041,37	5631,14	5659,02	5784,08	5613,76	5598,79	5502,46	5102,37

Çizelge 6.18. (devam) RMSE metriğine göre deneysel sonuçlar

10	5939,28	6041,37	5631,14	5660,08	5799,81	5613,80	5599,16	5510,44	5137,11
Ortalama	5939,28	6041,37	5631,14	5654,85	5761,88	5613,68	5597,96	5453,62	5066,57

RMSE metriğine göre yapılan deneysel çalışmalar sonucunda, RF'in 5939,28, SVM'in 6041,37, ARIMA'nın 5631,14, MLP'nin 5654,85, CNN'in 5761,88, GRU'nun 5613,68, LSTM'in 5597,96, CNN+GRU'nun 5453,62 ve HDLM'nin 5066,57 ortalama RMSE değerine sahip olduğu görülmüştür.

Çizelge 6.18'de ve Şekil 6.14'te HDLM'nin RF, SVM, ARIMA, MLP, CNN, GRU, LSTM ve CNN+GRU'ya göre RMSE metriği için karşılaştırmalı deneysel sonuçları görülmektedir.



Şekil 6.14. RMSE metriği için karşılaştırmalı deneysel sonuçlar

$$\text{HDLM'nin } \frac{\text{RF-HDLM}}{\text{RF}} \times 100, \frac{\text{SVR-HDLM}}{\text{SVR}} \times 100, \frac{\text{ARIMA-HDLM}}{\text{ARIMA}} \times 100, \frac{\text{MLP-HDLM}}{\text{MLP}} \times 100, \\ \frac{\text{CNN-HDLM}}{\text{CNN}} \times 100, \frac{\text{GRU-HDLM}}{\text{GRU}} \times 100, \frac{\text{LSTM-HDLM}}{\text{LSTM}} \times 100 \text{ ve } \frac{(\text{CNN+GRU})-\text{HDLM}}{(\text{CNN+GRU})} \times 100$$

eşitlikleri kullanılarak RMSE metriğine göre hesaplanan iyileşme oranları Çizelge 6.19'da görülmektedir.

Çizelge 6.19. RMSE metriğine göre HDLM'nin iyileşme oranları

	HDLM'nin en kötü değeri	HDLM'nin en iyi değeri	HDLM'nin ortalama değeri
RF'in değerine göre iyileşme oranı (%)	13,50	15,62	14,69
SVM'nin değerine göre iyileşme oranı (%)	14,96	17,04	16,13
ARIMA'nın değerine göre iyileşme oranı (%)	8,77	11,00	10,02
MLP'nin en kötü değerine göre iyileşme oranı (%)	9,23	11,46	10,48
MLP'nin en iyi değerine göre iyileşme oranı (%)	8,99	11,22	10,24
MLP'nin ortalama değerine göre iyileşme oranı (%)	9,17	11,37	10,40
CNN'in en kötü değerine göre iyileşme oranı (%)	11,42	13,59	12,64
CNN'in en iyi değerine göre iyileşme oranı (%)	10,06	12,26	11,29
CNN'in ortalama değerine göre iyileşme oranı (%)	10,93	13,02	12,06
GRU'nun en kötü değerine göre iyileşme oranı (%)	8,49	10,73	9,74
GRU'nun en iyi değerine göre iyileşme oranı (%)	8,48	10,72	9,74
GRU'nun ortalama değerine göre iyileşme oranı (%)	8,48	10,72	9,74
LSTM'in en kötü değerine göre iyileşme oranı (%)	8,25	10,49	9,51
LSTM'in en iyi değerine göre iyileşme oranı (%)	8,18	10,43	9,44

Çizelge 6.19. (devam) RMSE metriğine göre HDLM'nin iyileşme oranları

LSTM'in ortalama değerine göre iyileşme oranı (%)	8,23	10,47	9,49
CNN+GRU'nun en kötü değerine göre iyileşme oranı (%)	6,77	9,05	8,05
CNN+GRU'nun en iyi değerine göre iyileşme oranı (%)	3,98	6,34	5,30
CNN+GRU'nun ortalama değerine göre iyileşme oranı (%)	5,91	8,10	7,09

Yapılan 10 farklı deney sonucunda elde edilen RMSE değerlerine göre HDLM'nin RF, SVM, ARIMA, MLP, CNN, GRU, LSTM ve CNN+GRU'ya göre karşılaştırmalı bir analizi görülmektedir. HDLM'nin en kötü değeri RF'a göre %13,50, en iyi değeri %15,62 ve ortalama değeri ise %14,69 iyileşme sağlamıştır. HDLM'nin en kötü değeri SVM'e göre %14,96, en iyi değeri %17,04 ve ortalama değeri ise %16,13 iyileşme sağlamıştır. HDLM'nin en kötü değeri ARIMA'ya göre %8,77, en iyi değeri %11,00 ve ortalama değeri ise %10,02 iyileşme sağlamıştır.

HDLM'nin en kötü değeri MLP'nin en kötü değerine göre %9,23, MLP'nin en iyi değerine göre %8,99 ve MLP'nin ortalama değerine göre %9,17 iyileşme sağlamıştır. HDLM'nin en iyi değeri MLP'nin en kötü değerine göre %11,46, MLP'nin en iyi değerine göre %11,22 ve MLP'nin ortalama değerine göre %11,37 iyileşme sağlamıştır. HDLM'nin ortalama değeri MLP'nin en kötü değerine göre %10,48, MLP'nin en iyi değerine göre %10,24 ve MLP'nin ortalama değerine göre %10,40 iyileşme sağlamıştır.

HDLM'nin en kötü değeri CNN'in en kötü değerine göre %11,42, CNN'in en iyi değerine göre %10,06 ve CNN'in ortalama değerine göre %10,93 iyileşme sağlamıştır. HDLM'nin en iyi değeri CNN'in en kötü değerine göre %13,59, CNN'in en iyi değerine göre %12,26 ve CNN'in ortalama değerine göre %13,02 iyileşme sağlamıştır. HDLM'nin ortalama değeri CNN'in en kötü değerine göre %12,64, CNN'in en iyi değerine göre %11,29 ve CNN'in ortalama değerine göre %12,06 iyileşme sağlamıştır.

HDLM'nin en kötü değeri GRU'nun en kötü değerine göre %8,49, GRU'nun en iyi değerine göre %8,48 ve GRU'nun ortalama değerine göre %8,48 iyileşme sağlamıştır. HDLM'nin en iyi değeri GRU'nun en kötü değerine göre %10,73, GRU'nun en iyi değerine göre %10,72 ve GRU'nun ortalama değerine göre %10,72 iyileşme sağlamıştır. HDLM'nin ortalama değeri GRU'nun en kötü değerine göre %9,74, GRU'nun en iyi değerine göre %9,74 ve GRU'nun ortalama değerine göre %9,74 iyileşme sağlamıştır.

HDLM'nin en kötü değeri LSTM'in en kötü değerine göre %8,25, LSTM'in en iyi değerine göre %8,18 ve LSTM'in ortalama değerine göre %8,23 iyileşme sağlamıştır. HDLM'nin en iyi değeri LSTM'in en kötü değerine göre %10,49, LSTM'in en iyi değerine göre %10,43 ve LSTM'in ortalama değerine göre %10,47 iyileşme sağlamıştır. HDLM'nin ortalama değeri LSTM'in en kötü değerine göre %9,51, LSTM'in en iyi değerine göre %9,44 ve LSTM'in ortalama değerine göre %9,49 iyileşme sağlamıştır.

HDLM'nin en kötü değeri CNN+GRU'nun en kötü değerine göre %6,77, CNN+GRU'nun en iyi değerine göre %3,98 ve CNN+GRU'nun ortalama değerine göre %5,91 iyileşme sağlamıştır. HDLM'nin en iyi değeri CNN+GRU'nun en kötü değerine göre %9,05, CNN+GRU'nun en iyi değerine göre %6,34 ve CNN+GRU'nun ortalama değerine göre %8,10 iyileşme sağlamıştır. HDLM'nin ortalama değeri CNN+GRU'nun en kötü değerine göre %8,05, CNN+GRU'nun en iyi değerine göre %5,30 ve CNN+GRU'nun ortalama değerine göre %7,09 iyileşme sağlamıştır.

MAE metriğine göre deneysel sonuçlar Çizelge 6.20'de görülmektedir.

Çizelge 6.20. MAE metriğine göre deneysel sonuçlar

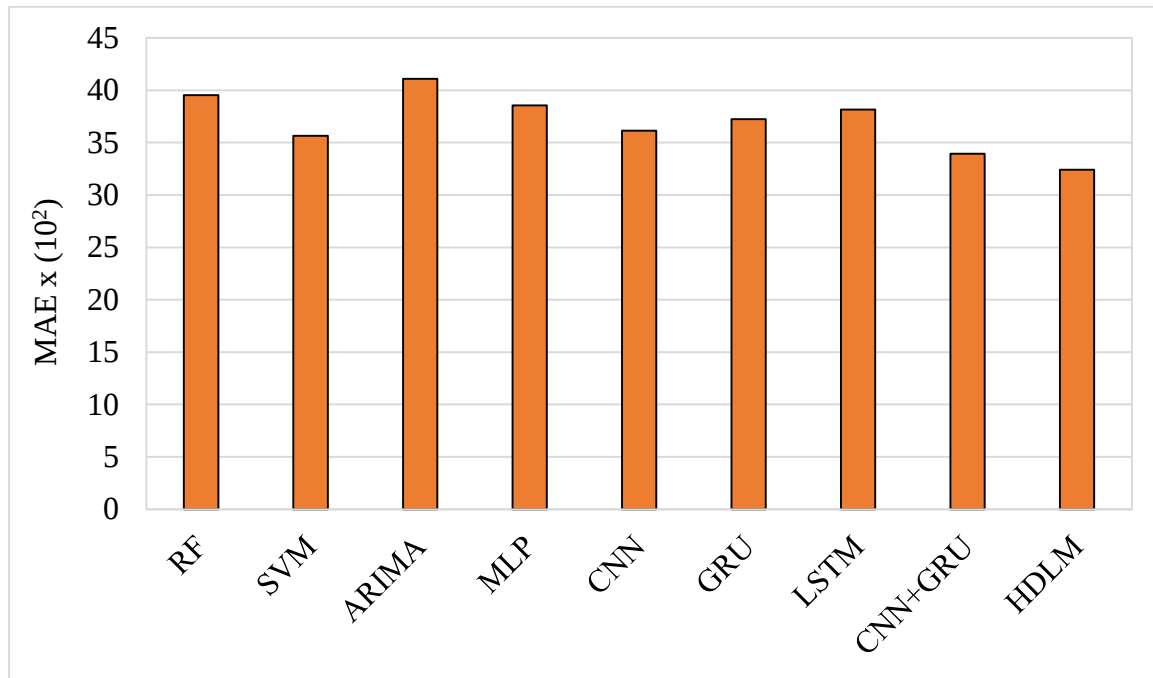
Test	RF	SVM	ARIMA	MLP	CNN	GRU	LSTM	CNN+GRU	HDLM
1	3952,24	3565,33	4108,62	3744,04	3598,19	3723,89	3795,77	3367,77	3181,06
2	3952,24	3565,33	4108,62	3863,33	3600,30	3723,99	3795,82	3368,03	3193,52
3	3952,24	3565,33	4108,62	3853,05	3600,50	3724,06	3795,99	3384,73	3201,07
4	3952,24	3565,33	4108,62	3859,49	3603,09	3724,22	3796,04	3399,24	3226,27
5	3952,24	3565,33	4108,62	3862,81	3603,59	3724,23	3796,24	3402,53	3229,41
6	3952,24	3565,33	4108,62	3865,45	3604,17	3724,24	3796,53	3404,72	3246,93
7	3952,24	3565,33	4108,62	3870,86	3614,73	3724,37	3796,91	3405,69	3261,21
8	3952,24	3565,33	4108,62	3872,78	3625,73	3724,44	3796,91	3406,90	3273,00
9	3952,24	3565,33	4108,62	3881,91	3630,49	3724,46	3890,87	3407,08	3287,64

Çizelge 6.20. (devam) MAE metriğine göre deneysel sonuçlar

10	3952,24	3565,33	4108,62	3883,68	3659,99	3724,60	3891,01	3407,71	3300,34
Ortalama	3952,24	3565,33	4108,62	3855,74	3614,07	3724,25	3815,20	3395,44	3240,04

MAE metriğine göre yapılan deneysel çalışmalar sonucunda, RF'in 3952,24, SVM'in 3565,33, ARIMA'nın 4108,62, MLP'nin 3855,74, CNN'in 3614,07, GRU'nun 3724,25, LSTM'in 3815,20, CNN+GRU'nun 3395,44 ve HDLM'nin 3240,04 ortalama MAE değerine sahip olduğu görülmüştür.

Çizelge 6.20'de ve Şekil 6.15'te HDLM'nin RF, SVM, ARIMA, MLP, CNN, GRU, LSTM ve CNN+GRU'ya göre MAE metriği için karşılaştırmalı deneysel sonuçları görülmektedir.



Şekil 6.15. MAE metriği için karşılaştırmalı deneysel sonuçlar

$$\text{HDLM'nin } \frac{\text{RF-HDLM}}{\text{RF}} \times 100, \frac{\text{SVR-HDLM}}{\text{SVR}} \times 100, \frac{\text{ARIMA-HDLM}}{\text{ARIMA}} \times 100, \frac{\text{MLP-HDLM}}{\text{MLP}} \times 100, \\ \frac{\text{CNN-HDLM}}{\text{CNN}} \times 100, \frac{\text{GRU-HDLM}}{\text{GRU}} \times 100, \frac{\text{LSTM-HDLM}}{\text{LSTM}} \times 100 \text{ ve } \frac{(\text{CNN+GRU})\text{-HDLM}}{(\text{CNN+GRU})} \times 100$$

eşitlikleri kullanılarak MAE metriğine göre hesaplanan iyileşme oranları Çizelge 6.21'de görülmektedir.

Çizelge 6.21. MAE metriğine göre HDLM'nin iyileşme oranları

	HDLM'nin en kötü değeri	HDLM'nin en iyi değeri	HDLM'nin ortalama değeri
RF'in değerine göre iyileşme oranı (%)	16,49	19,51	18,02
SVM'nin değerine göre iyileşme oranı (%)	7,43	10,77	9,12
ARIMA'nın değerine göre iyileşme oranı (%)	19,67	22,57	21,14
MLP'nin en kötü değerine göre iyileşme oranı (%)	15,02	18,09	16,57
MLP'nin en iyi değerine göre iyileşme oranı (%)	11,85	15,03	13,46
MLP'nin ortalama değerine göre iyileşme oranı (%)	14,83	17,49	15,96
CNN'in en kötü değerine göre iyileşme oranı (%)	9,82	13,08	11,47
CNN'in en iyi değerine göre iyileşme oranı (%)	8,27	11,59	9,95
CNN'in ortalama değerine göre iyileşme oranı (%)	8,71	11,98	10,34
GRU'nun en kötü değerine göre iyileşme oranı (%)	11,39	14,59	13,00
GRU'nun en iyi değerine göre iyileşme oranı (%)	11,37	14,57	12,99
GRU'nun ortalama değerine göre iyileşme oranı (%)	11,38	14,58	13,00
LSTM'in en kötü değerine göre iyileşme oranı (%)	15,18	18,24	16,72
LSTM'in en iyi değerine göre iyileşme oranı (%)	13,05	16,19	14,64

Çizelge 6.21. (devam) MAE metriğine göre HDLM'nin iyileşme oranları

LSTM'in ortalama değerine göre iyileşme oranı (%)	13,56	16,62	15,07
CNN+GRU'nun en kötü değerine göre iyileşme oranı (%)	3,15	6,65	4,92
CNN+GRU'nun en iyi değerine göre iyileşme oranı (%)	2,00	5,54	3,79
CNN+GRU'nun ortalama değerine göre iyileşme oranı (%)	2,82	6,31	4,57

Yapılan 10 farklı deney sonucunda elde edilen MAE değerlerine göre HDLM'nin RF, SVM, ARIMA, MLP, CNN, GRU, LSTM ve CNN+GRU'ya göre karşılaştırmalı bir analizi görülmektedir. HDLM'nin en kötü değeri RF'a göre %16,49, en iyi değeri %19,51 ve ortalama değeri ise %18,02 iyileşme sağlamıştır. HDLM'nin en kötü değeri SVM'e göre %7,43, en iyi değeri %10,77 ve ortalama değeri ise %9,12 iyileşme sağlamıştır. HDLM'nin en kötü değeri ARIMA'ya göre %19,67, en iyi değeri %22,57 ve ortalama değeri ise %21,14 iyileşme sağlamıştır.

HDLM'nin en kötü değeri MLP'nin en kötü değerine göre %15,02, MLP'nin en iyi değerine göre %11,85 ve MLP'nin ortalama değerine göre %14,83 iyileşme sağlamıştır. HDLM'nin en iyi değeri MLP'nin en kötü değerine göre %18,09, MLP'nin en iyi değerine göre %15,03 ve MLP'nin ortalama değerine göre %17,49 iyileşme sağlamıştır. HDLM'nin ortalama değeri MLP'nin en kötü değerine göre %16,57, MLP'nin en iyi değerine göre %13,46 ve MLP'nin ortalama değerine göre %15,96 iyileşme sağlamıştır.

HDLM'nin en kötü değeri CNN'in en kötü değerine göre %9,82, CNN'in en iyi değerine göre %8,27 ve CNN'in ortalama değerine göre %8,71 iyileşme sağlamıştır. HDLM'nin en iyi değeri CNN'in en kötü değerine göre %13,08, CNN'in en iyi değerine göre %11,59 ve CNN'in ortalama değerine göre %11,98 iyileşme sağlamıştır. HDLM'nin ortalama değeri CNN'in en kötü değerine göre %11,47, CNN'in en iyi değerine göre %9,95 ve CNN'in ortalama değerine göre %10,34 iyileşme sağlamıştır.

HDLM'nin en kötü değeri GRU'nun en kötü değerine göre %11,39, GRU'nun en iyi değerine göre %11,37 ve GRU'nun ortalama değerine göre %11,38 iyileşme sağlamıştır. HDLM'nin en iyi değeri GRU'nun en kötü değerine göre %14,59, GRU'nun en iyi değerine göre %14,57 ve GRU'nun ortalama değerine göre %14,58 iyileşme sağlamıştır. HDLM'nin ortalama değeri GRU'nun en kötü değerine göre %13,00, GRU'nun en iyi değerine göre %12,99 ve GRU'nun ortalama değerine göre %13,00 iyileşme sağlamıştır.

HDLM'nin en kötü değeri LSTM'in en kötü değerine göre %15,18, LSTM'in en iyi değerine göre %13,05 ve LSTM'in ortalama değerine göre %13,56 iyileşme sağlamıştır. HDLM'nin en iyi değeri LSTM'in en kötü değerine göre %18,24, LSTM'in en iyi değerine göre %16,19 ve LSTM'in ortalama değerine göre %16,62 iyileşme sağlamıştır. HDLM'nin ortalama değeri LSTM'in en kötü değerine göre %16,72, LSTM'in en iyi değerine göre %14,64 ve LSTM'in ortalama değerine göre %15,07 iyileşme sağlamıştır.

HDLM'nin en kötü değeri CNN+GRU'nun en kötü değerine göre %3,15, CNN+GRU'nun en iyi değerine göre %2,00 ve CNN+GRU'nun ortalama değerine göre %2,82 iyileşme sağlamıştır. HDLM'nin en iyi değeri CNN+GRU'nun en kötü değerine göre %6,65, CNN+GRU'nun en iyi değerine göre %5,54 ve CNN+GRU'nun ortalama değerine göre %6,31 iyileşme sağlamıştır. HDLM'nin ortalama değeri CNN+GRU'nun en kötü değerine göre %4,92, CNN+GRU'nun en iyi değerine göre %3,79 ve CNN+GRU'nun ortalama değerine göre %4,57 iyileşme sağlamıştır.

Deneyisel sonuçlar, HDLM'nin, RF, SVM, ARIMA, MLP, CNN, GRU, LSTM ve CNN+GRU'ya göre daha düşük MSE, RMSE ve MAE sonuçlarına sahip olduğunu göstermektedir. Bir sonraki suç işleme zamanının belirlenmesinde HDLM'nin diğer modellere göre daha başarılı ve etkin sonuçlar verdiği görülmüştür.

6.1.4. Acil çağrı veri kümesi

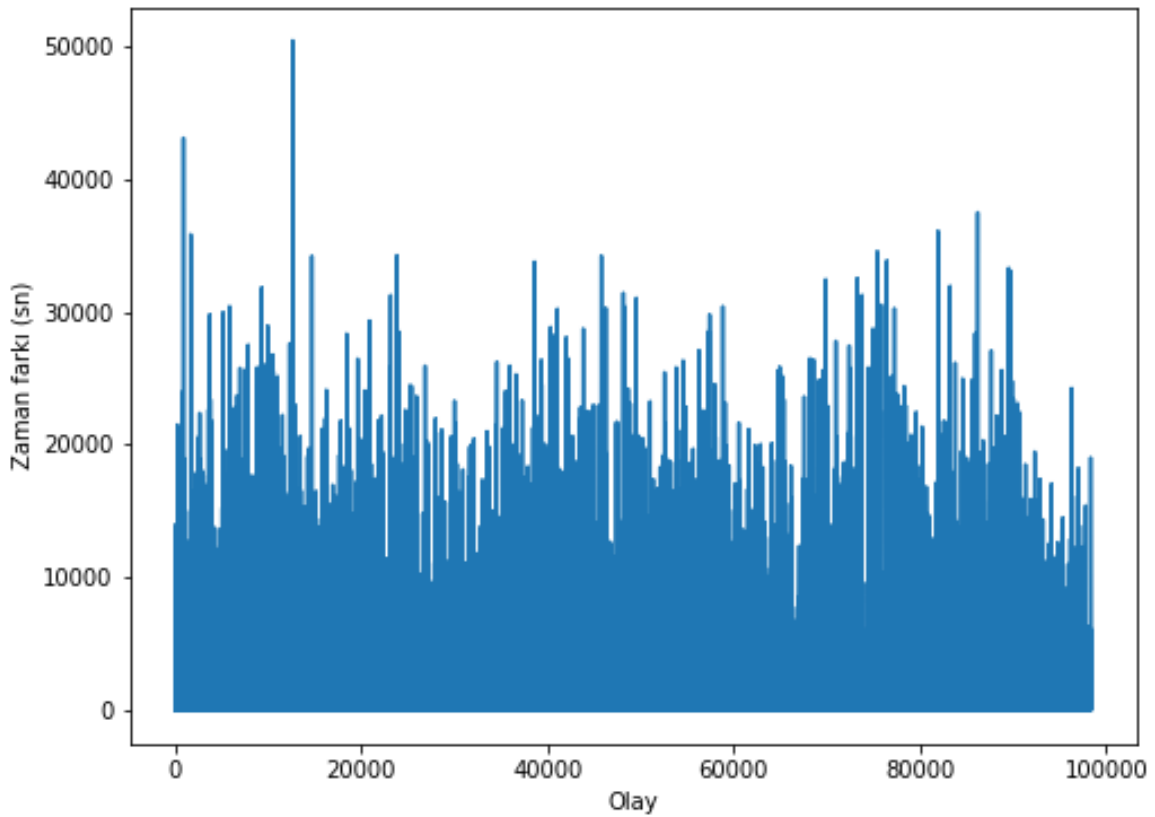
Acil çağrı veri kümesinde, trafik kazası çağrılarının meydana gelme zamanları arasındaki saniye cinsinden farklara göre oluşturulan verinin ilk 5 satırı Çizelge 6.22'de görülmektedir.

Çizelge 6.22. Trafik kazası çağrılarının meydana gelme zamanları arasındaki fark

Olay	Bir önceki trafik kazası çağrısından sonra geçen zaman (sn)
1	2259
2	178
3	182
4	911
5	170

Çizelge 6.22’de trafik kazası çağrıları arasındaki zaman farkı görülmektedir. Acil çağrı veri kümesi, e-ticaret veri kümesine göre nispeten daha düzenli, DDoS ve suç veri kümesine göre nispeten daha düzensiz bir örüntü izlemektedir.

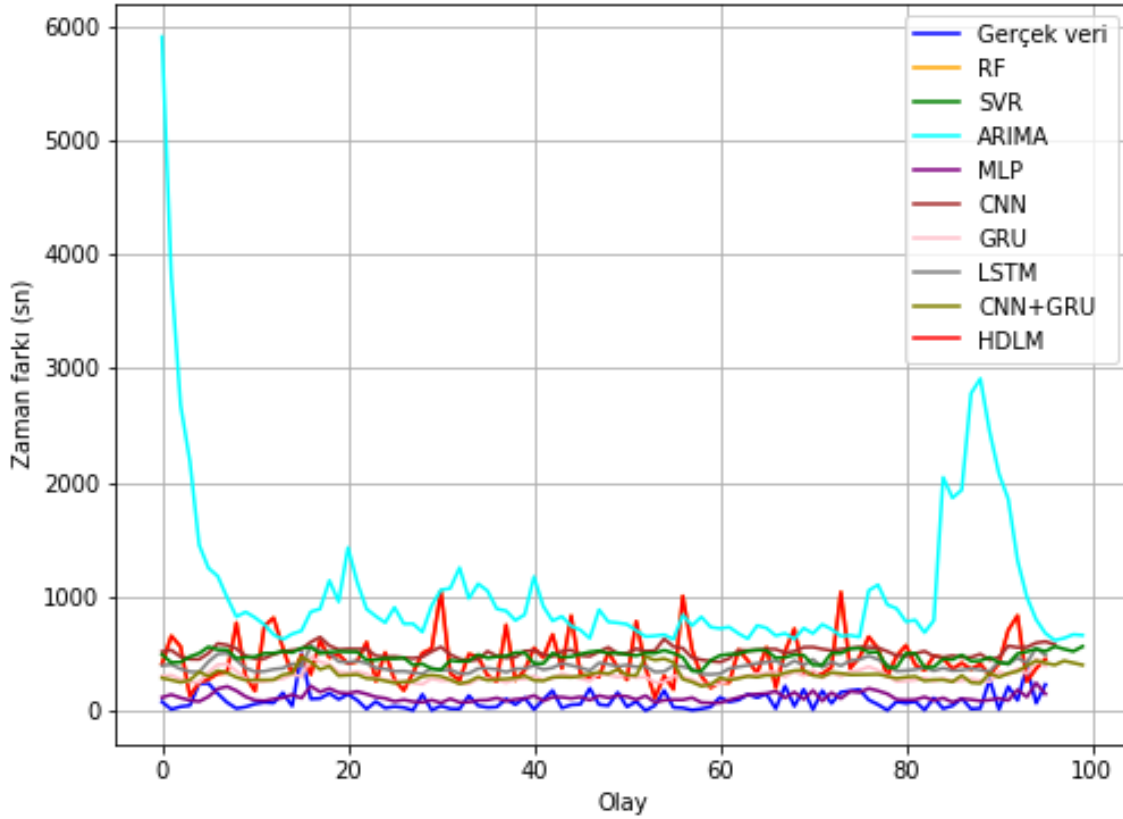
Trafik kazası çağrıları arasındaki saniye cinsinden zaman farkları Şekil 6.16’da görülmektedir.



Şekil 6.16. Trafik kazası çağrıları arasındaki zaman farkı

Şekil 6.16’da trafik kazası çağrıları arasındaki zaman farkı görülmektedir. Acil çağrı veri kümesi, 98401 trafik kazası çağrısı olayından oluşmaktadır. Veri kümesinde görülen sapmalar mevcuttur.

Şekil 6.17’de örnek olarak test verisinin 100 elemanı için deneysel sonuçlar görülmektedir.



Şekil 6.17. Test verisinin ilk 100 satırı için deneysel sonuçlar

Şekil 6.17’de test verisindeki ilk 100 veri örneği için deneysel sonuçlar görülmektedir. HDLM, diğer modellere kıyasla veri kümesindeki ani sapmaları daha başarılı bir şekilde belirleyerek daha yüksek tahmin doğruluğuna sahip olmuştur.

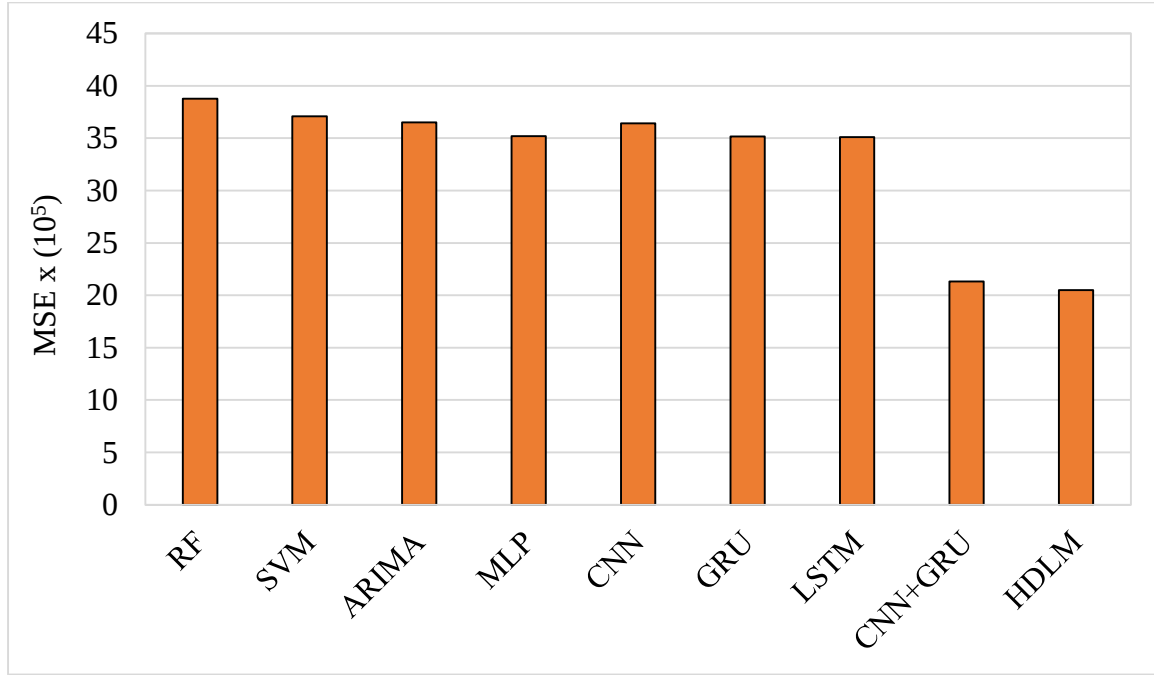
MSE metriğine göre deneysel sonuçlar Çizelge 6.23’te görülmektedir.

Çizelge 6.23. MSE metriğine göre deneysel sonuçlar

Test	RF	SVM	ARIMA	MLP	CNN	GRU	LSTM	CNN+GRU	HDLN
1	3876812,70	3709323,20	3652989,22	3510406,54	3632216,01	3516855,83	3492751,84	2119881,92	1994818,40
2	3876812,70	3709323,20	3652989,22	3515944,78	3635847,60	3516861,56	3492820,03	2120408,38	1999988,84
3	3876812,70	3709323,20	3652989,22	3517214,79	3637436,41	3516862,35	3514041,67	2125534,62	2022374,95
4	3876812,70	3709323,20	3652989,22	3517390,53	3637527,47	3516867,21	3514042,70	2128761,78	2026769,19
5	3876812,70	3709323,20	3652989,22	3520837,88	3638208,09	3516869,22	3514046,65	2129774,33	2060430,54
6	3876812,70	3709323,20	3652989,22	3523470,65	3639229,65	3516872,45	3514048,47	2134684,25	2066521,29
7	3876812,70	3709323,20	3652989,22	3523915,41	3642652,63	3516877,06	3514048,69	2134760,08	2073774,86
8	3876812,70	3709323,20	3652989,22	3526614,57	3649569,03	3516877,63	3514049,08	2138530,26	2083211,42
9	3876812,70	3709323,20	3652989,22	3527401,04	3650034,83	3516878,27	3514049,89	2142929,16	2087163,39
10	3876812,70	3709323,20	3652989,22	3532110,45	3650772,74	3516878,37	3514050,37	2144489,58	2095703,22
Ortalama	3876812,70	3709323,20	3652989,22	3521530,66	3641349,44	3516869,99	3509794,93	2131975,43	2051075,61

MSE metriğine göre yapılan deneysel çalışmalar sonucunda, RF'in 3876812,70, SVM'in 3709323,20, ARIMA'nın 3652989,22, MLP'nin 3521530,66, CNN'in 3641349,44, GRU'nun 3516869,99, LSTM'in 3509794,93, CNN+GRU'nun 2131975,43 ve HDLM'nin 2051075,61 ortalama MSE değerine sahip olduğu görülmüştür.

Çizelge 6.23'te ve Şekil 6.18'de HDLM'nin RF, SVM, ARIMA, MLP, CNN, GRU, LSTM ve CNN+GRU'ya göre MSE metriği için karşılaştırmalı deneysel sonuçları görülmektedir.



Şekil 6.18. MSE metriği için karşılaştırmalı deneysel sonuçlar

$$\text{HDLM'nin } \frac{\text{RF-HDLM}}{\text{RF}} \times 100, \frac{\text{SVR-HDLM}}{\text{SVR}} \times 100, \frac{\text{ARIMA-HDLM}}{\text{ARIMA}} \times 100, \frac{\text{MLP-HDLM}}{\text{MLP}} \times 100, \\ \frac{\text{CNN-HDLM}}{\text{CNN}} \times 100, \frac{\text{GRU-HDLM}}{\text{GRU}} \times 100, \frac{\text{LSTM-HDLM}}{\text{LSTM}} \times 100 \text{ ve } \frac{(\text{CNN+GRU})-\text{HDLM}}{(\text{CNN+GRU})} \times 100$$

eşitlikleri kullanılarak MSE metriğine göre hesaplanan iyileşme oranları Çizelge 6.24'te görülmektedir.

Çizelge 6.24. MSE metriğine göre HDLM'nin iyileşme oranları

	HDLM'nin en kötü değeri	HDLM'nin en iyi değeri	HDLM'nin ortalama değeri
RF'in değerine göre iyileşme oranı (%)	45,94	48,54	47,09
SVM'nin değerine göre iyileşme oranı (%)	43,50	46,22	44,70
ARIMA'nın değerine göre iyileşme oranı (%)	42,63	45,39	43,85
MLP'nin en kötü değerine göre iyileşme oranı (%)	40,66	43,52	41,93
MLP'nin en iyi değerine göre iyileşme oranı (%)	40,30	43,17	41,57
MLP'nin ortalama değerine göre iyileşme oranı (%)	40,61	43,35	41,75
CNN'in en kötü değerine göre iyileşme oranı (%)	42,59	45,35	43,81
CNN'in en iyi değerine göre iyileşme oranı (%)	42,30	45,07	43,53
CNN'in ortalama değerine göre iyileşme oranı (%)	42,55	45,21	43,67
GRU'nun en kötü değerine göre iyileşme oranı (%)	40,41	43,27	41,67
GRU'nun en iyi değerine göre iyileşme oranı (%)	40,40	43,27	41,67
GRU'nun ortalama değerine göre iyileşme oranı (%)	40,41	43,27	41,67
LSTM'in en kötü değerine göre iyileşme oranı (%)	40,36	43,23	41,63
LSTM'in en iyi değerine göre iyileşme oranı (%)	39,99	42,88	41,27

Çizelge 6.24. (devam) MSE metriğine göre HDLM'nin iyileşme oranları

LSTM'in ortalama değerine göre iyileşme oranı (%)	40,48	43,16	41,56
CNN+GRU'nun en kötü değerine göre iyileşme oranı (%)	2,27	6,97	4,35
CNN+GRU'nun en iyi değerine göre iyileşme oranı (%)	1,14	5,89	3,24
CNN+GRU'nun ortalama değerine göre iyileşme oranı (%)	1,71	6,43	3,79

Yapılan 10 farklı deney sonucunda elde edilen MSE değerlerine göre HDLM'nin RF, SVM, ARIMA, MLP, CNN, GRU, LSTM ve CNN+GRU'ya göre karşılaştırmalı bir analizi görülmektedir. HDLM'nin en kötü değeri RF'a göre %45,94, en iyi değeri %48,54 ve ortalama değeri ise %47,09 iyileşme sağlamıştır. HDLM'nin en kötü değeri SVM'e göre %43,50, en iyi değeri %46,22 ve ortalama değeri ise %44,70 iyileşme sağlamıştır. HDLM'nin en kötü değeri ARIMA'ya göre %42,63, en iyi değeri %45,39 ve ortalama değeri ise %43,85 iyileşme sağlamıştır.

HDLM'nin en kötü değeri MLP'nin en kötü değerine göre %40,66, MLP'nin en iyi değerine göre %40,30 ve MLP'nin ortalama değerine göre %40,61 iyileşme sağlamıştır. HDLM'nin en iyi değeri MLP'nin en kötü değerine göre %43,52, MLP'nin en iyi değerine göre %43,17 ve MLP'nin ortalama değerine göre %43,35 iyileşme sağlamıştır. HDLM'nin ortalama değeri MLP'nin en kötü değerine göre %41,93, MLP'nin en iyi değerine göre %41,57 ve MLP'nin ortalama değerine göre %41,75 iyileşme sağlamıştır.

HDLM'nin en kötü değeri CNN'in en kötü değerine göre %42,59, CNN'in en iyi değerine göre %42,30 ve CNN'in ortalama değerine göre %42,55 iyileşme sağlamıştır. HDLM'nin en iyi değeri CNN'in en kötü değerine göre %45,35, CNN'in en iyi değerine göre %45,07 ve CNN'in ortalama değerine göre %45,21 iyileşme sağlamıştır. HDLM'nin ortalama değeri CNN'in en kötü değerine göre %43,81, CNN'in en iyi değerine göre %43,53 ve CNN'in ortalama değerine göre %43,67 iyileşme sağlamıştır.

HDLM'nin en kötü değeri GRU'nun en kötü değerine göre %40,41, GRU'nun en iyi değerine göre %40,40 ve GRU'nun ortalama değerine göre %40,41 iyileşme sağlamıştır. HDLM'nin en iyi değeri GRU'nun en kötü değerine göre %43,27, GRU'nun en iyi değerine göre %43,27 ve GRU'nun ortalama değerine göre %43,27 iyileşme sağlamıştır. HDLM'nin ortalama değeri GRU'nun en kötü değerine göre %41,67, GRU'nun en iyi değerine göre %41,67 ve GRU'nun ortalama değerine göre %41,67 iyileşme sağlamıştır.

HDLM'nin en kötü değeri LSTM'in en kötü değerine göre %40,36, LSTM'in en iyi değerine göre %39,99 ve LSTM'in ortalama değerine göre %40,48 iyileşme sağlamıştır. HDLM'nin en iyi değeri LSTM'in en kötü değerine göre %43,23, LSTM'in en iyi değerine göre %42,88 ve LSTM'in ortalama değerine göre %43,16 iyileşme sağlamıştır. HDLM'nin ortalama değeri LSTM'in en kötü değerine göre %41,63, LSTM'in en iyi değerine göre %41,27 ve LSTM'in ortalama değerine göre %41,56 iyileşme sağlamıştır.

HDLM'nin en kötü değeri CNN+GRU'nun en kötü değerine göre %2,27, CNN+GRU'nun en iyi değerine göre %1,14 ve CNN+GRU'nun ortalama değerine göre %1,71 iyileşme sağlamıştır. HDLM'nin en iyi değeri CNN+GRU'nun en kötü değerine göre %6,97, CNN+GRU'nun en iyi değerine göre %5,89 ve CNN+GRU'nun ortalama değerine göre %6,43 iyileşme sağlamıştır. HDLM'nin ortalama değeri CNN+GRU'nun en kötü değerine göre %4,35, CNN+GRU'nun en iyi değerine göre %3,24 ve CNN+GRU'nun ortalama değerine göre %3,79 iyileşme sağlamıştır.

RMSE metriğine göre deneysel sonuçlar Çizelge 6.25'te görülmektedir.

Çizelge 6.25. RMSE metriğine göre deneysel sonuçlar

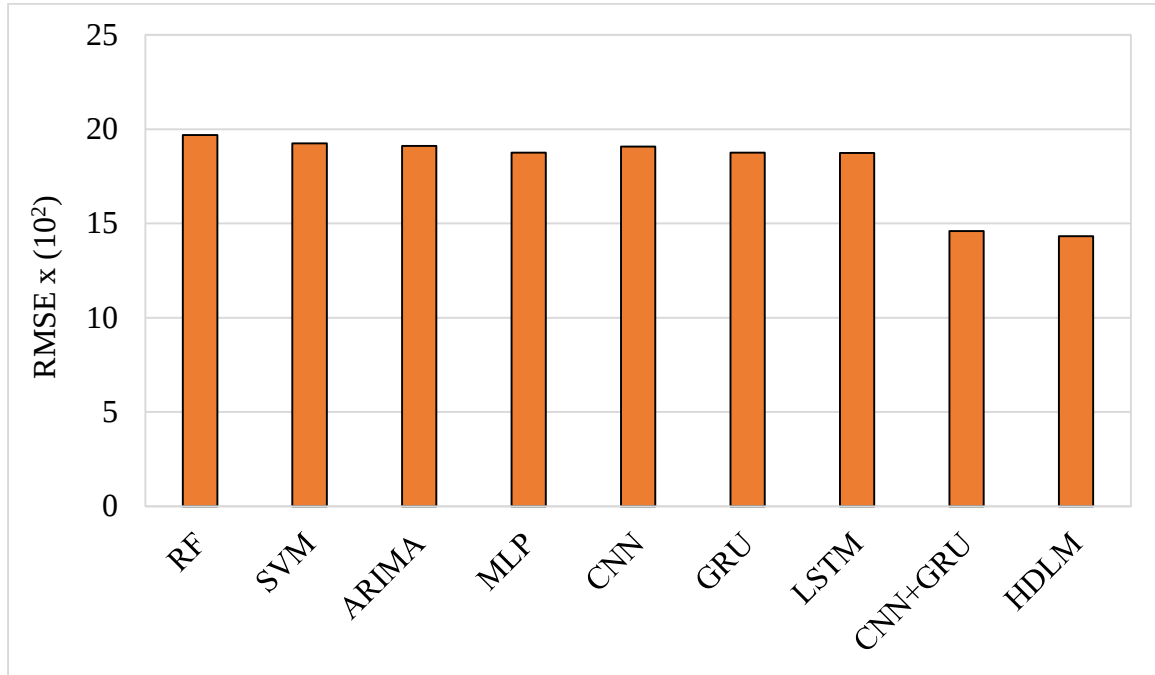
Test	RF	SVM	ARIMA	MLP	CNN	GRU	LSTM	CNN+GRU	HDLM
1	1968,96	1925,96	1911,27	1873,61	1905,83	1875,33	1868,89	1455,98	1412,38
2	1968,96	1925,96	1911,27	1875,09	1906,78	1875,33	1868,91	1456,16	1414,20
3	1968,96	1925,96	1911,27	1875,42	1907,20	1875,33	1874,58	1457,92	1422,10
4	1968,96	1925,96	1911,27	1875,47	1907,23	1875,33	1874,58	1459,02	1423,64
5	1968,96	1925,96	1911,27	1876,39	1907,40	1875,33	1874,58	1459,37	1435,41
6	1968,96	1925,96	1911,27	1877,09	1907,67	1875,33	1874,58	1461,05	1437,54
7	1968,96	1925,96	1911,27	1877,21	1908,57	1875,33	1874,58	1461,08	1440,06
8	1968,96	1925,96	1911,27	1877,93	1910,38	1875,33	1874,58	1462,37	1443,33
9	1968,96	1925,96	1911,27	1878,14	1910,50	1875,33	1874,58	1463,87	1444,70

Çizelge 6.25. (devam) RMSE metriğine göre deneysel sonuçlar

10	1968,96	1925,96	1911,27	1879,39	1910,69	1875,33	1874,58	1464,40	1447,65
Ortalama	1968,96	1925,96	1911,27	1876,57	1908,22	1875,33	1873,44	1460,12	1432,10

RMSE metriğine göre yapılan deneysel çalışmalar sonucunda, RF'in 1968,96, SVM'in 1925,96, ARIMA'nın 1911,27, MLP'nin 1876,57, CNN'in 1908,22, GRU'nun 1875,33, LSTM'in 1873,44, CNN+GRU'nun 1460,12 ve HDLM'nin 1432,10 ortalama RMSE değerine sahip olduğu görülmüştür.

Çizelge 6.25'te ve Şekil 6.19'da HDLM'nin RF, SVM, ARIMA, MLP, CNN, GRU, LSTM ve CNN+GRU'ya göre RMSE metriği için karşılaştırmalı deneysel sonuçları görülmektedir.



Şekil 6.19. RMSE metriğine göre karşılaştırmalı deneysel sonuçlar

$$\text{HDLM'nin } \frac{\text{RF-HDLM}}{\text{RF}} \times 100, \frac{\text{SVR-HDLM}}{\text{SVR}} \times 100, \frac{\text{ARIMA-HDLM}}{\text{ARIMA}} \times 100, \frac{\text{MLP-HDLM}}{\text{MLP}} \times 100, \\ \frac{\text{CNN-HDLM}}{\text{CNN}} \times 100, \frac{\text{GRU-HDLM}}{\text{GRU}} \times 100, \frac{\text{LSTM-HDLM}}{\text{LSTM}} \times 100 \text{ ve } \frac{(\text{CNN+GRU})\text{-HDLM}}{(\text{CNN+GRU})} \times 100$$

eşitlikleri kullanılarak RMSE metriğine göre hesaplanan iyileşme oranları Çizelge 6.26'da görülmektedir.

Çizelge 6.26. RMSE metriğine göre HDLM'nin iyileşme oranları

	HDLM'nin en kötü değeri	HDLM'nin en iyi değeri	HDLM'nin ortalama değeri
RF'in değerine göre iyileşme oranı (%)	26,47	28,26	27,26
SVM'nin değerine göre iyileşme oranı (%)	24,83	26,66	25,64
ARIMA'nın değerine göre iyileşme oranı (%)	24,25	26,10	25,07
MLP'nin en kötü değerine göre iyileşme oranı (%)	22,97	24,84	23,79
MLP'nin en iyi değerine göre iyileşme oranı (%)	22,73	24,61	23,56
MLP'nin ortalama değerine göre iyileşme oranı (%)	22,89	24,73	23,68
CNN'in en kötü değerine göre iyileşme oranı (%)	24,23	26,08	25,04
CNN'in en iyi değerine göre iyileşme oranı (%)	24,04	25,89	24,85
CNN'in ortalama değerine göre iyileşme oranı (%)	24,16	25,98	24,95
GRU'nun en kötü değerine göre iyileşme oranı (%)	22,80	24,68	23,63
GRU'nun en iyi değerine göre iyileşme oranı (%)	22,80	24,68	23,63
GRU'nun ortalama değerine göre iyileşme oranı (%)	22,80	24,68	23,63
LSTM'in en kötü değerine göre iyileşme oranı (%)	22,77	24,65	23,60
LSTM'in en iyi değerine göre iyileşme oranı (%)	22,53	24,42	23,37

Çizelge 6.26. (devam) RMSE metriğine göre HDLM'nin iyileşme oranları

LSTM'in ortalama değerine göre iyileşme oranı (%)	22,78	24,61	23,55
CNN+GRU'nun en kötü değerine göre iyileşme oranı (%)	1,14	3,55	2,20
CNN+GRU'nun en iyi değerine göre iyileşme oranı (%)	0,57	2,99	1,64
CNN+GRU'nun ortalama değerine göre iyileşme oranı (%)	0,85	3,26	1,91

Yapılan 10 farklı deney sonucunda elde edilen RMSE değerlerine göre HDLM'nin RF, SVM, ARIMA, MLP, CNN, GRU, LSTM ve CNN+GRU'ya göre karşılaştırmalı bir analizi görülmektedir. HDLM'nin en kötü değeri RF'a göre %26,47, en iyi değeri %28,26 ve ortalama değeri ise %27,26 iyileşme sağlamıştır. HDLM'nin en kötü değeri SVM'e göre %24,83, en iyi değeri %26,66 ve ortalama değeri ise %25,64 iyileşme sağlamıştır. HDLM'nin en kötü değeri ARIMA'ya göre %24,25, en iyi değeri %26,10 ve ortalama değeri ise %25,07 iyileşme sağlamıştır.

HDLM'nin en kötü değeri MLP'nin en kötü değerine göre %22,97, MLP'nin en iyi değerine göre %22,73 ve MLP'nin ortalama değerine göre %22,89 iyileşme sağlamıştır. HDLM'nin en iyi değeri MLP'nin en kötü değerine göre %24,84, MLP'nin en iyi değerine göre %24,61 ve MLP'nin ortalama değerine göre %24,73 iyileşme sağlamıştır. HDLM'nin ortalama değeri MLP'nin en kötü değerine göre %23,79, MLP'nin en iyi değerine göre %23,56 ve MLP'nin ortalama değerine göre %23,68 iyileşme sağlamıştır.

HDLM'nin en kötü değeri CNN'in en kötü değerine göre %24,23, CNN'in en iyi değerine göre %24,04 ve CNN'in ortalama değerine göre %24,16 iyileşme sağlamıştır. HDLM'nin en iyi değeri CNN'in en kötü değerine göre %26,08, CNN'in en iyi değerine göre %25,89 ve CNN'in ortalama değerine göre %25,98 iyileşme sağlamıştır. HDLM'nin ortalama değeri CNN'in en kötü değerine göre %25,04, CNN'in en iyi değerine göre %24,85 ve CNN'in ortalama değerine göre %24,95 iyileşme sağlamıştır.

HDLM'nin en kötü değeri GRU'nun en kötü değerine göre %22,80, GRU'nun en iyi değerine göre %22,80 ve GRU'nun ortalama değerine göre %22,80 iyileşme sağlamıştır. HDLM'nin en iyi değeri GRU'nun en kötü değerine göre %24,68, GRU'nun en iyi değerine göre %24,68 ve GRU'nun ortalama değerine göre %24,68 iyileşme sağlamıştır. HDLM'nin ortalama değeri GRU'nun en kötü değerine göre %23,63, GRU'nun en iyi değerine göre %23,63 ve GRU'nun ortalama değerine göre %23,63 iyileşme sağlamıştır.

HDLM'nin en kötü değeri LSTM'in en kötü değerine göre %22,77, LSTM'in en iyi değerine göre %22,53 ve LSTM'in ortalama değerine göre %22,78 iyileşme sağlamıştır. HDLM'nin en iyi değeri LSTM'in en kötü değerine göre %24,65, LSTM'in en iyi değerine göre %24,42 ve LSTM'in ortalama değerine göre %24,61 iyileşme sağlamıştır. HDLM'nin ortalama değeri LSTM'in en kötü değerine göre %23,60, LSTM'in en iyi değerine göre %23,37 ve LSTM'in ortalama değerine göre %23,55 iyileşme sağlamıştır.

HDLM'nin en kötü değeri CNN+GRU'nun en kötü değerine göre %1,14, CNN+GRU'nun en iyi değerine göre %0,57 ve CNN+GRU'nun ortalama değerine göre %0,85 iyileşme sağlamıştır. HDLM'nin en iyi değeri CNN+GRU'nun en kötü değerine göre %3,55, CNN+GRU'nun en iyi değerine göre %2,99 ve CNN+GRU'nun ortalama değerine göre %3,26 iyileşme sağlamıştır. HDLM'nin ortalama değeri CNN+GRU'nun en kötü değerine göre %2,20, CNN+GRU'nun en iyi değerine göre %1,64 ve CNN+GRU'nun ortalama değerine göre %1,91 iyileşme sağlamıştır.

MAE metriğine göre deneysel sonuçlar Çizelge 6.27'de görülmektedir.

Çizelge 6.27. MAE metriğine göre deneysel sonuçlar

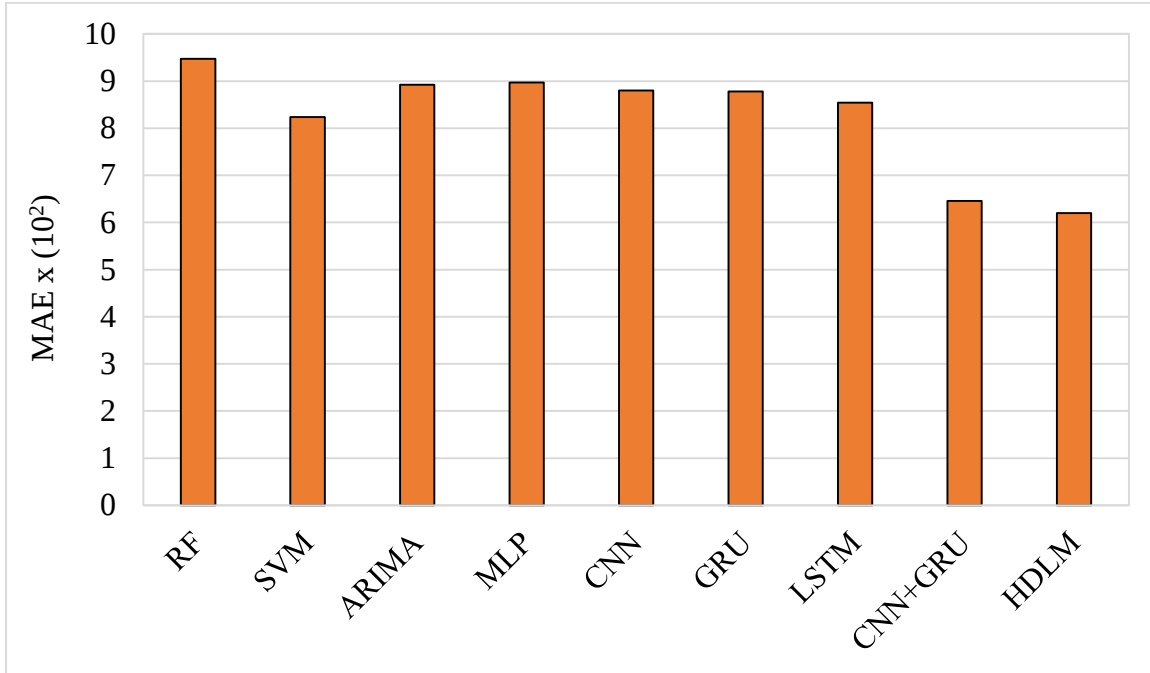
Test	RF	SVM	ARIMA	MLP	CNN	GRU	LSTM	CNN+GRU	HDLM
1	947,49	824,25	892,78	884,22	871,62	878,34	833,14	644,21	596,27
2	947,49	824,25	892,78	890,23	872,48	878,36	833,18	644,29	614,89
3	947,49	824,25	892,78	891,93	875,87	878,36	860,43	645,06	615,52
4	947,49	824,25	892,78	897,63	876,91	878,41	860,43	645,55	616,60
5	947,49	824,25	892,78	899,60	879,57	878,42	860,43	645,71	622,27
6	947,49	824,25	892,78	899,65	879,76	878,47	860,43	646,45	623,84
7	947,49	824,25	892,78	900,15	884,20	878,48	860,43	646,46	624,95

Çizelge 6.27. (devam) MAE metriğine göre deneysel sonuçlar

8	947,49	824,25	892,78	904,02	884,57	878,49	860,43	647,03	625,00
9	947,49	824,25	892,78	904,30	885,39	878,49	860,44	647,70	631,05
10	947,49	824,25	892,78	904,89	894,91	878,49	860,44	647,93	634,01
Ortalama	947,49	824,25	892,78	897,66	880,52	878,43	854,97	646,03	620,44

MAE metriğine göre yapılan deneysel çalışmalar sonucunda, RF'in 947,49, SVM'in 824,25, ARIMA'nın 892,78, MLP'nin 897,66, CNN'in 880,52, GRU'nun 878,43, LSTM'in 854,97, CNN+GRU'nun 646,03 ve HDLM'nin 620,44 ortalama MAE değerine sahip olduğu görülmüştür.

Çizelge 6.27'de ve Şekil 6.20'de HDLM'nin RF, SVM, ARIMA, MLP, CNN, GRU, LSTM ve CNN+GRU'ya göre MAE metriği için karşılaştırmalı deneysel sonuçları görülmektedir.



Şekil 6.20. MAE metriğine göre karşılaştırmalı deneysel sonuçlar

$$\begin{aligned}
 &\text{HDLM'nin } \frac{\text{RF-HDLM}}{\text{RF}} \times 100, \frac{\text{SVR-HDLM}}{\text{SVR}} \times 100, \frac{\text{ARIMA-HDLM}}{\text{ARIMA}} \times 100, \frac{\text{MLP-HDLM}}{\text{MLP}} \times 100, \\
 &\frac{\text{CNN-HDLM}}{\text{CNN}} \times 100, \frac{\text{GRU-HDLM}}{\text{GRU}} \times 100, \frac{\text{LSTM-HDLM}}{\text{LSTM}} \times 100 \text{ ve } \frac{(\text{CNN+GRU})\text{-HDLM}}{(\text{CNN+GRU})} \times 100
 \end{aligned}$$

eşitlikleri kullanılarak MAE metriğine göre hesaplanan iyileşme oranları Çizelge 6.28’de görülmektedir.

Çizelge 6.28. MAE metriğine göre HDLM’nin iyileşme oranları

	HDLM’nin en kötü değeri	HDLM’nin en iyi değeri	HDLM’nin ortalama değeri
RF’in değerine göre iyileşme oranı (%)	33,08	37,06	34,51
SVM’nin değerine göre iyileşme oranı (%)	23,08	27,65	24,72
ARIMA’nın değerine göre iyileşme oranı (%)	28,98	33,21	30,50
MLP’nin en kötü değerine göre iyileşme oranı (%)	29,93	34,10	31,43
MLP’nin en iyi değerine göre iyileşme oranı (%)	28,29	32,56	29,83
MLP’nin ortalama değerine göre iyileşme oranı (%)	29,81	33,57	30,88
CNN’in en kötü değerine göre iyileşme oranı (%)	29,15	33,37	30,67
CNN’in en iyi değerine göre iyileşme oranı (%)	27,26	31,59	28,81
CNN’in ortalama değerine göre iyileşme oranı (%)	28,28	32,28	29,53
GRU’nun en kötü değerine göre iyileşme oranı (%)	27,82	32,12	29,37
GRU’nun en iyi değerine göre iyileşme oranı (%)	27,81	32,11	29,36
GRU’nun ortalama değerine göre iyileşme oranı (%)	27,82	32,12	29,36

Çizelge 6.28. (devam) MAE metriğine göre HDLM'nin iyileşme oranları

LSTM'in en kötü değerine göre iyileşme oranı (%)	26,31	30,70	27,89
LSTM'in en iyi değerine göre iyileşme oranı (%)	23,90	28,43	25,52
LSTM'in ortalama değerine göre iyileşme oranı (%)	26,52	30,25	27,43
CNN+GRU'nun en kötü değerine göre iyileşme oranı (%)	2,14	7,97	4,24
CNN+GRU'nun en iyi değerine göre iyileşme oranı (%)	1,58	7,44	3,68
CNN+GRU'nun ortalama değerine göre iyileşme oranı (%)	1,86	7,70	3,96

Yapılan 10 farklı deney sonucunda elde edilen MAE değerlerine göre HDLM'nin RF, SVM, ARIMA, MLP, CNN, GRU, LSTM ve CNN+GRU'ya göre karşılaştırmalı bir analizi görülmektedir. HDLM'nin en kötü değeri RF'a göre %33,08, en iyi değeri %37,06 ve ortalama değeri ise %34,51 iyileşme sağlamıştır. HDLM'nin en kötü değeri SVM'e göre %23,08, en iyi değeri %27,65 ve ortalama değeri ise %24,72 iyileşme sağlamıştır. HDLM'nin en kötü değeri ARIMA'ya göre %28,98, en iyi değeri %33,21 ve ortalama değeri ise %30,50 iyileşme sağlamıştır.

HDLM'nin en kötü değeri MLP'nin en kötü değerine göre %29,93, MLP'nin en iyi değerine göre %28,29 ve MLP'nin ortalama değerine göre %29,81 iyileşme sağlamıştır. HDLM'nin en iyi değeri MLP'nin en kötü değerine göre %34,10, MLP'nin en iyi değerine göre %32,56 ve MLP'nin ortalama değerine göre %33,57 iyileşme sağlamıştır. HDLM'nin ortalama değeri MLP'nin en kötü değerine göre %31,43, MLP'nin en iyi değerine göre %29,83 ve MLP'nin ortalama değerine göre %30,88 iyileşme sağlamıştır.

HDLM'nin en kötü değeri CNN'in en kötü değerine göre %29,15, CNN'in en iyi değerine göre %27,26 ve CNN'in ortalama değerine göre %28,28 iyileşme sağlamıştır. HDLM'nin en iyi değeri CNN'in en kötü değerine göre %33,37, CNN'in en iyi değerine göre %31,59

ve CNN'in ortalama deęerine g re %32,28 iyileşme saęlamıştır. HDLM'nin ortalama deęeri CNN'in en k t  deęerine g re %30,67, CNN'in en iyi deęerine g re %28,81 ve CNN'in ortalama deęerine g re %29,53 iyileşme saęlamıştır.

HDLM'nin en k t  deęeri GRU'nun en k t  deęerine g re %27,82, GRU'nun en iyi deęerine g re %27,81 ve GRU'nun ortalama deęerine g re %27,82 iyileşme saęlamıştır. HDLM'nin en iyi deęeri GRU'nun en k t  deęerine g re %32,12, GRU'nun en iyi deęerine g re %32,11 ve GRU'nun ortalama deęerine g re %32,12 iyileşme saęlamıştır. HDLM'nin ortalama deęeri GRU'nun en k t  deęerine g re %29,37, GRU'nun en iyi deęerine g re %29,36 ve GRU'nun ortalama deęerine g re %29,36 iyileşme saęlamıştır.

HDLM'nin en k t  deęeri LSTM'in en k t  deęerine g re %26,31, LSTM'in en iyi deęerine g re %23,90 ve LSTM'in ortalama deęerine g re %26,52 iyileşme saęlamıştır. HDLM'nin en iyi deęeri LSTM'in en k t  deęerine g re %30,70, LSTM'in en iyi deęerine g re %28,43 ve LSTM'in ortalama deęerine g re %30,25 iyileşme saęlamıştır. HDLM'nin ortalama deęeri LSTM'in en k t  deęerine g re %27,89, LSTM'in en iyi deęerine g re %25,52 ve LSTM'in ortalama deęerine g re %27,43 iyileşme saęlamıştır.

HDLM'nin en k t  deęeri CNN+GRU'nun en k t  deęerine g re %2,14, CNN+GRU'nun en iyi deęerine g re %1,58 ve CNN+GRU'nun ortalama deęerine g re %1,86 iyileşme saęlamıştır. HDLM'nin en iyi deęeri CNN+GRU'nun en k t  deęerine g re %7,97, CNN+GRU'nun en iyi deęerine g re %7,44 ve CNN+GRU'nun ortalama deęerine g re %7,70 iyileşme saęlamıştır. HDLM'nin ortalama deęeri CNN+GRU'nun en k t  deęerine g re %4,24, CNN+GRU'nun en iyi deęerine g re %3,68 ve CNN+GRU'nun ortalama deęerine g re %3,96 iyileşme saęlamıştır.

Deneyisel sonular, HDLM'nin, RF, SVM, ARIMA, MLP, CNN, GRU, LSTM ve CNN+GRU'ya g re daha d ş k MSE, RMSE ve MAE sonularına sahip olduęunu g stermektedir. Bir sonraki acil aęrı zamanının belirlenmesinde HDLM'nin dięer modellere g re daha bařarılı ve etkin sonular verdięi g r lm şt r.

6.2. Bir Olayın Bir Sonraki Olma Zamanının Dakika Cinsinden Belirlenmesi

6.2.1. E-ticaret veri kümesi

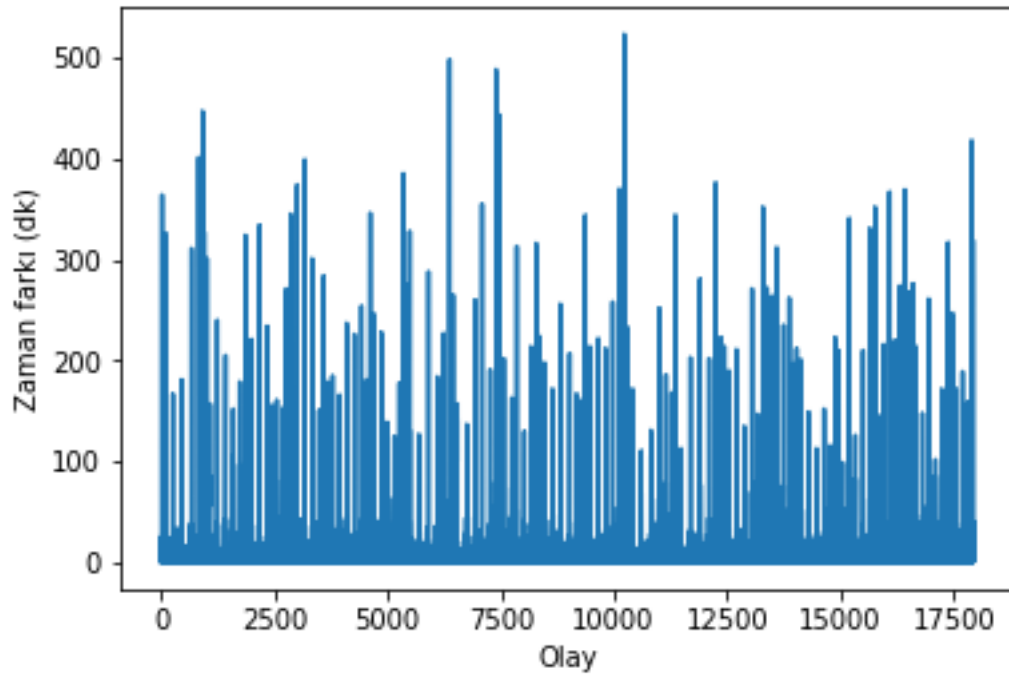
E-ticaret veri kümesi için satın alma olaylarının meydana gelme zamanları arasındaki dakika cinsinden farklara göre oluşturulan verinin ilk 5 satırı Çizelge 6.29’da görülmektedir.

Çizelge 6.29. Bir önceki satın almadan sonra geçen zaman

Olay	Bir önceki satın almadan sonra geçen zaman (dk)
1	7
2	26
3	5
4	23
5	0

Çizelge 6.29’da satın alma olaylarının meydana gelme zamanları arasındaki fark görülmektedir. E-ticaret veri kümesi 17939 satın alma olayından oluşmaktadır.

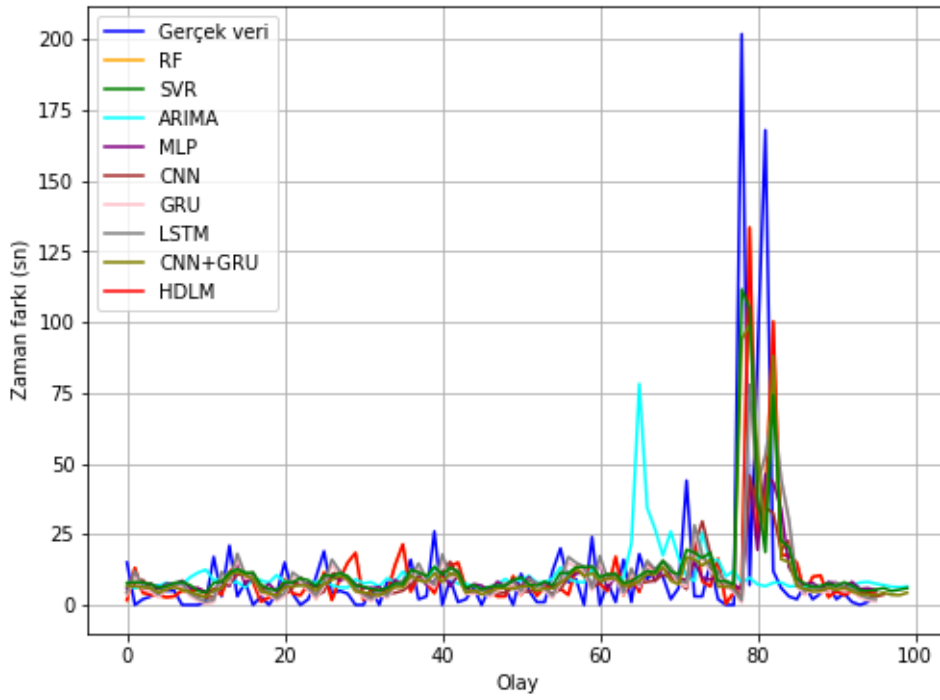
Satın alma olayları arasındaki dakika cinsinden zaman farkları Şekil 6.21’de görülmektedir.



Şekil 6.21. Satın alma olayları arasındaki dakika cinsinden zaman farkları

E-ticaret veri kümesindeki satın alma olayları arasındaki saniye cinsinden zaman farkları Şekil 6.21’de görülmektedir. Olayların dağılımındaki düzensizlikler nedeniyle meydana gelen sapmalar şekilde görülmektedir.

Şekil 6.22’de örnek olarak test verisinin 100 elemanı için deneysel sonuçlar görülmektedir.



Şekil 6.22. Test verisindeki ilk 100 veri örneği için deneysel sonuçlar

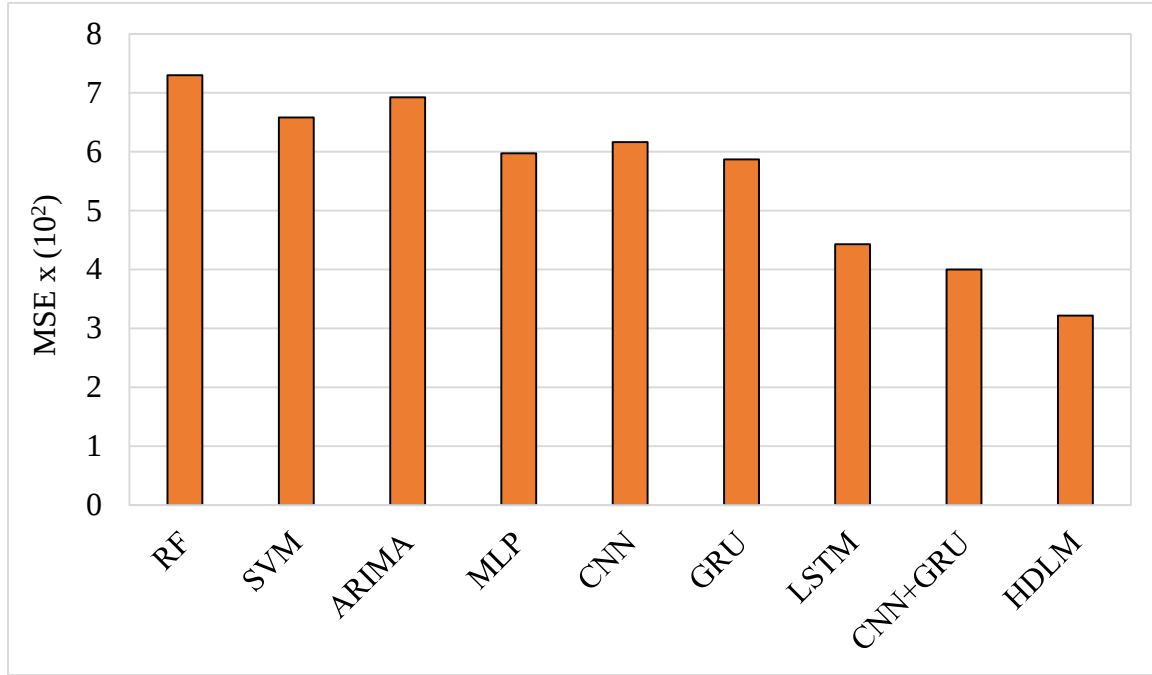
Şekil 6.22’de test verisindeki ilk 100 veri örneği için deneysel sonuçlar görülmektedir. HDLM, diğer modellere kıyasla veri kümesindeki ani sapmaları daha başarılı bir şekilde belirleyerek daha yüksek tahmin doğruluğuna sahip olmuştur.

MSE metriğine göre deneysel sonuçlar Çizelge 6.30’da ve Şekil 6.23’te görülmektedir.

Çizelge 6.30. MSE metriğine göre deneysel sonuçlar

Test	RF	SVM	ARIMA	MLP	CNN	GRU	LSTM	CNN+GRU	HDLM
1	730,59	658,46	692,39	597,48	588,55	584,79	415,10	397,83	294,98
2	730,59	658,46	692,39	597,48	593,37	584,83	429,63	398,11	308,13
3	730,59	658,46	692,39	597,49	597,27	585,21	433,22	398,74	309,54
4	730,59	658,46	692,39	597,49	598,23	585,64	438,63	398,95	310,48
5	730,59	658,46	692,39	597,49	616,35	586,66	444,10	400,27	310,60
6	730,59	658,46	692,39	597,49	621,12	588,28	445,90	400,42	322,67
7	730,59	658,46	692,39	597,49	625,38	590,22	449,91	400,99	329,06
8	730,59	658,46	692,39	597,49	626,88	590,23	454,57	401,84	335,92
9	730,59	658,46	692,39	597,50	629,55	590,57	459,34	402,55	345,58
10	730,59	658,46	692,39	597,50	669,36	591,00	466,52	403,69	355,46
Ortalama	730,59	658,46	692,39	597,49	616,60	587,74	443,69	400,33	322,24

MSE metriğine göre yapılan deneysel çalışmalar sonucunda, RF'in 730,59, SVM'in 658,46, ARIMA'nın 692,39, MLP'nin 597,49, CNN'in 616,60, GRU'nun 587,74, LSTM'in 443,69, CNN+GRU'nun 400,33 ve HDLM'nin 322,24 ortalama MSE değerine sahip olduğu görülmüştür.



Şekil 6.23. MSE metriğine göre karşılaştırmalı deneysel sonuçlar

$$\text{HDLM'nin } \frac{\text{RF-HDLM}}{\text{RF}} \times 100, \frac{\text{SVR-HDLM}}{\text{SVR}} \times 100, \frac{\text{ARIMA-HDLM}}{\text{ARIMA}} \times 100, \frac{\text{MLP-HDLM}}{\text{MLP}} \times 100, \\ \frac{\text{CNN-HDLM}}{\text{CNN}} \times 100, \frac{\text{GRU-HDLM}}{\text{GRU}} \times 100, \frac{\text{LSTM-HDLM}}{\text{LSTM}} \times 100 \text{ ve } \frac{(\text{CNN+GRU})\text{-HDLM}}{(\text{CNN+GRU})} \times 100$$

eşitlikleri kullanılarak MSE metriğine göre hesaplanan iyileşme oranları Çizelge 6.31'de görülmektedir.

Çizelge 6.31. MSE metriğine göre HDLM'nin iyileşme oranları

	HDLM'nin en kötü değeri	HDLM'nin en iyi değeri	HDLM'nin ortalama değeri
RF'in değerine göre iyileşme oranı (%)	51,34	59,62	55,89
SVM'nin değerine göre iyileşme oranı (%)	46,01	55,20	51,06
ARIMA'nın değerine göre iyileşme oranı (%)	48,66	57,39	53,45
MLP'nin en kötü değerine göre iyileşme oranı (%)	40,50	50,63	46,06
MLP'nin en iyi değerine göre iyileşme oranı (%)	40,50	50,62	46,06
MLP'nin ortalama değerine göre iyileşme oranı (%)	40,50	50,63	46,06
CNN'in en kötü değerine göre iyileşme oranı (%)	46,89	55,93	51,85
CNN'in en iyi değerine göre iyileşme oranı (%)	39,60	49,88	45,24
CNN'in ortalama değerine göre iyileşme oranı (%)	44,37	52,16	47,73
GRU'nun en kötü değerine göre iyileşme oranı (%)	39,85	50,08	45,47
GRU'nun en iyi değerine göre iyileşme oranı (%)	39,21	49,55	44,89
GRU'nun ortalama değerine göre iyileşme oranı (%)	39,72	49,81	45,17
LSTM'in en kötü değerine göre iyileşme oranı (%)	23,80	36,77	30,92
LSTM'in en iyi değerine göre iyileşme oranı (%)	14,36	28,93	22,37

Çizelge 6.31. (devam) MSE metriğine göre HDLM'nin iyileşme oranları

LSTM'in ortalama değerine göre iyileşme oranı (%)	21,25	33,51	27,37
CNN+GRU'nun en kötü değerine göre iyileşme oranı (%)	11,94	26,92	20,17
CNN+GRU'nun en iyi değerine göre iyileşme oranı (%)	10,65	25,85	19,00
CNN+GRU'nun ortalama değerine göre iyileşme oranı (%)	11,28	26,31	19,50

Yapılan 10 farklı deney sonucunda elde edilen MSE değerlerine göre HDLM'nin RF, SVM, ARIMA, MLP, CNN, GRU, LSTM ve CNN+GRU'ya göre karşılaştırmalı bir analizi görülmektedir. HDLM'nin en kötü değeri RF'a göre %51,34, en iyi değeri %59,62 ve ortalama değeri ise %55,89 iyileşme sağlamıştır. HDLM'nin en kötü değeri SVM'e göre %46,01, en iyi değeri %55,20 ve ortalama değeri ise %51,06 iyileşme sağlamıştır. HDLM'nin en kötü değeri ARIMA'ya göre %48,66, en iyi değeri %57,39 ve ortalama değeri ise %53,45 iyileşme sağlamıştır.

HDLM'nin en kötü değeri MLP'nin en kötü değerine göre %40,50, MLP'nin en iyi değerine göre %40,50 ve MLP'nin ortalama değerine göre %40,50 iyileşme sağlamıştır. HDLM'nin en iyi değeri MLP'nin en kötü değerine göre %50,63, MLP'nin en iyi değerine göre %50,62 ve MLP'nin ortalama değerine göre %50,63 iyileşme sağlamıştır. HDLM'nin ortalama değeri MLP'nin en kötü değerine göre %46,06, MLP'nin en iyi değerine göre %46,06 ve MLP'nin ortalama değerine göre %46,06 iyileşme sağlamıştır.

HDLM'nin en kötü değeri CNN'in en kötü değerine göre %46,89, CNN'in en iyi değerine göre %39,60 ve CNN'in ortalama değerine göre %44,37 iyileşme sağlamıştır. HDLM'nin en iyi değeri CNN'in en kötü değerine göre %55,93, CNN'in en iyi değerine göre %49,88 ve CNN'in ortalama değerine göre %52,16 iyileşme sağlamıştır. HDLM'nin ortalama değeri CNN'in en kötü değerine göre %51,85, CNN'in en iyi değerine göre %45,24 ve CNN'in ortalama değerine göre %47,73 iyileşme sağlamıştır.

HDLM'nin en kötü değeri GRU'nun en kötü değerine göre %39,85, GRU'nun en iyi değerine göre %39,21 ve GRU'nun ortalama değerine göre %39,72 iyileşme sağlamıştır. HDLM'nin en iyi değeri GRU'nun en kötü değerine göre %50,08, GRU'nun en iyi değerine göre %49,55 ve GRU'nun ortalama değerine göre %49,81 iyileşme sağlamıştır. HDLM'nin ortalama değeri GRU'nun en kötü değerine göre %45,47, GRU'nun en iyi değerine göre %44,89 ve GRU'nun ortalama değerine göre %45,17 iyileşme sağlamıştır.

HDLM'nin en kötü değeri LSTM'in en kötü değerine göre %23,80, LSTM'in en iyi değerine göre %14,36 ve LSTM'in ortalama değerine göre %21,25 iyileşme sağlamıştır. HDLM'nin en iyi değeri LSTM'in en kötü değerine göre %36,77, LSTM'in en iyi değerine göre %28,93 ve LSTM'in ortalama değerine göre %33,51 iyileşme sağlamıştır. HDLM'nin ortalama değeri LSTM'in en kötü değerine göre %30,92, LSTM'in en iyi değerine göre %22,37 ve LSTM'in ortalama değerine göre %27,37 iyileşme sağlamıştır.

HDLM'nin en kötü değeri CNN+GRU'nun en kötü değerine göre %11,94, CNN+GRU'nun en iyi değerine göre %10,65 ve CNN+GRU'nun ortalama değerine göre %11,28 iyileşme sağlamıştır. HDLM'nin en iyi değeri CNN+GRU'nun en kötü değerine göre %26,92, CNN+GRU'nun en iyi değerine göre %25,85 ve CNN+GRU'nun ortalama değerine göre %26,31 iyileşme sağlamıştır. HDLM'nin ortalama değeri CNN+GRU'nun en kötü değerine göre %20,17, CNN+GRU'nun en iyi değerine göre %19,00 ve CNN+GRU'nun ortalama değerine göre %19,50 iyileşme sağlamıştır.

RMSE metriğine göre deneysel sonuçlar Çizelge 6.32'de ve Şekil 6.24'te görülmektedir.

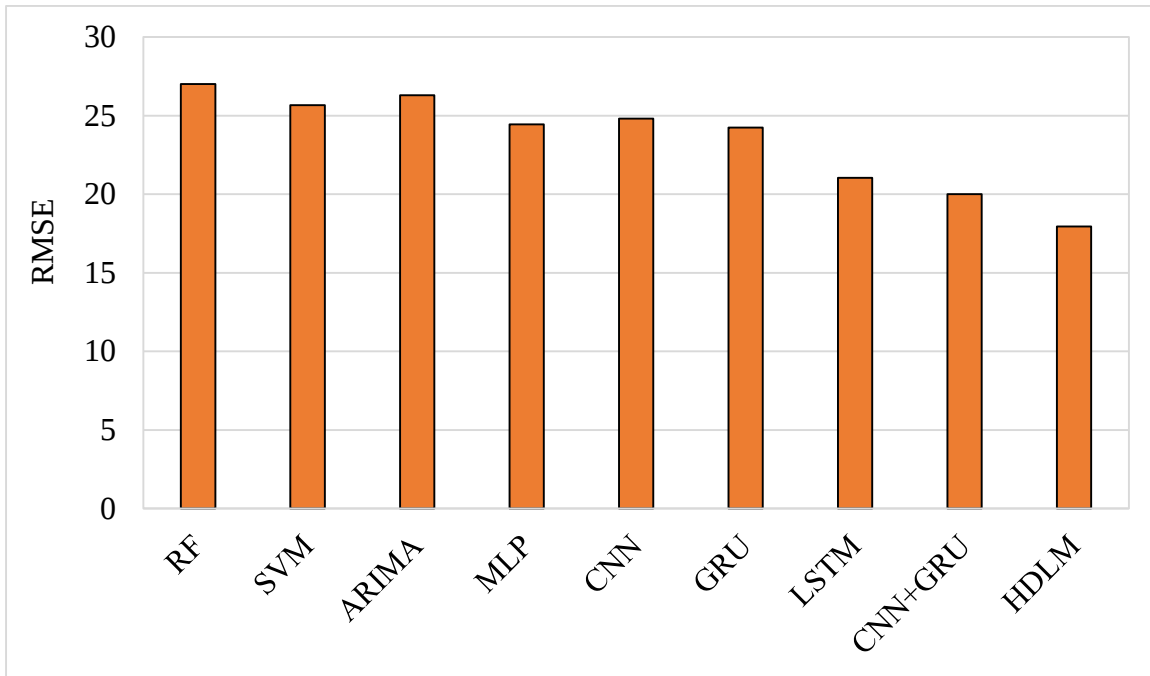
Çizelge 6.32. RMSE metriğine göre deneysel sonuçlar

Test	RF	SVM	ARIMA	MLP	CNN	GRU	LSTM	CNN+GRU	HDLM
1	27,02	25,66	26,31	24,44	24,26	24,18	20,37	19,94	17,17
2	27,02	25,66	26,31	24,44	24,35	24,18	20,72	19,95	17,55
3	27,02	25,66	26,31	24,44	24,43	24,19	20,81	19,96	17,59
4	27,02	25,66	26,31	24,44	24,46	24,20	20,94	19,97	17,62
5	27,02	25,66	26,31	24,44	24,82	24,22	21,07	20,00	17,62
6	27,02	25,66	26,31	24,44	24,92	24,25	21,11	20,01	17,96
7	27,02	25,66	26,31	24,44	25,00	24,29	21,21	20,02	18,14

Çizelge 6.32. (devam) RMSE metriğine göre deneysel sonuçlar

8	27,02	25,66	26,31	24,44	25,03	24,29	21,32	20,04	18,32
9	27,02	25,66	26,31	24,44	25,09	24,30	21,43	20,06	18,58
10	27,02	25,66	26,31	24,44	25,87	24,31	21,59	20,09	18,85
Ortalama	27,02	25,66	26,31	24,44	24,82	24,24	21,05	20,00	17,94

RMSE metriğine göre yapılan deneysel çalışmalar sonucunda, RF'ın 27,02, SVM'in 25,66, ARIMA'nın 26,31, MLP'nin 24,44, CNN'in 24,82, GRU'nun 24,24, LSTM'in 21,05, CNN+GRU'nun 20,00 ve HDLM'nin 17,94 ortalama RMSE değerine sahip olduğu görülmüştür.



Şekil 6.24. RMSE metriğine göre karşılaştırmalı deneysel sonuçlar

$$\text{HDLM'nin } \frac{\text{RF-HDLM}}{\text{RF}} \times 100, \frac{\text{SVR-HDLM}}{\text{SVR}} \times 100, \frac{\text{ARIMA-HDLM}}{\text{ARIMA}} \times 100, \frac{\text{MLP-HDLM}}{\text{MLP}} \times 100, \\ \frac{\text{CNN-HDLM}}{\text{CNN}} \times 100, \frac{\text{GRU-HDLM}}{\text{GRU}} \times 100, \frac{\text{LSTM-HDLM}}{\text{LSTM}} \times 100 \text{ ve } \frac{(\text{CNN+GRU})-\text{HDLM}}{(\text{CNN+GRU})} \times 100$$

eşitlikleri kullanılarak RMSE metriğine göre hesaplanan iyileşme oranları Çizelge 6.33'te görülmektedir.

Çizelge 6.33. RMSE metriğine göre HDLM'nin iyileşme oranları

	HDLM'nin en kötü değeri	HDLM'nin en iyi değeri	HDLM'nin ortalama değeri
RF'in değerine göre iyileşme oranı (%)	30,23	36,45	33,60
SVM'nin değerine göre iyileşme oranı (%)	26,53	33,08	30,08
ARIMA'nın değerine göre iyileşme oranı (%)	28,35	34,73	31,81
MLP'nin en kötü değerine göre iyileşme oranı (%)	22,87	29,74	26,59
MLP'nin en iyi değerine göre iyileşme oranı (%)	22,87	29,74	26,59
MLP'nin ortalama değerine göre iyileşme oranı (%)	22,87	29,74	26,59
CNN'in en kötü değerine göre iyileşme oranı (%)	27,13	33,62	30,65
CNN'in en iyi değerine göre iyileşme oranı (%)	22,30	29,22	26,05
CNN'in ortalama değerine göre iyileşme oranı (%)	24,62	30,83	27,72
GRU'nun en kötü değerine göre iyileşme oranı (%)	22,45	29,37	26,20
GRU'nun en iyi değerine göre iyileşme oranı (%)	22,04	28,99	25,80
GRU'nun ortalama değerine göre iyileşme oranı (%)	22,29	29,16	25,99
LSTM'in en kötü değerine göre iyileşme oranı (%)	12,69	20,47	16,90
LSTM'in en iyi değerine göre iyileşme oranı (%)	7,46	15,70	11,92

Çizelge 6.33. (devam) RMSE metriğine göre HDLM'nin iyileşme oranları

LSTM'in ortalama değerine göre iyileşme oranı (%)	10,83	18,45	14,80
CNN+GRU'nun en kötü değerine göre iyileşme oranı (%)	6,17	14,53	10,70
CNN+GRU'nun en iyi değerine göre iyileşme oranı (%)	5,46	13,89	10,03
CNN+GRU'nun ortalama değerine göre iyileşme oranı (%)	5,78	14,16	10,31

Yapılan 10 farklı deney sonucunda elde edilen RMSE değerlerine göre HDLM'nin RF, SVM, ARIMA, MLP, CNN, GRU, LSTM ve CNN+GRU'ya göre karşılaştırmalı bir analizi görülmektedir. HDLM'nin en kötü değeri RF'a göre %30,23, en iyi değeri %36,45 ve ortalama değeri ise %33,60 iyileşme sağlamıştır. HDLM'nin en kötü değeri SVM'e göre %26,53, en iyi değeri %33,08 ve ortalama değeri ise %30,08 iyileşme sağlamıştır. HDLM'nin en kötü değeri ARIMA'ya göre %28,35, en iyi değeri %34,73 ve ortalama değeri ise %31,81 iyileşme sağlamıştır.

HDLM'nin en kötü değeri MLP'nin en kötü değerine göre %22,87, MLP'nin en iyi değerine göre %22,87 ve MLP'nin ortalama değerine göre %22,87 iyileşme sağlamıştır. HDLM'nin en iyi değeri MLP'nin en kötü değerine göre %29,74, MLP'nin en iyi değerine göre %29,74 ve MLP'nin ortalama değerine göre %29,74 iyileşme sağlamıştır. HDLM'nin ortalama değeri MLP'nin en kötü değerine göre %26,59, MLP'nin en iyi değerine göre %26,59 ve MLP'nin ortalama değerine göre %26,59 iyileşme sağlamıştır.

HDLM'nin en kötü değeri CNN'in en kötü değerine göre %27,13, CNN'in en iyi değerine göre %22,30 ve CNN'in ortalama değerine göre %24,62 iyileşme sağlamıştır. HDLM'nin en iyi değeri CNN'in en kötü değerine göre %33,62, CNN'in en iyi değerine göre %29,22 ve CNN'in ortalama değerine göre %30,83 iyileşme sağlamıştır. HDLM'nin ortalama değeri CNN'in en kötü değerine göre %30,65, CNN'in en iyi değerine göre %26,05 ve CNN'in ortalama değerine göre %27,72 iyileşme sağlamıştır.

HDLM'nin en kötü değeri GRU'nun en kötü değerine göre %22,45, GRU'nun en iyi değerine göre %22,04 ve GRU'nun ortalama değerine göre %22,29 iyileşme sağlamıştır. HDLM'nin en iyi değeri GRU'nun en kötü değerine göre %29,37, GRU'nun en iyi değerine göre %28,99 ve GRU'nun ortalama değerine göre %29,16 iyileşme sağlamıştır. HDLM'nin ortalama değeri GRU'nun en kötü değerine göre %26,20, GRU'nun en iyi değerine göre %25,80 ve GRU'nun ortalama değerine göre %25,99 iyileşme sağlamıştır.

HDLM'nin en kötü değeri LSTM'in en kötü değerine göre %12,69, LSTM'in en iyi değerine göre %7,46 ve LSTM'in ortalama değerine göre %10,83 iyileşme sağlamıştır. HDLM'nin en iyi değeri LSTM'in en kötü değerine göre %20,47, LSTM'in en iyi değerine göre %15,70 ve LSTM'in ortalama değerine göre %18,45 iyileşme sağlamıştır. HDLM'nin ortalama değeri LSTM'in en kötü değerine göre %16,90, LSTM'in en iyi değerine göre %11,92 ve LSTM'in ortalama değerine göre %14,80 iyileşme sağlamıştır.

HDLM'nin en kötü değeri CNN+GRU'nun en kötü değerine göre %6,17, CNN+GRU'nun en iyi değerine göre %5,46 ve CNN+GRU'nun ortalama değerine göre %5,78 iyileşme sağlamıştır. HDLM'nin en iyi değeri CNN+GRU'nun en kötü değerine göre %14,53, CNN+GRU'nun en iyi değerine göre %13,89 ve CNN+GRU'nun ortalama değerine göre %14,16 iyileşme sağlamıştır. HDLM'nin ortalama değeri CNN+GRU'nun en kötü değerine göre %10,70, CNN+GRU'nun en iyi değerine göre %10,03 ve CNN+GRU'nun ortalama değerine göre %10,31 iyileşme sağlamıştır.

MAE metriğine göre deneysel sonuçlar Çizelge 6.34'te ve Şekil 6.25'te görülmektedir.

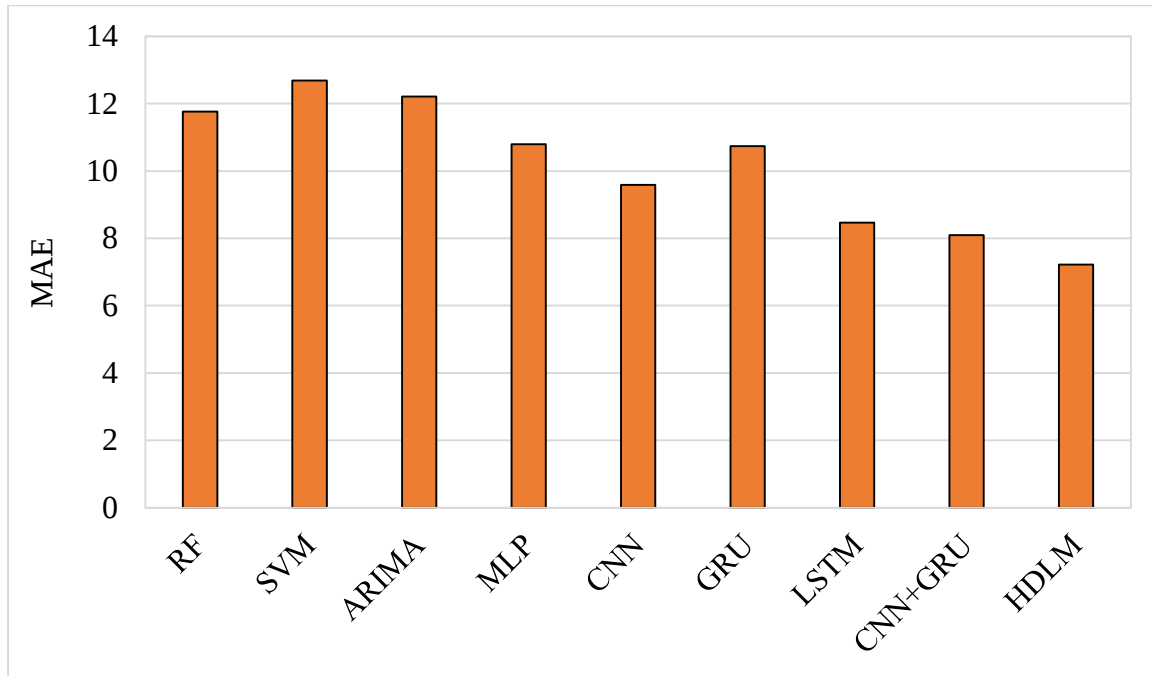
Çizelge 6.34. MAE metriğine göre deneysel sonuçlar

Test	RF	SVM	ARIMA	MLP	CNN	GRU	LSTM	CNN+GRU	HDLM
1	11,77	12,69	12,21	10,80	8,93	10,68	8,22	8,06	6,92
2	11,77	12,69	12,21	10,80	9,32	10,71	8,29	8,08	7,02
3	11,77	12,69	12,21	10,80	9,37	10,72	8,34	8,09	7,11
4	11,77	12,69	12,21	10,80	9,40	10,72	8,47	8,10	7,15
5	11,77	12,69	12,21	10,80	9,45	10,75	8,47	8,10	7,23
6	11,77	12,69	12,21	10,80	9,49	10,75	8,49	8,11	7,29
7	11,77	12,69	12,21	10,81	9,53	10,75	8,58	8,11	7,32

Çizelge 6.34. (devam) MAE metriğine göre deneysel sonuçlar

8	11,77	12,69	12,21	10,81	9,77	10,77	8,58	8,12	7,34
9	11,77	12,69	12,21	10,81	9,82	10,78	8,61	8,13	7,41
10	11,77	12,69	12,21	10,84	10,84	10,81	8,67	8,17	7,41
Ortalama	11,77	12,69	12,21	10,80	9,59	10,74	8,47	8,10	7,22

MAE metriğine göre yapılan deneysel çalışmalar sonucunda, RF'in 11,77, SVM'in 12,69, ARIMA'nın 12,21, MLP'nin 10,80, CNN'in 9,59, GRU'nun 10,74, LSTM'in 8,47, CNN+GRU'nun 8,10 ve HDLM'nin 7,22 ortalama MAE değerine sahip olduğu görülmüştür.



Şekil 6.25. MAE metriğine göre karşılaştırmalı deneysel sonuçlar

$$\text{HDLM'nin } \frac{\text{RF-HDLM}}{\text{RF}} \times 100, \frac{\text{SVR-HDLM}}{\text{SVR}} \times 100, \frac{\text{ARIMA-HDLM}}{\text{ARIMA}} \times 100, \frac{\text{MLP-HDLM}}{\text{MLP}} \times 100, \\ \frac{\text{CNN-HDLM}}{\text{CNN}} \times 100, \frac{\text{GRU-HDLM}}{\text{GRU}} \times 100, \frac{\text{LSTM-HDLM}}{\text{LSTM}} \times 100 \text{ ve } \frac{(\text{CNN+GRU})\text{-HDLM}}{(\text{CNN+GRU})} \times 100$$

eşitlikleri kullanılarak MAE metriğine göre hesaplanan iyileşme oranları Çizelge 6.35'te görülmektedir.

Çizelge 6.35. MAE metriğine göre HDLM'nin iyileşme oranları

	HDLM'nin en kötü değeri	HDLM'nin en iyi değeri	HDLM'nin ortalama değeri
RF'in değerine göre iyileşme oranı (%)	37,04	41,20	38,65
SVM'nin değerine göre iyileşme oranı (%)	41,60	45,46	43,10
ARIMA'nın değerine göre iyileşme oranı (%)	39,31	43,32	40,86
MLP'nin en kötü değerine göre iyileşme oranı (%)	31,64	36,16	33,39
MLP'nin en iyi değerine göre iyileşme oranı (%)	31,38	35,92	33,14
MLP'nin ortalama değerine göre iyileşme oranı (%)	31,45	35,96	33,19
CNN'in en kötü değerine göre iyileşme oranı (%)	31,64	36,16	33,39
CNN'in en iyi değerine göre iyileşme oranı (%)	17,02	22,50	19,14
CNN'in ortalama değerine göre iyileşme oranı (%)	24,43	27,85	24,72
GRU'nun en kötü değerine göre iyileşme oranı (%)	31,45	35,98	33,20
GRU'nun en iyi değerine göre iyileşme oranı (%)	30,61	35,20	32,39
GRU'nun ortalama değerine göre iyileşme oranı (%)	31,21	35,59	32,79
LSTM'in en kötü değerine göre iyileşme oranı (%)	14,53	20,18	16,72
LSTM'in en iyi değerine göre iyileşme oranı (%)	9,85	15,81	12,16

Çizelge 6.35. (devam) MAE metriğine göre HDLM'nin iyileşme oranları

LSTM'in ortalama değerine göre iyileşme oranı (%)	12,91	18,31	14,77
CNN+GRU'nun en kötü değerine göre iyileşme oranı (%)	9,30	15,29	11,62
CNN+GRU'nun en iyi değerine göre iyileşme oranı (%)	8,06	14,14	10,42
CNN+GRU'nun ortalama değerine göre iyileşme oranı (%)	8,64	14,64	10,94

Yapılan 10 farklı deney sonucunda elde edilen MAE değerlerine göre HDLM'nin RF, SVM, ARIMA, MLP, CNN, GRU, LSTM ve CNN+GRU'ya göre karşılaştırmalı bir analizi görülmektedir. HDLM'nin en kötü değeri RF'a göre %37,04, en iyi değeri %41,20 ve ortalama değeri ise %38,65 iyileşme sağlamıştır. HDLM'nin en kötü değeri SVM'e göre %41,60, en iyi değeri %45,46 ve ortalama değeri ise %43,10 iyileşme sağlamıştır. HDLM'nin en kötü değeri ARIMA'ya göre %39,31, en iyi değeri %43,32 ve ortalama değeri ise %40,86 iyileşme sağlamıştır.

HDLM'nin en kötü değeri MLP'nin en kötü değerine göre %31,64, MLP'nin en iyi değerine göre %31,38 ve MLP'nin ortalama değerine göre %31,45 iyileşme sağlamıştır. HDLM'nin en iyi değeri MLP'nin en kötü değerine göre %36,16, MLP'nin en iyi değerine göre %35,92 ve MLP'nin ortalama değerine göre %35,96 iyileşme sağlamıştır. HDLM'nin ortalama değeri MLP'nin en kötü değerine göre %33,39, MLP'nin en iyi değerine göre %33,14 ve MLP'nin ortalama değerine göre %33,19 iyileşme sağlamıştır.

HDLM'nin en kötü değeri CNN'in en kötü değerine göre %31,64, CNN'in en iyi değerine göre %17,02 ve CNN'in ortalama değerine göre %24,43 iyileşme sağlamıştır. HDLM'nin en iyi değeri CNN'in en kötü değerine göre %36,16, CNN'in en iyi değerine göre %22,50 ve CNN'in ortalama değerine göre %27,85 iyileşme sağlamıştır. HDLM'nin ortalama değeri CNN'in en kötü değerine göre %33,39, CNN'in en iyi değerine göre %19,14 ve CNN'in ortalama değerine göre %24,72 iyileşme sağlamıştır.

HDLM'nin en kötü değeri GRU'nun en kötü değerine göre %31,45, GRU'nun en iyi değerine göre %30,61 ve GRU'nun ortalama değerine göre %31,21 iyileşme sağlamıştır. HDLM'nin en iyi değeri GRU'nun en kötü değerine göre %35,98, GRU'nun en iyi değerine göre %35,20 ve GRU'nun ortalama değerine göre %35,59 iyileşme sağlamıştır. HDLM'nin ortalama değeri GRU'nun en kötü değerine göre %33,20, GRU'nun en iyi değerine göre %32,39 ve GRU'nun ortalama değerine göre %32,79 iyileşme sağlamıştır.

HDLM'nin en kötü değeri LSTM'in en kötü değerine göre %14,53, LSTM'in en iyi değerine göre %9,85 ve LSTM'in ortalama değerine göre %12,91 iyileşme sağlamıştır. HDLM'nin en iyi değeri LSTM'in en kötü değerine göre %20,18, LSTM'in en iyi değerine göre %15,81 ve LSTM'in ortalama değerine göre %18,31 iyileşme sağlamıştır. HDLM'nin ortalama değeri LSTM'in en kötü değerine göre %16,72, LSTM'in en iyi değerine göre %12,16 ve LSTM'in ortalama değerine göre %14,77 iyileşme sağlamıştır.

HDLM'nin en kötü değeri CNN+GRU'nun en kötü değerine göre %9,30, CNN+GRU'nun en iyi değerine göre %8,06 ve CNN+GRU'nun ortalama değerine göre %8,64 iyileşme sağlamıştır. HDLM'nin en iyi değeri CNN+GRU'nun en kötü değerine göre %15,29, CNN+GRU'nun en iyi değerine göre %14,14 ve CNN+GRU'nun ortalama değerine göre %14,64 iyileşme sağlamıştır. HDLM'nin ortalama değeri CNN+GRU'nun en kötü değerine göre %11,62, CNN+GRU'nun en iyi değerine göre %10,42 ve CNN+GRU'nun ortalama değerine göre %10,94 iyileşme sağlamıştır.

Deneyisel sonuçlar, HDLM'nin RF, SVM, ARIMA, MLP, CNN, GRU, LSTM ve CNN+GRU'ya göre daha düşük MSE, RMSE ve MAE sonuçlarına sahip olduğunu göstermektedir. Bir sonraki satın alma zamanının belirlenmesinde HDLM'nin diğer modellere göre daha başarılı ve etkin sonuçlar verdiği görülmüştür.

6.2.2. DDoS veri kümesi

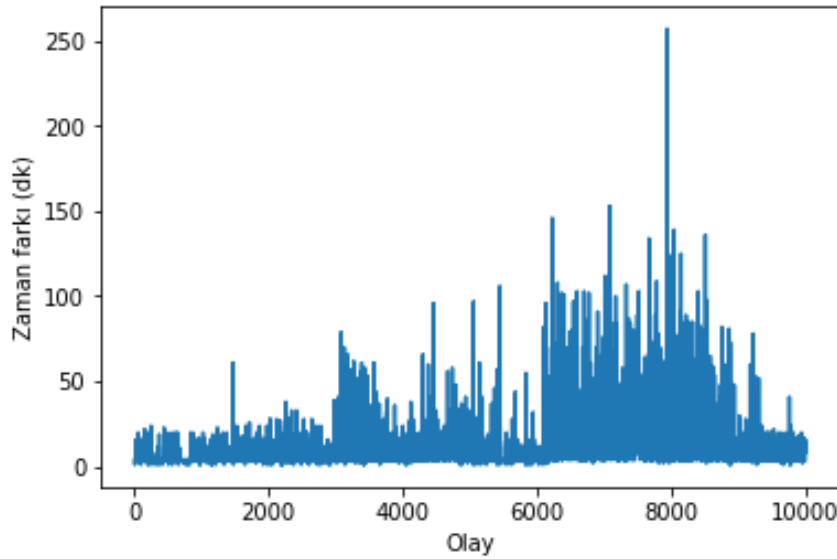
DDoS veri kümesi için saldırı olaylarının meydana gelme zamanları arasındaki dakika cinsinden farklara göre oluşturulan verinin ilk 5 satırı Çizelge 6.36'da görülmektedir.

Çizelge 6.36. Bir önceki saldırıdan sonra geçen zaman

Olay	Bir önceki saldırıdan sonra geçen zaman (dk)
1	4
2	2
3	4
4	1
5	4

Çizelge 6.36’da DDoS saldırıları arasındaki zaman farkı görülmektedir. DDoS veri kümesi, e-ticaret veri kümesine göre nispeten daha düzenli bir örüntü izlemektedir.

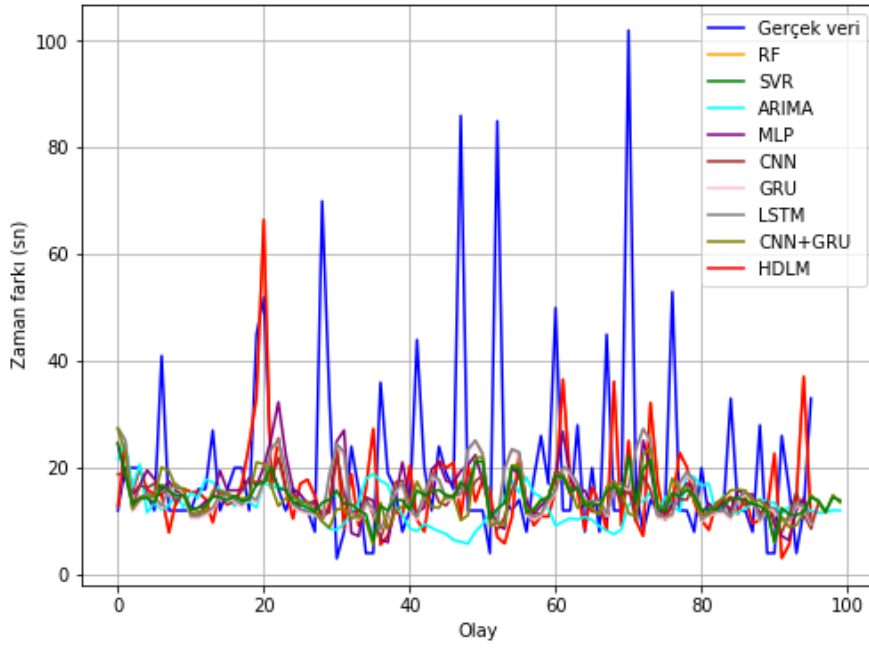
Satın alma olayları arasındaki dakika cinsinden zaman farkları Şekil 6.26’da görülmektedir.



Şekil 6.26. Saldırı olayları arasındaki dakika cinsinden zaman farkları

Şekil 6.26’da DDoS saldırıları arasındaki zaman farkı görülmektedir. DDoS veri kümesi 10000 DDoS saldırı olayından oluşmaktadır. Veri kümesinde nadiren de olsa görülen sapmalar mevcuttur.

Şekil 6.27’de örnek olarak test verisinin 100 elemanı için deneysel sonuçlar görülmektedir.



Şekil 6.27. Test verisindeki ilk 100 veri örneği için deneysel sonuçlar

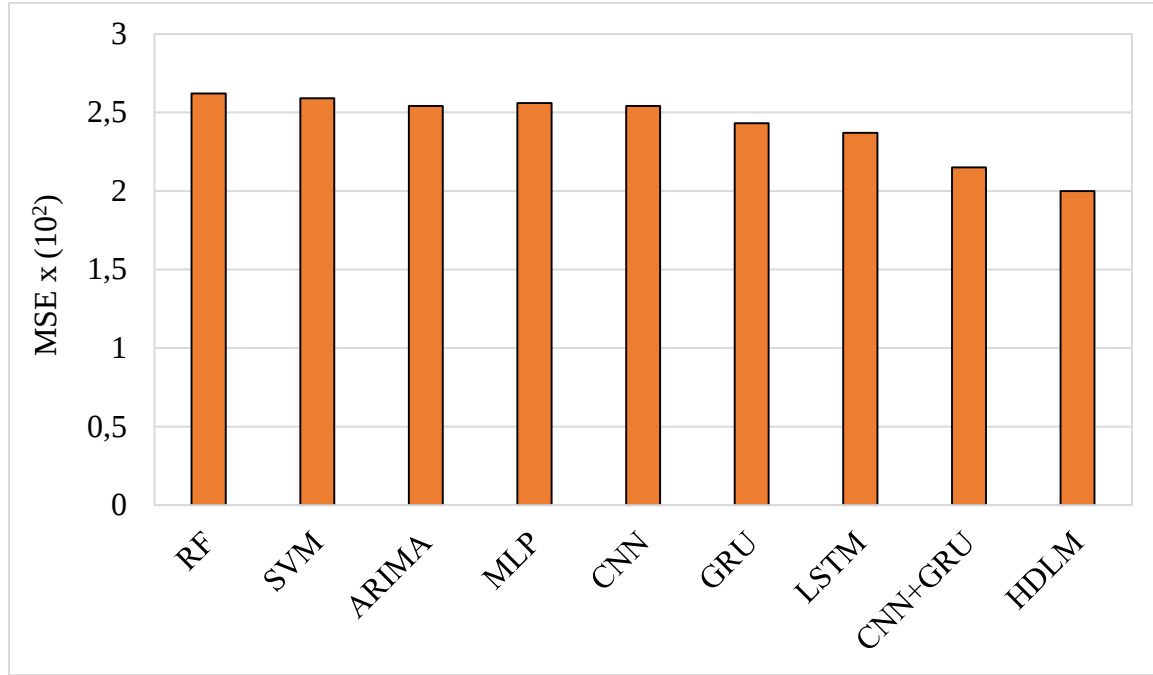
Şekil 6.27’de test verisindeki ilk 100 veri örneği için deneysel sonuçlar görülmektedir. HDLM, diğer modellere kıyasla veri kümesindeki ani sapmaları daha başarılı bir şekilde belirleyerek daha yüksek tahmin doğruluğuna sahip olmuştur.

MSE metriğine göre deneysel sonuçlar Çizelge 6.37’de ve Şekil 6.28’de görülmektedir.

Çizelge 6.37. MSE metriğine göre deneysel sonuçlar

Test	RF	SVM	ARIMA	MLP	CNN	GRU	LSTM	CNN+GRU	HDLM
1	262,70	259,09	254,38	250,08	252,45	238,65	237,46	213,06	199,25
2	262,70	259,09	254,38	251,68	252,99	239,26	237,52	213,52	199,64
3	262,70	259,09	254,38	251,89	253,64	240,60	237,64	213,73	199,72
4	262,70	259,09	254,38	253,44	253,73	243,35	237,69	214,08	200,09
5	262,70	259,09	254,38	255,43	253,77	243,48	237,89	215,36	200,48
6	262,70	259,09	254,38	256,71	254,59	243,48	238,03	215,67	200,77
7	262,70	259,09	254,38	256,74	254,63	243,49	238,06	216,65	200,83
8	262,70	259,09	254,38	260,81	254,71	243,71	238,21	217,37	201,07
9	262,70	259,09	254,38	262,50	255,30	247,20	238,24	218,05	202,49
10	262,70	259,09	254,38	268,39	255,30	250,69	238,54	220,07	203,09
Ortalama	262,70	259,09	254,38	256,76	254,11	243,39	237,92	215,75	200,74

MSE metriğine göre yapılan deneysel çalışmalar sonucunda, RF'in 262,70, SVM'in 259,09, ARIMA'nın 254,38, MLP'nin 256,76, CNN'in 254,11, GRU'nun 243,39, LSTM'in 237,92, CNN+GRU'nun 215,75 ve HDLM'nin 200,74 ortalama MSE değerine sahip olduğu görülmüştür.



Şekil 6.28. MSE metriğine göre karşılaştırmalı deneysel sonuçlar

$$\text{HDLM'nin } \frac{\text{RF-HDLM}}{\text{RF}} \times 100, \frac{\text{SVR-HDLM}}{\text{SVR}} \times 100, \frac{\text{ARIMA-HDLM}}{\text{ARIMA}} \times 100, \frac{\text{MLP-HDLM}}{\text{MLP}} \times 100, \\ \frac{\text{CNN-HDLM}}{\text{CNN}} \times 100, \frac{\text{GRU-HDLM}}{\text{GRU}} \times 100, \frac{\text{LSTM-HDLM}}{\text{LSTM}} \times 100 \text{ ve } \frac{(\text{CNN+GRU})\text{-HDLM}}{(\text{CNN+GRU})} \times 100$$

eşitlikleri kullanılarak MSE metriğine göre hesaplanan iyileşme oranları Çizelge 6.38'de görülmektedir.

Çizelge 6.38. MSE metriğine göre HDLM'nin iyileşme oranları

	HDLM'nin en kötü değeri	HDLM'nin en iyi değeri	HDLM'nin ortalama değeri
RF'in değerine göre iyileşme oranı (%)	22,69	24,15	23,58
SVM'nin değerine göre iyileşme oranı (%)	21,61	23,09	22,51
ARIMA'nın değerine göre iyileşme oranı (%)	20,16	21,67	21,08
MLP'nin en kötü değerine göre iyileşme oranı (%)	24,33	25,76	25,20
MLP'nin en iyi değerine göre iyileşme oranı (%)	18,78	20,32	19,72
MLP'nin ortalama değerine göre iyileşme oranı (%)	21,46	22,40	21,81
CNN'in en kötü değerine göre iyileşme oranı (%)	20,45	21,95	21,36
CNN'in en iyi değerine göre iyileşme oranı (%)	19,55	21,07	20,48
CNN'in ortalama değerine göre iyileşme oranı (%)	20,21	21,58	21,00
GRU'nun en kötü değerine göre iyileşme oranı (%)	18,98	20,51	19,92
GRU'nun en iyi değerine göre iyileşme oranı (%)	14,90	16,50	15,88
GRU'nun ortalama değerine göre iyileşme oranı (%)	16,88	18,13	17,52
LSTM'in en kötü değerine göre iyileşme oranı (%)	14,86	16,47	15,84
LSTM'in en iyi değerine göre iyileşme oranı (%)	14,47	16,09	15,46

Çizelge 6.38. (devam) MSE metriğine göre HDLM'nin iyileşme oranları

LSTM'in ortalama değerine göre iyileşme oranı (%)	14,67	16,25	15,62
CNN+GRU'nun en kötü değerine göre iyileşme oranı (%)	7,71	9,46	8,78
CNN+GRU'nun en iyi değerine göre iyileşme oranı (%)	4,67	6,48	5,78
CNN+GRU'nun ortalama değerine göre iyileşme oranı (%)	5,94	7,65	6,95

Yapılan 10 farklı deney sonucunda elde edilen MSE değerlerine göre HDLM'nin RF, SVM, ARIMA, MLP, CNN, GRU, LSTM ve CNN+GRU'ya göre karşılaştırmalı bir analizi görülmektedir. HDLM'nin en kötü değeri RF'a göre %22,69, en iyi değeri %24,15 ve ortalama değeri ise %23,58 iyileşme sağlamıştır. HDLM'nin en kötü değeri SVM'e göre %21,61, en iyi değeri %23,09 ve ortalama değeri ise %22,51 iyileşme sağlamıştır. HDLM'nin en kötü değeri ARIMA'ya göre %20,16, en iyi değeri %21,67 ve ortalama değeri ise %21,08 iyileşme sağlamıştır.

HDLM'nin en kötü değeri MLP'nin en kötü değerine göre %24,33, MLP'nin en iyi değerine göre %18,78 ve MLP'nin ortalama değerine göre %21,46 iyileşme sağlamıştır. HDLM'nin en iyi değeri MLP'nin en kötü değerine göre %25,76, MLP'nin en iyi değerine göre %20,32 ve MLP'nin ortalama değerine göre %22,40 iyileşme sağlamıştır. HDLM'nin ortalama değeri MLP'nin en kötü değerine göre %25,20, MLP'nin en iyi değerine göre %19,72 ve MLP'nin ortalama değerine göre %21,81 iyileşme sağlamıştır.

HDLM'nin en kötü değeri CNN'in en kötü değerine göre %20,45, CNN'in en iyi değerine göre %19,55 ve CNN'in ortalama değerine göre %20,21 iyileşme sağlamıştır. HDLM'nin en iyi değeri CNN'in en kötü değerine göre %21,95, CNN'in en iyi değerine göre %21,07 ve CNN'in ortalama değerine göre %21,58 iyileşme sağlamıştır. HDLM'nin ortalama değeri CNN'in en kötü değerine göre %21,36, CNN'in en iyi değerine göre %20,48 ve CNN'in ortalama değerine göre %21,00 iyileşme sağlamıştır.

HDLM'nin en kötü değeri GRU'nun en kötü değerine göre %18,98, GRU'nun en iyi değerine göre %14,90 ve GRU'nun ortalama değerine göre %16,88 iyileşme sağlamıştır. HDLM'nin en iyi değeri GRU'nun en kötü değerine göre %20,51, GRU'nun en iyi değerine göre %16,50 ve GRU'nun ortalama değerine göre %18,13 iyileşme sağlamıştır. HDLM'nin ortalama değeri GRU'nun en kötü değerine göre %19,92, GRU'nun en iyi değerine göre %15,88 ve GRU'nun ortalama değerine göre %17,52 iyileşme sağlamıştır.

HDLM'nin en kötü değeri LSTM'in en kötü değerine göre %14,86, LSTM'in en iyi değerine göre %14,47 ve LSTM'in ortalama değerine göre %14,67 iyileşme sağlamıştır. HDLM'nin en iyi değeri LSTM'in en kötü değerine göre %16,47, LSTM'in en iyi değerine göre %16,09 ve LSTM'in ortalama değerine göre %16,25 iyileşme sağlamıştır. HDLM'nin ortalama değeri LSTM'in en kötü değerine göre %15,84, LSTM'in en iyi değerine göre %15,46 ve LSTM'in ortalama değerine göre %15,62 iyileşme sağlamıştır.

HDLM'nin en kötü değeri CNN+GRU'nun en kötü değerine göre %7,71, CNN+GRU'nun en iyi değerine göre %4,67 ve CNN+GRU'nun ortalama değerine göre %5,94 iyileşme sağlamıştır. HDLM'nin en iyi değeri CNN+GRU'nun en kötü değerine göre %9,46, CNN+GRU'nun en iyi değerine göre %6,48 ve CNN+GRU'nun ortalama değerine göre %7,65 iyileşme sağlamıştır. HDLM'nin ortalama değeri CNN+GRU'nun en kötü değerine göre %8,78, CNN+GRU'nun en iyi değerine göre %5,78 ve CNN+GRU'nun ortalama değerine göre %6,95 iyileşme sağlamıştır.

RMSE metriğine göre deneysel sonuçlar Çizelge 6.39'da ve Şekil 6.29'da görülmektedir.

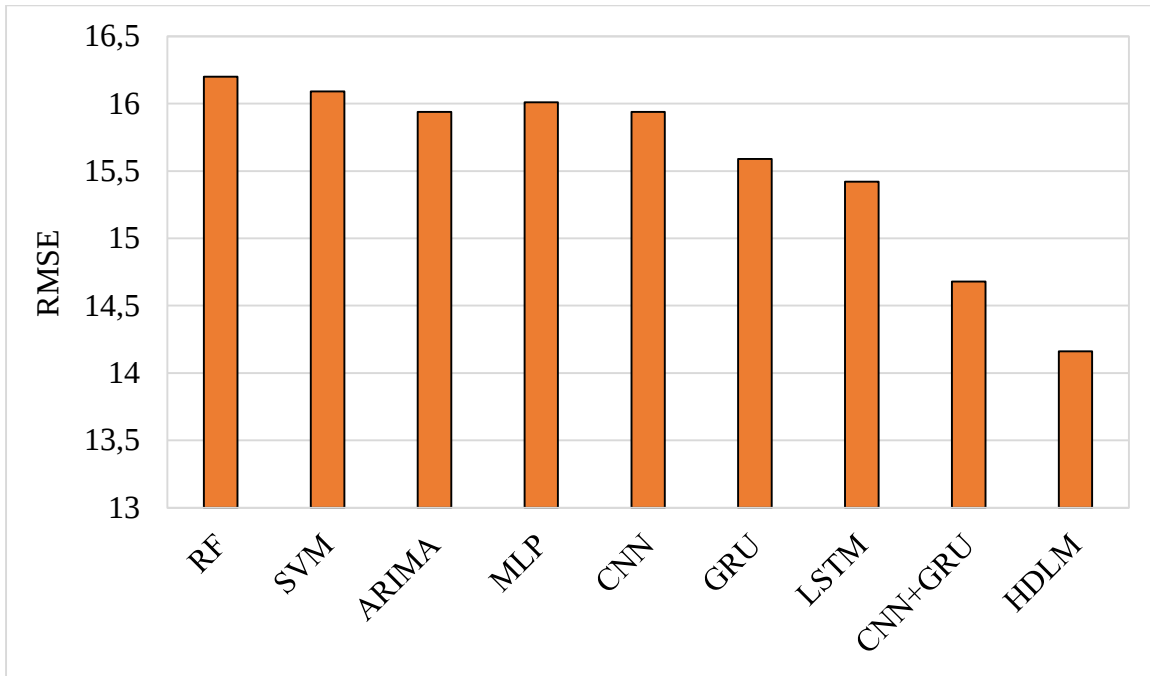
Çizelge 6.39. RMSE metriğine göre deneysel sonuçlar

Test	RF	SVM	ARIMA	MLP	CNN	GRU	LSTM	CNN+GRU	HDLM
1	16,20	16,09	15,94	15,81	15,89	15,45	15,41	14,59	14,11
2	16,20	16,09	15,94	15,86	15,91	15,47	15,41	14,61	14,12
3	16,20	16,09	15,94	15,87	15,93	15,51	15,41	14,61	14,13
4	16,20	16,09	15,94	15,91	15,93	15,60	15,41	14,63	14,14
5	16,20	16,09	15,94	15,98	15,93	15,60	15,42	14,67	14,15
6	16,20	16,09	15,94	16,02	15,96	15,60	15,42	14,68	14,16
7	16,20	16,09	15,94	16,02	15,96	15,60	15,42	14,71	14,17

Çizelge 6.39. (devam) RMSE metriğine göre deneysel sonuçlar

8	16,20	16,09	15,94	16,14	15,96	15,61	15,43	14,74	14,17
9	16,20	16,09	15,94	16,20	15,98	15,72	15,43	14,76	14,23
10	16,20	16,09	15,94	16,38	15,98	15,83	15,44	14,83	14,25
Ortalama	16,20	16,09	15,94	16,01	15,94	15,59	15,42	14,68	14,16

RMSE metriğine göre yapılan deneysel çalışmalar sonucunda, RF'ın 16,20, SVM'in 16,09, ARIMA'nın 15,94, MLP'nin 16,01, CNN'in 15,94, GRU'nun 15,59, LSTM'in 15,42, CNN+GRU'nun 14,68 ve HDLM'nin 14,16 ortalama RMSE değerine sahip olduğu görülmüştür.



Şekil 6.29. RMSE metriğine göre karşılaştırmalı deneysel sonuçlar

$$\text{HDLM'nin } \frac{\text{RF-HDLM}}{\text{RF}} \times 100, \frac{\text{SVR-HDLM}}{\text{SVR}} \times 100, \frac{\text{ARIMA-HDLM}}{\text{ARIMA}} \times 100, \frac{\text{MLP-HDLM}}{\text{MLP}} \times 100, \\ \frac{\text{CNN-HDLM}}{\text{CNN}} \times 100, \frac{\text{GRU-HDLM}}{\text{GRU}} \times 100, \frac{\text{LSTM-HDLM}}{\text{LSTM}} \times 100 \text{ ve } \frac{(\text{CNN+GRU})-\text{HDLM}}{(\text{CNN+GRU})} \times 100$$

eşitlikleri kullanılarak RMSE metriğine göre hesaplanan iyileşme oranları Çizelge 6.40'ta görülmektedir.

Çizelge 6.40. RMSE metriğine göre HDLM'nin iyileşme oranları

	HDLM'nin en kötü değeri	HDLM'nin en iyi değeri	HDLM'nin ortalama değeri
RF'in değerine göre iyileşme oranı (%)	12,03	12,90	12,57
SVM'nin değerine göre iyileşme oranı (%)	11,43	12,30	11,97
ARIMA'nın değerine göre iyileşme oranı (%)	10,60	11,48	11,14
MLP'nin en kötü değerine göre iyileşme oranı (%)	13,00	13,85	13,53
MLP'nin en iyi değerine göre iyileşme oranı (%)	9,86	10,75	10,41
MLP'nin ortalama değerine göre iyileşme oranı (%)	11,18	11,91	11,58
CNN'in en kötü değerine göre iyileşme oranı (%)	10,82	11,70	11,37
CNN'in en iyi değerine göre iyileşme oranı (%)	10,32	11,20	10,86
CNN'in ortalama değerine göre iyileşme oranı (%)	10,65	11,49	11,16
GRU'nun en kötü değerine göre iyileşme oranı (%)	9,98	10,86	10,53
GRU'nun en iyi değerine göre iyileşme oranı (%)	7,76	8,67	8,33
GRU'nun ortalama değerine göre iyileşme oranı (%)	8,73	9,54	9,20
LSTM'in en kötü değerine göre iyileşme oranı (%)	7,70	8,61	8,27
LSTM'in en iyi değerine göre iyileşme oranı (%)	7,52	8,43	8,09

Çizelge 6.40. (devam) RMSE metriğine göre HDLM'nin iyileşme oranları

LSTM'in ortalama değerine göre iyileşme oranı (%)	7,59	8,49	8,15
CNN+GRU'nun en kötü değerine göre iyileşme oranı (%)	3,91	4,85	4,49
CNN+GRU'nun en iyi değerine göre iyileşme oranı (%)	2,33	3,28	2,92
CNN+GRU'nun ortalama değerine göre iyileşme oranı (%)	2,96	3,90	3,54

Yapılan 10 farklı deney sonucunda elde edilen RMSE değerlerine göre HDLM'nin RF, SVM, ARIMA, MLP, CNN, GRU, LSTM ve CNN+GRU'ya göre karşılaştırmalı bir analizi görülmektedir. HDLM'nin en kötü değeri RF'a göre %12,03, en iyi değeri %12,90 ve ortalama değeri ise %12,57 iyileşme sağlamıştır. HDLM'nin en kötü değeri SVM'e göre %11,43, en iyi değeri %12,30 ve ortalama değeri ise %11,97 iyileşme sağlamıştır. HDLM'nin en kötü değeri ARIMA'ya göre %10,60, en iyi değeri %11,48 ve ortalama değeri ise %11,14 iyileşme sağlamıştır.

HDLM'nin en kötü değeri MLP'nin en kötü değerine göre %13,00, MLP'nin en iyi değerine göre %9,86 ve MLP'nin ortalama değerine göre %11,18 iyileşme sağlamıştır. HDLM'nin en iyi değeri MLP'nin en kötü değerine göre %13,85, MLP'nin en iyi değerine göre %10,75 ve MLP'nin ortalama değerine göre %11,91 iyileşme sağlamıştır. HDLM'nin ortalama değeri MLP'nin en kötü değerine göre %13,53, MLP'nin en iyi değerine göre %10,41 ve MLP'nin ortalama değerine göre %11,58 iyileşme sağlamıştır.

HDLM'nin en kötü değeri CNN'in en kötü değerine göre %10,82, CNN'in en iyi değerine göre %10,32 ve CNN'in ortalama değerine göre %10,65 iyileşme sağlamıştır. HDLM'nin en iyi değeri CNN'in en kötü değerine göre %11,70, CNN'in en iyi değerine göre %11,20 ve CNN'in ortalama değerine göre %11,49 iyileşme sağlamıştır. HDLM'nin ortalama değeri CNN'in en kötü değerine göre %11,37, CNN'in en iyi değerine göre %10,86 ve CNN'in ortalama değerine göre %11,16 iyileşme sağlamıştır.

HDLM'nin en kötü değeri GRU'nun en kötü değerine göre %9,98, GRU'nun en iyi değerine göre %7,76 ve GRU'nun ortalama değerine göre %8,73 iyileşme sağlamıştır. HDLM'nin en iyi değeri GRU'nun en kötü değerine göre %10,86, GRU'nun en iyi değerine göre %8,67 ve GRU'nun ortalama değerine göre %9,54 iyileşme sağlamıştır. HDLM'nin ortalama değeri GRU'nun en kötü değerine göre %10,53, GRU'nun en iyi değerine göre %8,33 ve GRU'nun ortalama değerine göre %9,20 iyileşme sağlamıştır.

HDLM'nin en kötü değeri LSTM'in en kötü değerine göre %7,70, LSTM'in en iyi değerine göre %7,52 ve LSTM'in ortalama değerine göre %7,59 iyileşme sağlamıştır. HDLM'nin en iyi değeri LSTM'in en kötü değerine göre %8,61, LSTM'in en iyi değerine göre %8,43 ve LSTM'in ortalama değerine göre %8,49 iyileşme sağlamıştır. HDLM'nin ortalama değeri LSTM'in en kötü değerine göre %8,27, LSTM'in en iyi değerine göre %8,09 ve LSTM'in ortalama değerine göre %8,15 iyileşme sağlamıştır.

HDLM'nin en kötü değeri CNN+GRU'nun en kötü değerine göre %3,91, CNN+GRU'nun en iyi değerine göre %2,33 ve CNN+GRU'nun ortalama değerine göre %2,96 iyileşme sağlamıştır. HDLM'nin en iyi değeri CNN+GRU'nun en kötü değerine göre %4,85, CNN+GRU'nun en iyi değerine göre %3,28 ve CNN+GRU'nun ortalama değerine göre %3,90 iyileşme sağlamıştır. HDLM'nin ortalama değeri CNN+GRU'nun en kötü değerine göre %4,49, CNN+GRU'nun en iyi değerine göre %2,92 ve CNN+GRU'nun ortalama değerine göre %3,54 iyileşme sağlamıştır.

MAE metriğine göre deneysel sonuçlar Çizelge 6.41'de ve Şekil 6.30'da görülmektedir.

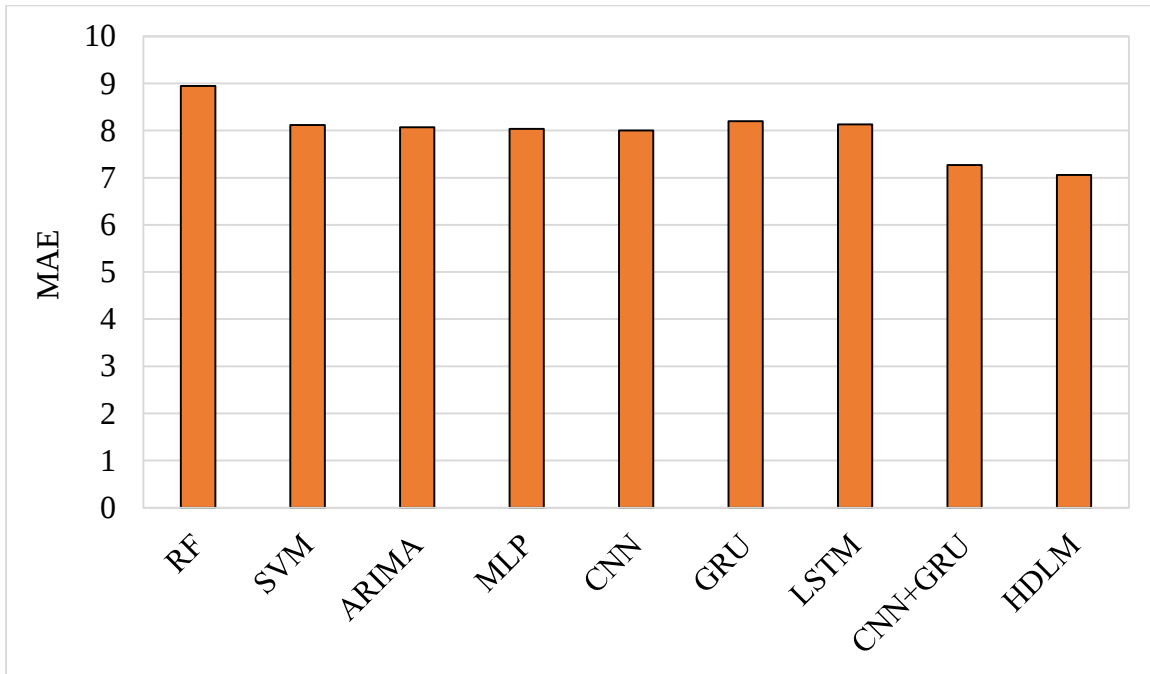
Çizelge 6.41. MAE metriğine göre deneysel sonuçlar

Test	RF	SVM	ARIMA	MLP	CNN	GRU	LSTM	CNN+GRU	HDLM
1	8,95	8,12	8,07	8,01	7,82	8,13	8,10	7,23	7,05
2	8,95	8,12	8,07	8,03	7,89	8,18	8,11	7,25	7,05
3	8,95	8,12	8,07	8,03	7,92	8,18	8,12	7,25	7,05
4	8,95	8,12	8,07	8,04	7,94	8,18	8,12	7,26	7,06
5	8,95	8,12	8,07	8,04	8,00	8,18	8,14	7,26	7,06
6	8,95	8,12	8,07	8,06	8,01	8,21	8,14	7,27	7,06
7	8,95	8,12	8,07	8,06	8,03	8,21	8,14	7,28	7,06

Çizelge 6.41. (devam) MAE metriğine göre deneysel sonuçlar

8	8,95	8,12	8,07	8,06	8,07	8,22	8,14	7,30	7,07
9	8,95	8,12	8,07	8,08	8,15	8,23	8,16	7,30	7,07
10	8,95	8,12	8,07	8,08	8,19	8,31	8,16	7,31	7,11
Ortalama	8,95	8,12	8,07	8,04	8,00	8,20	8,13	7,27	7,06

MAE metriğine göre yapılan deneysel çalışmalar sonucunda, RF'ın 8,95, SVM'in 8,12, ARIMA'nın 8,07, MLP'nin 8,04, CNN'in 8,00, GRU'nun 8,20, LSTM'in 8,13, CNN+GRU'nun 7,27 ve HDLM'nin 7,06 ortalama MAE değerine sahip olduğu görülmüştür.



Şekil 6.30. MAE metriğine göre karşılaştırmalı deneysel sonuçlar

$$\text{HDLM'nin } \frac{\text{RF-HDLM}}{\text{RF}} \times 100, \frac{\text{SVR-HDLM}}{\text{SVR}} \times 100, \frac{\text{ARIMA-HDLM}}{\text{ARIMA}} \times 100, \frac{\text{MLP-HDLM}}{\text{MLP}} \times 100, \\ \frac{\text{CNN-HDLM}}{\text{CNN}} \times 100, \frac{\text{GRU-HDLM}}{\text{GRU}} \times 100, \frac{\text{LSTM-HDLM}}{\text{LSTM}} \times 100 \text{ ve } \frac{(\text{CNN+GRU})\text{-HDLM}}{(\text{CNN+GRU})} \times 100$$

eşitlikleri kullanılarak MAE metriğine göre hesaplanan iyileşme oranları Çizelge 6.42'de görülmektedir.

Çizelge 6.42. MAE metriğine göre HDLM'nin iyileşme oranları

	HDLM'nin en kötü değeri	HDLM'nin en iyi değeri	HDLM'nin ortalama değeri
RF'in değerine göre iyileşme oranı (%)	20,55	21,22	21,07
SVM'nin değerine göre iyileşme oranı (%)	12,43	13,17	13,00
ARIMA'nın değerine göre iyileşme oranı (%)	11,89	12,63	12,46
MLP'nin en kötü değerine göre iyileşme oranı (%)	12,00	12,74	12,57
MLP'nin en iyi değerine göre iyileşme oranı (%)	11,23	11,98	11,81
MLP'nin ortalama değerine göre iyileşme oranı (%)	11,72	12,41	12,23
CNN'in en kötü değerine göre iyileşme oranı (%)	13,18	13,91	13,74
CNN'in en iyi değerine göre iyileşme oranı (%)	9,07	9,84	9,66
CNN'in ortalama değerine göre iyileşme oranı (%)	11,40	11,89	11,72
GRU'nun en kötü değerine göre iyileşme oranı (%)	14,44	15,16	14,99
GRU'nun en iyi değerine göre iyileşme oranı (%)	12,54	13,28	13,11
GRU'nun ortalama değerine göre iyileşme oranı (%)	13,44	14,05	13,88
LSTM'in en kötü değerine göre iyileşme oranı (%)	12,86	13,60	13,43
LSTM'in en iyi değerine göre iyileşme oranı (%)	12,22	12,96	12,79

Çizelge 6.42. (devam) MAE metriğine göre HDLM'nin iyileşme oranları

LSTM'in ortalama değerine göre iyileşme oranı (%)	12,62	13,31	13,14
CNN+GRU'nun en kötü değerine göre iyileşme oranı (%)	2,73	3,55	3,36
CNN+GRU'nun en iyi değerine göre iyileşme oranı (%)	1,65	2,48	2,29
CNN+GRU'nun ortalama değerine göre iyileşme oranı (%)	2,22	3,03	2,84

Yapılan 10 farklı deney sonucunda elde edilen MAE değerlerine göre HDLM'nin RF, SVM, ARIMA, MLP, CNN, GRU, LSTM ve CNN+GRU'ya göre karşılaştırmalı bir analizi görülmektedir. HDLM'nin en kötü değeri RF'a göre %20,55, en iyi değeri %21,22 ve ortalama değeri ise %21,07 iyileşme sağlamıştır. HDLM'nin en kötü değeri SVM'e göre %12,43, en iyi değeri %13,17 ve ortalama değeri ise %13,00 iyileşme sağlamıştır. HDLM'nin en kötü değeri ARIMA'ya göre %11,89, en iyi değeri %12,63 ve ortalama değeri ise %12,46 iyileşme sağlamıştır.

HDLM'nin en kötü değeri MLP'nin en kötü değerine göre %12,00, MLP'nin en iyi değerine göre %11,23 ve MLP'nin ortalama değerine göre %11,72 iyileşme sağlamıştır. HDLM'nin en iyi değeri MLP'nin en kötü değerine göre %12,74, MLP'nin en iyi değerine göre %11,98 ve MLP'nin ortalama değerine göre %12,41 iyileşme sağlamıştır. HDLM'nin ortalama değeri MLP'nin en kötü değerine göre %12,57, MLP'nin en iyi değerine göre %11,81 ve MLP'nin ortalama değerine göre %12,23 iyileşme sağlamıştır.

HDLM'nin en kötü değeri CNN'in en kötü değerine göre %13,18, CNN'in en iyi değerine göre %9,07 ve CNN'in ortalama değerine göre %11,40 iyileşme sağlamıştır. HDLM'nin en iyi değeri CNN'in en kötü değerine göre %13,91, CNN'in en iyi değerine göre %9,84 ve CNN'in ortalama değerine göre %11,89 iyileşme sağlamıştır. HDLM'nin ortalama değeri CNN'in en kötü değerine göre %13,74, CNN'in en iyi değerine göre %9,66 ve CNN'in ortalama değerine göre %11,72 iyileşme sağlamıştır.

HDLM'nin en kötü değeri GRU'nun en kötü değerine göre %14,44, GRU'nun en iyi değerine göre %12,54 ve GRU'nun ortalama değerine göre %13,44 iyileşme sağlamıştır. HDLM'nin en iyi değeri GRU'nun en kötü değerine göre %15,16, GRU'nun en iyi değerine göre %13,28 ve GRU'nun ortalama değerine göre %14,05 iyileşme sağlamıştır. HDLM'nin ortalama değeri GRU'nun en kötü değerine göre %14,99, GRU'nun en iyi değerine göre %13,11 ve GRU'nun ortalama değerine göre %13,88 iyileşme sağlamıştır.

HDLM'nin en kötü değeri LSTM'in en kötü değerine göre %12,86, LSTM'in en iyi değerine göre %12,22 ve LSTM'in ortalama değerine göre %12,62 iyileşme sağlamıştır. HDLM'nin en iyi değeri LSTM'in en kötü değerine göre %13,60, LSTM'in en iyi değerine göre %12,96 ve LSTM'in ortalama değerine göre %13,31 iyileşme sağlamıştır. HDLM'nin ortalama değeri LSTM'in en kötü değerine göre %13,43, LSTM'in en iyi değerine göre %12,79 ve LSTM'in ortalama değerine göre %13,14 iyileşme sağlamıştır.

HDLM'nin en kötü değeri CNN+GRU'nun en kötü değerine göre %2,73, CNN+GRU'nun en iyi değerine göre %1,65 ve CNN+GRU'nun ortalama değerine göre %2,22 iyileşme sağlamıştır. HDLM'nin en iyi değeri CNN+GRU'nun en kötü değerine göre %3,55, CNN+GRU'nun en iyi değerine göre %2,48 ve CNN+GRU'nun ortalama değerine göre %3,03 iyileşme sağlamıştır. HDLM'nin ortalama değeri CNN+GRU'nun en kötü değerine göre %3,36, CNN+GRU'nun en iyi değerine göre %2,29 ve CNN+GRU'nun ortalama değerine göre %2,84 iyileşme sağlamıştır.

Deneyisel sonuçlar, HDLM'nin RF, SVM, ARIMA, MLP, CNN, GRU, LSTM ve CNN+GRU'ya göre daha düşük MSE, RMSE ve MAE sonuçlarına sahip olduğunu göstermektedir. Bir sonraki saldırı zamanının belirlenmesinde HDLM'nin diğer modellere göre daha başarılı ve etkin sonuçlar verdiği görülmüştür.

6.2.3. Suç veri kümesi

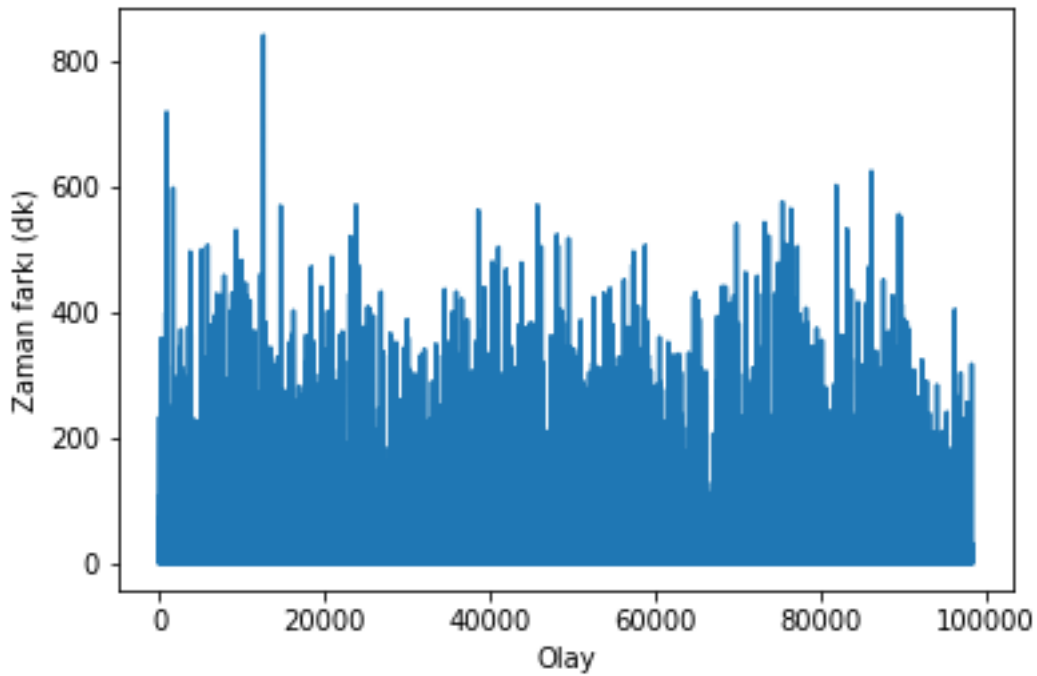
Suç veri kümesi için çete saldırısı olaylarının meydana gelme zamanları arasındaki dakika cinsinden farklara göre oluşturulan verinin ilk 5 satırı Çizelge 6.43'te görülmektedir.

Çizelge 6.43. Bir önceki çete saldırısından sonra geçen zaman

Olay	Bir önceki çete saldırısından sonra geçen zaman (dk)
1	37
2	2
3	3
4	15
5	2

Çizelge 6.43'te çete saldırısı olayları arasındaki zaman farkı görülmektedir. Suç veri kümesi, DDoS veri kümesine göre daha nispeten daha düzensiz, e-ticaret veri kümesine göre nispeten daha düzenli bir örüntü izlemektedir.

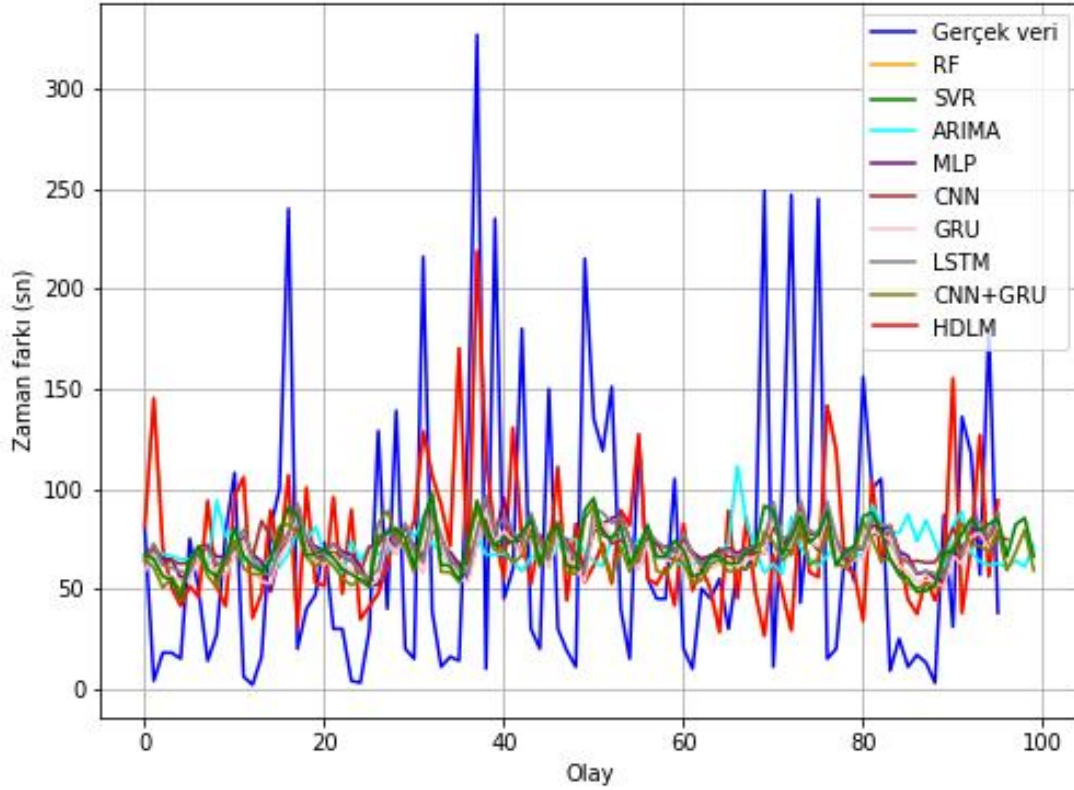
Çete saldırısı olayları arasındaki dakika cinsinden zaman farkları Şekil 6.31'de görülmektedir.



Şekil 6.31. Çete saldırısı olayları arasındaki dakika cinsinden zaman farkları

Şekil 6.31’de çete saldırısı olayları arasındaki zaman farkı görülmektedir. Suç veri kümesi, 41967 çete saldırısı olayından oluşmaktadır. Veri kümesinde nadiren de olsa görülen sapmalar mevcuttur.

Şekil 6.32’de örnek olarak test verisinin 100 elemanı için deneysel sonuçlar görülmektedir.



Şekil 6.32. Test verisindeki ilk 100 veri örneği için deneysel sonuçlar

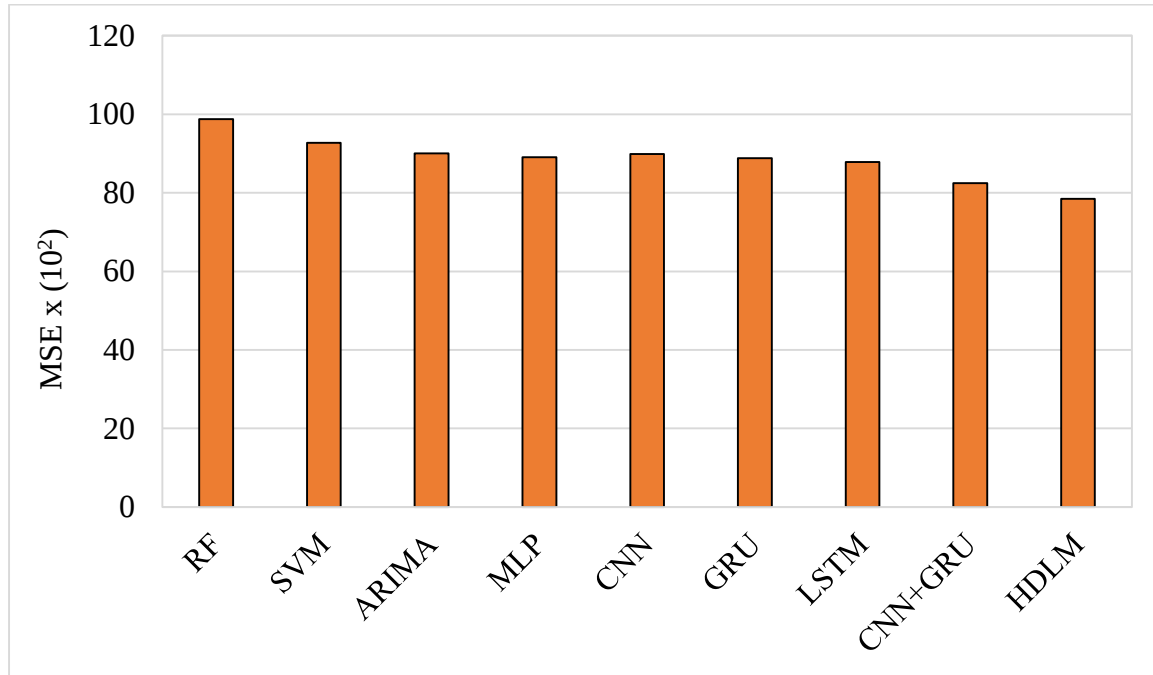
Şekil 6.32’de test verisindeki ilk 100 veri örneği için deneysel sonuçlar görülmektedir. HDLM, diğer modellere kıyasla veri kümesindeki ani sapmaları daha başarılı bir şekilde belirleyerek daha yüksek tahmin doğruluğuna sahip olmuştur.

MSE metriğine göre deneysel sonuçlar Çizelge 6.44’te ve Şekil 6.33’te görülmektedir.

Çizelge 6.44. MSE metriğine göre deneysel sonuçlar

Test	RF	SVM	ARIMA	MLP	CNN	GRU	LSTM	CNN+GRU	HDLM
1	9871,44	9274,94	8999,17	8906,02	8968,29	8876,92	8784,10	8190,39	7807,44
2	9871,44	9274,94	8999,17	8906,13	8970,17	8877,05	8784,12	8193,82	7822,94
3	9871,44	9274,94	8999,17	8906,13	8978,05	8877,52	8784,18	8225,46	7825,37
4	9871,44	9274,94	8999,17	8906,19	8982,84	8877,52	8784,18	8236,81	7830,52
5	9871,44	9274,94	8999,17	8906,22	8985,58	8877,75	8784,19	8247,58	7835,03
6	9871,44	9274,94	8999,17	8906,25	8986,03	8877,80	8784,22	8249,54	7845,65
7	9871,44	9274,94	8999,17	8906,29	8986,53	8877,81	8784,23	8250,55	7857,68
8	9871,44	9274,94	8999,17	8906,34	8989,23	8878,02	8784,25	8273,45	7886,79
9	9871,44	9274,94	8999,17	8906,34	8995,60	8878,09	8784,33	8293,61	7887,36
10	9871,44	9274,94	8999,17	8906,35	9004,53	8878,31	8784,45	8314,91	7900,66
Ortalama	9871,44	9274,94	8999,17	8906,22	8984,68	8877,67	8784,22	8247,61	7849,94

MSE metriğine göre yapılan deneysel çalışmalar sonucunda, RF'ın 9871,44, SVM'in 9274,94, ARIMA'nın 8999,17, MLP'nin 8906,22, CNN'in 8984,68, GRU'nun 8877,67, LSTM'in 8784,22, CNN+GRU'nun 8247,61 ve HDLM'nin 7849,94 ortalama MSE değerine sahip olduğu görülmüştür.



Şekil 6.33. MSE metriğine göre karşılaştırmalı deneysel sonuçlar

$$\text{HDLM'nin } \frac{\text{RF-HDLM}}{\text{RF}} \times 100, \frac{\text{SVR-HDLM}}{\text{SVR}} \times 100, \frac{\text{ARIMA-HDLM}}{\text{ARIMA}} \times 100, \frac{\text{MLP-HDLM}}{\text{MLP}} \times 100, \\ \frac{\text{CNN-HDLM}}{\text{CNN}} \times 100, \frac{\text{GRU-HDLM}}{\text{GRU}} \times 100, \frac{\text{LSTM-HDLM}}{\text{LSTM}} \times 100 \text{ ve } \frac{(\text{CNN+GRU})\text{-HDLM}}{(\text{CNN+GRU})} \times 100$$

eşitlikleri kullanılarak MSE metriğine göre hesaplanan iyileşme oranları Çizelge 6.45'te görülmektedir.

Çizelge 6.45. MSE metriğine göre HDLM'nin iyileşme oranları

	HDLM'nin en kötü değeri	HDLM'nin en iyi değeri	HDLM'nin ortalama değeri
RF'in değerine göre iyileşme oranı (%)	19,96	20,90	20,47
SVM'nin değerine göre iyileşme oranı (%)	14,81	15,82	15,36
ARIMA'nın değerine göre iyileşme oranı (%)	12,20	13,24	12,77
MLP'nin en kötü değerine göre iyileşme oranı (%)	11,29	12,33	11,86
MLP'nin en iyi değerine göre iyileşme oranı (%)	11,28	12,33	11,85
MLP'nin ortalama değerine göre iyileşme oranı (%)	11,29	12,33	11,86
CNN'in en kötü değerine göre iyileşme oranı (%)	12,25	13,29	12,82
CNN'in en iyi değerine göre iyileşme oranı (%)	11,90	12,94	12,47
CNN'in ortalama değerine göre iyileşme oranı (%)	12,08	13,10	12,62
GRU'nun en kötü değerine göre iyileşme oranı (%)	11,01	12,06	11,58
GRU'nun en iyi değerine göre iyileşme oranı (%)	10,99	12,04	11,56

Çizelge 6.45. (devam) MSE metriğine göre HDLM'nin iyileşme oranları

GRU'nun ortalama değerine göre iyileşme oranı (%)	11,00	12,05	11,57
LSTM'in en kötü değerine göre iyileşme oranı (%)	10,06	11,12	10,63
LSTM'in en iyi değerine göre iyileşme oranı (%)	10,05	11,11	10,63
LSTM'in ortalama değerine göre iyileşme oranı (%)	10,05	11,11	10,63
CNN+GRU'nun en kötü değerine göre iyileşme oranı (%)	4,98	6,10	5,59
CNN+GRU'nun en iyi değerine göre iyileşme oranı (%)	3,53	4,67	4,15
CNN+GRU'nun ortalama değerine göre iyileşme oranı (%)	4,23	5,33	4,82

Yapılan 10 farklı deney sonucunda elde edilen MSE değerlerine göre HDLM'nin RF, SVM, ARIMA, MLP, CNN, GRU, LSTM ve CNN+GRU'ya göre karşılaştırmalı bir analizi görülmektedir. HDLM'nin en kötü değeri RF'a göre %19,96, en iyi değeri %20,90 ve ortalama değeri ise %20,47 iyileşme sağlamıştır. HDLM'nin en kötü değeri SVM'e göre %14,81, en iyi değeri %15,82 ve ortalama değeri ise %15,36 iyileşme sağlamıştır. HDLM'nin en kötü değeri ARIMA'ya göre %12,20, en iyi değeri %13,24 ve ortalama değeri ise %12,77 iyileşme sağlamıştır.

HDLM'nin en kötü değeri MLP'nin en kötü değerine göre %11,29, MLP'nin en iyi değerine göre %11,28 ve MLP'nin ortalama değerine göre %11,29 iyileşme sağlamıştır. HDLM'nin en iyi değeri MLP'nin en kötü değerine göre %12,33, MLP'nin en iyi değerine göre %12,33 ve MLP'nin ortalama değerine göre %12,33 iyileşme sağlamıştır. HDLM'nin ortalama değeri MLP'nin en kötü değerine göre %11,86, MLP'nin en iyi değerine göre %11,85 ve MLP'nin ortalama değerine göre %11,86 iyileşme sağlamıştır.

HDLM'nin en kötü değeri CNN'in en kötü değerine göre %12,25, CNN'in en iyi değerine göre %11,90 ve CNN'in ortalama değerine göre %12,08 iyileşme sağlamıştır. HDLM'nin en iyi değeri CNN'in en kötü değerine göre %13,29, CNN'in en iyi değerine göre %12,94 ve CNN'in ortalama değerine göre %13,10 iyileşme sağlamıştır. HDLM'nin ortalama değeri CNN'in en kötü değerine göre %12,82, CNN'in en iyi değerine göre %12,47 ve CNN'in ortalama değerine göre %12,62 iyileşme sağlamıştır.

HDLM'nin en kötü değeri GRU'nun en kötü değerine göre %11,01, GRU'nun en iyi değerine göre %10,99 ve GRU'nun ortalama değerine göre %11,00 iyileşme sağlamıştır. HDLM'nin en iyi değeri GRU'nun en kötü değerine göre %12,06, GRU'nun en iyi değerine göre %12,04 ve GRU'nun ortalama değerine göre %12,05 iyileşme sağlamıştır. HDLM'nin ortalama değeri GRU'nun en kötü değerine göre %11,58, GRU'nun en iyi değerine göre %11,56 ve GRU'nun ortalama değerine göre %11,57 iyileşme sağlamıştır.

HDLM'nin en kötü değeri LSTM'in en kötü değerine göre %10,06, LSTM'in en iyi değerine göre %10,05 ve LSTM'in ortalama değerine göre %10,05 iyileşme sağlamıştır. HDLM'nin en iyi değeri LSTM'in en kötü değerine göre %11,12, LSTM'in en iyi değerine göre %11,11 ve LSTM'in ortalama değerine göre %11,11 iyileşme sağlamıştır. HDLM'nin ortalama değeri LSTM'in en kötü değerine göre %10,63, LSTM'in en iyi değerine göre %10,63 ve LSTM'in ortalama değerine göre %10,63 iyileşme sağlamıştır.

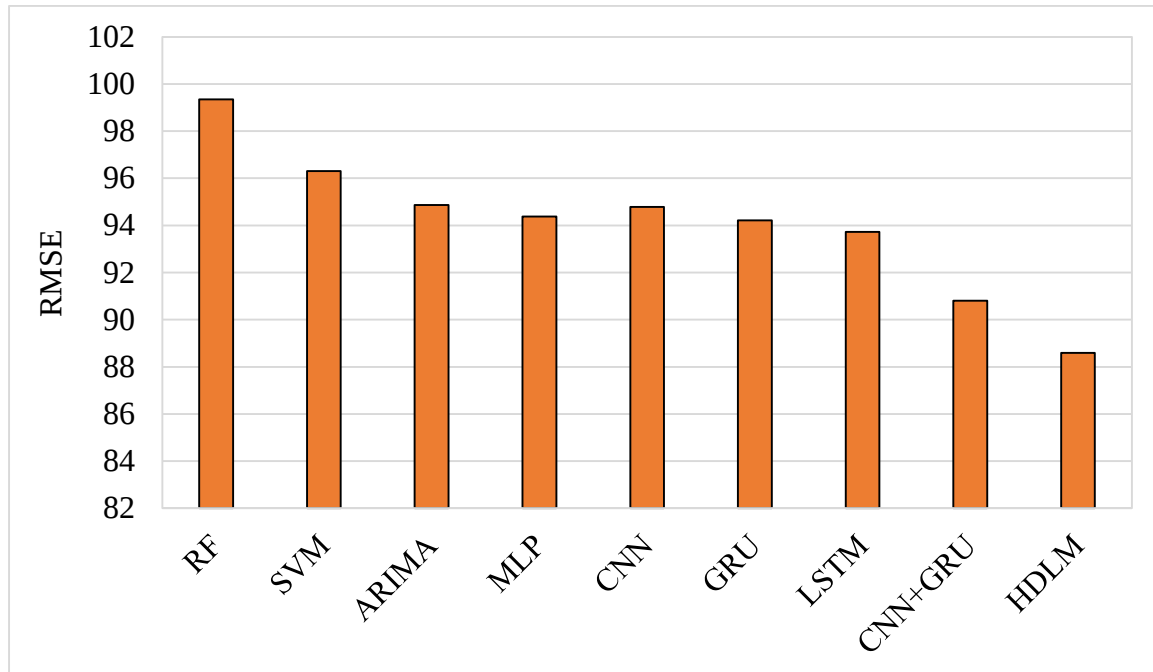
HDLM'nin en kötü değeri CNN+GRU'nun en kötü değerine göre %4,98, CNN+GRU'nun en iyi değerine göre %3,53 ve CNN+GRU'nun ortalama değerine göre %4,23 iyileşme sağlamıştır. HDLM'nin en iyi değeri CNN+GRU'nun en kötü değerine göre %6,10, CNN+GRU'nun en iyi değerine göre %4,67 ve CNN+GRU'nun ortalama değerine göre %5,33 iyileşme sağlamıştır. HDLM'nin ortalama değeri CNN+GRU'nun en kötü değerine göre %5,59, CNN+GRU'nun en iyi değerine göre %4,15 ve CNN+GRU'nun ortalama değerine göre %4,82 iyileşme sağlamıştır.

RMSE metriğine göre deneysel sonuçlar Çizelge 6.46'da ve Şekil 6.34'te görülmektedir.

Çizelge 6.46. RMSE metriğine göre deneysel sonuçlar

Test	RF	SVM	ARIMA	MLP	CNN	GRU	LSTM	CNN+GRU	HDLM
1	99,35	96,30	94,86	94,37	94,70	94,22	93,72	90,50	88,35
2	99,35	96,30	94,86	94,37	94,71	94,22	93,72	90,51	88,44
3	99,35	96,30	94,86	94,37	94,75	94,22	93,72	90,69	88,46
4	99,35	96,30	94,86	94,37	94,77	94,22	93,72	90,75	88,49
5	99,35	96,30	94,86	94,37	94,79	94,22	93,72	90,81	88,51
6	99,35	96,30	94,86	94,37	94,79	94,22	93,72	90,82	88,57
7	99,35	96,30	94,86	94,37	94,79	94,22	93,72	90,83	88,64
8	99,35	96,30	94,86	94,37	94,81	94,22	93,72	90,95	88,80
9	99,35	96,30	94,86	94,37	94,84	94,22	93,72	91,06	88,81
10	99,35	96,30	94,86	94,37	94,89	94,22	93,73	91,18	88,88
Ortalama	99,35	96,30	94,86	94,37	94,78	94,22	93,72	90,81	88,59

RMSE metriğine göre yapılan deneysel çalışmalar sonucunda, RF'ın 99,35, SVM'in 96,30, ARIMA'nın 94,86, MLP'nin 94,37, CNN'in 94,78, GRU'nun 94,22, LSTM'in 93,72, CNN+GRU'nun 90,81 ve HDLM'nin 88,59 ortalama RMSE değerine sahip olduğu görülmüştür.



Şekil 6.34. RMSE metriğine göre karşılaştırmalı deneysel sonuçlar

$$\text{HDLM'nin } \frac{\text{RF-HDLM}}{\text{RF}} \times 100, \frac{\text{SVR-HDLM}}{\text{SVR}} \times 100, \frac{\text{ARIMA-HDLM}}{\text{ARIMA}} \times 100, \frac{\text{MLP-HDLM}}{\text{MLP}} \times 100, \\ \frac{\text{CNN-HDLM}}{\text{CNN}} \times 100, \frac{\text{GRU-HDLM}}{\text{GRU}} \times 100, \frac{\text{LSTM-HDLM}}{\text{LSTM}} \times 100 \text{ ve } \frac{(\text{CNN+GRU})\text{-HDLM}}{(\text{CNN+GRU})} \times 100$$

eşitlikleri kullanılarak RMSE metriğine göre hesaplanan iyileşme oranları Çizelge 6.47’de görülmektedir.

Çizelge 6.47. RMSE metriğine göre HDLM’nin iyileşme oranları

	HDLM’nin en kötü değeri	HDLM’nin en iyi değeri	HDLM’nin ortalama değeri
RF’in değerine göre iyileşme oranı (%)	10,53	11,07	10,82
SVM’nin değerine göre iyileşme oranı (%)	7,70	8,25	8,00
ARIMA’nın değerine göre iyileşme oranı (%)	6,30	6,86	6,60
MLP’nin en kötü değerine göre iyileşme oranı (%)	5,81	6,37	6,11
MLP’nin en iyi değerine göre iyileşme oranı (%)	5,81	6,37	6,11
MLP’nin ortalama değerine göre iyileşme oranı (%)	5,81	6,37	6,11
CNN’in en kötü değerine göre iyileşme oranı (%)	6,33	6,89	6,63
CNN’in en iyi değerine göre iyileşme oranı (%)	6,14	6,70	6,44
CNN’in ortalama değerine göre iyileşme oranı (%)	6,23	6,78	6,52
GRU’nun en kötü değerine göre iyileşme oranı (%)	5,66	6,23	5,97
GRU’nun en iyi değerine göre iyileşme oranı (%)	5,66	6,23	5,97

Çizelge 6.47. (devam) RMSE metriğine göre HDLM'nin iyileşme oranları

GRU'nun ortalama değerine göre iyileşme oranı (%)	5,66	6,23	5,97
LSTM'in en kötü değerine göre iyileşme oranı (%)	5,17	5,73	5,47
LSTM'in en iyi değerine göre iyileşme oranı (%)	5,16	5,72	5,46
LSTM'in ortalama değerine göre iyileşme oranı (%)	5,16	5,73	5,46
CNN+GRU'nun en kötü değerine göre iyileşme oranı (%)	2,52	3,10	2,83
CNN+GRU'nun en iyi değerine göre iyileşme oranı (%)	1,79	2,37	2,10
CNN+GRU'nun ortalama değerine göre iyileşme oranı (%)	2,13	2,70	2,43

Yapılan 10 farklı deney sonucunda elde edilen RMSE değerlerine göre HDLM'nin RF, SVM, ARIMA, MLP, CNN, GRU, LSTM ve CNN+GRU'ya göre karşılaştırmalı bir analizi görülmektedir. HDLM'nin en kötü değeri RF'a göre %10,53, en iyi değeri %11,07 ve ortalama değeri ise %10,82 iyileşme sağlamıştır. HDLM'nin en kötü değeri SVM'e göre %7,70, en iyi değeri %8,25 ve ortalama değeri ise %8,00 iyileşme sağlamıştır. HDLM'nin en kötü değeri ARIMA'ya göre %6,30, en iyi değeri %6,86 ve ortalama değeri ise %6,60 iyileşme sağlamıştır.

HDLM'nin en kötü değeri MLP'nin en kötü değerine göre %5,81, MLP'nin en iyi değerine göre %5,81 ve MLP'nin ortalama değerine göre %5,81 iyileşme sağlamıştır. HDLM'nin en iyi değeri MLP'nin en kötü değerine göre %6,37, MLP'nin en iyi değerine göre %6,37 ve MLP'nin ortalama değerine göre %6,37 iyileşme sağlamıştır. HDLM'nin ortalama değeri MLP'nin en kötü değerine göre %6,11, MLP'nin en iyi değerine göre %6,11 ve MLP'nin ortalama değerine göre %6,11 iyileşme sağlamıştır.

HDLM'nin en kötü değeri CNN'in en kötü değerine göre %6,33, CNN'in en iyi değerine göre %6,14 ve CNN'in ortalama değerine göre %6,23 iyileşme sağlamıştır. HDLM'nin en iyi değeri CNN'in en kötü değerine göre %6,89, CNN'in en iyi değerine göre %6,70 ve CNN'in ortalama değerine göre %6,78 iyileşme sağlamıştır. HDLM'nin ortalama değeri CNN'in en kötü değerine göre %6,63, CNN'in en iyi değerine göre %6,44 ve CNN'in ortalama değerine göre %6,52 iyileşme sağlamıştır.

HDLM'nin en kötü değeri GRU'nun en kötü değerine göre %5,66, GRU'nun en iyi değerine göre %5,66 ve GRU'nun ortalama değerine göre %5,66 iyileşme sağlamıştır. HDLM'nin en iyi değeri GRU'nun en kötü değerine göre %6,23, GRU'nun en iyi değerine göre %6,23 ve GRU'nun ortalama değerine göre %6,23 iyileşme sağlamıştır. HDLM'nin ortalama değeri GRU'nun en kötü değerine göre %5,97, GRU'nun en iyi değerine göre %5,97 ve GRU'nun ortalama değerine göre %5,97 iyileşme sağlamıştır.

HDLM'nin en kötü değeri LSTM'in en kötü değerine göre %5,17, LSTM'in en iyi değerine göre %5,16 ve LSTM'in ortalama değerine göre %5,16 iyileşme sağlamıştır. HDLM'nin en iyi değeri LSTM'in en kötü değerine göre %5,73, LSTM'in en iyi değerine göre %5,72 ve LSTM'in ortalama değerine göre %5,73 iyileşme sağlamıştır. HDLM'nin ortalama değeri LSTM'in en kötü değerine göre %5,47, LSTM'in en iyi değerine göre %5,46 ve LSTM'in ortalama değerine göre %5,46 iyileşme sağlamıştır.

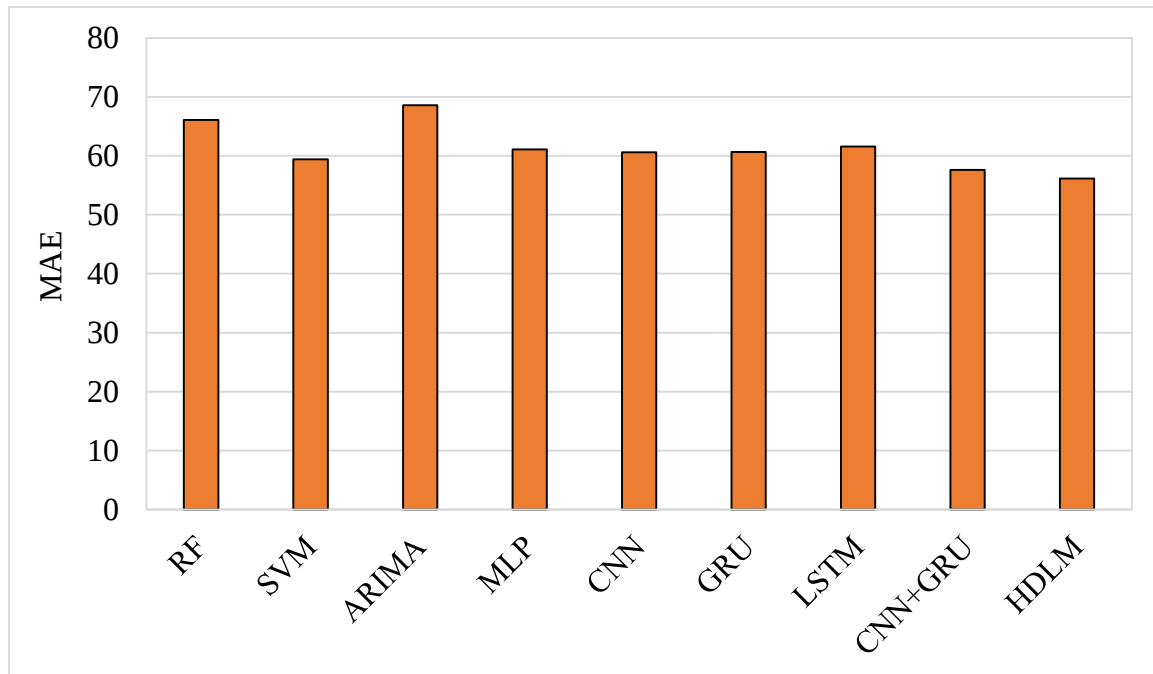
HDLM'nin en kötü değeri CNN+GRU'nun en kötü değerine göre %2,52, CNN+GRU'nun en iyi değerine göre %1,79 ve CNN+GRU'nun ortalama değerine göre %2,13 iyileşme sağlamıştır. HDLM'nin en iyi değeri CNN+GRU'nun en kötü değerine göre %3,10, CNN+GRU'nun en iyi değerine göre %2,37 ve CNN+GRU'nun ortalama değerine göre %2,70 iyileşme sağlamıştır. HDLM'nin ortalama değeri CNN+GRU'nun en kötü değerine göre %2,83, CNN+GRU'nun en iyi değerine göre %2,10 ve CNN+GRU'nun ortalama değerine göre %2,43 iyileşme sağlamıştır.

MAE metriğine göre deneysel sonuçlar Çizelge 6.48'de ve Şekil 6.35'te görülmektedir.

Çizelge 6.48. MAE metriğine göre deneysel sonuçlar

Test	RF	SVM	ARIMA	MLP	CNN	GRU	LSTM	CNN+GRU	HDLM
1	66,09	59,43	68,60	61,08	60,52	60,67	61,57	57,47	56,05
2	66,09	59,43	68,60	61,08	60,52	60,67	61,57	57,50	56,06
3	66,09	59,43	68,60	61,08	60,55	60,67	61,57	57,54	56,07
4	66,09	59,43	68,60	61,08	60,57	60,67	61,57	57,54	56,12
5	66,09	59,43	68,60	61,08	60,60	60,67	61,57	57,56	56,14
6	66,09	59,43	68,60	61,08	60,61	60,67	61,57	57,58	56,15
7	66,09	59,43	68,60	61,08	60,61	60,68	61,57	57,67	56,17
8	66,09	59,43	68,60	61,08	60,62	60,68	61,57	57,73	56,19
9	66,09	59,43	68,60	61,08	60,66	60,68	61,57	57,89	56,21
10	66,09	59,43	68,60	61,08	60,70	60,68	61,57	57,94	56,26
Ortalama	66,09	59,43	68,60	61,08	60,59	60,67	61,57	57,64	56,14

MAE metriğine göre yapılan deneysel çalışmalar sonucunda, RF'ın 66,09, SVM'in 59,43, ARIMA'nın 68,60, MLP'nin 61,08, CNN'in 60,59, GRU'nun 60,67, LSTM'in 61,57, CNN+GRU'nun 57,64 ve HDLM'nin 56,14 ortalama MAE değerine sahip olduğu görülmüştür.



Şekil 6.35. MAE metriğine göre karşılaştırmalı deneysel sonuçlar

$$\text{HDLM'nin } \frac{\text{RF-HDLM}}{\text{RF}} \times 100, \frac{\text{SVR-HDLM}}{\text{SVR}} \times 100, \frac{\text{ARIMA-HDLM}}{\text{ARIMA}} \times 100, \frac{\text{MLP-HDLM}}{\text{MLP}} \times 100, \\ \frac{\text{CNN-HDLM}}{\text{CNN}} \times 100, \frac{\text{GRU-HDLM}}{\text{GRU}} \times 100, \frac{\text{LSTM-HDLM}}{\text{LSTM}} \times 100 \text{ ve } \frac{(\text{CNN+GRU})\text{-HDLM}}{(\text{CNN+GRU})} \times 100$$

eşitlikleri kullanılarak MAE metriğine göre hesaplanan iyileşme oranları Çizelge 6.49'da görülmektedir.

Çizelge 6.49. MAE metriğine göre HDLM'nin iyileşme oranları

	HDLM'nin en kötü değeri	HDLM'nin en iyi değeri	HDLM'nin ortalama değeri
RF'in değerine göre iyileşme oranı (%)	14,87	15,19	15,05
SVM'nin değerine göre iyileşme oranı (%)	5,33	5,68	5,53
ARIMA'nın değerine göre iyileşme oranı (%)	17,98	18,29	18,16
MLP'nin en kötü değerine göre iyileşme oranı (%)	7,89	8,23	8,08
MLP'nin en iyi değerine göre iyileşme oranı (%)	7,89	8,23	8,08
MLP'nin ortalama değerine göre iyileşme oranı (%)	7,89	8,23	8,08
CNN'in en kötü değerine göre iyileşme oranı (%)	7,31	7,66	7,50
CNN'in en iyi değerine göre iyileşme oranı (%)	7,03	7,38	7,23
CNN'in ortalama değerine göre iyileşme oranı (%)	7,16	7,50	7,35
GRU'nun en kötü değerine göre iyileşme oranı (%)	7,28	7,63	7,47
GRU'nun en iyi değerine göre iyileşme oranı (%)	7,26	7,61	7,46

Çizelge 6.49. (devam) MAE metriğine göre HDLM'nin iyileşme oranları

GRU'nun ortalama değerine göre iyileşme oranı (%)	7,27	7,62	7,46
LSTM'in en kötü değerine göre iyileşme oranı (%)	8,62	8,96	8,81
LSTM'in en iyi değerine göre iyileşme oranı (%)	8,62	8,96	8,81
LSTM'in ortalama değerine göre iyileşme oranı (%)	8,62	8,96	8,81
CNN+GRU'nun en kötü değerine göre iyileşme oranı (%)	2,89	3,26	3,10
CNN+GRU'nun en iyi değerine göre iyileşme oranı (%)	2,10	2,47	2,31
CNN+GRU'nun ortalama değerine göre iyileşme oranı (%)	2,40	2,76	2,60

Yapılan 10 farklı deney sonucunda elde edilen MAE değerlerine göre HDLM'nin RF, SVM, ARIMA, MLP, CNN, GRU, LSTM ve CNN+GRU'ya göre karşılaştırmalı bir analizi görülmektedir. HDLM'nin en kötü değeri RF'a göre %14,87, en iyi değeri %15,19 ve ortalama değeri ise %15,05 iyileşme sağlamıştır. HDLM'nin en kötü değeri SVM'e göre %5,33, en iyi değeri %5,68 ve ortalama değeri ise %5,53 iyileşme sağlamıştır. HDLM'nin en kötü değeri ARIMA'ya göre %17,98, en iyi değeri %18,29 ve ortalama değeri ise %18,16 iyileşme sağlamıştır.

HDLM'nin en kötü değeri MLP'nin en kötü değerine göre %7,89, MLP'nin en iyi değerine göre %7,89 ve MLP'nin ortalama değerine göre %7,89 iyileşme sağlamıştır. HDLM'nin en iyi değeri MLP'nin en kötü değerine göre %8,23, MLP'nin en iyi değerine göre %8,23 ve MLP'nin ortalama değerine göre %8,23 iyileşme sağlamıştır. HDLM'nin ortalama değeri MLP'nin en kötü değerine göre %8,08, MLP'nin en iyi değerine göre %8,08 ve MLP'nin ortalama değerine göre %8,08 iyileşme sağlamıştır.

HDLM'nin en kötü değeri CNN'in en kötü değerine göre %7,31, CNN'in en iyi değerine göre %7,03 ve CNN'in ortalama değerine göre %7,16 iyileşme sağlamıştır. HDLM'nin en iyi değeri CNN'in en kötü değerine göre %7,66, CNN'in en iyi değerine göre %7,38 ve CNN'in ortalama değerine göre %7,50 iyileşme sağlamıştır. HDLM'nin ortalama değeri CNN'in en kötü değerine göre %7,50, CNN'in en iyi değerine göre %7,23 ve CNN'in ortalama değerine göre %7,35 iyileşme sağlamıştır.

HDLM'nin en kötü değeri GRU'nun en kötü değerine göre %7,28, GRU'nun en iyi değerine göre %7,26 ve GRU'nun ortalama değerine göre %7,27 iyileşme sağlamıştır. HDLM'nin en iyi değeri GRU'nun en kötü değerine göre %7,63, GRU'nun en iyi değerine göre %7,61 ve GRU'nun ortalama değerine göre %7,62 iyileşme sağlamıştır. HDLM'nin ortalama değeri GRU'nun en kötü değerine göre %7,47, GRU'nun en iyi değerine göre %7,46 ve GRU'nun ortalama değerine göre %7,46 iyileşme sağlamıştır.

HDLM'nin en kötü değeri LSTM'in en kötü değerine göre %8,62, LSTM'in en iyi değerine göre %8,62 ve LSTM'in ortalama değerine göre %8,62 iyileşme sağlamıştır. HDLM'nin en iyi değeri LSTM'in en kötü değerine göre %8,96, LSTM'in en iyi değerine göre %8,96 ve LSTM'in ortalama değerine göre %8,96 iyileşme sağlamıştır. HDLM'nin ortalama değeri LSTM'in en kötü değerine göre %8,81, LSTM'in en iyi değerine göre %8,81 ve LSTM'in ortalama değerine göre %8,81 iyileşme sağlamıştır.

HDLM'nin en kötü değeri CNN+GRU'nun en kötü değerine göre %2,89, CNN+GRU'nun en iyi değerine göre %2,10 ve CNN+GRU'nun ortalama değerine göre %2,40 iyileşme sağlamıştır. HDLM'nin en iyi değeri CNN+GRU'nun en kötü değerine göre %3,26, CNN+GRU'nun en iyi değerine göre %2,47 ve CNN+GRU'nun ortalama değerine göre %2,76 iyileşme sağlamıştır. HDLM'nin ortalama değeri CNN+GRU'nun en kötü değerine göre %3,10, CNN+GRU'nun en iyi değerine göre %2,31 ve CNN+GRU'nun ortalama değerine göre %2,60 iyileşme sağlamıştır.

Deneyisel sonuçlar, HDLM'nin RF, SVM, ARIMA, MLP, CNN, GRU, LSTM ve CNN+GRU'ya göre daha düşük MSE, RMSE ve MAE sonuçlarına sahip olduğunu

göstermektedir. Bir sonraki suç işleme zamanının belirlenmesinde HDLM'nin diğer modellere göre daha başarılı ve etkin sonuçlar verdiği görülmüştür.

6.2.4. Acil çağrı veri kümesi

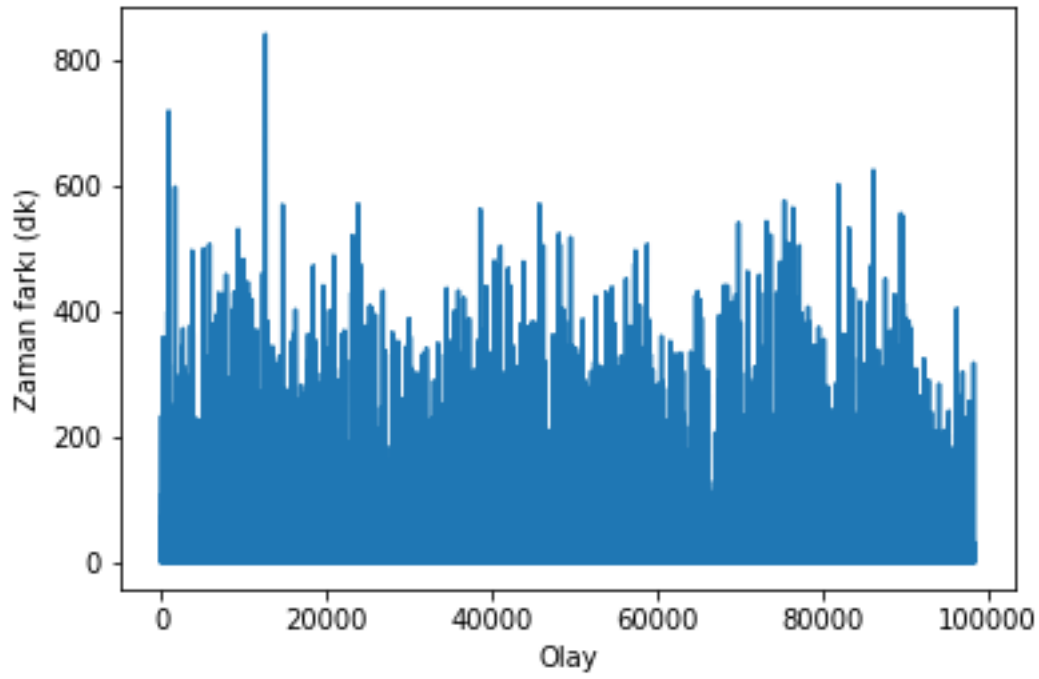
Acil çağrı veri kümesi için trafik kazası olaylarının meydana gelme zamanları arasındaki dakika cinsinden farklara göre oluşturulan verinin ilk 5 satırı Çizelge 6.50'de görülmektedir.

Çizelge 6.50. Bir önceki trafik kazasından sonra geçen zaman (dk)

Olay	Bir önceki trafik kazasından sonra geçen zaman (dk)
1	37
2	2
3	3
4	15
5	2

Çizelge 6.50'de trafik kazası çağrıları arasındaki zaman farkı görülmektedir. Acil çağrı veri kümesi, e-ticaret veri kümesine göre nispeten daha düzenli, DDoS ve suç veri kümesine göre nispeten daha düzensiz bir örüntü izlemektedir.

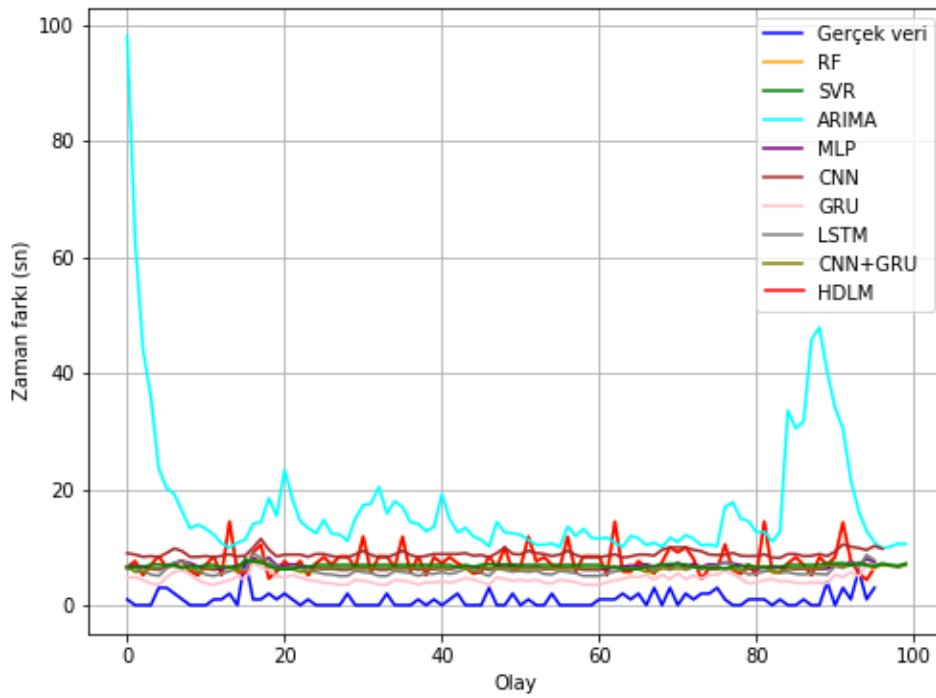
Trafik kazası olayları arasındaki dakika cinsinden zaman farkları Şekil 6.36'da görülmektedir.



Şekil 6.36. Trafik kazası olayları arasındaki dakika cinsinden zaman farkları

Şekil 6.36’da trafik kazası çağrıları arasındaki zaman farkı görülmektedir. Acil çağrı veri kümesi, 98401 trafik kazası çağrısı olayından oluşmaktadır. Veri kümesinde görülen sapmalar mevcuttur.

Şekil 6.37’de örnek olarak test verisinin 100 elemanı için deneysel sonuçlar görülmektedir.



Şekil 6.37. Test verisindeki ilk 100 veri örneği için deneysel sonuçlar

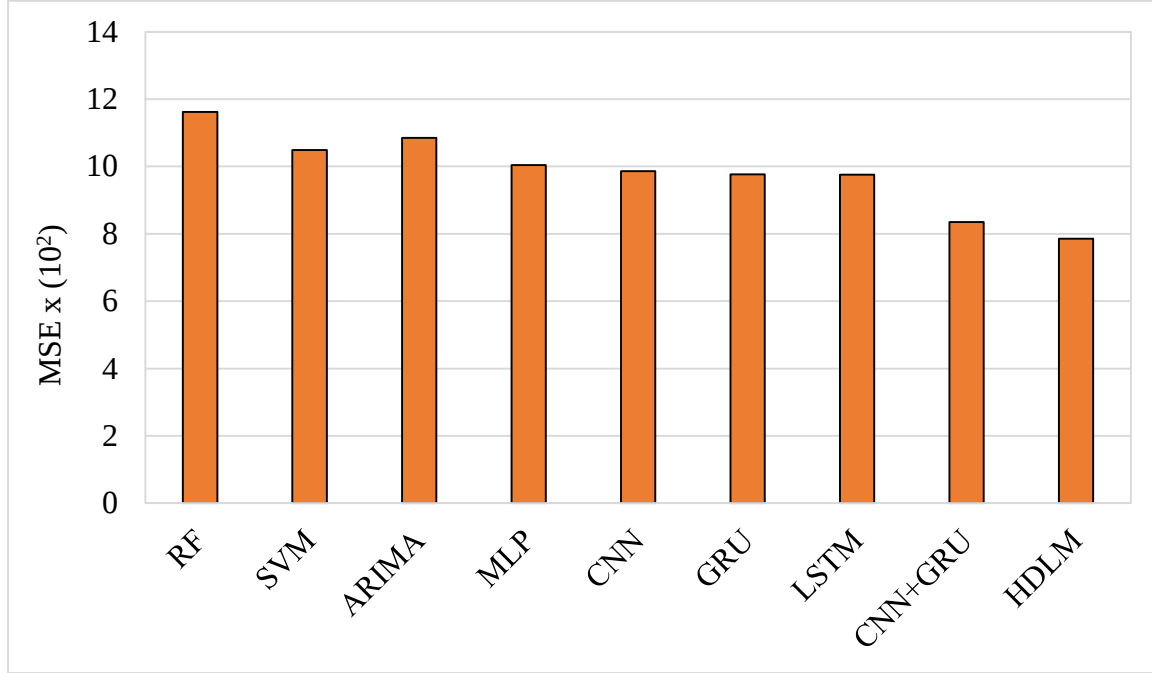
Şekil 6.37’de test verisindeki ilk 100 veri örneği için deneysel sonuçlar görülmektedir. HDLM, diğer modellere kıyasla veri kümesindeki ani sapmaları daha başarılı bir şekilde belirleyerek daha yüksek tahmin doğruluğuna sahip olmuştur.

MSE metriğine göre deneysel sonuçlar Çizelge 6.51’de ve Şekil 6.38’de görülmektedir.

Çizelge 6.51. MSE metriğine göre deneysel sonuçlar

Test	RF	SVM	ARIMA	MLP	CNN	GRU	LSTM	CNN+GRU	HDLM
1	1162,70	1049,27	1085,43	992,86	978,49	973,86	975,66	820,19	781,93
2	1162,70	1049,27	1085,43	998,43	980,39	974,82	976,02	821,52	783,50
3	1162,70	1049,27	1085,43	1001,20	982,50	975,40	976,03	825,56	784,99
4	1162,70	1049,27	1085,43	1001,45	982,98	975,73	976,03	829,49	785,98
5	1162,70	1049,27	1085,43	1001,53	985,23	976,17	976,04	831,08	786,90
6	1162,70	1049,27	1085,43	1001,85	985,70	976,50	976,05	831,89	787,04
7	1162,70	1049,27	1085,43	1009,29	988,42	976,75	976,05	832,78	787,10
8	1162,70	1049,27	1085,43	1013,19	991,53	977,57	976,06	847,45	787,16
9	1162,70	1049,27	1085,43	1013,19	992,69	978,26	976,06	853,28	788,29
10	1162,70	1049,27	1085,43	1015,77	996,90	978,57	976,09	856,84	788,36
Ortalama	1162,70	1049,27	1085,43	1004,87	986,48	976,36	976,00	835,00	786,12

MSE metriğine göre yapılan deneysel çalışmalar sonucunda, RF'in 1162,70, SVM'in 1049,27, ARIMA'nın 1085,43, MLP'nin 1004,87, CNN'in 986,48, GRU'nun 976,36, LSTM'in 976,00, CNN+GRU'nun 835,00 ve HDLM'nin 786,12 ortalama MSE değerine sahip olduğu görülmüştür.



Şekil 6.38. MSE metriğine göre karşılaştırmalı deneysel sonuçlar

$$\text{HDLM'nin } \frac{\text{RF-HDLM}}{\text{RF}} \times 100, \frac{\text{SVR-HDLM}}{\text{SVR}} \times 100, \frac{\text{ARIMA-HDLM}}{\text{ARIMA}} \times 100, \frac{\text{MLP-HDLM}}{\text{MLP}} \times 100, \\ \frac{\text{CNN-HDLM}}{\text{CNN}} \times 100, \frac{\text{GRU-HDLM}}{\text{GRU}} \times 100, \frac{\text{LSTM-HDLM}}{\text{LSTM}} \times 100 \text{ ve } \frac{(\text{CNN+GRU})\text{-HDLM}}{(\text{CNN+GRU})} \times 100$$

eşitlikleri kullanılarak MSE metriğine göre hesaplanan iyileşme oranları Çizelge 6.52'de görülmektedir.

Çizelge 6.52. MSE metriğine göre HDLM'nin iyileşme oranları

	HDLM'nin en kötü değeri	HDLM'nin en iyi değeri	HDLM'nin ortalama değeri
RF'in değerine göre iyileşme oranı (%)	32,19	32,74	32,38
SVM'nin değerine göre iyileşme oranı (%)	24,86	25,47	25,07
ARIMA'nın değerine göre iyileşme oranı (%)	27,36	27,96	27,57
MLP'nin en kötü değerine göre iyileşme oranı (%)	22,38	23,02	22,60
MLP'nin en iyi değerine göre iyileşme oranı (%)	20,59	21,24	20,82
MLP'nin ortalama değerine göre iyileşme oranı (%)	21,80	22,18	21,76
CNN'in en kötü değerine göre iyileşme oranı (%)	20,91	21,56	21,14
CNN'in en iyi değerine göre iyileşme oranı (%)	19,43	20,08	19,65
CNN'in ortalama değerine göre iyileşme oranı (%)	20,24	20,73	20,31
GRU'nun en kötü değerine göre iyileşme oranı (%)	19,43	20,09	19,66
GRU'nun en iyi değerine göre iyileşme oranı (%)	19,04	19,70	19,27
GRU'nun ortalama değerine göre iyileşme oranı (%)	19,30	19,91	19,48
LSTM'in en kötü değerine göre iyileşme oranı (%)	19,23	19,89	19,46
LSTM'in en iyi değerine göre iyileşme oranı (%)	19,19	19,85	19,42

Çizelge 6.52. (devam) MSE metriğine göre HDLM'nin iyileşme oranları

LSTM'in ortalama değerine göre iyileşme oranı (%)	19,23	19,88	19,45
CNN+GRU'nun en kötü değerine göre iyileşme oranı (%)	7,99	8,74	8,25
CNN+GRU'nun en iyi değerine göre iyileşme oranı (%)	3,88	4,66	4,15
CNN+GRU'nun ortalama değerine göre iyileşme oranı (%)	5,68	6,35	5,85

Yapılan 10 farklı deney sonucunda elde edilen MSE değerlerine göre HDLM'nin RF, SVM, ARIMA, MLP, CNN, GRU, LSTM ve CNN+GRU'ya göre karşılaştırmalı bir analizi görülmektedir. HDLM'nin en kötü değeri RF'a göre %32,19, en iyi değeri %32,74 ve ortalama değeri ise %32,38 iyileşme sağlamıştır. HDLM'nin en kötü değeri SVM'e göre %24,86, en iyi değeri %25,47 ve ortalama değeri ise %25,07 iyileşme sağlamıştır. HDLM'nin en kötü değeri ARIMA'ya göre %27,36, en iyi değeri %27,96 ve ortalama değeri ise %27,57 iyileşme sağlamıştır.

HDLM'nin en kötü değeri MLP'nin en kötü değerine göre %22,38, MLP'nin en iyi değerine göre %20,59 ve MLP'nin ortalama değerine göre %21,80 iyileşme sağlamıştır. HDLM'nin en iyi değeri MLP'nin en kötü değerine göre %23,02, MLP'nin en iyi değerine göre %21,24 ve MLP'nin ortalama değerine göre %22,18 iyileşme sağlamıştır. HDLM'nin ortalama değeri MLP'nin en kötü değerine göre %22,60, MLP'nin en iyi değerine göre %20,82 ve MLP'nin ortalama değerine göre %21,76 iyileşme sağlamıştır.

HDLM'nin en kötü değeri CNN'in en kötü değerine göre %20,91, CNN'in en iyi değerine göre %19,43 ve CNN'in ortalama değerine göre %20,24 iyileşme sağlamıştır. HDLM'nin en iyi değeri CNN'in en kötü değerine göre %21,56, CNN'in en iyi değerine göre %20,08 ve CNN'in ortalama değerine göre %20,73 iyileşme sağlamıştır. HDLM'nin ortalama değeri CNN'in en kötü değerine göre %21,14, CNN'in en iyi değerine göre %19,65 ve CNN'in ortalama değerine göre %20,31 iyileşme sağlamıştır.

HDLM'nin en kötü değeri GRU'nun en kötü değerine göre %19,43, GRU'nun en iyi değerine göre %19,04 ve GRU'nun ortalama değerine göre %19,30 iyileşme sağlamıştır. HDLM'nin en iyi değeri GRU'nun en kötü değerine göre %20,09, GRU'nun en iyi değerine göre %19,70 ve GRU'nun ortalama değerine göre %19,91 iyileşme sağlamıştır. HDLM'nin ortalama değeri GRU'nun en kötü değerine göre %19,66, GRU'nun en iyi değerine göre %19,27 ve GRU'nun ortalama değerine göre %19,48 iyileşme sağlamıştır.

HDLM'nin en kötü değeri LSTM'in en kötü değerine göre %19,23, LSTM'in en iyi değerine göre %19,19 ve LSTM'in ortalama değerine göre %19,23 iyileşme sağlamıştır. HDLM'nin en iyi değeri LSTM'in en kötü değerine göre %19,89, LSTM'in en iyi değerine göre %19,85 ve LSTM'in ortalama değerine göre %19,88 iyileşme sağlamıştır. HDLM'nin ortalama değeri LSTM'in en kötü değerine göre %19,46, LSTM'in en iyi değerine göre %19,42 ve LSTM'in ortalama değerine göre %19,45 iyileşme sağlamıştır.

HDLM'nin en kötü değeri CNN+GRU'nun en kötü değerine göre %7,99, CNN+GRU'nun en iyi değerine göre %3,88 ve CNN+GRU'nun ortalama değerine göre %5,68 iyileşme sağlamıştır. HDLM'nin en iyi değeri CNN+GRU'nun en kötü değerine göre %8,74, CNN+GRU'nun en iyi değerine göre %4,66 ve CNN+GRU'nun ortalama değerine göre %6,35 iyileşme sağlamıştır. HDLM'nin ortalama değeri CNN+GRU'nun en kötü değerine göre %8,25, CNN+GRU'nun en iyi değerine göre %4,15 ve CNN+GRU'nun ortalama değerine göre %5,85 iyileşme sağlamıştır.

RMSE metriğine göre deneysel sonuçlar Çizelge 6.53'te ve Şekil 6.39'da görülmektedir.

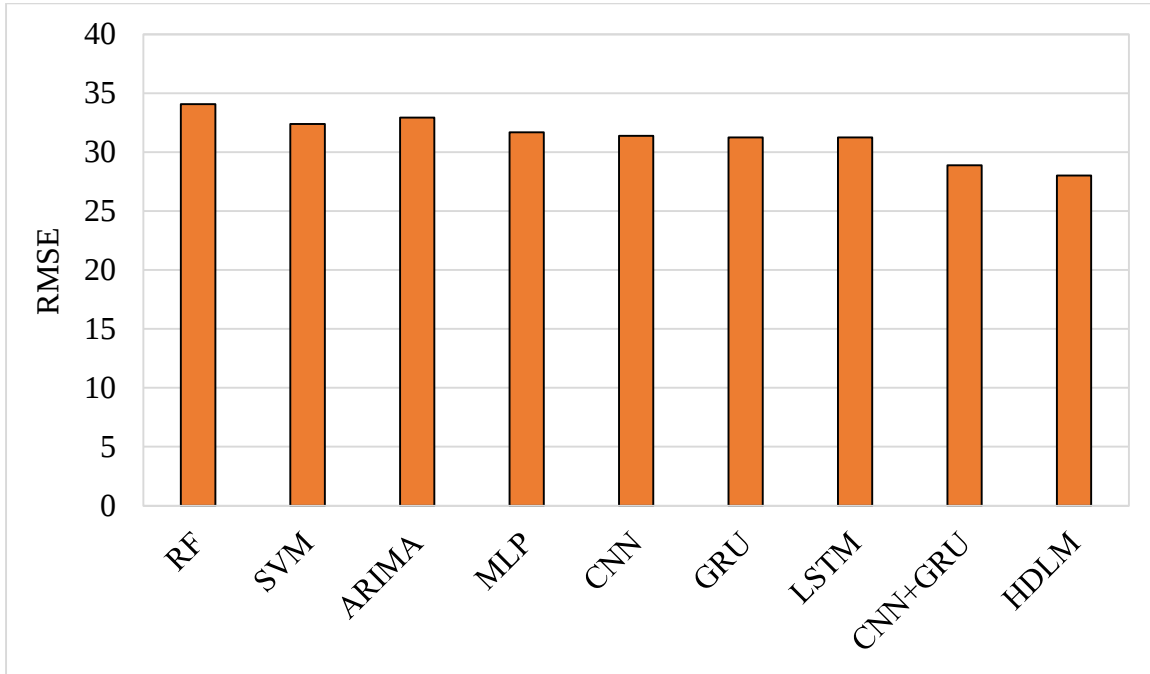
Çizelge 6.53. RMSE metriğine göre deneysel sonuçlar

Test	RF	SVM	ARIMA	MLP	CNN	GRU	LSTM	CNN+GRU	HDLM
1	34,09	32,39	32,94	31,51	31,28	31,20	31,12	28,63	27,96
2	34,09	32,39	32,94	31,59	31,31	31,22	31,15	28,66	27,99
3	34,09	32,39	32,94	31,64	31,34	31,23	31,18	28,73	28,01
4	34,09	32,39	32,94	31,64	31,35	31,23	31,19	28,80	28,03
5	34,09	32,39	32,94	31,64	31,38	31,24	31,23	28,82	28,05
6	34,09	32,39	32,94	31,65	31,39	31,24	31,24	28,84	28,05
7	34,09	32,39	32,94	31,76	31,43	31,25	31,28	28,85	28,05

Çizelge 6.53. (devam) RMSE metriğine göre deneysel sonuçlar

8	34,09	32,39	32,94	31,83	31,48	31,26	31,33	29,11	28,05
9	34,09	32,39	32,94	31,83	31,50	31,27	31,35	29,21	28,07
10	34,09	32,39	32,94	31,87	31,57	31,28	31,41	29,27	28,07
Ortalama	34,09	32,39	32,94	31,69	31,40	31,24	31,24	28,89	28,03

RMSE metriğine göre yapılan deneysel çalışmalar sonucunda, RF'ın 34,09, SVM'in 32,39, ARIMA'nın 32,94, MLP'nin 31,69, CNN'in 31,40, GRU'nun 31,24, LSTM'in 31,24, CNN+GRU'nun 28,89 ve HDLM'nin 28,03 ortalama RMSE değerine sahip olduğu görülmüştür.



Şekil 6.39. RMSE metriğine göre karşılaştırmalı deneysel sonuçlar

$$\text{HDLM'nin } \frac{\text{RF-HDLM}}{\text{RF}} \times 100, \frac{\text{SVR-HDLM}}{\text{SVR}} \times 100, \frac{\text{ARIMA-HDLM}}{\text{ARIMA}} \times 100, \frac{\text{MLP-HDLM}}{\text{MLP}} \times 100, \\ \frac{\text{CNN-HDLM}}{\text{CNN}} \times 100, \frac{\text{GRU-HDLM}}{\text{GRU}} \times 100, \frac{\text{LSTM-HDLM}}{\text{LSTM}} \times 100 \text{ ve } \frac{(\text{CNN+GRU})\text{-HDLM}}{(\text{CNN+GRU})} \times 100$$

eşitlikleri kullanılarak RMSE metriğine göre hesaplanan iyileşme oranları Çizelge 6.54'te görülmektedir.

Çizelge 6.54. RMSE metriğine göre HDLM'nin iyileşme oranları

	HDLM'nin en kötü değeri	HDLM'nin en iyi değeri	HDLM'nin ortalama değeri
RF'in değerine göre iyileşme oranı (%)	17,65	17,98	17,76
SVM'nin değerine göre iyileşme oranı (%)	13,33	13,67	13,45
ARIMA'nın değerine göre iyileşme oranı (%)	14,78	15,11	14,89
MLP'nin en kötü değerine göre iyileşme oranı (%)	11,92	12,26	12,03
MLP'nin en iyi değerine göre iyileşme oranı (%)	10,91	11,26	11,03
MLP'nin ortalama değerine göre iyileşme oranı (%)	11,50	11,78	11,55
CNN'in en kötü değerine göre iyileşme oranı (%)	11,08	11,43	11,20
CNN'in en iyi değerine göre iyileşme oranı (%)	10,26	10,61	10,38
CNN'in ortalama değerine göre iyileşme oranı (%)	10,65	10,96	10,73
GRU'nun en kötü değerine göre iyileşme oranı (%)	10,26	10,61	10,38
GRU'nun en iyi değerine göre iyileşme oranı (%)	10,03	10,38	10,15
GRU'nun ortalama değerine göre iyileşme oranı (%)	10,16	10,50	10,27
LSTM'in en kötü değerine göre iyileşme oranı (%)	10,63	10,98	10,75
LSTM'in en iyi değerine göre iyileşme oranı (%)	9,80	10,15	9,91

Çizelge 6.54. (devam) RMSE metriğine göre HDLM'nin iyileşme oranları

LSTM'in ortalama değerine göre iyileşme oranı (%)	10,21	10,52	10,28
CNN+GRU'nun en kötü değerine göre iyileşme oranı (%)	4,09	4,47	4,22
CNN+GRU'nun en iyi değerine göre iyileşme oranı (%)	1,95	2,34	2,08
CNN+GRU'nun ortalama değerine göre iyileşme oranı (%)	2,87	3,22	2,97

Yapılan 10 farklı deney sonucunda elde edilen RMSE değerlerine göre HDLM'nin RF, SVM, ARIMA, MLP, CNN, GRU, LSTM ve CNN+GRU'ya göre karşılaştırmalı bir analizi görülmektedir. HDLM'nin en kötü değeri RF'a göre %17,65, en iyi değeri %17,98 ve ortalama değeri ise %17,76 iyileşme sağlamıştır. HDLM'nin en kötü değeri SVM'e göre %13,33, en iyi değeri %13,67 ve ortalama değeri ise %13,45 iyileşme sağlamıştır. HDLM'nin en kötü değeri ARIMA'ya göre %14,78, en iyi değeri %15,11 ve ortalama değeri ise %14,89 iyileşme sağlamıştır.

HDLM'nin en kötü değeri MLP'nin en kötü değerine göre %11,92, MLP'nin en iyi değerine göre %10,91 ve MLP'nin ortalama değerine göre %11,50 iyileşme sağlamıştır. HDLM'nin en iyi değeri MLP'nin en kötü değerine göre %12,26, MLP'nin en iyi değerine göre %11,26 ve MLP'nin ortalama değerine göre %11,78 iyileşme sağlamıştır. HDLM'nin ortalama değeri MLP'nin en kötü değerine göre %12,03, MLP'nin en iyi değerine göre %11,03 ve MLP'nin ortalama değerine göre %11,55 iyileşme sağlamıştır.

HDLM'nin en kötü değeri CNN'in en kötü değerine göre %11,08, CNN'in en iyi değerine göre %10,26 ve CNN'in ortalama değerine göre %10,65 iyileşme sağlamıştır. HDLM'nin en iyi değeri CNN'in en kötü değerine göre %11,43, CNN'in en iyi değerine göre %10,61 ve CNN'in ortalama değerine göre %10,96 iyileşme sağlamıştır. HDLM'nin ortalama değeri CNN'in en kötü değerine göre %11,20, CNN'in en iyi değerine göre %10,38 ve CNN'in ortalama değerine göre %10,73 iyileşme sağlamıştır.

HDLM'nin en kötü değeri GRU'nun en kötü değerine göre %10,26, GRU'nun en iyi değerine göre %10,03 ve GRU'nun ortalama değerine göre %10,16 iyileşme sağlamıştır. HDLM'nin en iyi değeri GRU'nun en kötü değerine göre %10,61, GRU'nun en iyi değerine göre %10,38 ve GRU'nun ortalama değerine göre %10,50 iyileşme sağlamıştır. HDLM'nin ortalama değeri GRU'nun en kötü değerine göre %10,38, GRU'nun en iyi değerine göre %10,15 ve GRU'nun ortalama değerine göre %10,27 iyileşme sağlamıştır.

HDLM'nin en kötü değeri LSTM'in en kötü değerine göre %10,63, LSTM'in en iyi değerine göre %9,80 ve LSTM'in ortalama değerine göre %10,21 iyileşme sağlamıştır. HDLM'nin en iyi değeri LSTM'in en kötü değerine göre %10,98, LSTM'in en iyi değerine göre %10,15 ve LSTM'in ortalama değerine göre %10,52 iyileşme sağlamıştır. HDLM'nin ortalama değeri LSTM'in en kötü değerine göre %10,75, LSTM'in en iyi değerine göre %9,91 ve LSTM'in ortalama değerine göre %10,28 iyileşme sağlamıştır.

HDLM'nin en kötü değeri CNN+GRU'nun en kötü değerine göre %4,09, CNN+GRU'nun en iyi değerine göre %1,95 ve CNN+GRU'nun ortalama değerine göre %2,87 iyileşme sağlamıştır. HDLM'nin en iyi değeri CNN+GRU'nun en kötü değerine göre %4,47, CNN+GRU'nun en iyi değerine göre %2,34 ve CNN+GRU'nun ortalama değerine göre %3,22 iyileşme sağlamıştır. HDLM'nin ortalama değeri CNN+GRU'nun en kötü değerine göre %4,22, CNN+GRU'nun en iyi değerine göre %2,08 ve CNN+GRU'nun ortalama değerine göre %2,97 iyileşme sağlamıştır.

MAE metriğine göre deneysel sonuçlar Çizelge 6.55'te ve Şekil 6.40'ta görülmektedir.

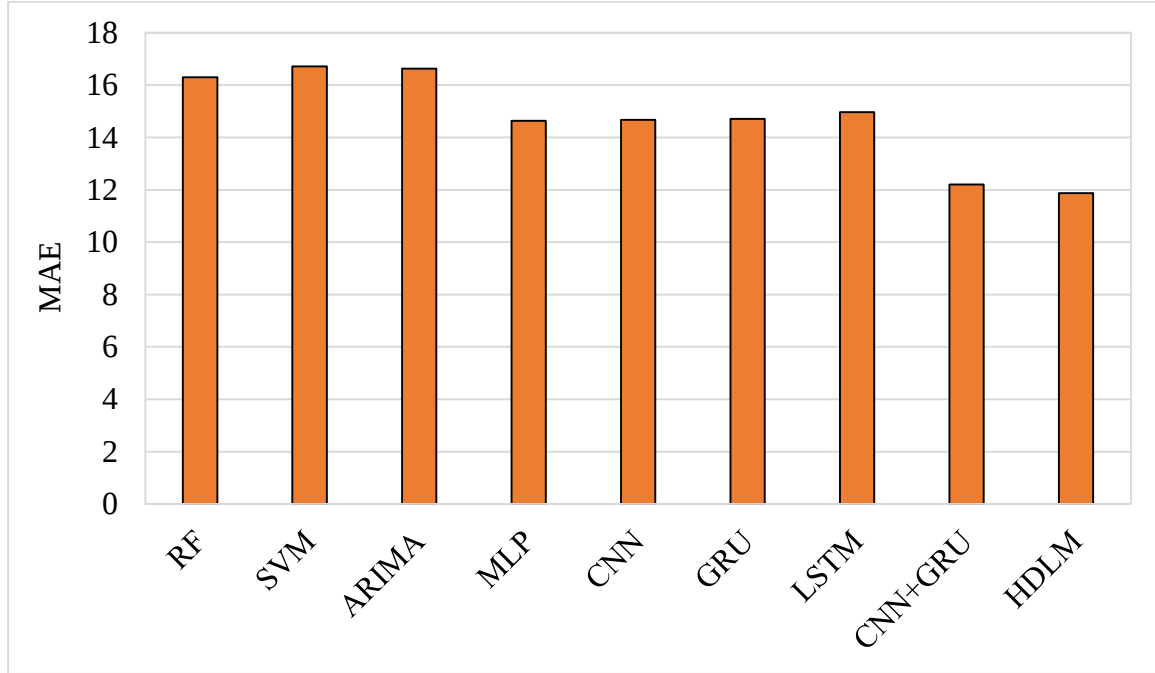
Çizelge 6.55. MAE metriğine göre deneysel sonuçlar

Test	RF	SVM	ARIMA	MLP	CNN	GRU	LSTM	CNN+GRU	HDLM
1	16,30	16,71	16,63	14,60	14,51	14,70	13,85	12,07	11,80
2	16,30	16,71	16,63	14,60	14,54	14,71	14,28	12,08	11,81
3	16,30	16,71	16,63	14,61	14,54	14,71	14,46	12,08	11,82
4	16,30	16,71	16,63	14,61	14,67	14,71	14,47	12,08	11,88
5	16,30	16,71	16,63	14,61	14,68	14,71	14,57	12,09	11,89
6	16,30	16,71	16,63	14,61	14,68	14,71	15,25	12,09	11,89

Çizelge 6.55. (devam) MAE metriğine göre deneysel sonuçlar

7	16,30	16,71	16,63	14,61	14,71	14,71	15,43	12,10	11,89
8	16,30	16,71	16,63	14,69	14,72	14,72	15,56	12,44	11,91
9	16,30	16,71	16,63	14,74	14,79	14,72	15,69	12,45	11,92
10	16,30	16,71	16,63	14,74	14,87	14,72	16,06	12,55	11,94
Ortalama	16,30	16,71	16,63	14,64	14,67	14,71	14,96	12,20	11,87

MAE metriğine göre yapılan deneysel çalışmalar sonucunda, RF'ın 16,30, SVM'in 16,71, ARIMA'nın 16,63, MLP'nin 14,64, CNN'in 14,67, GRU'nun 14,71, LSTM'in 14,96, CNN+GRU'nun 12,20 ve HDLM'nin 11,87 ortalama MAE değerine sahip olduğu görülmüştür.



Şekil 6.40. MAE metriğine göre karşılaştırmalı deneysel sonuçlar

$$\text{HDLM'nin } \frac{\text{RF-HDLM}}{\text{RF}} \times 100, \frac{\text{SVR-HDLM}}{\text{SVR}} \times 100, \frac{\text{ARIMA-HDLM}}{\text{ARIMA}} \times 100, \frac{\text{MLP-HDLM}}{\text{MLP}} \times 100, \\ \frac{\text{CNN-HDLM}}{\text{CNN}} \times 100, \frac{\text{GRU-HDLM}}{\text{GRU}} \times 100, \frac{\text{LSTM-HDLM}}{\text{LSTM}} \times 100 \text{ ve } \frac{(\text{CNN+GRU})\text{-HDLM}}{(\text{CNN+GRU})} \times 100$$

eşitlikleri kullanılarak MAE metriğine göre hesaplanan iyileşme oranları Çizelge 6.56'da görülmektedir.

Çizelge 6.56. MAE metriğine göre HDLM'nin iyileşme oranları

	HDLM'nin en kötü değeri	HDLM'nin en iyi değeri	HDLM'nin ortalama değeri
RF'ın değerine göre iyileşme oranı (%)	26,74	27,60	27,14
SVM'nin değerine göre iyileşme oranı (%)	28,54	29,38	28,93
ARIMA'nın değerine göre iyileşme oranı (%)	28,20	29,04	28,59
MLP'nin en kötü değerine göre iyileşme oranı (%)	18,99	19,94	19,43
MLP'nin en iyi değerine göre iyileşme oranı (%)	18,21	19,17	18,66
MLP'nin ortalama değerine göre iyileşme oranı (%)	18,50	19,40	18,89
CNN'in en kötü değerine göre iyileşme oranı (%)	19,70	20,64	20,14
CNN'in en iyi değerine göre iyileşme oranı (%)	17,71	18,67	18,15
CNN'in ortalama değerine göre iyileşme oranı (%)	18,82	19,56	19,05
GRU'nun en kötü değerine göre iyileşme oranı (%)	18,88	19,83	19,32
GRU'nun en iyi değerine göre iyileşme oranı (%)	18,77	19,72	19,21
GRU'nun ortalama değerine göre iyileşme oranı (%)	18,85	19,79	19,28
LSTM'in en kötü değerine göre iyileşme oranı (%)	25,65	26,52	26,05
LSTM'in en iyi değerine göre iyileşme oranı (%)	13,79	14,80	14,25

Çizelge 6.56. (devam) MAE metriğine göre HDLM'nin iyileşme oranları

LSTM'in ortalama değerine göre iyileşme oranı (%)	21,81	21,13	20,63
CNN+GRU'nun en kötü değerine göre iyileşme oranı (%)	4,86	5,97	5,37
CNN+GRU'nun en iyi değerine göre iyileşme oranı (%)	1,07	2,23	1,61
CNN+GRU'nun ortalama değerine göre iyileşme oranı (%)	2,17	3,30	2,68

Yapılan 10 farklı deney sonucunda elde edilen MAE değerlerine göre HDLM'nin RF, SVM, ARIMA, MLP, CNN, GRU, LSTM ve CNN+GRU'ya göre karşılaştırmalı bir analizi görülmektedir. HDLM'nin en kötü değeri RF'a göre %26,74, en iyi değeri %27,60 ve ortalama değeri ise %27,14 iyileşme sağlamıştır. HDLM'nin en kötü değeri SVM'e göre %28,54, en iyi değeri %29,38 ve ortalama değeri ise %28,93 iyileşme sağlamıştır. HDLM'nin en kötü değeri ARIMA'ya göre %28,20, en iyi değeri %29,04 ve ortalama değeri ise %28,59 iyileşme sağlamıştır.

HDLM'nin en kötü değeri MLP'nin en kötü değerine göre %18,99, MLP'nin en iyi değerine göre %18,21 ve MLP'nin ortalama değerine göre %18,50 iyileşme sağlamıştır. HDLM'nin en iyi değeri MLP'nin en kötü değerine göre %19,94, MLP'nin en iyi değerine göre %19,17 ve MLP'nin ortalama değerine göre %19,40 iyileşme sağlamıştır. HDLM'nin ortalama değeri MLP'nin en kötü değerine göre %19,43, MLP'nin en iyi değerine göre %18,66 ve MLP'nin ortalama değerine göre %18,89 iyileşme sağlamıştır.

HDLM'nin en kötü değeri CNN'in en kötü değerine göre %19,70, CNN'in en iyi değerine göre %17,71 ve CNN'in ortalama değerine göre %18,82 iyileşme sağlamıştır. HDLM'nin en iyi değeri CNN'in en kötü değerine göre %20,64, CNN'in en iyi değerine göre %18,67 ve CNN'in ortalama değerine göre %19,56 iyileşme sağlamıştır. HDLM'nin ortalama değeri CNN'in en kötü değerine göre %20,14, CNN'in en iyi değerine göre %18,15 ve CNN'in ortalama değerine göre %19,05 iyileşme sağlamıştır.

HDLM'nin en kötü değeri GRU'nun en kötü değerine göre %18,88, GRU'nun en iyi değerine göre %18,77 ve GRU'nun ortalama değerine göre %18,85 iyileşme sağlamıştır. HDLM'nin en iyi değeri GRU'nun en kötü değerine göre %19,83, GRU'nun en iyi değerine göre %19,72 ve GRU'nun ortalama değerine göre %19,79 iyileşme sağlamıştır. HDLM'nin ortalama değeri GRU'nun en kötü değerine göre %19,32, GRU'nun en iyi değerine göre %19,21 ve GRU'nun ortalama değerine göre %19,28 iyileşme sağlamıştır.

HDLM'nin en kötü değeri LSTM'in en kötü değerine göre %25,65, LSTM'in en iyi değerine göre %13,79 ve LSTM'in ortalama değerine göre %21,81 iyileşme sağlamıştır. HDLM'nin en iyi değeri LSTM'in en kötü değerine göre %26,52, LSTM'in en iyi değerine göre %14,80 ve LSTM'in ortalama değerine göre %21,13 iyileşme sağlamıştır. HDLM'nin ortalama değeri LSTM'in en kötü değerine göre %26,05, LSTM'in en iyi değerine göre %14,25 ve LSTM'in ortalama değerine göre %20,63 iyileşme sağlamıştır.

HDLM'nin en kötü değeri CNN+GRU'nun en kötü değerine göre %4,86, CNN+GRU'nun en iyi değerine göre %1,07 ve CNN+GRU'nun ortalama değerine göre %2,17 iyileşme sağlamıştır. HDLM'nin en iyi değeri CNN+GRU'nun en kötü değerine göre %5,97, CNN+GRU'nun en iyi değerine göre %2,23 ve CNN+GRU'nun ortalama değerine göre %3,30 iyileşme sağlamıştır. HDLM'nin ortalama değeri CNN+GRU'nun en kötü değerine göre %5,37, CNN+GRU'nun en iyi değerine göre %1,61 ve CNN+GRU'nun ortalama değerine göre %2,68 iyileşme sağlamıştır.

Deneyisel sonuçlar, HDLM'nin RF, SVM, ARIMA, MLP, CNN, GRU, LSTM ve CNN+GRU'ya göre daha düşük MSE, RMSE ve MAE sonuçlarına sahip olduğunu göstermektedir. Bir sonraki acil çağrı zamanının belirlenmesinde HDLM'nin diğer modellere göre daha başarılı ve etkin sonuçlar verdiği görülmüştür.

7. SONUÇLAR

Bu tez çalışmasında, ileriye dönük bir olayın olma zamanının tahmini için CNN ve LSTM'i birlikte kullanarak hibrit bir derin öğrenme mimarisi geliştirilmiştir. HDLM'de, CNN özellik çıkarımı için, LSTM ise öğrenme ve tahmin için kullanılmıştır. HDLM, e-ticaret kullanıcı davranış verileri, DDoS saldırı verileri, suç verileri ve acil çağrı verileri olmak üzere 4 farklı veri kümesine uygulanarak ileriye dönük zaman tahmini yapmak hedeflenmiştir. Olaylar arasındaki zaman farkları saniye ve dakika cinsinden hesaplanmıştır. Her iki senaryo için HDLM, RF, SVM, ARIMA, MLP, CNN, GRU, LSTM ve CNN+GRU ile uygulamalı olarak karşılaştırılmıştır. Olaylar arasındaki zaman farkları, saniye cinsinden ve dakika cinsinden alınarak karşılaştırmalar yapılmıştır.

Olaylar arasındaki zaman farklarının saniye cinsinden hesaplandığı senaryoda e-ticaret veri kümesi için, ortalama MSE değerlerine göre HDLM'nin RF'a göre %30,77, SVM'e göre %28,45, ARIMA'ya göre %29,19, MLP'ye göre %27,41, CNN'e göre %28,58, GRU'ya göre %22,15, LSTM'e göre %18,26 ve CNN+GRU'ya göre %10,47 iyileşme sağlayarak daha başarılı olduğu görülmüştür. Ortalama RMSE değerlerine göre HDLM'nin RF'a göre %16,80, SVM'e göre %15,42, ARIMA'ya göre %15,85, MLP'ye göre %14,80, CNN'e göre %15,49, GRU'ya göre %11,77, LSTM'e göre %9,59 ve CNN+GRU'ya göre %5,38 iyileşme sağlayarak daha başarılı olduğu görülmüştür. Ortalama MAE değerlerine göre ise HDLM'nin RF'a göre %25,97, SVM'e göre %11,01, ARIMA'ya göre %25,27, MLP'ye göre %24,44, CNN'e göre %24,77, GRU'ya göre %25,31, LSTM'e göre %23,53 ve CNN+GRU'ya göre %4,13 iyileşme sağlayarak daha başarılı olduğu görülmüştür.

DDoS saldırı veri kümesi için, ortalama MSE değerlerine göre HDLM'nin RF'a göre %29,25, SVM'e göre %30,63, ARIMA'ya göre %26,32, MLP'ye göre %27,74, CNN'e göre %26,91, GRU'ya göre %25,60, LSTM'e göre %25,27 ve CNN+GRU'ya göre %10,06 iyileşme sağlayarak daha başarılı olduğu görülmüştür. Ortalama RMSE değerlerine göre HDLM'nin RF'a göre %15,89, SVM'e göre %16,71, ARIMA'ya göre %14,16, MLP'ye göre %15,00, CNN'e göre %14,51, GRU'ya göre %13,74, LSTM'e göre %13,55 ve CNN+GRU'ya göre %5,16 iyileşme sağlayarak daha başarılı olduğu görülmüştür. Ortalama MAE değerlerine göre ise HDLM'nin RF'a göre %22,06, SVM'e göre %16,24, ARIMA'ya

göre %15,80, MLP'ye göre %22,05, CNN'e göre %19,20, GRU'ya göre %16,59, LSTM'e göre %14,61 ve CNN+GRU'ya göre %3,46 iyileşme sağlayarak daha başarılı olduğu görülmüştür.

Suç veri kümesi için, ortalama MSE değerlerine göre HDLM'nin RF'a göre %27,22, SVM'e göre %29,66, ARIMA'ya göre %19,04, MLP'ye göre %19,71, CNN'e göre %22,67, GRU'ya göre %18,53, LSTM'e göre %18,07 ve CNN+GRU'ya göre %13,69 iyileşme sağlayarak daha başarılı olduğu görülmüştür. Ortalama RMSE değerlerine göre HDLM'nin RF'a göre %14,69, SVM'e göre %16,13, ARIMA'ya göre %10,02, MLP'ye göre %10,40, CNN'e göre %12,06, GRU'ya göre %9,74, LSTM'e göre %9,49 ve CNN+GRU'ya göre %7,09 iyileşme sağlayarak daha başarılı olduğu görülmüştür. Ortalama MAE değerlerine göre ise HDLM'nin RF'a göre %18,02, SVM'e göre %9,12, ARIMA'ya göre %21,14, MLP'ye göre %15,96, CNN'e göre %10,34, GRU'ya göre %13,00, LSTM'e göre %15,07 ve CNN+GRU'ya göre %4,57 iyileşme sağlayarak daha başarılı olduğu görülmüştür.

Acil çağrı veri kümesi için, ortalama MSE değerlerine göre HDLM'nin RF'a göre %47,09, SVM'e göre %44,70, ARIMA'ya göre %43,85, MLP'ye göre %41,75, CNN'e göre %43,67, GRU'ya göre %41,67, LSTM'e göre %41,56 ve CNN+GRU'ya göre %3,79 iyileşme sağlayarak daha başarılı olduğu görülmüştür. Ortalama RMSE değerlerine göre HDLM'nin RF'a göre %27,26, SVM'e göre %25,64, ARIMA'ya göre %25,07, MLP'ye göre %23,68, CNN'e göre %24,95, GRU'ya göre %23,63, LSTM'e göre %23,55 ve CNN+GRU'ya göre %1,91 iyileşme sağlayarak daha başarılı olduğu görülmüştür. Ortalama MAE değerlerine göre ise HDLM'nin RF'a göre %34,51, SVM'e göre %24,72, ARIMA'ya göre %30,50, MLP'ye göre %30,88, CNN'e göre %29,53, GRU'ya göre %29,36, LSTM'e göre %27,43 ve CNN+GRU'ya göre %3,96 iyileşme sağlayarak daha başarılı olduğu görülmüştür.

Olaylar arasındaki zaman farklarının dakika cinsinden hesaplandığı senaryoda e-ticaret veri kümesi için, ortalama MSE sonuçlarına göre HDLM'nin RF'a göre %55,89, SVM'e göre %51,06, ARIMA'ya göre %53,45, MLP'ye göre %46,06, CNN'e göre %47,73, GRU'ya göre %45,17, LSTM'e göre %27,37 ve CNN+GRU'ya göre %19,50 iyileşme sağlayarak daha başarılı olduğu görülmüştür. Ortalama RMSE sonuçlarına göre HDLM'nin RF'a göre %33,60, SVM'e göre %30,08, ARIMA'ya göre %31,81, MLP'ye göre %26,59, CNN'e göre

%27,72, GRU'ya göre %25,99, LSTM'e göre %14,80 ve CNN+GRU'ya göre %10,31 iyileşme sağlayarak daha başarılı olduğu görülmüştür. Ortalama MAE sonuçlarına göre ise HDLM'nin RF'a göre %38,65, SVM'e göre %43,10, ARIMA'ya göre %40,86, MLP'ye göre %33,19, CNN'e göre %24,72, GRU'ya göre %32,79, LSTM'e göre %14,77 ve CNN+GRU'ya göre %10,94 iyileşme sağlayarak daha başarılı olduğu görülmüştür.

DDoS saldırı veri kümesi için, ortalama MSE sonuçlarına göre HDLM'nin RF'a göre %23,58, SVM'e göre %22,51, ARIMA'ya göre %21,08, MLP'ye göre %21,81, CNN'e göre %21,00, GRU'ya göre %17,52, LSTM'e göre %15,62 ve CNN+GRU'ya göre %6,95 iyileşme sağlayarak daha başarılı olduğu görülmüştür. Ortalama RMSE sonuçlarına göre HDLM'nin RF'a göre %12,57, SVM'e göre %11,97, ARIMA'ya göre %11,14, MLP'ye göre %11,58, CNN'e göre %11,16, GRU'ya göre %9,20, LSTM'e göre %8,15 ve CNN+GRU'ya göre %3,54 iyileşme sağlayarak daha başarılı olduğu görülmüştür. Ortalama MAE sonuçlarına göre ise HDLM'nin RF'a göre %21,07, SVM'e göre %13,00, ARIMA'ya göre %12,46, MLP'ye göre %12,23, CNN'e göre %11,72, GRU'ya göre %13,88, LSTM'e göre %13,14 ve CNN+GRU'ya göre %2,84 iyileşme sağlayarak daha başarılı olduğu görülmüştür.

Suç veri kümesi için, ortalama MSE sonuçlarına göre HDLM'nin RF'a göre %20,47, SVM'e göre %15,36, ARIMA'ya göre %12,77, MLP'ye göre %11,86, CNN'e göre %12,62, GRU'ya göre %11,57, LSTM'e göre %10,63 ve CNN+GRU'ya göre %4,82 iyileşme sağlayarak daha başarılı olduğu görülmüştür. Ortalama RMSE sonuçlarına göre HDLM'nin RF'a göre %10,82, SVM'e göre %8,00, ARIMA'ya göre %6,60, MLP'ye göre %6,11, CNN'e göre %6,52, GRU'ya göre %5,97, LSTM'e göre %5,46 ve CNN+GRU'ya göre %2,43 iyileşme sağlayarak daha başarılı olduğu görülmüştür. Ortalama MAE sonuçlarına göre ise HDLM'nin RF'a göre %15,05, SVM'e göre %5,53, ARIMA'ya göre %18,16, MLP'ye göre %8,08, CNN'e göre %7,35, GRU'ya göre %7,46, LSTM'e göre %8,81 ve CNN+GRU'ya göre %2,60 iyileşme sağlayarak daha başarılı olduğu görülmüştür.

Acil çağrı veri kümesi için, ortalama MSE sonuçlarına göre HDLM'nin RF'a göre %32,38, SVM'e göre %25,07, ARIMA'ya göre %27,57, MLP'ye göre %21,76, CNN'e göre %20,31, GRU'ya göre %19,48, LSTM'e göre %19,45 ve CNN+GRU'ya göre %5,85 iyileşme

sağlayarak daha başarılı olduğu görülmüştür. Ortalama RMSE sonuçlarına göre HDLM'nin RF'a göre %17,76, SVM'e göre %13,45, ARIMA'ya göre %14,89, MLP'ye göre %11,55, CNN'e göre %10,73, GRU'ya göre %10,27, LSTM'e göre %10,28 ve CNN+GRU'ya göre %2,97 iyileşme sağlayarak daha başarılı olduğu görülmüştür. Ortalama MAE sonuçlarına göre ise HDLM'nin RF'a göre %27,14, SVM'e göre %28,93, ARIMA'ya göre %28,59, MLP'ye göre %18,89, CNN'e göre %19,05, GRU'ya göre %19,28, LSTM'e göre %20,63 ve CNN+GRU'ya göre %2,68 iyileşme sağlayarak daha başarılı olduğu görülmüştür.

Saniye ve dakika cinsinde yapılan deneysel çalışmalardan elde edilen sonuçların birbiriyle tutarlı olduğu görülmüştür. CNN ve LSTM mimarileri birleştirilerek oluşturulan hibrit derin öğrenme mimarisinin, olayların bir sonraki olma zamanının önceden tahmin edilmesinde literatürdeki benzerlerine göre daha başarılı olduğu deneysel çalışmalardan görülmüştür.

Sonraki çalışmalarda ikiden fazla mimarinin birleşimiyle hibrit mimariler geliştirilerek tahmin doğruluğunun artırılmasına yönelik çalışmalar yapılabilir.

KAYNAKLAR

1. Ricci, F., Rokach, L. and Shapira, B. (2011). *Introduction to recommender systems handbook*. ABD: Springer, 1-35.
2. Cheng, H. T., Koc, L., Harmsen, J., Shaked, T., Chandra, T., Aradhye, H. and Anil, R. (2016). *Wide & deep learning for recommender systems*. In Proceedings of the 1st workshop on deep learning for recommender systems, 7-10, New York, ABD.
3. Aggarwal, C. C. (2016). *Recommender systems*. ABD: Springer, 29-44.
4. Zhang, S., Yao, L., Sun, A. and Tay, Y. (2019). Deep learning based recommender system: A survey and new perspectives. *ACM Computing Surveys (CSUR)*, 52(1), 1-38.
5. Singhal, A., Sinha, P. and Pant, R. (2017). Use of deep learning in modern recommendation system: A summary of recent works. *arXiv preprint arXiv:1712.07525*.
6. Musto, C., Greco, C., Suglia, A. and Semeraro, G. (2016). *Ask Me Any Rating: A Content-based Recommender System based on Recurrent Neural Networks*. IIR 2016 7th Italian Information Retrieval Workshop, 7–10, Venezia, Italy.
7. Martínez-Álvarez, F., Troncoso, A., Asencio-Cortés, G. and Riquelme, J. C. (2015). A survey on data mining techniques applied to electricity-related time series forecasting. *Energies*, 8(11), 13162-13193.
8. Granger, C. W. J. and Newbold, P. (2014). *Forecasting economic time series*. ABD: Academic Press, 1-44.
9. Adhikari, R. and Agrawal, R. K. (2013). *An introductory study on time series modeling and forecasting*. Almany: Lambert Academic Publishing, 12-31.
10. Montgomery, D. C., Jennings, C. L. and Kulahci, M. (2015). *Introduction to time series analysis and forecasting*. ABD: John Wiley & Sons, 231-299.
11. Chatfield, C. (2016). *The analysis of time series: An introduction*. ABD: CRC press, 33-59.
12. Box, G. E., Jenkins, G. M., Reinsel, G. C. and Ljung, G. M. (2015). *Time series analysis: forecasting and control*. ABD: John Wiley & Sons, 19-47.
13. Siami-Namini, S. and Namin, A. S. (2018). Forecasting Economics and Financial Time Series: ARIMA vs. LSTM. *arXiv Learning*, 1-19.
14. Claveria, O. and Torra, S. (2014). Forecasting tourism demand to Catalonia: Neural networks vs. time series models. *Economic Modelling*, 36, 220-228.

15. Chujai, P., Kerdprasop, N. and Kerdprasop, K. (2013). *Time series analysis of household electric consumption with ARIMA and ARMA models*. Proceedings of the International MultiConference of Engineers and Computer Scientists, 1, 295-300, Hong Kong, Çin.
16. Radziukynas, V. and Klementavicius, A. (2014). *Short-term wind speed forecasting with ARIMA model*. Power and Electrical Engineering of Riga Technical University (RTUCON), 2014 55th International Scientific Conference, 145-149, Riga, Letonya.
17. Kane, M. J., Price, N., Scotch, M. and Rabinowitz, P. (2014). Comparison of ARIMA and Random Forest time series models for prediction of avian influenza H5N1 outbreaks. *BMC bioinformatics*, 15(1), 276.
18. Fu, R., Zhang, Z. and Li, L. (2016). *Using LSTM and GRU neural network methods for traffic flow prediction*. Chinese Association of Automation (YAC), Youth Academic Annual Conference, 324-328, Wuhan, Çin.
19. Amini, M. H., Kargarian, A. and Karabasoglu, O. (2016). ARIMA-based decoupled time series forecasting of electric vehicle charging demand for stochastic power system operation. *Electric Power Systems Research*, 140, 378-390.
20. Cesario, E., Catlett, C. and Talia, D. (2016). *Forecasting crimes using autoregressive models*. 2016 IEEE 14th Intl Conf on Dependable, Autonomic and Secure Computing, 795-802, Auckland, Yeni Zelanda.
21. Kang, H. W. and Kang, H. B. (2017). Prediction of crime occurrence from multi-modal data using deep learning. *PloS One*, 12(4), 1-19.
22. Kouziokas, G. N. (2017). The application of artificial intelligence in public administration for forecasting high crime risk transportation areas in urban environment. *Transportation research procedia*, 24, 467-473.
23. Lin, T., Guo, T. and Aberer, K. (2017). *Hybrid neural networks for learning the trend in time series*. Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, 2273-2279, Melbourne, Avustralya.
24. Yang, Y., Gao, W. and Guo, C. (2017). *Aero-Engine Lubricating Oil Metal Content Prediction Using Non-stationary Time Series ARIMA Model*. Computational Intelligence and Design (ISCID), 2017 10th International Symposium, 2, 51-54, Hangzhou, Çin.
25. Zheng, J., Xu, C., Zhang, Z. and Li, X. (2017). *Electric load forecasting in smart grids using long-short-term-memory based recurrent neural network*. Information Sciences and Systems (CISS) 2017 51st Annual Conference, 1-6, Baltimore, ABD.

26. Sobreiro, P., Martinho, D. and Pratas, A. (2018). *Sales forecast in an IT company using time series*. 2018 13th Iberian Conference on Information Systems and Technologies (CISTI), 1-5, Caceres, İspanya.
27. McNally, S., Roche, J. and Caton, S. (2018). *Predicting the price of Bitcoin using Machine Learning*. Parallel, Distributed and Network-based Processing (PDP), 2018 26th Euromicro International Conference, 339-343, Cambridge, İngiltere.
28. Wang, H., Wang, K., Zhao, H. and Yue, Y. (2018). *Prediction of User Behavior in Smart Home Based on Improved ARIMA Model*. 2018 IEEE International Conference on Mechatronics and Automation (ICMA), 298-302, Changchun, Çin.
29. Tian, C., Ma, J., Zhang, C. and Zhan, P. (2018). A Deep Neural Network Model for Short-Term Load Forecast Based on Long Short-Term Memory Network and Convolutional Neural Network. *Energies*, 11(12), 3493.
30. Stec, A. and Klabjan, D. (2018). Forecasting Crime with Deep Learning. *arXiv preprint arXiv:1806.01486*.
31. Sülo, I., Keskin, S. R., Dogan, G. and Brown, T. (2019). *Energy Efficient Smart Buildings: LSTM Neural Networks for Time Series Prediction*. 2019 International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML), 18-22, Istanbul, Türkiye.
32. Livieris, I. E., Pintelas, E. and Pintelas, P. (2020). A CNN–LSTM model for gold price time-series forecasting. *Neural Computing and Applications*, 1-10.
33. Ma, T., Antoniou, C. and Toledo, T. (2020). Hybrid machine learning algorithm and statistical time series model for network-wide traffic forecast. *Transportation Research Part C: Emerging Technologies*, 111, 352-372.
34. Khusro, S., Ali, Z. and Ullah, I. (2016). Recommender systems: issues, challenges, and research opportunities. *Information Science and Applications (ICISA)*, 1179-1189.
35. Jannach, D., Zanker, M., Felfernig, A. and Friedrich, G. (2010). *Recommender systems: An introduction*. İngiltere: Cambridge University Press, 11-80.
36. Li, H., Hong, R., Lian, D., Wu, Z., Wang, M. and Ge, Y. (2016). *A Relaxed Ranking-Based Factor Model for Recommender System from Implicit Feedback*. Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, 1683-1689.
37. Lu, J., Wu, D., Mao, M., Wang, W. and Zhang, G. (2015). Recommender system application developments: A survey. *Decision Support Systems*, 74, 12-32.
38. Koren, Y. and Bell, R. (2015). *Advances in collaborative filtering*. In *Recommender systems handbook*, ABD: Springer, Boston, 77-118.

39. Yang, Z., Wu, B., Zheng, K., Wang, X. and Lei, L. (2016). A survey of collaborative filtering-based recommender systems for mobile internet applications. *IEEE Access*, 4, 3273-3287.
40. Nilashi, M., bin Ibrahim, O., Ithnin, N. and Sarmin, N. H. (2015). A multi-criteria collaborative filtering recommender system for the tourism domain using Expectation Maximization (EM) and PCA–ANFIS. *Electronic Commerce Research and Applications*, 14(6), 542-562.
41. Thorat, P. B., Goudar, R. M. and Barve, S. (2015). Survey on collaborative filtering, content-based filtering and hybrid recommendation system. *International Journal of Computer Applications*, 110(4), 31-36.
42. Isinkaye, F. O., Folajimi, Y. O. and Ojokoh, B. A. (2015). Recommendation systems: Principles, methods and evaluation. *Egyptian Informatics Journal*, 16(3), 261-273.
43. Lee, K. and Lee, K. (2015). Escaping your comfort zone: A graph-based recommender system for finding novel recommendations among relevant items. *Expert Systems with Applications*, 42(10), 4851-4858.
44. Masthoff, J. (2015). *Group recommender systems: aggregation, satisfaction and group attributes*. In *recommender systems handbook*. ABD: Springer, 743-776.
45. Nguyen, T. T., Harper, F. M., Terveen, L. and Konstan, J. A. (2018). User personality and user satisfaction with recommender systems. *Information Systems Frontiers*, 20(6), 1173-1189.
46. Herlocker, J., Konstan, J., Borchers, A. and Riedl, J. *An algorithmic framework for performing collaborative filtering*. Proceedings of the 22nd International ACM SIGIR Conference on Research and Development in Information Retrieval, 230–237, California, ABD.
47. Alhamid, M. F., Rawashdeh, M., Al Osman, H., Hossain, M. S. and El Saddik, A. (2015). Towards context-sensitive collaborative media recommender system. *Multimedia Tools and Applications*, 74(24), 11399-11428.
48. Lee, S and Choi, J. (2017). Enhancing user experience with conversational agent for movie recommendation: Effects of self-disclosure and reciprocity. *International Journal of Human-Computer Studies*, 103, 95-105.
49. Ma, T., Zhou, J., Tang, M., Tian, Y., Al-Dhelaan, A., Al-Rodhaan, M. and Lee, S. (2015). Social network and tag sources based augmenting collaborative recommender system. *IEICE transactions on Information and Systems*, 98(4), 902-910.
50. Khusro, S., Ali, Z. and Ullah, I. (2016). *Recommender systems: issues, challenges, and research opportunities*. Information Science and Applications (ICISA) 2016, 1179-1189, Ho Chi Minh, Vietnam.

51. Colombo-Mendoza, L. O., Valencia-García, R., Rodríguez-González, A., Alor-Hernández, G. and Samper-Zapater, J. J. (2015). RecomMetz: A context-aware knowledge-based mobile recommender system for movie showtimes. *Expert Systems with Applications*, 42(3), 1202-1222.
52. Bobadilla, J., Ortega, F., Hernando, A. and Gutiérrez, A. (2013). Recommender systems survey. *Knowledge-based systems*, 46, 109-132.
53. Sikka, R., Dhankhar, A. and Rana, C. (2012). A survey paper on e-learning recommender system. *International Journal of Computer Applications*, 47(9), 27-30.
54. Al-Shamri, M. Y. H. (2014). Power coefficient as a similarity measure for memory-based collaborative recommender systems. *Expert Systems with Applications*, 41(13), 5680-5688.
55. Bellogín, A., Castells, P. and Cantador, I. (2014). Neighbor selection and weighting in user-based collaborative filtering: A performance prediction approach. *ACM Transactions on the Web (TWEB)*, 8(2), 1-30.
56. Pirasteh, P., Jung, J. J. and Hwang, D. (2014). *Item-based collaborative filtering with attribute correlation: a case study on movie recommendation*. Asian conference on intelligent information and database systems, 245-252, Ladkrabang, Tayland.
57. Ebadi, A. and Krzyzak, A. (2016). A hybrid multi-criteria hotel recommender system using explicit and implicit feedbacks. *International Journal of Computer and Information Engineering*, 10(8), 1450-1458.
58. Volkovs, M. and Yu, G. W. (2015). *Effective latent models for binary feedback in recommender systems*. Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval, 313-322, Santiago, Şili.
59. Jawaheer, G., Weller, P. and Kostkova, P. (2014). Modeling user preferences in recommender systems: A classification framework for explicit and implicit user feedback. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 4(2), 1-26.
60. Núñez-Valdéz, E. R., Lovelle, J. M. C., Martínez, O. S., García-Díaz, V., de Pablos, P. O. and Marín, C. E. M. (2012). *Implicit feedback techniques on recommender systems applied to electronic books*. *Computers in Human Behavior*, 28(4), 1186-1193.
61. Park, Y., Oh, J. and Yu, H. (2017). RecTime: Real-Time recommender system for online broadcasting. *Information Sciences*, 409, 1-16.
62. Chandramouli, B., Levandoski, J.J, Eldawy, A. and Mokbel M.F. (2011). *StreamRec: a real-time recommender system*. Proceedings of the 2011 ACM SIGMOD international conference on management of data (SIGMOD'11), 1243–1246, Atina, Yunanistan.

63. Chen, C., Yin, H., Yao, J. and Cui, B. (2013). *Terec: A temporal recommender system over tweet stream*. Proceedings of the VLDB Endowment, 6(12), 1254-1257, Trento, Italy.
64. Werner, S. and Lommatzsch, A. (2014). *Optimizing and Evaluating Stream-based News Recommendation Algorithms*. CLEF Working Notes, 813-824, Toulouse, Fransa.
65. Ludmann, C. A. (2015). *Online Recommender Systems based on Data Stream Management Systems*. Proceedings of the 9th ACM Conference on Recommender Systems, 391-394, Vienna, Avusturya.
66. Lommatzsch, A. and Albayrak, S. (2015). *Real-time recommendations for user-item streams*. Proceedings of the 30th Annual ACM Symposium on Applied Computing, 1039-1046, Salamanca, İspanya.
67. Subbian, K., Aggarwal, C. and Hegde, K. (2016). *Recommendations for streaming data*. Proceedings of the 25th ACM International on Conference on Information and Knowledge Management, 2185-2190, Indiana, ABD.
68. Chang, S., Zhang, Y., Tang, J., Yin, D., Chang, Y., Hasegawa-Johnson, M. A. and Huang, T. S. (2016). *Streaming Recommender Systems*. *arXiv preprint arXiv:1607.06182*.
69. Mehmood, F., Ahmad, S. and Kim, D. (2019). Design and Development of a Real-Time Optimal Route Recommendation System Using Big Data for Tourists in Jeju Island. *Electronics*, 8(5), 506.
70. PhridviRaj, M. S. B. and GuruRao, C. V. (2014). Data mining—past, present and future—a typical survey on data streams. *Procedia Technology*, 12, 255-263.
71. Gama, J. (2010). *Knowledge discovery from data streams*. ABD: CRC Press, 171-189.
72. Kamburugamuve, S., Fox, G., Leake, D. and Qiu, J. (2013). *Survey of Streaming Data Algorithms*. Tech. Rep., 1-9, Indiana University, Indiana, İngiltere.
73. Ahmed, N. K., Duffield, N., Willke, T. and Rossi, R. A. (2017). On sampling from massive graph streams. *arXiv preprint arXiv:1703.02625*.
74. Aggarwal, C. C. (2007). *Data streams: models and algorithms (Vol. 31)*. ABD: Springer Science & Business Media, 1-85.
75. Ankleshwaria, T. B. and Dhobi, J. S. (2014). Mining Data Streams: A Survey. *International Journal*, 2(2), 379-386.
76. Chatfield, C. (2016). *The analysis of time series: An introduction*. ABD: CRC press, 1-15.

77. Martínez-Álvarez, F., Troncoso, A., Asencio-Cortés, G. and Riquelme, J. C. (2015). A survey on data mining techniques applied to electricity-related time series forecasting. *Energies*, 8(11), 13162-13193.
78. Weigend, A. S. (2018). *Time series prediction: forecasting the future and understanding the past*. İngiltere: Routledge, 11-18.
79. Karakoyun, E. S. and Cibikdiken, A. O. (2018). *Comparison of arima time series model and lstm deep learning algorithm for bitcoin price forecasting*. 13th multidisciplinary academic conference, 171-180, Prag, Çek Cumhuriyeti.
80. Valipour, M., Banihabib, M. E. and Behbahani, S. M. R. (2013). Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir. *Journal of hydrology*, 476, 433-441.
81. Wang, W. C., Chau, K. W., Xu, D. M. and Chen, X. Y. (2015). Improving forecasting accuracy of annual runoff time series using ARIMA based on EEMD decomposition. *Water Resources Management*, 29(8), 2655-2675.
82. Louppe, G. (2014). Understanding random forests: From theory to practice. *arXiv preprint arXiv:1407.7502*.
83. Sapankevych, N. I. and Sankar, R. (2009). Time series prediction using support vector machines: a survey. *IEEE Computational Intelligence Magazine*, 4(2), 24-38.
84. Långkvist, M., Karlsson, L. and Loutfi, A. (2014). A review of unsupervised feature learning and deep learning for time-series modeling. *Pattern Recognition Letters*, 42, 11-24.
85. Gamboa, J.C.B. (2017). Deep learning for time-series analysis. *arXiv preprint arXiv:1701.01887*.
86. Khashei, M. and Bijari, M. (2010). An artificial neural network (p, d, q) model for time series forecasting. *Expert Systems with applications*, 37(1), 479-489.
87. Karayiannis, N. and Venetsanopoulos, A. N. (2013). *Artificial neural networks: learning algorithms, performance evaluation, and applications*. ABD: Springer Science & Business Media, 9-46.
88. Graves, A., Mohamed, A. R. and Hinton, G. (2013). *Speech recognition with deep recurrent neural networks*. Acoustics, speech and signal processing (icassp), 2013 ieee international conference, 6645-6649, Vancouver, Kanada.
89. Zhang, Y., Sun, Y., Phillips, P., Liu, G., Zhou, X. and Wang, S. (2016). A multilayer perceptron based smart pathological brain detection system by fractional Fourier entropy. *Journal of medical systems*, 40(7), 173.

90. Ramchoun, H., Idrissi, M. A. J., Ghanou, Y. and Ettaouil, M. (2016). Multilayer Perceptron: Architecture Optimization and Training. *IJIMAI*, 4(1), 26-30.
91. Bayat, F.M., Prezioso, M., Chakrabarti, B., Nili, H., Kataeva, I. and Strukov, D. (2018). Implementation of multilayer perceptron network with highly uniform passive memristive crossbar circuits. *Nature communications*, 9(1), 1-7.
92. Esteves, J.T., de Souza Rolim, G. and Ferraudo, A.S. (2019). Rainfall prediction methodology with binary multilayer perceptron neural networks. *Climate Dynamics*, 52(3-4), 2319-2331.
93. Huang, C. J. and Kuo, P. H. (2018). A deep cnn-lstm model for particulate matter (PM_{2.5}) forecasting in smart cities. *Sensors*, 18(7), 1-22.
94. Yan, K., Wang, X., Du, Y., Jin, N., Huang, H. and Zhou, H. (2018). Multi-Step Short-Term Power Consumption Forecasting with a Hybrid Deep Learning Strategy. *Energies*, 11(11), 3089.
95. Rana, R. (2016). Gated recurrent unit (GRU) for emotion classification from noisy speech. *arXiv preprint arXiv:1612.07778*.
96. Kim, J. and Kim, H. (2016). *Classification performance using gated recurrent unit recurrent neural network on energy disaggregation*. 2016 International Conference on Machine Learning and Cybernetics (ICMLC), 1, 105-110, Adelaide, Avustralya.
97. Malhotra, P., Vig, L., Shroff, G. and Agarwal, P. (2015). *Long short term memory networks for anomaly detection in time series*. European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, Bruges, Belçika.
98. Pascanu, R., Mikolov, T. and Bengio, Y. (2013). *On the difficulty of training recurrent neural networks*. International Conference on Machine Learning, 1310-1318, Georgia, ABD.
99. İnternet: Retailrocket recommender system dataset. *Kaggle*. URL: <https://www.kaggle.com/retailrocket/ecommerce-dataset>, Son Erişim Tarihi: 23.09.2018.
100. İnternet: Attack Dataset. *Hacking and Countermeasure Research Lab*. URL: <http://ocslab.hksecurity.net/Datasets>, Son Erişim Tarihi: 10.01.2019.
101. İnternet: Baltimore Crime Dataset. *Data World*. URL: <https://data.world/data-society/city-of-baltimore-crime-data>, Son Erişim Tarihi: 01.05.2019.
102. İnternet: Montgomery country emergency call dataset. *Kaggle*. URL: <https://www.kaggle.com/mchirico/montcoalert>, Son Erişim Tarihi: 23.08.2019.

ÖZGEÇMİŞ

Kişisel Bilgiler

Soyadı, adı : UTKU, Anıl
 Uyuğu : T.C.
 Doğum tarihi ve yeri : 19.07.1989, Elazığ
 Medeni hali : Evli
 Telefon : 0 (546) 654 36 01
 e-mail : anilutku@msn.com



Eğitim

Derece	Eğitim Birimi	Mezuniyet Tarihi
Doktora	Gazi Üniversitesi / Bilgisayar Mühendisliği	Devam Ediyor
Yüksek lisans	Gazi Üniversitesi / Bilgisayar Mühendisliği	2015
Lisans	Kocaeli Üniversitesi / Bilgisayar Mühendisliği	2010
Lise	Karakoçan Lisesi	2005

İş Deneyimi

Yıl	Yer	Görev
2013-Halen	Gazi Üniversitesi	Araştırma Görevlisi
2011-2013	Tunceli Üniversitesi	Araştırma Görevlisi

Yabancı Dil

İngilizce

Yayınlar

1. Utku, A. and Akcayol, M.A. (2020). Deep learning based new prediction model for the next purchase. *Advances in Electrical and Computer Engineering*, 20(2), 35–44.
2. Utku, A. ve Akcayol, M.A. (basımda). Derin Öğrenme Tabanlı Model ile Bir Olayın Sonraki Olma Zamanının Tahmini. *Politeknik Dergisi*.
3. Ozer, E.G., Karapinar, I.N., Busbug, S., Turan, S., Utku, A. and Akcayol, M.A. (2020).

4. Deep Learning based a New Model for Video Captioning. *International Journal of Advanced Computer Science And Applications*, 11(3), 514–519.
5. Utku, A. and Akcayol, M.A. (2019). A comprehensive study on stream data characterization, generation and analytics. *Erzincan University Journal of Science and Technology*, 12(1), 379–410.
6. Akcayol, M.A., Utku, A., Aydoğan Duman, E. and Mutlu, B. (2018). A weighted multi attribute based recommender system using extended user behavior analysis. *Electronic Commerce Research and Applications*, 28, 86–93.
7. Utku, A. and Akcayol, M.A. (2018). A comprehensive review on the use of big data in recommendation systems. *International Journal of Advances in Engineering and Pure Sciences*, 30(4), 339–357.
8. Utku, A. and Doğru, İ.A. (2017). Permission based detection system for android malware. *Gazi Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi*, 32(4), 1015–1024.
9. Utku, A. and Akcayol, M.A. (2017). A Comparative and Comprehensive Review for Learning and Adaptive Recommendation Systems. *Erciyes University Journal of Natural and Applied Sciences*, 33(3), 13-34.
10. Kansu, Y., Mutlu, B., Utku, A. and Akcayol, M.A. (2017). An Efficient Multi Threaded Web Crawler Using HashMaps. *Journal of Advances in Computer Networks (JACN)*, 5(2), 65–69.
11. Utku, A. and Doğru, İ.A. (2016). Mobil Kötücül Yazılımlar ve Güvenlik Çözümleri Üzerine Bir İnceleme. *Gazi Üniversitesi Fen Bilimleri Dergisi*, 4(2), 49-64.
12. Utku, A., Karacan, H., Yıldız, O. and Akcayol, M.A. (2015). Implementation of a New Recommendation System Based on Decision Tree Using Implicit Relevance Feedback. *Journal of Software*, 10(12), 1367–1374.
13. Karamollaoğlu, H., Doğru, İ.A. and Utku, A. (2017). *Identification of shares containing offensive charge in social media*. 2017 25th Signal Processing and Communications Applications Conference (SIU), Antalya, Türkiye.
14. Yılmaz, M., Mutlu, B., Utku, A. and Akcayol, M.A. (2018). *Deep Learning Approach for Author Verification Problem on Twitter*. SIU 2018, İzmir, Türkiye.
15. Utku, A., Doğru, İ.A. and Akcayol, M.A. (2018). *Permission Based Android Malware Detection With Multilayer Perceptron*. SIU 2018, İzmir, Türkiye.
16. Utku, A. and Doğru, İ.A. (2017). *Malware detection system based on machine learning methods for Android operating systems*. 2017 25th Signal Processing and Communications Applications Conference (SIU), Antalya, Türkiye.
17. Utku, A., E. Aydoğan Duman, E., Mutlu, B. and Akcayol, M.A. (2017). *A New Recommender System Based on Multiple Parameters and Extended User Behavior*

Analysis. Proceedings of the 9th International Conference on Information Management and Engineering-ICIME 2017, Barcelona, İspanya.

18. Utku, A., Doğru, İ.A. and Akcayol, M.A. (2018). *Decision Tree Based Android Malware Detection System*. SİU 2018, İzmir, Turkey.
19. Karamollaoğlu, H., Doğru, İ.A., Dörterler, M., Utku, A. and Yıldız, O. (2018). *Sentiment Analysis on Turkish Social Media Shares through Lexicon Based Approach*. 2018 3rd International Conference on Computer Science and Engineering (UBMK), Sarajevo, Bosna Hersek.

Hobiler

Yüzme, Dijital oyunlar



GAZİ GELECEKTİR..