



ARAÇ VERİLERİ İLE TRAFİK HIZ TAHMİN VE MAHREMİYET MODELİ

Murat AKIN

DOKTORA TEZİ

BİLGİSAYAR MÜHENDİSLİĞİ ANA BİLİM DALI

**GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

TEMMUZ 2021

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.

Murat AKIN

08/07/2021

ARAÇ VERİLERİ İLE TRAFİK HIZ TAHMİN VE MAHREMİYET MODELİ
(Doktora Tezi)

Murat AKIN

GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Temmuz 2021

ÖZET

Bu tez çalışmasında; trafik parametrelerinin tahmini ve yörünge mahremiyetinde kullanılabilecek üç yeni ve özgün model geliştirilmiş, Ankara ilinde bulunan iki ana güzergâhta yoğun ve orta yoğun trafik zaman dilimlerinde test edilmiş ve başarılı sonuçlar elde edilmiştir. İlk önerilen trafik hız tahmin modelinde, trafik hız tahmini için filtreleme, veri temizliği ve harita eşleme aşamasından geçirilen araç verileri, OPTICS algoritması ile yoğunluk tabanlı olarak kümelenmiş ve böylece her yola ait karakteristik özellikler elde edilmiş, yol karakteristik bilgilerinin tekrarlı örüntüleri dikkate alınarak LSTM ve GRU sinir ağ modelleriyle hız tahmini yapılmış, Eskişehir ve İstanbul yollarında bu modeller test edilmiş ve RMSE hata değerleri km/s cinsinden başarıımı en yüksek olarak yoğun trafikte 7,814, orta yoğun trafikte 4,372 elde edilmiştir. Mahremiyet korumalı ikinci tahmin modeli, önerilen birinci modeli temel alarak ve modele ara bir aşama olarak diferansiyel mahremiyet (DM) yöntemi uygulanarak geliştirilmiş, DM yöntemi hız tahmini yapan bir modele ilk kez uyarlanmış, iki farklı veri kümesiyle Eskişehir ve İstanbul yollarında test edilmiş, orijinal küme hız özneliği ile kıyaslanmış, başarıımı en yüksek RMSE hata değerleri İstanbul yolunda yoğun trafikte 6,127 km/s ve orta yoğun trafikte 8,789 km/s olarak ölçülmüştür. Üçüncü model; yörünge bilgilerinin yayınlanmasında mahremiyeti sağlayacak pencere tabanlı ve DM'nin uygulandığı bir model olup önceki iki modelde olduğu gibi Eskişehir ve İstanbul yollarında test edilmiş, en fazla GPS sinyali üreten örnek araçların epsilon ve pencere boyutunun 2 olduğu değerlerde, sırasıyla, x-y koordinatları RMSE hata değerleri sırasıyla 0,00031973 ve 0,00013009 ile 0,00011643 ve 0,00008601 derece olarak elde edilmiştir. 3191873 numaralı TÜBİTAK proje kapsamında geliştirilmiş olan bu çalışmanın gerçek ortama uyarlama çalışmaları devam etmektedir. Sonuç olarak; önerilen modellerin gelecekte yapılacak olan gerçek zamanlı trafik parametrelerinin tahmininde farklı ve yeni çözümlerin geliştirilmesine katkılar sağlayacağı, gerçek uygulamalarda kullanılabileceği, mahremiyet odaklı çalışmalara yön vereceği, farklı veri yayınlama alanlarında uygulanabileceği değerlendirilmektedir.

Bilim Kodu : 92432

Anahtar Kelimeler : Trafik bilgi sistemleri, araç trafik verisi, trafik parametreleri, trafik tahmin modeli, mahremiyet korumalı sistemler, diferansiyel mahremiyet, OPTICS, LSTM, GRU, sinir ağları

Sayfa Adedi : 110

Danışman : Prof. Dr. Şeref SAĞIROĞLU

TRAFFIC SPEED PREDICTION AND PRIVACY MODEL WITH VEHICLE DATA

(Ph.D. Thesis)

Murat AKIN

GAZİ UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

July 2021

ABSTRACT

In this thesis study, three novel and new models have been developed for the prediction of traffic parameters and providing trajectory privacy, tested in heavy and medium-heavy traffic time zones on two main roads in Ankara, and successful results have been obtained. In the first proposed traffic speed prediction model, the vehicle data, which were passed through the filtering, data cleaning and map-matching processes for the traffic speed prediction, were density-based clustered using OPTICS algorithm and thus characteristic features of each of roads were generated, predictions were made with LSTM and GRU neural network models considering the recurrent patterns of road characteristic information, these models were tested on Eskisehir and Istanbul roads, and RMSE error values in km/h were obtained with the highest performance of 7,814 in heavy traffic and 4,372 in medium heavy traffic, respectively. The privacy-preserved second prediction model was developed based on the first proposed model with applying differential privacy (DP) method to the model an intermediate stage, DP method was adopted for the first time to a speed prediction model, was tested on Eskisehir and Istanbul roads with two different datasets compared with the original cluster speed features, the RMSE error values with the highest performance were measured as 6,127 km/h in heavy traffic and 8,789 km/h in medium heavy traffic on the Istanbul road. The third one introduces a new privacy-preserving model which is window-based integrated with DP applied to ensure privacy when publishing the trajectory information. Like the previous two models, this model was also tested on Eskisehir and Istanbul roads, the RMSE error values of x-y coordinates were obtained as 0,00031973 and 0,00013009 degrees, 0,00011643 and 0,00008601 degrees, respectively, for the sample vehicles producing the highest GPS signal at the value of epsilon and window size 2. The proposed models in this thesis were developed within the scope of TÜBİTAK project given number 3191873 and, improvements in the studies still continue in order to apply for real-time implementation. As a result, it is concluded that the proposed models might offer different and new solutions in the prediction of real-time traffic parameters for future studies and can be used in real applications, will guide privacy-oriented studies, and can be applied in different data publishing research fields.

Science Code : 92432

Key Words : Traffic information system, vehicle traffic data, traffic parameters, traffic prediction model, privacy preserving systems, differential privacy, OPTICS, LSTM, GRU, neural network

Page Number : 110

Supervisor : Prof. Dr. Şeref SAĞIROĞLU

TEŞEKKÜR

Bu tez çalışmamda büyük emekleri olan kıymetli hocam ve doktora tez danışmanım Sn. Prof. Dr. Şeref SAĞIROĞLU'na, Tez İzleme Komitemde bulunan ve değerli katkı ve desteklerini her daim sunan Sn. Prof. Dr. Ali YAZICI ve Sn. Dr. Öğr. Üyesi Mehmet DEMİRCİ hocalarıma ve bugünlere gelmemde büyük emeği olan değerli büyüğüm Sn. Prof. Dr. Metin GÜRÜ'ye ve aileme çok teşekkür ederim. Ayrıca, Gazi Üniversitesi Büyük Veri ve Bilgi Güvenliği Laboratuvarına (BIDISEC) sunduğu altyapı, Başarsoft Bilgi Sistemleri A.Ş.'ye tez çalışmalarımın bir kısmını firmalarında yapmama verdikleri imkân, sağladıkları trafik verileri, 3191873 numaralı TÜBİTAK projesinin bir kısmı olan çalışma kapsamında sağladıkları destek ve BiSoft firmasına kullanıma sundukları görselleştirme aracı için şükranlarımı sunarım.

İÇİNDEKİLER

	Sayfa
ÖZET	iv
ABSTRACT.....	v
TEŞEKKÜR.....	vi
İÇİNDEKİLER	vii
ÇİZELGELERİN LİSTESİ.....	x
ŞEKİLLERİN LİSTESİ.....	xii
SİMGELER VE KISALTMALAR.....	xv
1. GİRİŞ.....	1
2. TRAFİK YÖNETİM EKOSİSTEMİ.....	7
2.1. Akıllı Ulaşım Sistemleri.....	7
2.1.1. Gelişmiş trafik yönetim sistemleri (GTYS)	8
2.1.2. Gelişmiş yolcu bilgi sistemleri (GYBS).....	9
2.1.3. Gelişmiş toplu taşıma sistemleri (GTTS).....	9
2.1.4. Ulaşım ücretlendirme sistemleri (UÜS).....	10
2.2. Trafik Veri Kaynakları	11
2.2.1. Mobil cihazlar	12
2.2.2. Mikrodalga sensörleri.....	12
2.2.3. Döngü detektörleri.....	12
2.2.4. Video kameralar	13
2.2.5. Özel araç ağları.....	13
2.3. Trafik Parametreleri ve Tahmini.....	13
2.3.1. Trafik akışı üzerine yapılan çalışmalar	15
2.3.2. Trafik yoğunluğu üzerine yapılan çalışmalar	17

Sayfa

2.3.3. Trafik hızı üzerine yapılan çalışmalar	19
2.4. Veri Mahremiyeti	23
2.4.1. Konum mahremiyeti.....	25
2.4.2. Konum mahremiyeti sağlamak için yöntemler	26
3. MATERYAL METOT	29
3.1. Kullanılan Veri Kümesi	29
3.2. Görselleştirme Araçları ve Kullanılan Platform	31
3.3. OPTICS Algoritması	31
3.4. Tekrarlı Sinir Ağı (RNN).....	33
3.5. Diferansiyel Mahremiyet	34
3.6. Hata Metrikleri	35
3.7. Haversine Uzaklığı.....	36
4. TRAFİK HIZ TAHMİN MODELİ	37
4.1. Trafik Verilerini Okuma	40
4.2. Veri Filtreleme	40
4.3. Harita Eşleme ve Yön Seçimi	42
4.4. Araç Kümelerinin Oluşturulması	43
4.5. Aykırı Kümelerin ve Verilerin Temizlenmesi	43
4.6. Küme Özniteliklerinin Çıkarımı	44
4.7. Akış Oluşturma	44
4.8. Görselleştirme	45
4.9. Hız Tahmini ve Sonuçları Karşılaştırma.....	45
4.10. Deneysel Sonuçlar.....	46
4.10.1. VK1 için elde edilen sonuçlar	50
4.10.2. VK2 için elde edilen sonuçlar	55

Sayfa

5. MAHREMİYET KORUMALI TRAFİK HIZ TAHMİN MODELİ.....	61
5.1. Veri Mahremiyet İhlalleri ve Saldırıları.....	62
5.2. Veri İhlal ve Saldırılarından Korunma Yöntemleri	63
5.3. Model Aşamaları ve Değerlendirme	65
5.3.1. VK1 için elde edilen deneysel sonuçlar	66
5.3.2. VK2 için elde edilen deneysel sonuçlar	70
6. VERİ YAYINLAMA İÇİN ÖNERİLEN MAHREMİYET KORUMALI YÖRÜNGE MODELİ.....	75
6.1. Pencere Tabanlı Gruplama.....	76
6.2. Grupların Ortalamasını Alma.....	77
6.3. Ortalamalara Gürültü Ekleme	77
6.4. Örnek Araç için Test Sonuçları.....	80
6.4.1. İstanbul yolu test sonuçları.....	80
6.4.2. Eskişehir yolu test sonuçları.....	83
6.5. Veri Kümesindeki Tüm Araçlar için Elde Edilen Sonuçlar.....	87
7. DEĞERLENDİRME VE SONUÇLAR.....	89
KAYNAKLAR	97
ÖZGEÇMİŞ	109

ÇİZELGELERİN LİSTESİ

Çizelge	Sayfa
Çizelge 2.1. Temel trafik parametreleri üzerine çalışmalar	22
Çizelge 3.1. Araçlar için üretilen öznitelikler ve açıklamaları	31
Çizelge 3.2. Araç GPS kayıtlarının örnek bir gösterimi	31
Çizelge 4.1. Model için kullanılan veri kümelerinin özellikleri	40
Çizelge 4.2. Seçilen yollara ait SHP dosya bilgilerine ait örnek gösterim	43
Çizelge 4.3. Küme Öznitelikleri	44
Çizelge 4.4. Eskişehir yolu güzergâh dışı araçlardan örnek gösterim	49
Çizelge 4.5. VK1-Eskişehir yol güzergâhı için örnek küme öznitelikleri	50
Çizelge 4.6. VK1-İstanbul yol güzergâhı için örnek küme öznitelikleri	50
Çizelge 4.7. VK1-Başlangıç noktalarına göre en yakın 5 küme ve uzaklık değerleri	52
Çizelge 4.8. VK1-Trafik hız tahmini için hata değerleri	55
Çizelge 4.9. VK2-Eskişehir yol güzergâhı için örnek küme öznitelikleri	56
Çizelge 4.10. VK2-İstanbul yol güzergâhı için örnek küme öznitelikleri	56
Çizelge 4.11. VK2 için Trafik hız tahmini için hata değerleri.....	58
Çizelge 5.1. VK1-Eskişehir yolu örnek kümelere ait orijinal ve mahremiyet uygulanan öznitelikler.....	67
Çizelge 5.2. VK1- İstanbul yolu örnek kümelere ait orijinal ve mahremiyet uygulanan öznitelikler.....	67
Çizelge 5.3. VK1 için oluşan bütün kümelere ait hız ve arka araç X-Y koordinatı RMSE değerleri	67
Çizelge 5.4. Mahremiyet korumalı tahmin modelinde VK1 için üretilen hız tahminleri hata değerleri	69
Çizelge 5.5. VK2-Eskişehir yolu örnek kümelere ait orijinal ve mahremiyet uygulanan öznitelikler.....	70
Çizelge 5.6. VK2-İstanbul yolu örnek kümelere ait orijinal ve mahremiyet uygulanan öznitelikler.....	70

Çizelge	Sayfa
Çizelge 5.7. VK2 için oluşan bütün kümelere ait hız ve arka araç X-Y koordinatı RMSE değerleri	71
Çizelge 5.8. Mahremiyet korumalı tahmin modelinde VK2 için üretilen hız tahminleri hata değerleri	72
Çizelge 6.1. Model için kullanılan veri detayları.....	80
Çizelge 6.2. İstanbul yolu - Örnek araç için elde edilen sonuçlar	81
Çizelge 6.3. Eskişehir yolu - Örnek araç için elde edilen sonuçlar	85
Çizelge 6.4. İstanbul yolu - Tüm araçlar için elde edilen ortalama RMSE değerleri.....	87
Çizelge 6.5. Eskişehir yolu - Tüm araçlar için elde edilen ortalama RMSE değerleri...	87

ŞEKİLLERİN LİSTESİ

Şekil	Sayfa
Şekil 2.1. İBB Ulaşım Yönetim Merkezi.....	8
Şekil 2.2. Trafik için veri kaynakları	11
Şekil 2.3. Trafik Akış-Yoğunluk ve Hız ilişkisi	14
Şekil 3.1. Tüm Türkiye’de gezinen araçların örnek konum verileri.....	30
Şekil 3.2. Ankara iline ait örnek konum verileri.....	30
Şekil 3.3. DBSCAN ve OPTICS algoritmalarının örnek kümeleme işlemi	32
Şekil 3.4. Derin öğrenmede kullanılan diferansiyel mahremiyet yöntemleri	35
Şekil 4.1. Kısa zamanlı trafik hız tahmin model süreci ve öne çıkan aşamalar.....	37
Şekil 4.2. VK1 - Eskişehir yolu filtrelenen araç konum noktaları.....	41
Şekil 4.3. VK1 - İstanbul yolu filtrelenen araç konum noktaları.....	42
Şekil 4.4. Eskişehir ve İstanbul yolları SHP çizimleri.....	42
Şekil 4.5. Trafik hız tahmini için uygulanan sinir ağı modeli	46
Şekil 4.6. Eskişehir Yolu ağ bilgisi ile eşlenen noktaların görünümü	47
Şekil 4.7. İstanbul Yolu ağ bilgisi ile eşlenen noktaların görünümü	47
Şekil 4.8. Eskişehir Yolu - Elde Edilen Yoğunluk Tabanlı Kümeler	48
Şekil 4.9. İstanbul Yolu - Elde Edilen Yoğunluk Tabanlı Kümeler	48
Şekil 4.10. Eskişehir yolu güzergâh dışı örnek araçlar.....	49
Şekil 4.11. VK1-Eskişehir yolunda oluşan kümelerin araç sayı, hız ve yoğunluk grafiği.....	51
Şekil 4.12. VK1- İstanbul yolunda oluşan kümelerin araç sayı, hız ve yoğunluk grafiği.....	51
Şekil 4.13. VK1-İstanbul Yolu Anlık Trafik Hız Gösterimi	52
Şekil 4.14. VK1-Eskişehir Yolu Anlık Trafik Hız Gösterimi	52

Şekil	Sayfa
Şekil 4.15. VK1-Eskişehir Yolu LSTM hız tahmin sonuçları	53
Şekil 4.16. VK1-Eskişehir Yolu GRU hız tahmin sonuçları	53
Şekil 4.17. VK1-İstanbul Yolu LSTM hız tahmin sonuçları	54
Şekil 4.18. VK1-İstanbul Yolu GRU hız tahmin sonuçları	54
Şekil 4.19. VK2- Eskişehir Yolunda oluşan kümelerin araç sayı, hız ve yoğunluk grafiği.....	55
Şekil 4.20. VK2- İstanbul Yolunda oluşan kümelerin araç sayı, hız ve yoğunluk grafiği.....	56
Şekil 4.21. VK2-Eskişehir Yolu LSTM hız tahmin sonuçları	57
Şekil 4.22. VK2-Eskişehir Yolu GRU hız tahmin sonuçları	57
Şekil 4.23. VK2-İstanbul Yolu LSTM hız tahmin sonuçları	58
Şekil 4.24. VK2-İstanbul Yolu GRU hız tahmin sonuçları	58
Şekil 5.1. Mahremiyet korumalı trafik hız tahmin modeli.....	66
Şekil 5.2. VK1-Eskişehir yolu DM küme öznitelikleri ile LSTM model sonucu.....	68
Şekil 5.3. VK1-Eskişehir yolu DM küme öznitelikleri ile GRU model sonucu.....	68
Şekil 5.4. VK1-İstanbul yolu DM küme öznitelikleri ile LSTM model sonucu.....	69
Şekil 5.5. VK1-İstanbul yolu DM küme öznitelikleri ile GRU model sonucu.....	69
Şekil 5.6. VK2-Eskişehir yolu DM küme öznitelikleri ile LSTM model sonucu.....	71
Şekil 5.7. VK2-Eskişehir yolu DM küme öznitelikleri ile GRU model sonucu.....	71
Şekil 5.8. VK2-İstanbul yolu DM küme öznitelikleri ile LSTM model sonucu.....	72
Şekil 5.9. VK2-İstanbul yolu DM küme öznitelikleri ile GRU model sonucu.....	72
Şekil 6.1. Yörünge verilerinde ifşa saldırısı.....	76
Şekil 6.2. Yörünge verisine pencere yapısının uygulanması	77
Şekil 6.3. Pencere içindeki konum noktalarının ortalamasıyla T' yörünge verisi oluşturma.....	77

Şekil	Sayfa
Şekil 6.4. T' yörünge verisine gürültü ekleyerek T'' anonim yörünge verisini oluşturma.....	78
Şekil 6.5. Modelde uygulanan genel iş akış adımları	80
Şekil 6.6. İstanbul yolu - Örnek aracın ortalama ve anonimleştirilmiş X değerleri ($\epsilon=2$).....	81
Şekil 6.7. İstanbul yolu - Örnek aracın ortalama ve anonimleştirilmiş Y değerleri ($\epsilon=2$).....	82
Şekil 6.8. İstanbul yolu - Örnek aracın X, Y ve zaman boyutunda GPS sinyal grafiği..	82
Şekil 6.9. İstanbul yolu - Örnek aracın orijinal yörünge verisi ($T_{\text{örnek}}$).....	82
Şekil 6.10. İstanbul yolu - Örnek aracın ortalama alınmış yörünge verisi ($T'_{\text{örnek}}$)	83
Şekil 6.11. İstanbul yolu örnek aracın DM uygulanmış yörünge verisi ($T''_{\text{örnek}}$)	83
Şekil 6.12. Eskişehir yolu - Örnek aracın ortalama ve anonimleştirilmiş X değerleri ($\epsilon=2$).....	84
Şekil 6.13. Eskişehir yolu - Örnek aracın ortalama ve anonimleştirilmiş Y değerleri ($\epsilon=2$).....	84
Şekil 6.14. Eskişehir yolu - Örnek aracın X, Y ve zaman boyutunda GPS sinyal grafiği	84
Şekil 6.15. Eskişehir yolu - Örnek aracın orijinal yörünge verisi ($T_{\text{örnek}}$).....	85
Şekil 6.16. Eskişehir yolu - Örnek aracın ortalama alınmış yörünge verisi ($T'_{\text{örnek}}$)	86
Şekil 6.17. Eskişehir yolu örnek aracın DM uygulanmış yörünge verisi ($T''_{\text{örnek}}$)	86
Şekil 7.1 Yoğun trafik veri kümesi için hız tahmin hata değerleri	92
Şekil 7.2 Orta yoğun trafik veri kümesi için hız tahmin hata değerleri	92
Şekil 7.3 Orijinal küme ve DM uygulanmış küme hızları arasında RMSE değerleri.....	93
Şekil 7.4 Yoğun trafik veri kümesi için orijinal küme ve DM uygulanmış küme özniteliklerinden üretilen tahmin RMSE değerleri	93
Şekil 7.5 Yörünge mahremiyet modeli örnek araç X-Y koordinatı RMSE değerleri.....	94
Şekil 7.6 Yörünge mahremiyet modeli örnek araç X-Y koordinatı RMSE değerleri.....	94

SİMGELER VE KISALTMALAR

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

Kısaltmalar	Açıklamalar
AA	Araştırma Araçları
ARIMA	Hareketli Ortalama Zaman Serisi
AUC	Eğri Altında Kalan Alan
AUS	Akıllı Ulaşım Sistemleri
CNN	Evrişimsel Sinir Ağları
DBSCAN	Yoğunluk Tabanlı Mekânsal Kümeleme Yöntemi
DD	Döngü Detektörü
DL	Derin Öğrenme
DM	Diferansiyel Mahremiyet
FCD	Hareketli Araç Verisi
GDPR	Uluslararası Veri Koruma Yasası
GLR	Genelleştirilmiş Olasılık Oranı
GNN	Çizge Sinir Ağları
GPS	Küresel Konumlama Sistemi
GRU	Tekrarlı Kapı Ünitesi
GTYS	Gelişmiş Trafik Yönetim Sistemleri
GTTS	Gelişmiş Toplu Taşıma Sistemleri
GYBS	Gelişmiş Yolcu Bilgi Sistemleri
HMM	Gizli Markov Modeli
HV1	Birinci Ham Veri
HV2	İkinci Ham Veri
IBB	İstanbul Büyükşehir Belediyesi
KNN	K-en Yakın Komşuluk
KTS	Konum Tabanlı Servisler
KVKK	Kişisel Verileri Koruma Kanunu
LGP	Yerel Gauss İşlemi
LSTM	Uzun-Kısa Zamanlı Hafıza

Kısaltmalar**Açıklamalar**

MAE	Ortalama Mutlak Hata
MAPE	Ortalama Mutlak Yüzdesel Hata
MRE	Ortalama Göreceli Hata
MPE	Ortalama Yüzdesel Hata
MS	Mikrodalga Sensörü
MSE	Ortalama Karekök Hata
NN	Sinir Ağı
OPTICS	Küme Yapısı Tanımı için Sıralı Noktalar Yöntemi
PCA	Temel Bileşenler Analizi
POI	İlgi Noktası
RNN	Tekrarlı Sinir Ağları
RMSE	Ortalama Karekök Sapma Hatası
RRMSE	Göreceli RMSE
SHP	Yol Geometrik Dosyası
SVR	Destek Vektör Regresyonu
SVM	Destek Vektör Makinesi
TMS	Trafik Mesaj Servisi
UÜS	Ulaşım Ücretlendirme Sistemleri
VAPE	Varyans Mutlak Yüzdesel Hata
VANET	Araçlar Arası Ağ
VK	Video Kamera
VK1	Birinci Veri Kümesi
VK2	İkinci Veri Kümesi
V2V	Araçtan Araca İletişim Ağı
V2I	Araçtan Altyapıya İletişim Ağı
WMAPE	Ağırlıklı MAPE

1. GİRİŞ

Sanayileşmenin ve kentsel yaşam kalitesinin artmasından dolayı şehir merkezlerinde nüfus artmaktadır. Bununla birlikte, büyük şehirler ve gelişmekte olan yerleşim alanlarında kullanılan ve trafiğe katılan araç sayıları da artmakta, trafiğe katılan araçların oluşturduğu yoğunluk ulaşım problemlerini beraberinde getirmektedir. Türkiye İstatistik Kurumu verilerine göre 2020 yılı Ekim ayında 23 milyonu geçmiştir ve bu dönemdeki araç sayısı geçtiğimiz yıla göre yüzde ellinin üzerinde artış göstermiştir [1].

Artan araç sayısına bağlı olarak mevcut ulaşım altyapısı aynı oranda gelişmemekte, şehir trafiğinde araçlar sınırlı ulaşım ağı kullanmaktadır. Devlet ve kamu kurumları şehir içi ve şehirlerarası karayolu ulaşım ağını iyileştirmekte, gereklilik halinde yeni yollar yapmakta fakat ulaşım ağının geliştirilmesi artan araç sayısındaki talepleri karşılayamamakta, şehirlerdeki trafik yoğunluğu her geçen gün artmaktadır. Bunun sonucunda, ulaşım yönetimi ile ilgilenen kurumlar trafik sıkışıklığının kontrolünden, sürücü güvenliği ve çevresel etkiye kadar olan konulara çözüm üretmeye ve iyileştirmeye çalışmaktadır [2].

Gelişen teknolojiye paralel trafik kaynaklarının artması sayesinde trafik araştırmaları ve mühendisliği, ekonomik koşullar dikkate alınarak sorunsuz, güvenli trafik uygulamaları geliştirmek için trafik davranışlarını incelemektedir [3]. Trafikin davranışlarını ve karakteristiğini anlamak, bir trafik ağını tanımlamak için trafik parametreleri tanımlanmış ve temel parametreler olarak akış, yoğunluk ve hız olarak ifade edilmiştir [4,5]. Trafikin temel parametreleri kullanılarak trafik sistemleri kullanıcılara yol akış bilgisini belirli renklerle göstermekte, uygun rotaları alternatifleriyle birlikte sunmakta, yol akış bilgisine göre seyahat süresini de hesaplayabilmekte ve böylelikle kullanıcılar günlük seyahatlerini planlamaktadır.

Kullanıcıların günlük seyahat planlama için kullandıkları mobil uygulamalar konum tabanlı servisler çalıştırırken, bu tür uygulamaları geliştiren kurum ya da şirketler kişilerin kullandıkları cihazlardaki bu servislerden elde ettikleri konum verilerini kendi sistemlerinde depolamakta ve işlemektedir. Çoğu zaman servis sağlayıcılar tarafından işlenen bu veriler daha fazla değer üretilmesi, analizler yapılması amacıyla araştırmacılarla paylaşılabilmekte, çeşitli sorgular için genel özellik bilgileri verilebilmekte bunun sonucunda bağlantı

saldırıları ve arka plan bilgisiyle veri sahipleri hakkında çeşitli çıkarımlar yapılarak kimlikleri açığa çıkarabilmektedir [6]. Dolayısıyla trafik sistemlerinin tahmin üretmesinin yanı sıra, veri mahremiyeti kapsamında konum mahremiyeti ve kişiye özel yörünge verisinin mahremiyeti dikkate alınarak çözümler geliştirilmesi gerekmektedir.

Problem Tanımı:

Trafik sistemlerinin ihtiyaç duyduğu trafik parametrelerinin tahmini kapsamlı bir araştırma konusu olmakla birlikte trafik tahmini hava durumu, olaylar, yol kazaları, veri toplama gibi birçok dış etkenden etkilenen zorlu bir süreçtir. Trafik tahmin çalışmalarında bahsedilen dış etkenlere de bağımlı olmadan tahmin doğruluğunu arttıracak sistemlere ihtiyaç vardır. Trafik tahmini üzerine yapılan çalışmalarda simülasyon tabanlı sentetik verilerle tahminler üretilmiş veya paylaşılan, herkese açık az sayıda gerçek verilerle veri odaklı çözümler geliştirilmiştir. Veri odaklı tahmin çalışmaları döngü detektörü, mikrodalga sensörü, kamera ve radar cihazı gibi yola yerleştirilen sabit veri kaynaklarından elde edilen verilerle yapılmakta fakat bu veri kaynaklarının her yola yerleştirilmesi hem büyük bir maliyet getirmekte hem de bu kaynaklarından elde edilen veriler ile sadece o yola ait tahminler üretilmektedir. Veri odaklı tahmin modelleri için ayrıca farklı zaman dilimlerinde hava durumu, kaza bildirimi gibi ek olaylara da ihtiyaç duyulmaktadır. Bu kapsamda ele alınan problem için trafik hız tahmin modeli, “istenilen yol bölümünde” ve “kısa zamanlı tahminde ek bir olaya ihtiyaç duymadan “kendi değişkenliği” içinde tahmin üretmektedir.

İnternetin yaygınlaşması ve teknolojinin gelişmesiyle katılımcı cihazlar trafik yönetim sistemlerine veri sağlarken katılımcılara ait kişisel veriler ilgili kurumların sunucularında saklanmaktadır. Elde edilen ve saklanan verilerden anlamlı, faydalı sonuçlar çıkarılsa da kişisel verilerin mahremiyetin garanti edildiği uygulamaların geliştirilmesi gerekmektedir. Yapılan tez çalışması kapsamında geliştirilen yeni bir tahmin modelinin yanı sıra ayrıca mahremiyet korumalı trafik tahmin modeli ve konum bilgisinden elde edilebilen yörünge bilgisinin mahremiyetini sağlama adına geliştirilen model yeni ve özgün birer model olarak literatüre kazanım olacaktır.

Tez çalışmasının amacı için belirlenen araştırma soruları ve hipotezler şu şekildedir;

Araştırma Sorusu - 1: Ayrıştırılan ve filtrelenen araç konum verileri, yakınlık ve yoğunluğa göre kümelenip anlık trafik yoğunluğunun belirlenmesi ile yol karakteristiği oluşturulabilir mi? Yol karakteristiği oluşturulursa ve tekrarlayan örüntüler dikkate alınırsa kısa zamanlı trafik tahmini yapılabilir mi?

Hipotez - 1:

Bir bölgeye ait araç konum verileri ayrıştırılarak, filtrelenerek ve araçların yoğunluk ve yakınlık durumuna göre kümelenecek anlık trafik yoğunluğu tespit edilen yolun karakteristiği oluşturulur ve tekrarlayan örüntüler dikkate alınarak kısa zamanlı trafik tahmini yapılır.

Araştırma Sorusu - 2: Trafik tahmin modelinde yol karakteristiği oluşturulan verilere mahremiyet yöntemi uygulanırsa mahremiyeti korunmuş verilerle orijinal küme özniteliklerine yakın tahmin yapılır mı?

Hipotez - 2:

Geliştirilen tahmin modelinin bir aşamasında konum mahremiyet yöntemi uygulanırsa mahremiyet dikkate alınarak orijinal küme özniteliklerine yakın trafik tahmini yapılır.

Araştırma Sorusu - 3: Dağıtık GPS araç verileri zamana bilgisi ile sıralandığında elde edilen yörünge bilgilerine ait konum verilerine ardışık olarak gruplanarak mahremiyet yöntemi uygulanırsa veri yayınlama için yörünge bilgilerinin mahremiyeti sağlanır mı?

Hipotez - 3:

Araç yörünge verilerine ait konum bilgileri ardışık olarak gruplanarak konum bilgilerinin ortalaması alınır ve gürültü eklenirse yörünge verilerinin mahremiyeti sağlanır.

Tezin Amacı:

Tez kapsamında yapılan çalışmada;

- Kısa bir zaman periyodu içinde, veri odaklı trafik hız tahmin modeli geliştirmek,

- Geliştirilen modelde, dağıtık verinin sunduğu fırsat ile istenilen bölge/yol için filtreleme yaparak ilgilenilen yol segmentinde yoğunluk tabanlı kümelerle yol karakteristiği oluşturarak anlık trafik durumunu belirlemek ve yol renklendirmesini yapmak,
- Karakteristiği oluşan yollarda tekrarlayan örüntüler dikkate alınarak kısa zamanlı trafik hız tahmini üretmek,
- Kısa zamanlı trafik tahmini yapılırken veri mahremiyeti kapsamında konum ve tanımlayıcı bilgi mahremiyeti sağlamak,
- Mahremiyeti sağlanan verilerle birlikte istatistiksel sorgu gerçekleştiren kişilere araç kimlik bilgileri ayırt edilemeyecek şekilde sonuçlar döndürmek,
- Mahremiyet korumalı olarak trafik hız tahmini yapmak,
- Araç verilerine ait konum verilerinin zaman bilgisiyle sıralandığında oluşan yörünge bilgilerine mahremiyet yöntemi uygulanarak araç yörünge bilgilerinin mahrem hale getirmek amaçlanmıştır.

Tezin Önemi:

Trafik akış ve yoğunluk durumunun doğru şekilde gösterimi ile yolculuk yapacak kişiler seyahat güzergâhlarını belirleyebilmekte, seyahat sürelerini hesaplayabilmekte ve alternatif rotaları listeleyebilmektedir. Kullanıcılara bu kolaylıkları sağlamak için ortaya çıkan akıllı ulaşım sistemleri (AUS) mevcut ulaştırma sistemlerinin gelişmesini, seyahat güvenliğinin artırılması ve seyahat eden kişilere birçok alternatif sunma gibi özellikleri sağlamaktadır. AUS, gelişmiş yolcu bilgi sistemleri, otomatik sürüş sistemleri, rota belirleme sistemleri, filo yönetimi, araçtan araca iletişim sistemleri gibi teknolojileri kapsamaktadır. Katılımcı cihazlardan elde edilen verilerle hızla büyüyen bu alanda, trafik akışını anlık hesaplayacak, dış etkilerin ve ani durumların etkisini en aza indirecek güçlü karar sistemlerine ihtiyaç duyulurken bir yandan mahremiyet odaklı çözümlerin geliştirilmesi beklenmektedir. Trafik için tahmin üretirken yüksek tahmin doğruluğu sağlayacak sistemlerle beraber mahremiyet korumalı sistemlerin geliştirilmesi trafik yönetimi için son derece önem arz etmektedir.

Tez kapsamında yapılan çalışmalar sonucunda yukarıda bahsedilen ihtiyaçlara yönelik olarak üç yeni ve özgün model geliştirilerek literatüre kazandırılmış, bunlardan ilkinde trafik yönetim sistemleri için yeni bir trafik tahmin modeli geliştirilmiş, ikinci olarak trafik tahmin

modeli içerisinde mahremiyet yöntemi uygulanarak elde edilen tahmin sonuçlarının mahremiyeti sağlanmış, son olarak konum bilgilerinden elde edilebilen yörünge bilgilerinin mahremiyetini sağlayacak bir model önerilmiştir.

Hazırlanan tez çalışmasının akışı şu şekildedir;

- Birinci bölümde, problem tanımı yapılmış, çalışma için araştırma soruları ve hipotezler verilmiş, tezin amacı, önemi sunulmuştur.
- İkinci bölümde akıllı ulaşım sistemleri ve çeşitleri anlatılmış, trafik için veri kaynakları örneklendirilmiş, temel trafik parametreleri açıklanmış ve trafik parametreleri üzerine yapılan literatür çalışmaları sunulmuştur.
- Üçüncü bölümde geliştirilen modeller için kullanılan veri kümesinin detayları, özellikleri verilmiş ve model aşamalarında yer alan yöntemler açıklanmıştır.
- Dördüncü bölümde dokuz aşamadan oluşan yeni ve özgün bir trafik hız tahmin modeli ve modele ait her aşama detaylıca anlatılmış, Ankara iline ait iki ana güzergâh olan Eskişehir ve İstanbul Yollarına ait araç verileri ile test edilerek sonuçları değerlendirilmiştir.
- Beşinci bölümde; dördüncü bölümde geliştirilen trafik hız tahmin modeli mahremiyet korumalı olarak ele alınmış, dokuz aşamalı modele ek olarak öznitelik oluşumu sırasında diferansiyel mahremiyet uygulanarak mahremiyet korumalı trafik hız tahmin modeli geliştirilmiştir. Öznitelik bilgileri trafik yol karakteristik bilgisini oluşturacağından bu bilgilerin istatistiki koruma altına alınması mahremiyeti sağlarken, mahremiyet korumalı istatistiki bilginin paylaşılması trafik yöneten sistemler ve trafik araştırmacıları için kolaylık sağlayacak, trafik tahmin modeli mahremiyet korumalı tahminler üretecektir.
- Altıncı bölümde araç konum verilerinin zamana göre sıralanması ile elde edilebilen yörünge bilgilerinin, mahremiyeti için pencere tabanlı ve diferansiyel mahremiyet yöntemi kullanarak veri yayınlama için bir model önerilmiş, Eskişehir ve İstanbul Yollarında oluşan araç yörünge verilerinde test edilmiş, sonuçlarını sunularak değerlendirilmiştir.
- Son bölümde ise tez kapsamında yapılan çalışmaların ve önerilen modellerin değerlendirilmesi yapılmış, literatür katkıları sunulmuş, gelecek çalışmalara oluşturabilecek projeksiyonlar verilmiş, çalışma sonuçlarından elde edilen grafikler yorumlanmış ve değerlendirilmiştir.



2. TRAFİK YÖNETİM EKOSİSTEMİ

Bu bölümde, tez kapsamında geliştirilecek olan çalışmaların katkılarının daha iyi anlaşılması için Akıllı Ulaşım Sistemleri alt başlıklarıyla açıklanmış, trafik sistemleri için veri kaynakları, trafik tahmin parametreleri, bu parametrelerin tahmin çeşitleri detaylı anlatılmıştır.

2.1. Akıllı Ulaşım Sistemleri

Trafik yönetim sistemleri şehirleri yönetmek ve hayatı kolaylaştırmak için akıllı şehir uygulamaları geliştirilmektedir. Bir şehrin akıllı olması, şehirden elde edilen temel bilgilerin anlaşılması, analiz edilmesi, birleştirilmesi ve bu bilgilerin şehri yönetmek için bilgi ve iletişim teknolojilerinde kullanılması olarak ifade edilmektedir [7]. Akıllı şehir uygulamaları aynı zamanda günlük yaşam, çevre koruma ve toplum güvenliği gibi konuların yanı sıra ticari ve endüstriyel faaliyetleri de düzenlemektedir. Akıllı şehirlerin amaçları arasında şehir sakinlerinin hayatını düzenleyen kararların alınmasını sağlayacak akıllı ulaşım sistemleri ve akıllı şehir yönetim sistemlerini oluşturmak hedeflenmiştir. Akıllı şehirler için ulaşım sistemleri oluşturmak üzerine çalışmalar yoğunlaşmakta ve adına akıllı ulaşım sistemi adı verilen sistemler geliştirilmektedir. Akıllı ulaşım sistemleri (AUS), ekonomik performansı, güvenliği, hareketi ve çevresel sürdürülebilirliği geliştirmek için bilgi ve iletişim teknolojilerinin ulaşım altyapısı ile bütünleştirilmesi olarak adlandırılmaktadır [8]. AUS'ın belediyelerin ve kamu kurumların şehirleri daha iyi yönetmesi, araçların daha güvenli seyahat etmesi, işe giden vatandaşların seyahat süresini düzenlemesi veya daha düşük yakıt tüketimi gibi konular için fayda sağlanması amaçlanmıştır [9]. Bu sistemlerin uygulanması pahalı olsa da yakıt tüketiminin azaltılması, kazaların ve kazalı ölüm oranlarının düşürülmesi, emniyet teşkilatı, itfaiye gibi kamu kurum hizmet etkinliğinin artmasından dolayı maliyetini çok daha fazlasıyla karşılayabilmektedir. Araç trafiği, toplu ulaşım, yük taşımacılığı ve hatta yaya trafiği dâhil olmak üzere çeşitli seyahat taleplerindeki artışın bir sonucu olarak ulaşımına bağlı trafik sıkışıklığı, kazalar ve kirlilik sorunları gibi ciddi problemleri çözen AUS'ın etkin çalışabilmesi için veri toplama, veri analizi ve bilgi iletim aşamalarının eksiksiz ve dikkatli yapılması gerekmektedir [10].

Ülkemizde özellikle kamu kurumları AUS üzerine sürekli geliřtirmeler yapmakta bařta büyükřehirler olmak üzere kendi bünyelerinde merkezler oluřturmaktadır. Örnek olması adına İBB ulařım yönetim merkezi řekil 2.1 ile görölmektedir.



řekil 2.1. İBB Ulařım Yönetim Merkezi [11]

AUS bilgi ve ulařım sistemlerinin entegre edildiđi, birlikte kullanıldıđı geniş kapsamlı sistemler olmasından dolayı farklı alt sistemlere ayrılmıřtır. Akıllı ulařım sistemleri akıllı ulařım ve akıllı araçlar olarak kategorik ayrılırken [12], kullanım alanlarına göre geliřmiř ulařım yönetim sistemleri, geliřmiř yolcu bilgi sistemleri, geliřmiř araç kontrol sistemleri, ticari araç yönetimi, geliřmiř toplu tařıma yönetimi, geliřmiř řehir ulařım sistemleri olarak ele alınmaktadır [13]. Farklı kaynaklarda farklı sınıflandırmalar olsa temel olarak alıřmalar geliřmiř trafik yönetimi, geliřmiř yolcu bilgi sistemleri, geliřmiř toplu tařıma sistemleri ve akıllı ulařım ücretlendirme üzerinedir [14] ve alt bařlıklarda açıklanmıřtır.

2.1.1. Geliřmiř trafik yönetim sistemleri (GTYS)

GTYS son yıllarda farklı yöntemlerle arařtırılan ve etkin bir řekilde kullanılan önemli bir AUS alt dalıdır [15,16]. Trafik polis teřkilatı ve trafiđi yöneten belediye vb. kurumlar tarafından kullanılan GTYS trafiđin akıřını izlemekte ve zamanında kararlar alarak trafiđi yönetmektedir. Yönetilen sistemler sayesinde trafiđin yönlendirilmesi, sinyalizasyon düzenlenmesi, yol alıřmaları gibi farklı konularda kararlar alınmakta, yolcu ve sürücüler

için elektronik işaretçilerle bilgiler sağlanmaktadır. Tüm bu veri akışı ve bilgilendirmeler için trafik ağını oluşturan karayolu ağı ve altyapısına yerleştirilen detektörler, sensörler, kamera görüntüleri, araçtan araca-aracıdan altyapıya sistemlerden elde edilen bilgiler kullanılmaktadır.

2.1.2. Gelişmiş yolcu bilgi sistemleri (GYBS)

GYBS, seyahat eden kişilerin hayatını kolaylaştıran ve seyahat sürelerini verimli kullanılmasını sağlayan önemli bir AUS uygulama alanıdır [17]. Bu sistemler kullanıcılara yolculuğa ne zaman çıkacakları, gidecekleri alternatif yolları veya seyahatin özel araç-toplu taşıma gibi seçenekleri internet, mobil cihaz, radyo gibi farklı teknolojilerle sunmaktadır. GYBS, seyahat öncesinde farklı karayolu alternatifleri sunarken seyahat esnasında sağladığı bilgilerle ve gerekli yönlendirmelerle gezinme süresini azaltabilmekte, trafiğin olağan akışında tıkanıklıkları da önleyebilmektedir. Yönlendirmeler ve bilgilendirmelerle güzergâh hakkında bilgisi olmayan yolcular için oldukça kullanışlı iken her yolcu için güzergâh üzerindeki ani değişimleri bilgilendirmesinden dolayı da oldukça önemlidir. Ayrıca tekrar eden trafik koşullarının önceden bilinmesi de yolcular için seyahat sürelerini azaltabilmektedir. GYBS sistemlerde gerçek zamanlı trafik bilgilerine göre rota seçimi yapan yolcular doğru yönlendirilmediği takdirde şehir trafiği de olumsuz anlamda etkilenmektedir [18].

2.1.3. Gelişmiş toplu taşıma sistemleri (GTTS)

Kent sakinlerini toplu taşıma kullanımını teşvik etmek şehir trafiğini etkin bir biçimde azaltmakta, belediye ve özel kurumlar kentlerde yaşayan ve özellikle çalışan insanlar için toplu ulaşım araçları sağlamaktadır. Toplu ulaşımın yönetimi ve kontrolünü sağlayan GTTS, trafik yönetim sistemlerinin bir parçası olarak toplu ulaşım ile ilgili bilgileri sağlamakla sorumludur [19,20]. GTTS, çalışanların işe gidiş geliş süreleri düzenlenirken, trafik sıkışıklığını ve karbon emisyonu azaltarak güvenliği, güvenilirliği ve verimliliği arttırmayı amaçlamaktadır. Ayrıca toplu taşıma araçlarına yerleştirilen sensörler ile gerçek zamanlı araç takibi, kaza/olay bildirimi, sürücü ve araç izleme gibi trafik bilgi sistemlerinin ihtiyacı olan bilgiyi de sağlamaktadır.

2.1.4. Ulaşım ücretlendirme sistemleri (UÜS)

Ulaşım ücretlendirme sistemleri, belirli saatler arasında ücretlendirme yaparak trafiği düzenlemeyi amaçlamaktadır. Belirlenmiş bölgelerdeki trafik, elektronik algılayıcılar ile yol ücretlendirme yapılarak düzenlenirken bazı durumlarda dinamik ücretlendirme de yapılmaktadır. Ulaşım ücretlendirmenin dinamik yapılması, araçların yönlendirilmesi, otoyol/park ücretlerinin en aza indirilmesi, elektrikli araç şarj istasyonlarının doluluk oranları, otopark sahiplerinin/şarj istasyonlarının kar oranı artırma gibi konuları da ele almaktadır [21].

AUS sınıflandırma konuları açısından daha da çeşitlendirilebilir ve detaylandırılabilir. Farklı başlıklar üzerinden ele alınan ulaşım sistemleri teoride kalmaması ve uygulamaya geçilmesi için Avrupa Birliği (AB) tarafından bir aksiyon planı da hazırlanmıştır. AB'nin hazırladığı aksiyon planına göre akıllı ulaşım sistemleri;

- Yol, trafik ve seyahat verisinin verimli şekilde kullanımını,
- Trafik ve nakliye yönetiminin sürekliliğini,
- Karayolu emniyeti ve güvenliğini,
- Araç ve ulaşım altyapı entegrasyonu,
- Veri güvenliği-korunmasını,
- Avrupa AUS koordinasyonu,

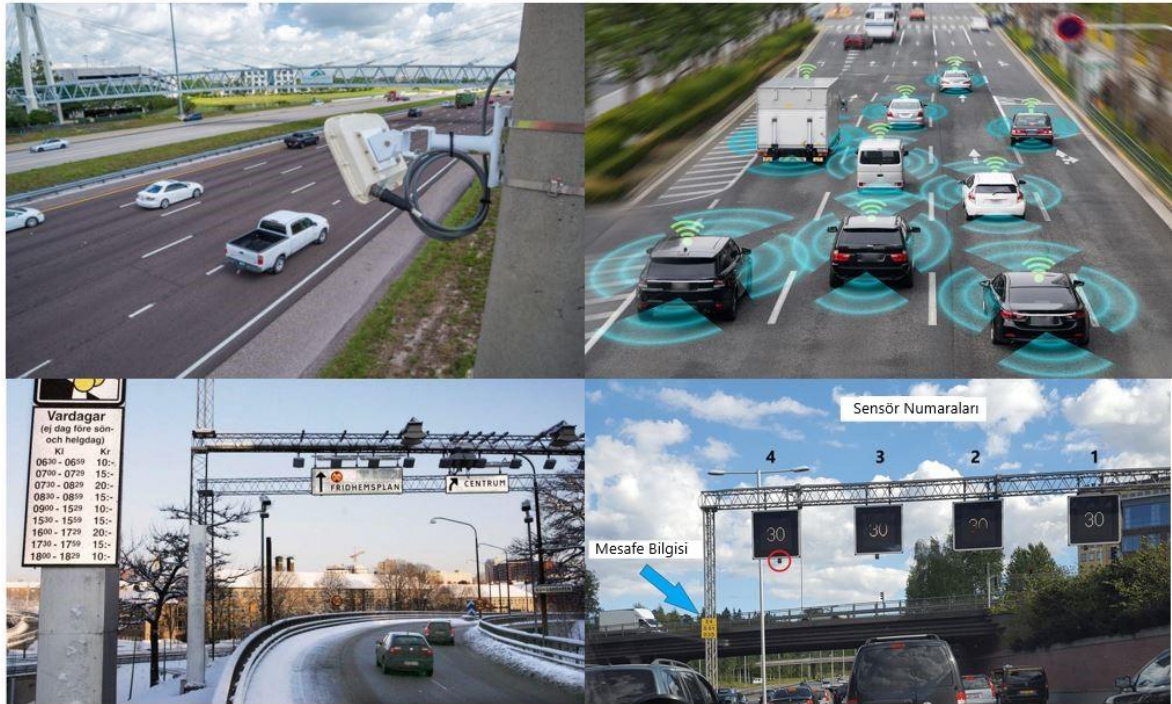
konularını ele almalıdır [22].

AUS için uygulanan politikaların temelinde, ulaşım ve seyahatin daha verimli, daha güvenli ve daha temiz çevresel şartlarla gerçekleşmesi vardır. Hazırlanan aksiyon planıyla farklı ulaşım sistemlerinin ve altyapılarının bütünleştirilmesi, karayolu trafiğinin koordine edilmesi ve hızlandırılması hedeflenmiştir. Trafiğin koordinasyonu ve yönetimi için elde edilen verinin türü, kalitesi ve verimli kullanımı oldukça önemlidir. Trafikte elde edilen veriler, geçmiş örüntülerin tanımlanması, anlık bilgilendirmenin yapılması ve gelecek planların hazırlanmasında kullanılmaktadır. Veriler elde edildikten sonra ilgili trafik koşul ve parametreleri tahmin edilerek gerçek zamanlı ve güvenilir bir trafik yönetimi

sağlanmaktadır. Takip eden başlıkta trafik için veri sağlayan başlıca veri kaynakları açıklanmış ve örneklendirilmiştir.

2.2. Trafik Veri Kaynakları

Günümüz güncel teknolojileri kullanan akıllı şehir uygulamaları şehri yöneten ve düzenleyen kurumların yönetim etkinliğini arttırmaktadır. Belediye ve devlet kurumları, bilgi ve iletişim araçları ile şehir sakinlerine hayatı kolaylaştıracak servisleri sağlarken aynı zamanda şehirleri izlemek, istatistiki bilgi tutmak ve şehirle ilgili çıkarımlar yapabilmek için veri toplamaktadır. Akıllı şehir uygulamaları için farklı veri toplama araçları mevcuttur ve veri toplamak için araçlar; mobil cihazlar, mikrodalga sensörü, döngü detektörü, video kamera sistemleri, araç özel ağları, radar cihazları olarak sıralanabilir [23]. Tüm bu araçlar etkin olarak trafik takibinde kullanılırken bir yandan da maliyet getirmektedir. Trafiğin takibi için daha az maliyetli RFID teknolojisi ve kablosuz sensörler de kullanılmakta [24,25] bunun yanı sıra internetin oldukça yaygınlaşması ile IoT cihazlardan da faydalanılmaktadır [26]. Trafik yönetim sistemleri için yaygın kullanılan bazı veri kaynakları alt başlıklarda açıklanmış ve örnek veri kaynakları Şekil 2.2 ile gösterilmiştir.



Şekil 2.2. Trafik için veri kaynakları [27-30]

2.2.1. Mobil cihazlar

İnternete bağı hareketli cihazların trafikte seyahat eden araçlarda kullanılması sonucu elde edilen veriler sürekli çoğalmakta ve trafik için büyük miktarda veri sağlamaktadır. Mobil ve hareketli cihazların sağladığı küresel konumlama verisi (GPS) kullanarak araçlar kolaylıkla takip edilebilmekte, coğrafi bilgi sistemleri ve haritalar ile birleştirilen hareket verileri sayesinde etkin çözümler üretilmektedir. Mobil verilerden elde edilen konum, hız, zaman gibi bilgiler trafik akış, yoğunluk, rota gibi trafik parametrelerinin belirlenmesinde ve tahmininde önemli rol oynamaktadır [31].

2.2.2. Mikrodalga sensörleri

Hareket eden cisimleri algılamak için kullanılan mikrodalga sensörleri, temel olarak hedef bölgeye mikrodalga sinyali gönderip tekrar alarak hareketli ve durağan nesneleri tespit etmektedir. Eğer bir araç tespit alanında ise geri dönüş sinyali güçlenerek eşik değeri aşar ve araç tespit edilir. Trafiği engellemeden ve ucuz maliyetinden dolayı birçok kurum tarafından tercih edilen sensörler trafiğin izlenmesi için kavşaklarda ve karayolunda belirli noktalara yerleştirilmektedir ve trafiğin temel parametreleri ölçülmektedir [32].

2.2.3. Döngü detektörleri

Bir döngü detektörü, yalıtılmış bir veya birkaç telin yola yerleştirilmesi ve manyetik alan oluşturması ile araç tespiti yapmaktadır. Bu tespit sistemi belirlenmiş noktalarda, kavşaklarda veya bariyerli sistemlerde kullanılmakta, istenilen araç boyutunda ayarlanabilmektedir. Döngü detektörlerinde veriler, araçların hızını, kayıt zamanını, geçen araç sayısını, kendi kimlik ve konum bilgisi ile birleştirilerek saklanır. Video kamera ve radar cihazlara nazaran daha ucuz maliyetlidir ve yol boyunca istenilen aralıklarla uygulanabilmektedir. Döngü detektörleri sıcaklık, titreşim ve karayolu yapısal değişimlerden kolaylıkla etkilenmekte, bağlantı hataları ve yola yerleştirilen kabloların yalıtımının yeterli yapılamamasından dolayı da problemler yaşanabilmektedir [33].

2.2.4. Video kameralar

Şehrin belirli noktalarına konumlandırılan video kamera teknolojileri ile araçların ve yolların durumları gözlemlenebilmektedir. Video izleme ile trafiğin akışı, yoğunluğu güncellenebilmekte, araç tipi ve hızı tespit edilebilmekte, araç plaka okuma sistemleri ile güvenlik sağlanabilmektedir. Tüm bu işlemler araçların sabit olarak izlenmesi ile değil araçları ve farklı hareketli cisimleri ayırt edecek, takip edecek sistemler ile mümkündür. Video kameralarla farklı araç izleme ve ölçüm sistemleri geliştirilse de kamera görüntü sistemleri için gün içindeki farklı zaman dilimleri ve farklı hava koşulları da göz önünde bulundurulmalıdır [34].

2.2.5. Özel araç ağları

Özel araç ağları, kablosuz iletişim yoluyla araçtan araca haberleşmeyi gerçekleştiren ve her aracı bir ağ elemanı gibi kullanan bir tür araç iletişim ağıdır. Kablosuz ağ sensörlerin daha yapısal hale getirildiği özel araç ağlarında (VANET) araçlar kendi aralarında (V2V) haberleşebildiği gibi, karayolu kenarında sabitlenmiş bir ünite ile iletişim (V2I) sağlayabilmektedir. V2V iletişimde araçlar hız, konum, hareket yönü gibi bilgileri değişim mesajları ile gönderirken V2I ağlarda araçlar benzer bilgileri sabit yol ünitesine iletmektedir. Literatürde özel araç ağları için üretilen trafik verisi hareketli araç verisi (FCD) olarak ifade edilmekte, araçlara gömülü cihaz ve sistemlerle elde edilmektedir. Özel araç ağından elde edilen veriler trafik tahmini ve kaza-acil durum bildirimleri için kullanılmaktadır [35,36].

Trafik izlemek için birçok veri kaynağı mevcuttur ve kaynakların sağladığı veri ile trafiği izlemenin temel amacı, devlet kurumu, ticari firma ve şehir sakini gibi tüm paydaşlar için daha iyi karar sistemleri oluşturmaktır. Trafik karar sistemleri için trafik parametrelerinin izlenmesi, geçmiş verilerin incelenmesi ve ileriye yönelik doğru tahminler yapılması gerekmektedir. Karar sistemleri için ele alınan ve tahmin çalışmalarında kullanılan temel trafik parametreleri takip eden başlıkta anlatılmıştır.

2.3. Trafik Parametreleri ve Tahmini

Trafik araştırmaları ve mühendisliği, ekonomik koşullar dikkate alınarak sorunsuz, güvenli trafik uygulamaları geliştirmek için trafik davranışlarını incelemektedir [3]. Trafiğin

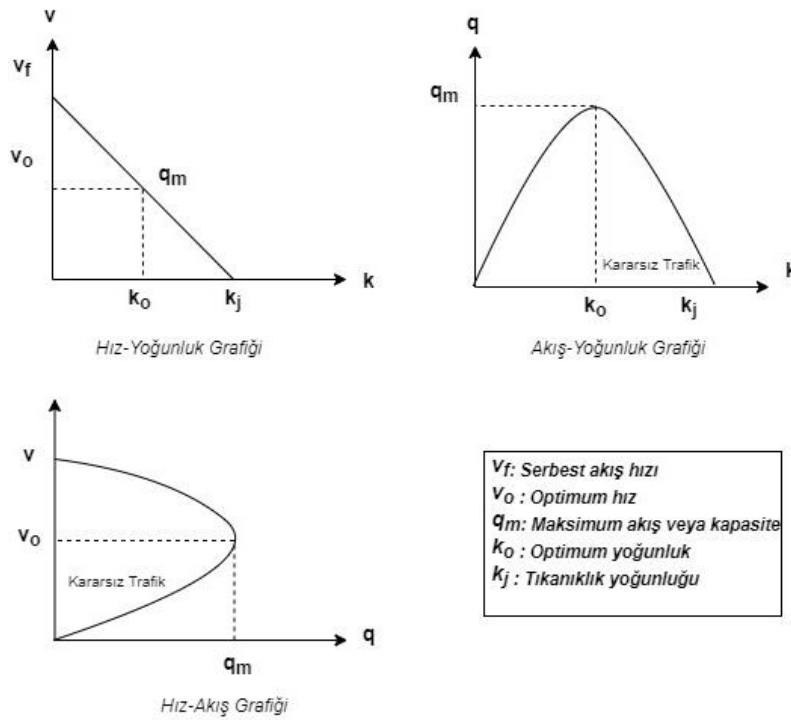
davranışını ve karakteristiğini anlamak, bir trafik ağını tanımlamak için trafik parametreleri kullanılmakta ve temel parametreler olarak akış, yoğunluk ve hız olarak ifade edilmektedir [4,5]. Temel parametreler homojen ve akan bir trafikte Eş. 2.1 ile hesaplanabilmektedir [37,38].

$$q = k.v \quad (2.1)$$

Denklemde;

- Bir noktadan birim zamanda geçen araç sayısı: Trafik akışı (q),
- Belirli bir uzunluktaki yolda ölçülen araç sayısı: Trafik yoğunluğu (k),
- Belirli bir zamanda ölçülen düzlemin araç ortalama hızı: Trafik hızı (v) ile ifade edilmiştir.

Temel trafik parametreleri farklı denklemler ile de modellenebilirken birbirleri ile olan basit ilişkisi Şekil 2.3’de grafiklerle gösterilmiştir.



Şekil 2.3. Trafik Akış-Yoğunluk ve Hız ilişkisi [39]

Trafik yönetimi ve kontrolü için yapılan tahmin çalışmaları için farklı sınıflandırmalar yapılmıştır. Zamansal olarak kısa ve uzun dönem olarak sınıflandırılan çalışmalar [40] model tabanlı ve veri odaklı yaklaşımlar olarak da sınıflandırılmıştır [41,42]. Model tabanlı

yaklaşımlar simülasyon yoluyla yol ağını matematiksel ve hesaplamalı olarak modelleyip sürücü ve araçların davranışlarını analiz ederken veri odaklı çalışmalar tahmin için herhangi bir ön kabul yapmadan sadece mevcut veri türüne bağlı olarak sonuç üreterek tahmin yapmaktadır. Tez kapsamında incelenen tahmin çalışmaları zaman aralığına ve veri türüne göre bakılmaksızın trafik parametrelerine göre sınıflandırılmış, yapılan çalışmalar incelenmiş ve alt başlıklarda sunulmuştur.

2.3.1. Trafik akışı üzerine yapılan çalışmalar

Sabit bir noktadan ölçülen ve birim zamanda geçen araç sayısı olarak ifade edilen trafik akışını ölçmede genellikle döngü detektörü, mikro dalga sensörü gibi bir yol segmentine yerleştirilen veri kaynaklarından elde edilen trafik verisi kullanılmaktadır. Trafik akışının tahmini trafik yönetimi ve kontrolünde önemli bir rol oynarken trafik güvenliğinin artırılmasına, trafik verimliliğine, trafik sıkışıklığının azaltılmasına konularında önemli ölçüde fayda sağlamaktadır. Karayolu ağındaki doğru trafik akışı tahmini AUS'nin trafik yönetim kararlarının önceden alınmasında ve verimli olarak uygulanmasında önemli bilgiler sunmaktadır [43]. Trafik çalışmaları için trafik akış tahmini çok eskiden beri araştırma konusu olup öncelikli bir çalışma alanıdır. Trafik akış tahmini için literatürde de birçok çalışma yapılmış, bu çalışmalardan öne çıkanlar liste halinde verilmiştir.

- Guo vd. [44], kısa zamanlı bir trafik akış tahmini yapmak için yeni bir model önermiştir. İki aşamadan oluşan model, veri ön işleminde spektrum analizi kullanılmış, Gri Sistem Modeli ile tahmin üretmiştir. Zaman serisi ARIMA yöntemi ile kıyaslanan modelde tahmin için Londra şehir merkezinde bulunan yaklaşık 6000 döngü detektörlerinden elde edilen veri kullanılmış ve şehrin hem normal hem de olay bulunan trafik durumları gözlemlenmiştir.
- Yuan vd. [45], kümeleme analizine dayalı bir yapay sinir ağı yaklaşımı önererek kısa zamanlı trafik akış tahmini gerçekleştirmiştir. Önerilen yaklaşımda akış tahmini gerçekleştirmeden önce trafik örüntülerini gruplamak için kümeleme yöntemi kullanılmış, kümeleme sonucuna göre farklı veri tabanları oluşturulmuş, uygun veri tabanı seçilerek geri yayımlı sinir ağı yöntemi ile trafik tahmini yapılmıştır. Çalışmada 4500 döngü detektöründen elde edilen trafik verisi kullanılmıştır.
- Zhao vd. [46], büyük bir trafik veri kümesinde trafik akışına ait kuralların çıkarılmasının zorluğundan bahsetmiş, akışa ait mekan-zaman ilişkilerinin kullanılmamasından dolayı

tahmin güçlüğünü vurgulamıştır. Yazarlar yollara ait hem gelecek hem de geçmiş kavşak bilgilerini tamamıyla kullanabilen çift yönlü regresyon sinir ağı modeli önermişler ve daha önceki modellerin zaman serilerine bağlı olarak bir önceki zamandan bir sonraki tahmin üretme eksikliğine çözüm bulduklarını ifade etmişlerdir. Kısa zamanlı trafik akış tahmini yapılan çalışmada, kavşaklardaki izleme noktalarından elde edilen verilerde kavşak ismi, yönü gibi kavşaklara ait bilgilerin yanı sıra araçlara ait plaka, araç rengi ve şerit bilgisi gibi bilgiler kullanılmıştır.

- Loan vd. [47], trafik akış tahmini için mekânsal ve zamansal özellikleri dikkate alan bir derin öğrenme modeli önermiştir. Yapılacak olan tahmin doğruluğunu sağlama konusunda şehir yol ağlarının çok karmaşık yapıda olabileceğini göz önünde bulundurarak yollara uygun mekan-zaman özelliklerin çıkarılması gerekliliğini vurgulamıştır. Çalışmada 30 yol segmenti ve 24 kavşak verisi kullanılmış, veri kümesinde bulunan yollara ait bilgilerle mekânsal özellikler, zaman aralıkları ile zamansal özellikler çıkarılarak tahmin üretilmiştir.
- Duan vd. [48], trafik verisinin akışkan özellikleri olduğunu fakat veri için mesafe, hızlanma ve eylemsizlik gibi mekanik özelliklerinin çıkarılması ile daha doğru tahminler üretileceğini öne sürmüştür. Mekanik özellikleri çıkarılan trafik verisi ve gri sistem modeli ile kısa zamanlı akış tahmini gerçekleştirilmiş, 28 km uzunluğunda yol segmentinde bulunan dokuz adet döngü detektöründen elde edilen veri kullanılmıştır.
- Polson ve Sokolov [49], trafik akışının serbest akış, kırıma, düzelleme ve tıkanıklık durumları arasında keskin lineer olmayan geçişlerine çözüm üreten yeni bir derin öğrenme modeli geliştirilmiştir. İlk aşamada mekan-zamansal ilişkiler tanımlanmış, sonrasında lineer olmayan ilişkiler modellenmiştir. Illinois ulaşım departmanına ait 21 döngü detektörü verisi kullanılan çalışmada belirlenen yol segmentinde ülke çapında iki büyük sosyal etkinlik zaman dilimi seçilerek kısa zamanlı tahmin gerçekleştirilmiştir.
- Zhao vd. [50], kısa zamanlı trafik akışı için LSTM tabanlı derin öğrenme modeli önermiştir. Önerilen modelde ilk olarak yollar arası ilişkilerin belirlendiği çıkış-hedef nokta matrisleri oluşturulmuştur. Oluşan matrisler girdi değeri olarak sinir ağına verilmiş ve trafik tahmini için iki boyutu mekan-zamansal özellikleri tanımlayan çok katmanlı bir LSTM ağı modellenmiştir. Çalışma için Pekin Trafik Bürosu tarafından sağlanan ve yollara yerleştirilen 500 gözlem noktasının kamera ve radar verisi kullanılmıştır.
- Qu vd. [51], kavşak noktalarındaki tıkanma problemine yönelik çalışmada, akış tahminlerinin 1-5 dakika arasında yapıldığını ve bu sürenin kısa olduğunu belirterek tahmin çalışmaların 30 dakika sonrasında yapılması gerektiğini ifade etmiştir. Çalışmada

tarihsel trafik akış verilerinin benzer örüntülerini ve aralarındaki ilişkileri bulmak için KNN ve kümeleme yöntemleri kullanılmış, geri beslemeli ve Elman sinir ağları ile tahmin üretilmiştir. Çin'in Jinan şehrine ait trafik verileri, araç sayısı, kuyruk uzunluğu, sinyal zamanı ve zaman bilgisi özelliklerini içermektedir.

- Habtemichel ve Cetin [52], kısa zamanlı trafik tahmini gerçekleştirmek için benzer trafik davranışlarını tanımlayan ve bu örüntülerden öğrenerek tahmin üreten bir model geliştirilmiştir. Trafik örüntüleri tarihsel veriler içinden KNN yöntemiyle belirlenmiş, komşu yakınlık için korelasyon ve Öklid mesafe uzaklığı kullanılmıştır. Londra ve ABD'nin çeşitli eyaletlerine ait farklı yollara yerleştirilen sensör verisi kullanılan çalışmada tahmin aralığı için 15 dakikalık zaman dilimi belirlenmiştir.
- Li vd. [53], zaman serileri ve lineer regresyon yöntemleri gibi parametrik yöntemlerin önceden kabul yaptıklarını, parametrik olmayan modellerin sabit olmayan dağıtılmış veride zayıf tahmin sonuçları verdiğini belirterek zayıf yönleri iyileştirecek birleştirilmiş bir yapı önermiştir. Önerilen modelde hem mekânsal hem de zamansal bilgiler göz önünde bulundurulmuş, lineer trafik davranışı için ARIMA, lineer olmayan davranış için SVR yöntemini birleştirilerek kullanılmıştır.

2.3.2. Trafik yoğunluğu üzerine yapılan çalışmalar

Trafik yoğunluğu tahmini yoğun nüfus ve araç hareketliliğinin olduğu şehirlerde trafiği kontrol etmek adına önemli bir rol oynamaktadır. AUS için yoğunluk tahmini yapılarak akıllı yönlendirme, yol ve güzergâh planlama, ücretlendirme gibi konularda kullanıcılara yardımcı olunmaktadır. Fakat trafik yönetim sistemlerine sağlanan yoğunluk tahmini çalışmalarında kullanılan veri sağlayan kaynakların yerleşimi ve bakımı için yüksek maliyet gerekmektedir. Bu konuda çalışmalar sürekli devam etmektedir ve çalışmalar aşağıda liste halinde verilmiştir.

- Nam vd. [54], trafik yoğunluğunu tahmin etmek için LSTM sinir ağı yöntemini kullanmış, birbirine bağlı ve otonom araçlara ait trafiği PARAMICS simülasyon programı ile üreterek yöntemi test etmiştir. Çalışmada trafik yoğunluğu serbest akan, geçişli ve tıkanık olmak üzere 3 farklı durum için tahmin edilmiştir.
- Raj vd. [55], yolcu bilgi sistemleri tarafından seyahat eden kişilere sağlanan yoğunluk bilgisinin kişilerin yolculuk öncesi ve esnasında karar almalarına doğrudan etki ettiğini belirtmiş, bu kapsamda yapılan çalışma ile Hindistan'da belirlenen yola yerleştirilmiş

sensör verilerinden trafik yoğunluğu tahmin edilmiştir. Trafik yoğunluğu KNN ve yapay sinir ağı ile tahmin edilirken trafikte bulunan motosiklet, otobüs ve araba gibi farklı türde araçların tümü dikkate alınmıştır.

- Da Silva vd. [56], hareketli mobil verisi kullanarak trafik yoğunluğu ve tıkanıklığını yoğunluk tabanlı kümeleme yöntemi ile belirlemiştir. Yapılan çalışma ile büyük veride çalışma maliyetini düşürdüklerini, sokak isimleriyle veri kümesini bölümlere ayırdıklarını ve birleştirilen verinin tekrarlanan örüntülere ihtiyaç duymadığını ifade etmişlerdir.
- Aljamal vd. [57], lineer ve lineer olmayan trafik durumlarında özel filtreleme yaklaşımı kullanarak trafik yoğunluğunu tahmin etmiştir. Video görüntülerinin maliyetinin çok olduğu ve hareketli araç verisinin kullanım gerekliliği vurgulanan çalışmada özel araç ağında araç verileri simülasyonla üretilmiş, belirlenen sinyalizasyon noktalarından geçişler izlenmiş, kullanılan yaklaşım Kalman filtreleme yöntemiyle kıyaslanmıştır.
- Zeroual vd. [58], trafiğin serbest akış değişimi, kaza, trafik ışığı gibi farklı dinamiklerini göz önüne alarak ve istatistiksel GLR yöntemi kullanarak gözlem tabanlı trafik yoğunluğu izleme yöntemi önermiş ve yoğunluk tahmini yapmıştır. Ayrıca gözlem sisteminden çıkan verilere gürültülü ortamda başarıyı arttırmak için filtreleme işlemi uygulanmıştır. Çalışmanın değerlendirmesi için Amerika Kaliforniya eyaletinde bulunan iki farklı yol segmentinden elde edilen döngü detektör verileri kullanılmıştır.
- Abbas vd. [30], Stockholm şehrinde yollara yerleştirilen mikrodalga sensör verisi kullanarak ve LSTM sinir ağı yöntemi uygulayarak kısa zamanlı trafik yoğunluğu tahmini gerçekleştirmiştir. Yoğunluk tahmini için yollar belirlenen sensör sayısının segmentlere ayrılmış, her sensörün komşu sensörü arasındaki bağlantılar kullanarak segmentlere ait mekan-zaman özellikleri belirlenmiştir. Ayrıca okunan sensör verisi trafik araç sayısını ve hızını ölçerken, yoğunluk değerleri genelleştirilmiş akış-yoğunluk denkleminde çıkarılmıştır.
- Alkouz ve Aghbari [59], İngilizce ve Arapça dillerinde kullanılan sosyal medya hesaplarından elde edilen verilerle trafik yoğunluğu ve tıkanıklığını belirlemek için SNSJam adını verdikleri modeli geliştirmiştir. Geliştirilen modelde trafik olaylarını bildiren sosyal medya gönderileri sınıflandırılmış, gönderilere ait GPS bilgilerine göre harita üzerinde gösterilmiş ve lineer regresyon yöntemi ile tahmin yapılmıştır.
- Hu vd. [60], trafik yoğunluğunu belirlemek için Amerika ve Çin'de otoyollara yerleştirilen kamera görüntülerini sınıflandırmaya çalışmıştır. Öncelikle görüntüler hücrelere bölünmüş, her hücrenin ön plan ve arka plan görüntülerini ayırarak ilgilenen

bölge bilgisi tanımlanmış, görüntüye ait örüntü bilgileri histogram teknikleri çıkarılmış ve PCA boyut azaltma teknikleri uygulanmıştır. Bu işlemlerden görüntüler SVM tekniği ile sınıflandırılarak trafik hafif, orta ve yoğun olarak ayrılmıştır.

- Chung ve Sohn [61], video kamera görüntüleri ve CNN derin öğrenme yöntemi kullanarak trafik yoğunluğu tahmini yapmıştır. CNN yöntemi uygulanmadan önce araçları tespit etmek için eşik değerleri tanımlanmış, tespit edilen araçlar belirlenen bölgede sayılarak yoğunluk tahmini gerçekleştirilmiştir.
- Goves vd. [62], gerçek zamana yakın olan bir zaman diliminde yapılacak trafik tahmininin ulaşım sisteminde oluşabilecek tıkanıklığı azaltabileceğini, büyük miktarda oluşan veriler ve bunlardan elde edilecek örüntülerin karmaşık ve lineer olmayan problemler için bir çözüm fırsatı olduğunu ileri sürmüştür. Çalışmada yapay sinir ağı kullanılarak tarihsel ve mevcut verilerle 15 dakika sonrası için yoğunluk tahmini yapılmış, tahmin için İngiltere Manchester şehrinde farklı kara yollarına yerleştirilen döngü detektör verileri kullanılmıştır.

2.3.3. Trafik hızı üzerine yapılan çalışmalar

Ulaşım sistemlerinde trafik hız tahmini bir başka önemli parametredir. Hız tahmini üzerine yapılan çalışmaların çoğunluğu tahmin problemini zaman serisi olarak ele almıştır ve trafik veri kaynaklarının çeşitlendirilmesi ile farklı yaklaşımlarda mevcuttur. Hız tahmini üzerine çalışmalar aşağıda listelenmiştir.

- Cui vd. [63], bir yol segmenti üzerindeki trafik durumunun öncül ve devam eden yol bağlantılarıyla yakın konumda olan darboğazdan etkilendiğini, komşu segmentlere ait özniteliklerinin iyi şekilde tanımlanmasının trafik tahmini için son derece önemli olduğunu ifade etmiştir. Çalışmada geniş bir yol ağı çizge (graf) şeklinde tanımlanmış, her yola ait komşu yolların özniteliklerin mekânsal ve zamansal özelliklerinin belirlenmesi için dalga tabanlı tekrarlı bir sinir ağı modeli önerilmiştir. Belirlenen yollar için hız tahmini yapılan modelde iki farklı veri kümesi kullanılmış, veriler döngü detektörü ve hareketli araçlara ait verilerden oluşmaktadır.
- Niu vd. [64], bir yol üzerindeki araç hız tahmininin zorlu bir süreç olduğunu ve yollar üzerindeki araç hızlarının mekânsal, zamansal ve gizli olarak çeşitli faktörlerden etkilendiğini ifade etmiştir. Yaptıkları çalışmada L-U Net ismini verdikleri iki aşamadan oluşan yeni bir model önererek şehir ölçeğinde hız tahmini gerçekleştirmiştir. L-Net

aşamasında LSTM yöntemiyle zamansal, U-Net aşamasında mekânsal özellikler çıkarılmıştır. Model için Çin Chengdu şehrinde gezinen taksilere ait GPS verileri kullanılırken, her aracın hız değeri iki GPS noktası arasındaki mesafenin zaman değerine bölümü hesaplanmış ve kabul yapılmıştır.

- Le vd. [65], Gaussian yöntemiyle yapılan önceki çalışmaların akış ve hız üzerine olduğunu, kübik çalışma zamanından dolayı büyük trafik verisini ölçeklendirmedeğini ileri sürmüştür. Çalışma zamanının etkisini azaltmak için yerel Gauss yöntemi kullanılmış, matris ayrıştırma yöntemiyle yakın segmentler ve bu segmentlere ait ortalama hız değerleri gruplanmış, ilgili segmentlerin yakın zamandaki gözetlenemeyen ve verisi olmayan yollara ait hız tahmini üretmiştir. Yapılan çalışmada trafik mesaj sistemi olarak bilinen radyo frekans sisteminden elde edilen veriler kullanılmıştır.
- Ma vd. [66], mikrodalga sensör verisi kullanarak hız tahmini yapmak için yeni bir LSTM sinir ağı modeli önermiştir. Çalışmada kullanılmak üzere Çin Pekin şehrinde bulunan iki sensör arasından bir aylık trafik verisi elde edilmiş, zaman aralığı için en uygun zaman penceresini belirleyen önerilen tahmin modeli SVM, ARIMA ve Kalman Filtresi gibi farklı yöntemlerle kıyaslanmıştır.
- Li vd. [67], derin öğrenme yöntemlerinin eğitilmesi için büyük miktarda veriye ihtiyaç duyduğunu ayrıca büyük miktarda hesaplama maliyeti olduğunu ileri sürerek daha az karmaşıklığa sahip bir model önermiştir. Lineer regresyonun genelleştirilmiş formu olan kısmi en küçük kare metoduna dayalı olan model ile trafik hızının daha hızlı ve doğru şekilde tahmin edilmesi amaçlanmıştır. Çin Xian şehrinde gezinen taksi verileri kullanılan çalışmada kısa bir yol segmenti için her beş dakikalık dilimde kısa zamanlı tahmin gerçekleştirilmiştir.
- Cui vd. [68], tek ve çift yönlü LSTM mimarisi önererek ve mekan-zamansal veriden zaman bağımlılıklarını çıkararak trafik hız tahmini yapmıştır. Tahmin için Amerika Seattle bölgesinde birbirine bağlı dört farklı yola ait döngü detektörü verileri ve Kaliforniya ulaşım ajansından elde edilen sensör verileri kullanılmıştır. Çalışmada ayrıca trafik verilerinde bulunan eksik verilerin yerini doldurmak için önerilen modelde işlem gerçekleştirilmiştir.
- Zhao vd. [69], tekrar etmeyen trafik koşullarında ve trafiğin yoğun olduğu yollarda tek bir araç tipi ağır vasıta araçlarına ait hız bilgileri ile tahmin gerçekleştirmiştir. Çalışmada Çin Pekin şehri çevre otoyolunda iki farklı yol segmenti ele alınarak bu segmentlere ait bir aylık veride ilk yirmi gün eğitim verisi seçilmiş, takip eden gün bazında tahmin edilmiştir. Seçilen segmentlerde gezinen araçların GPS bilgileri kullanılmış, konum

bilgileri harita eşleme yöntemi ile yollara yerleştirilmiş, tekrarlı sinir ağı GRU yöntemi ile tahmin yapılmış ve GRU yönteminde farklı optimize fonksiyonları da denenmiştir.

- Jiang ve Fei [70], sinir ağı ve gizli Markov yönteminden oluşan iki seviyeli bir trafik hız tahmin modeli önermiştir. Araçların sürüş rotasını dikkate alan çalışma, ilk aşamada sinir ağı kullanarak tarihsel verilere dayalı ilgili yola ait hız tahmini yapılırken ikinci aşamada trafik hızı ve araç hızları arasındaki istatistiksel ilişkiler çıkarılmış, gizli Markov modeliyle araçlara özgü hızlar da tahmin edilmiştir. Tahmin için Luxemburg'da belirlenen yollar için SUMO simülasyon aracıyla hareketli araç verileri üretilmiş ve kullanılmıştır.
- Kim vd. [71], kısa zamanlı trafik tahmini yapmak için ölçeklenebilir derin bir sinir ağı modeli önermiştir. Büyük trafik verisinde mekan-zamansal ilişkilerinin çıkarılmasının büyük bir hesaplama maliyeti getirdiğini veya özel bir yol ağı için özel tasarım gerektiğini vurgulayan yazarlar, yapısal RNN oluşturmak için yol ağı ile tekrar sinir ağlarını birleştirmiştir. Yapısal RNN oluşturulmasıyla birbirine komşu yollar arasındaki etkileşim ve zaman ilişkileri çıkarılmıştır. Çalışmada İspanya-Santander şehrinden toplanan döngü detektör verileri kullanılmış ve bu veriler Avrupa Birliği projesi kapsamında paylaşılmıştır.
- Jeon ve Hong [72], her yola ait mekan-zamansal özelliklerin farklı olmasından dolayı trafik örüntülerinin çeşitliliğini vurgulamış, bunun için kullanılacak tarihsel verinin doğru seçimiyle tahmin doğruluğunu arttıracak ileri sürmüştür. Yapılan çalışmada her bir yol bağlantısı için en uygun veri seçimini yapacak istatistiksel model önerilmiş, ilk etapta değişim, korelasyon analizi ve Monte Carlo simülasyonuna dayalı üç adımlı filtreleme tarihsel verideki aykırı verileri tespit için kullanılmış sonrasında kriter seçim yöntemleri ile tahmin yapılmıştır. Çalışma için kullanılan veriler Çin Busan şehrinde bulunan segmentlere ait araç sayısı ve ortalama hız bilgilerinden oluşmaktadır.
- Ru vd. [73], trafik tahmini üzerine yapılan çalışmalarda büyük miktarda ve sonuçları etkileyen eksik verinin göz ardı edildiğini ayrıca büyük ölçekte trafik verisinin toplamının maliyetli olduğunu belirterek kritik segment seçimi çözümü önererek trafik tahmini ve çıkarımını amaçlamıştır. Kritik segmentleri belirlemek için farklı yollardaki hareketli araçların durumunun birbiri ile ilişkili olduğu ve karşılıklı olarak birbirini temsil ettiğini kabul ederek dikkat mekanizmasını kullanan çizge sinir ağına dayalı bir yaklaşım tasarlanmıştır. Çalışmada Çin Xi'an şehrinde 75 yol segmentinden elde edilen GPS verisi kullanılmış, iki saatlik tarihsel veri ile 5 dakika sonrası trafik hızı tahmin edilmiştir.

Çizelge 2.1. Temel trafik parametreleri üzerine çalışmalar

Referans	Trafik Parametresi	Veri Kaynağı (*)	Yöntem (**)	Doğrulama (***)
[44]	Akış	DD	ARIMA	MPE, MAPE, RMSE
[45]	Akış	MS	NN	MAPE
[46]	Akış	VK	Regresyon ve NN	MSE
[47]	Akış	DD	DL	MAE, RMSE, WMAPE
[48]	Akış	DD	Gri Sistem	MAPE
[49]	Akış	DD	DL	MSE
[50]	Akış	VK, MS	LSTM	MAE, MSE, MRE
[51]	Akış	MS	KNN ve NN	RMSE
[52]	Akış	MS	KNN	MAE, MAPE, RMSE
[53]	Akış	DD	ARIMA ve SVR	MAPE, VAPE
[54]	Yoğunluk	Simülasyon	LSTM	RMSE
[55]	Yoğunluk	MS	KNN ve NN	MAPE
[56]	Yoğunluk	Radar	DBSCAN	Zamansal
[57]	Yoğunluk	Simülasyon	Kalman Filtresi	RRMSE
[58]	Yoğunluk	DD	İstatistiksel-GLR	RMSE, MAPE
[30]	Yoğunluk	MS	LSTM	RMSE, MAE
[59]	Yoğunluk	GPS-Tweet	Lineer Regresyon	RMSE
[60]	Yoğunluk	VK	PCA-SVM	AUC
[61]	Yoğunluk	VK	CNN	RMSE, MAE
[62]	Yoğunluk	DD	NN	RMSE
[63]	Hız	DD-AA	RNN	MAE, MAPE, RMSE
[64]	Hız	GPS	LSTM	MSE
[65]	Hız	TMS	LGP	MAPE
[66]	Hız	MS	LSTM	MAPE, MSE
[67]	Hız	GPS	DL	RMSE, MAPE
[68]	Hız	DD-MS	LSTM	MAE, MAPE, RMSE
[69]	Hız	GPS	LSTM-GRU	MAPE, RMSE
[70]	Hız	Simülasyon	NN-HMM	RMSE, MAPE
[71]	Hız	DD	RNN	RMSE

Çizelge 2.1. (devam) Temel trafik parametreleri üzerine çalışmalar

[72]	Hız	MS	Monte Carlo	RMSE
[73]	Hız	GPS	GNN	MAPE
* DD: Döngü Detektörü, MS: Mikrodalga Sensör, VK: Video Kamera, AA: Araştırma Araçları, TMS: Trafik Mesaj Servisi ** NN: Sinir Ağları, ARIMA: Hareketli Ortalama Zaman Serisi, DL: Derin Öğrenme, LSTM: Uzun-Kısa Zamanlı Hafıza, GRU: Tekrarlı Kapı Ünitesi, SVR: Destek Vektör Regresyonu, GLR: Genelleştirilmiş Olasılık Oranı, PCA: Temel Bileşenler Analizi, SVM: Destek Vektör Makinesi, LGP: Yerel Gauss İşlemi, HMM: Gizli Markov Model, GNN: Çizge Sinir Ağları *** MAE: Ortalama Mutlak Hata, MSE: Ortalama Karekök Hata, MRE: Ortalama Bağlı Hata, MPE: Ortalama Yüzdesel Hata, MAPE: Ortalama Mutlak yüzdesel Hata, WMAPE: Ağırlıklı MAPE, VAPE: Varyans Mutlak Yüzdesel Hata, RMSE: Ortalama Karekök Sapma Hatası, RRMSE: Bağlı RMSE, AUC: Eğri Altında Kalan Alan				

Çalışmanın bu bölümünde trafik parametreleri açıklanmış, temel trafik parametreleri olan trafik akışı, yoğunluğu ve hızı için yapılan tahmin çalışmalarından örnekler verilmiştir. Bu çalışmaların özeti Çizelge 2.1’de gösterilmiştir. Takip eden başlıkta ise veri mahremiyeti ve trafik için elde edilen verilerde bulunan konum verisi için mahremiyet açıklanarak literatür örnekleri sunulmuştur.

2.4. Veri Mahremiyeti

Literatürde ilk olarak yalnız bırakılma hakkı [74] olarak tanımlanan mahremiyet kavramı, kişiye ait bilgilerin nasıl toplanacağı, kullanılacağı hakkında ayrıcalıkların belirlenmesi ve bir birey veya grup hakkındaki bilgilerin verilen kurum ve kişiler dışında kimse tarafından bilinmeme, ayırt edilememe durumu olarak da ifade edilmektedir [75]. Mahremiyet kavramı içerisinde farklı konuları barındırmakta ve her konu için farklı koruma yaklaşımı bulunmaktadır. Mahremiyet kavramını Global İnternet Özgürlük Kampanya (The Global Internet Liberty Campaign) raporu dört kategoriye ayırmıştır [76]. Bunlar sırasıyla bilgi mahremiyeti, fiziksel mahremiyet, iletişim mahremiyeti ve bölgesel mahremiyet olup aşağıda kısaca açıklanmıştır.

- Veri mahremiyeti, kredi kartı veya hasta verileri gibi kişisel bilgileri içeren verilerin toplanmasına ve işlenmesindeki kuralların oluşturulması,
 - Fiziksel mahremiyet, insanların fiziksel olarak üst arama ve ilaç testi gibi eylemlere karşı korunması,
 - İletişim mahremiyeti, e-posta, telefon ve diğer iletişim türleri için güvenlik ve gizliliğinin sağlanması ve son olarak,
 - Bölgesel mahremiyeti, çalışma ortamları ve toplumsal alanlar gibi alanlara giriş ihlallerine karşı düzenlemeler getirilmesi,
- gibi konuları kapsamaktadır.

Son yıllarda, tıbbi araştırma ve bakım, telekomünikasyon, gelişmiş ulaşım sistemleri ve finansal uygulamalardaki yeni gelişmeler her bireyin ürettiği bilgi seviyesini önemli ölçüde arttırırken, çeşitli sektörlerde ve farklı alanlarda faaliyet gösteren kurumlar verilerini dijitalleştirmekte ve elektronik olarak kullanılabilir hale getirmektedir [77]. Kurumlar müşterilerinden veya kullanıcılarından toplanan, dijitalleşen verilerini araştırma, analiz ve karar sistemleri oluşturmak için kullanmaktadır. Kişiler hakkında bilgi toplama, analiz etme ve yayımlama ile bilgi teknolojisinin artan karmaşıklığı birleşince mahremiyet konusunda endişeler artmış ve mahremiyet çalışmaları hız kazanmıştır.

Veri mahremiyeti ve mahremiyet ihlalini önleme adına yapılan çalışmalar yaygın olarak tıp alanındadır ve çalışma konusu hasta verileridir. Patil ve Seshadri yaptıkları çalışmada [78], tıbbi verilerin dijitalleşmesiyle sağlık sektöründeki verinin büyük hacimlere ulaştığını, artan verinin doğru yönetilmesi ve gerekli çıkarımların yapılması sonucunda hastalık teşhisi, hastalığın tedavisi, hasta kayıt yönetimi içine önemli bir maliyet ve zaman tasarrufu yapılabildiği vurgulanmıştır. Ancak, büyük verinin yönetimi tıp sektörü için fırsat sunsa da verinin mahremiyeti ve güvenliğinin ilgilenilmesi gereken önemli bir sorun olduğunu da yapılan çalışmada vurgulanmıştır.

Mahremiyet ihlalleri sadece tıp alanında değil kişiyi tanımlayacak verilerin saklandığı her alanda gerçekleşmekte, veri toplama ve analiz çalışmaları ile birlikte mahremiyet ihlalleri de istikrarlı bir şekilde artış göstermektedir. Verizon firmasının yaptığı araştırmayla eğitimden finans, sigorta, perakende, üretim sektörlerine uzanan geniş iş kollarında gerçekleşen veri ihlalleri ve sayıları raporlanmıştır [79]. 2018 yılına ait raporda 53000 üzerinde olay, 2216 onaylanmış veri ihlali tespit edilmiştir. Veri ihlalleri kaynakların artması ve internetin

yaygınlaşması ile artmaya devam ederken, elde edilen bilgiler ile kişi kimlik numarası, adı, soyadı, hastalığı gibi mahrem bilgiler öğrenilerek kişinin bağlantılı olduğu insan ve çevre hakkında bilgi edinilerek mahremiyet ihlali geniş boyutlara ulaşacaktır. Kişiyeye ait hassas veri ve bilgiler çeşitlendirilebilmekte, konum mahremiyeti de bu kapsamda ele alınmaktadır ve takip eden başlıkta konum mahremiyeti açıklanmıştır.

2.4.1. Konum mahremiyeti

Bireyi doğrudan veya dolaylı olarak kimliğini tanımlayabilecek her türlü veri kişisel veri kapsamında kabul edilmektedir. Sağlık-hastalık verileri, ticari veriler, demografik veriler, maaş verileri, fotoğraf ve video gibi veriler kişisel ve hassas veri olarak ifade edilirken “konum verisi” de bu kapsamda yer almaktadır [80]. Konum verisinin mahremiyeti bir kişinin mevcut veya geçmiş konum bilgisinin başka taraflarca öğrenilmesinin engellenmesi ve tespit edilememesi olarak tanımlanmıştır [81]. Konum verisi için mahremiyet, yasal olarak ülkemizde KVKK [82] ve Avrupa’da GDPR [83] ile kişisel veri mahremiyeti kapsamında değerlendirilmekte ve konum verisi yayınlayan/kullanan kurumlar, kamu ve özel sektör kuruluşlarıyla paylaşan tüzel ve özel kişiler belirtilen yasal düzenlemelere tabidir. Yasal düzenlemeye tabi olan konum verilerini günümüzde kullanan birçok cihaz oldukça yaygınlaşmış ve konum tabanlı servisler de buna bağlı olarak popüler hale gelmiştir. Konum tabanlı servis (KTS) sağlayıcılar, kullanıcının konum bilgisini kullanarak yerel bölgede arama, trafik durumu görüntüleme, rota planlama ve konum yakınında POI listeleme gibi farklı çeşitte hizmetleri kullanıcıya sunmaktadır. KTS bu gibi faydalı hizmetleri sunarken kaydedilen kullanıcı konum bilgileriyle kullanıcının yaşadığı çevre, komşuları hakkında bilgi veya varsa çocuklarının gittiği okul gibi birçok mahrem veri çıkarımına sebep olmaktadır. Krumm yaptığı çalışmada [84], kullanıcının son konumu ile kullanıcının yaşadığı evi yaklaşık 60 metre uzaklıkta belirleyebilmiştir.

KTS kullanıcıları bir şehir etrafında gezinirken ve gezindiği bölgede yakındaki şehre ait mekânları aramak gibi işlemlerle servisi sıklıkla kullandığında durum daha ciddi bir hal almaktadır, çünkü kayıt alınan konum geçmişleri özel konum tespitini kolaylaştırmaktadır. Ayrıca KTS sağlayıcılar kullanıcı konumu dikkate alarak yakınındaki POI noktaları (restoran, süper market, otopark) belirleyebilmekte [85], sürekli olarak servis sağlayıcıdan hizmet alan ve sorgulamalar yapan kullanıcılar için kayıt altına alınan her nokta onun için bir rota oluşturmaktadır. Oluşan rotalar, KTS kullanıcılarına verilen hizmetlerin dışında aynı

zamanda ulaşım ve trafik gezinme ve alışveriş önerileri gibi gerçek hayat uygulamalarında kullanılmaktadır [86]. Rota bilgisinin değerli ve çıkarım yapılacak bilgiler içermesinden dolayı KTS bu bilgiyi üçüncü parti kurumlar ile paylaşmaktadır. Bilginin paylaşılması ve bir zaman periyodunda kullanıcının konumunun sürekli kaydedilmesi ile güvenilir olmayan bir servis sayesinde, kullanıcıya ait inanç, hastalık bilgisi veya sosyal ilişkilerini direkt veya dolaylı yoldan ortaya çıkarabilmektedir ve kişisel mahremiyet bu noktada oldukça savunmasızdır. Konum mahremiyetinin sağlanması kişisel veri mahremiyeti için oldukça önemli rol oynamaktadır.

Katılımcı mobil cihazlarda kullanılan uygulamalar ve oluşturdıkları rota verileri, trafik izleme, pazarlama analizi ve şehir planlaması gibi çeşitli uygulamaların yayılması ve geliştirilmesine katkılar sağlamaktadır. Bununla birlikte, rota verileri için uygun veri temizleme işlemi yapılmadan yayınlamak ve kullanmak, kullanıcıların gizliliğine yönelik ciddi tehditler oluşturmaktadır. Yasal düzenlemelerle korunmaya çalışılan konum verisi üzerine oldukça kapsamlı çalışmalar yapılmaktadır. Mevcut çalışmalar, mahremiyeti bir şekilde elde sağlasa da, veri kümesi içindeki yörüngeler veya bazı konum noktaların kaldırılarak temizlenmesi gerekmektedir [87]. Dolayısıyla yapılan çalışmalarda mahremiyet garanti edilirken, birleştirme ve temizleme işlemi verinin kullanılabilirliğini azaltmaktadır.

2.4.2. Konum mahremiyeti sağlamak için yöntemler

Konum mahremiyetini sağlamak için farklı yöntemlerin uygulandığı birçok çalışma yapılmıştır ve bu alanda çalışmalar halen sürmektedir. Bu çalışmalarda esas alınan nokta, saldırgan ve gerçekleştirdiği ihlal için verinin nasıl elde edildiği, saldırının nasıl gerçekleştiği, hangi bilgilerin ele geçtiği, hedefin kim olduğudur [88]. Konum bilgisini elde eden saldırgan arka plan bilgisine sahip olarak veya farklı veri tabanları arasında bağlantılar kurarak hedef kişinin kimlik ve konumunu tespit edebilmektedir.

Hara vd. konum mahremiyetini sağlamayı 3 başlık altında toplamıştır [89].

- Bunlardan ilki, kullanıcı ve servis sağlayıcı arasında olup kullanıcının ayırt edilemeyecek şekilde gizlenmesidir. Bu yöntemde mekânsal gizleme yapılırken genelleme ve baskılama teknikleri kullanılmaktadır. Yöntem her kullanıcının konumunu diğer kayıtların konumu ile karıştırmakta ve mahremiyeti garanti etmektedir.

- İkinci yöntem şaşırtma tekniği olarak bilinmekte, kullanıcı konumunun yaklaşık bir konumla değiştirilmektedir. Fakat bu yöntemde konum değiştirme esnasında kullanıcı yaya ise bina veya ilgi noktası (POI) noktası ile değiştirilmeli, araç ise bina konum dışında bir yerde konum eklenmelidir.
- Üçüncü yöntemde ise sahte kullanıcılar üretilerek yanıltıcı konumlar sisteme eklenmektedir. Kullanıcı servis sağlayıcıya POI sorgusu gönderdiğinde servis sahte konumlar için cevap döndürse de kullanıcı kendi bölgesi için cevabı almış olacaktır.

Konum verisinin kime ait olduğu, kullanıcıya nasıl bir servis sağlandığı gibi detaylar konum mahremiyeti sağlamak için yöntemleri belirlemektedir. Tez çalışmalarında kullanılan gerçek araç verileri GPS verilerini içermekte ve bu veriler zamana göre sıralandığında yörünge bilgileri oluşmaktadır. Yörünge ve hareketli araç konum bilgilerinin mahremiyeti üzerine çalışmalar incelendiğinde;

- Abul vd. [90], k-anonimlik yöntemiyle hareket eden araçlar için konum verilerini zamansal sırada ele alarak, aynı zaman diliminde olan ve belirli bir delta değeri içerisindeki konum noktalarını k komşuluk değerinde anonim hale getirmiştir.
- Huo vd. [91], hareketli araç verileri için nerelerden geçtiğinden ziyade nereleri ziyaret ettiğinin arka plan bilgisiyle daha ilişkili olduğunu, kalıcı noktaların genelleştirilmesi ile mahremiyet sağladıklarını ifade etmiştir. Araç konum noktalarını için durağan ve hareketli olarak tanımlayan araştırmacılar haritayı bölerek kalıcı noktaları farklı bölgelere yerleştirmiştir.
- Zhang vd. [92], hem dinamik hem de konum seçme mekanizması kullanarak k-anonimlik yöntemi ile araç yörünge verilerinin mahremiyetini sağlama modeli önermiştir. Çalışmada sentetik araç verileri kullanılmış, her araç için yörünge boyunca KTS sorguları gönderilmiştir.
- Ren ve Tang [93], trafik yoğunluk tespiti için kesin konum noktalara ihtiyaç duyulmadığını ve haritaları bitmap düzeyinde soyutlayarak mahremiyet sağlanabileceğini vurgulamıştır. EgeoIndis adını verdikleri yeni yöntemde lineer programlama ve taksi yörünge verileri kullanılmıştır.
- Li vd. [87], mevcut çalışmaların diferansiyel mahremiyet başarımına tam anlamıyla ulaşmadığını, bunun nedeninin sınır değeri olmadan rastgele gürültü eklediğini yörünge birleştirme metotlarının veri kümesindeki bazı yörüngeleri çıkardığını, böylece veri

kullanılabilirliğinin azaldığını ifade etmişlerdir. Araştırmacılar bu sorunların çözümü için sınırlı bir gürültü ekleme ve yörünge birleştirme metodu önermişlerdir.

- Qi vd. [94], akıllı ulaşım sistemlerinde yapılan trafik akış tahminin tek bir merkezde yapılması nedeniyle güvenlik açıkları oluşabileceğinin, federatif öğrenme yöntemiyle veri transferi olmadan makine öğrenme yöntemlerinin ağırlık değerlerinin güncellenmesi sonucu güvenlik sağlanabileceğini ileri sürmüştür. Çalışmada ayrıca federatif öğrenmede tek bir kontrol noktasının düşmesinin de bir sorun olduğu, blok zincir yöntemi kullanarak sorunun çözülebileceğini de vurgulanmıştır. Fakat akıllı ulaşım sistemlerinde veri kaynağının çeşitliliğinin artması blok zincir ve ağırlıkların her araçtan güncellenmesi federatif öğrenmenin en büyük problemi olan maliyet sorunu aşırı derecede arttıracaktır. Dolayısıyla önerilen yöntem mahremiyeti sağlarken ulaşım sistemleri için doğru tahmin üretimi için yeterli araç sayısı ulaşamayacaktır ki çalışmada test için 10 araç seçilmiştir.

Mevcut veri kümesi farklı kurumların filolarına ait farklı türdeki filo araç verilerden oluşmakta, konum-zaman ve araç kimlik bilgisini içermektedir. Veri kümesi içerisindeki kurumlara ait filo bilgilerinden kurum araçlarının yörünge bilgileri oluşturulabilmektedir. Oluşan yörüngeler ile kurum hakkında araç-kişi pazarlama bilgileri veya hassas çıkarımlar yapılabilir. Dolayısıyla araç konum bilgilerinin mahremiyeti sağlanması gerekmekte fakat trafik yönetim sistemleri için konum verisi kimlik verisine ulaşılmadan kullanılabilir. Tez kapsamında yapılacak çalışmalarda mahremiyeti sağlanan veri öznetelikleriyle trafik hız tahmini yapılacak ve konum bilgisinin zaman bilgisi ile sıralanmasından elde edilebilen yörünge bilgileri için mahremiyet sağlanacaktır.

Bu bölümde trafik ekosistemini oluşturan bileşenler detaylıca anlatılmış, akıllı ulaşım sistemlerinin çeşitleri sunulmuş, trafik için veri kaynakları örneklendirilmiştir. Farklı trafik kaynaklarından elde edilen verilerle yapılan trafik akış, yoğunluk ve hız üzerine yapılan literatür çalışmaları incelenmiş ve ayrıca veri mahremiyeti ve bu kapsamda ele alınan konum ve yörünge bilgilerinin mahremiyeti üzerine çalışmalarda sunulmuştur. Takip eden bölümde tez çalışması kapsamında geliştirilen tahmin ve mahremiyet modellerinde kullanılan araç verilerinin detayları verilmiş ve kullanılan yöntemler anlatılmıştır.

3. MATERİYAL METOT

Bu bölümde geliştirilen üç modelde kullanılan veri kümesi hakkında bilgiler verilmiş, model geliştirmede kullanılan görselleştirme araçları, platformlar anlatılmış, model aşamalarında uygulanan metotlar açıklanmıştır.

3.1. Kullanılan Veri Kümesi

Tez çalışmasında Başarsoft Bilgi Teknolojileri A.Ş. firmasının sağladığı ve mobil cihazlardan elde edilen “gerçek araç verileri” kullanılmıştır. Araç verilerini sağlayan Başarsoft firması; Başar Bilgisayar adıyla 1997 yılında kurulmuştur ve kurulduğu tarihten itibaren coğrafi bilgi sistemleri üzerine faaliyet göstermektedir. Firma ülkemizde birçok kurumun ihtiyaç duyduğu şehir, mahalle, sokak gibi detayların tümünü içeren sayısallaştırılmış harita altyapısını hazırlamaktadır. Yurtiçinde farklı firmaların navigasyon ve araç takip sistemlerinde, su, elektrik, telekomünikasyon altyapı projelerinde hizmet verirken, yurtdışında Google, Yandex gibi şirketlere de harita sağlamaktadır [95]. Başarsoft firması uzun süredir farklı tipte şirket filo araçlarına yerleştirilen araç takip sistemlerinden canlı ve anlık olarak araç verileri elde etmekte ve saklamaktadır. Saklanan araç verileri hizmet ağının genişlemesiyle de gün geçtikçe artmakta, büyük bir trafik verisini oluşturmaktadır. Büyük trafik verisinden anlamlı sonuçlar çıkarılması ve büyük verinin yönetimi için planlanan tez çalışması, literatür incelendiğinde dağıtık konum verileriyle tahmin çalışması ve mahremiyet odaklı yaklaşımların bu alanda eksik olduğu, dağıtık verilerle çalışmanın fırsat olduğu fark edilerek mahremiyet odaklı tahmin modeli üzerine evrilmiştir. Dağıtık olan araç verileri sayesinde bölge, alan filtrelemeleri ve yol eşlemeleri yapılarak bu alanda istenilen yol ağında mahremiyet odaklı tahmin yürütülebileceği, daha sonrasında geliştirilen modeller büyük veri teknolojisi kullanarak uygulanabileceği öngörülmektedir.

Başarsoft firmasının elde ettiği ve kendi bünyesinde sakladığı veriler tüm Türkiye’ye gezinen farklı tipte şirket filo araçlarına ait verilerdir ve bu veriler firma tarafından Ankara iline ait olmak üzere filtrelenmiştir. Tüm Türkiye’ye ve Ankara iline ait araçların konum bilgilerini içeren örnek gösterimler Şekil 3.1 ve 3.2’de görülmektedir.



Şekil 3.1. Tüm Türkiye’de gezinen araçların örnek konum verileri



Şekil 3.2. Ankara iline ait örnek konum verileri

Tahmin modeli ve konum mahremiyeti için kullanılan veri kümesi ise farklı zaman dilimlerinde ele alınarak seçilmiştir. Seçilen kümelerden örnek bir veri kümesi Ankara şehrine ait 27 Ocak 2020 tarihli 14:59-15:59 arasında araç verilerinin bulunduğu bir saatlik veridir. Örnek veri kümesi GPS sinyalleri içermektedir ve herhangi bir filtreleme işlemi uygulanmadan 205MB boyutunda olup yaklaşık 3 milyon 300 bin kayıt içermektedir. Kullanılan veri kümelerinde bulunan araç GPS sinyalleri farklı periyotlarda kaydedilmektedir ve kaydedilen verilerin belirli bir frekansı yoktur. Araçların ürettiği sinyaller için altı öznitelik mevcuttur. Araçlar tarafından üretilen öznitelikler ile bunların

açıklamaları Çizelge 3.1’de ve araçlara ait GPS kayıtlarının örnek bir gösterimi Çizelge 3.2’de verilmiştir.

Çizelge 3.1. Araçlar tarafından üretilen öznitelikler ve açıklamaları

Öznitelik sayısı	Araçlar tarafından üretilen öznitelikler	Açıklama
1	Araç ID	Navigasyon cihazını kullanan araçlara ait benzersiz bir numara
2	Zaman Damgası	Tarih ve saat formatında olan zaman bilgisi
3	Enlem Bilgisi	X bilgisi Enlem Koordinatı
4	Boylam Bilgisi	Y bilgisi Boylam Koordinatı
5	Araç Hızı	Hız bilgisi aracın anlık ölçülen hızı
6	Açısal Yön	Açısal Yön bilgisi 0° kuzey yönünü gösterecek şekilde aracın koordinat düzlemindeki yönü

Çizelge 3.2. Araç GPS kayıtlarının örnek bir gösterimi

ID	Zaman	Y	X	Hız	Açısal Yön
fce*****	15.01.2020 15:00	39,954361	32,571415	35	40
fcf*****	15.01.2020 15:00	39,902203	32,765202	50	171
84C*****	15.01.2020 15:00	39,920960	32,993813	101	236
129*****	15.01.2020 15:00	39,894039	32,803215	49	270
DA1*****	15.01.2020 15:01	39,949560	32,597176	33	279
cb4*****	15.01.2020 15:01	39,960018	32,686771	65	322
cb5*****	15.01.2020 15:01	39,802967	32,393925	57	202

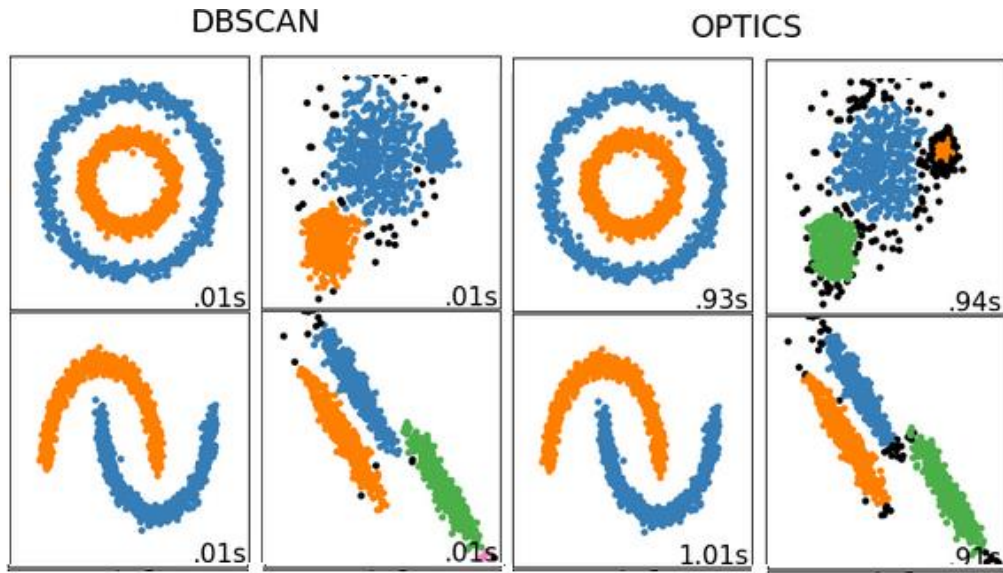
3.2. Görselleştirme Araçları ve Kullanılan Platform

Tez kapsamında yapılan tüm çalışmalar Python dili ile gerçekleştirilmiş, araç verileri Holoviews ve Geoviews kütüphaneleri ile harita üzerinde görselleştirilmiştir. Holoviews veri analizi ve görselleştirme yapmak için kullanılırken, Geoviews ise coğrafi, meteorolojik veri kümelerini keşfetmeyi ve görselleştirmeyi kolaylaştıran açık kaynaklı Python kütüphaneleridir [96,97]. Görselleştirme işlemlerinde kullanılan haritaların altyapısını sağlayan OpenStreetMap uygulaması, açık kaynaklı dünya çapında harita oluşturma projesidir [98].

3.3. OPTICS Algoritması

OPTICS genişletilmiş bir DBSCAN algoritması olarak çalışmaktadır [99]. DBSCAN algoritmasında rastgele seçilen bir p noktası çekirdek nesne olarak seçilmektedir ve girdi

parametresi olarak çekirdek nesnenin komşu nesnelere yakınlık yarıçapı ε (epsilon) ve kaç üye komşuluğuna bakılacağı $MinPts$ değerlerini kullanılmaktadır. OPTICS algoritmasında ise “üretilen uzaklık” ε değerinden küçük ε_i değerleri dikkate alınmaktadır. OPTICS algoritmasının DBSCAN algoritmasından farkı küme üyeliklerinin ilk etapta belirlenmemesidir. Küme üyelerinin hemen belirlenmesi yerine her bir nesnenin sırası oluşturulmakta, çekirdek mesafe ve uygun bir erişilebilirlik mesafe bilgisi kaydedilmektedir. Veri sırası elde edilen nesnelerden çekirdek nesne olarak belirlenen nesneler için en az sayıdaki yakın komşuluğuna bakılarak kümeleme işlemi yapılmaktadır. Örnek bir kümeleme işlemi Şekil 3.3’de görülmektedir.



Şekil 3.3. DBSCAN ve OPTICS algoritmalarının örnek kümeleme işlemi [100]

OPTICS algoritmasının çekirdek mesafesi belirleme eşitliği Eş. 1 ile tanımlanmaktadır. Bir p noktasının ε uzaklığındaki komşuluğu $N_\varepsilon(p)$ ile ifade edilirken, $MinPts$ değerinden küçük ise tanımsız kabul edilmiştir. Eğer bu uzaklık en az noktaya eşit ve büyük ise komşu noktalara uzaklıklar $MinPts$ -mesafe olarak ele alınmış ve an yakın komşu eleman sonraki çekirdek nokta seçilmiştir [101].

$$\begin{cases} \text{Eğer yakınlık } (N_\varepsilon(p)) < MinPts, & \text{Tanımsız} \\ \text{Diğer durumlarda} & , \quad MinPts - \text{mesafe } (p) \end{cases} \quad (3.1)$$

3.4. Tekrarlı Sinir Ağı (RNN)

RNN yöntemi sıralı veriler için dahili hafızayı kullanarak bir tekrar oluşturan güçlü bir derin sinir ağıdır [102]. RNN yönteminde hafıza hücreleri bir zincir oluşturmaktadır ve giriş dizisindeki her eleman için aynı görevi gerçekleştirmektedir. Yöntemin güçlü olması giriş ve çıkış değerleri arasındaki eşleşmeyi doğrudan öğrenmesinden ileri gelmektedir. Fakat RNN yöntemi eğimin yok olması (vanishing gradient) ve çok yüksek değerlere çıkması (exploding gradient) probleminden etkilenmektedir ve sorunu çözmek için Kısa-Uzun Zamanlı Hafıza (LSTM) yöntemi önerilmiştir [103,104]. LSTM yönteminde hücreler kendi durumlarını zamana karşı korurken uzun zaman bağımlılıklarını da öğrenmektedir. Giriş elemanları $x = (x_1, x_2, x_3, \dots, x_n)$ olan modelde t zaman aralığında LSTM modeline ait matematiksel ifadeler Eş. 3.2-3.6 ile görülmektedir. Eş. 3.2’de t anındaki giriş kapısı (i_t), yine t anındaki ağırlık matrisi (W_{ix_t}) $t-1$ anındaki ağırlık matrisi ($W_{ih_{t-1}}$) ve ön ağırlık (b_i) değerinin toplamı aktivasyon fonksiyonu (σ) ile işleme alındığında oluşmaktadır. Aynı işlemler Eş. 3.3-Eş.3.4 denklemlerinde uygulanarak t anındaki unutma (f_t), ve çıkış (o_t) değerleri elde edilmektedir. Eş. 3.5 ile hücre durumu (c_t), hücre ağırlık matrisleri ($W_{cx_t}, W_{ic_{t-1}}$) ve ön ağırlık (b_c) toplamının $\tanh$ fonksiyon değeri, önceki hücre durumunun (c_{t-1}) t anındaki giriş değeri (i_t) toplamı ve t anındaki unutma değerinin (f_t) “Hadamard” çarpımıyla hesaplanmaktadır. Hadamard çarpımı eleman bazında çarpma işlemi olarak bilinmektedir ve her iki matrisin aynı derece elemanını çarparak yeni matrisi oluşturmaktadır. Sonraki hücreye giriş değeri olacak ve öğrenmeyi sağlayacak gizli katman, t anındaki çıkış değerinin (o_t) yine t anındaki hücre durum $\tanh$ değerinin Hadamard çarpımıyla oluşmaktadır ve Eş. 3.6’da tanımlanmıştır.

$$i_t = \sigma (W_{ix_t} + W_{ih_{t-1}} + b_i) \quad (3.2)$$

$$f_t = \sigma (W_{fx_t} + W_{fh_{t-1}} + b_f) \quad (3.3)$$

$$o_t = \sigma (W_{ox_t} + W_{oh_{t-1}} + b_o) \quad (3.4)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh (W_{cx_t} + W_{ic_{t-1}} + b_c) \quad (3.5)$$

$$h_t = o_t \odot \tanh (c_t) \quad (3.6)$$

Öğrenme süreci hızlı olan ve eğimin yok olması problemine karşı başarılı başka bir yöntem Tekrarlı Kapı Ünitesi (GRU) ayrı hafıza hücresi bulundurmeyen LSTM farklı bir türüdür.

GRU basitleştirilmiş LSTM olarak ifade edilebilir ve LSTM gibi üç kapıya değil güncelleme kapısı (z) ve sıfırlama kapısı (r) olarak iki kapıya sahiptir.

3.5. Diferansiyel Mahremiyet

Diferansiyel mahremiyet yöntemi Dwork ve arkadaşları tarafından önerilmiş [105], günümüzde halen farklı alanlarda sıklıkla uygulanmaktadır. Gürültü ekleme metoduna dayanan yöntem özellikle istatistiksel veri tabanları üzerine yapılan analizlerde üst seviyede koruma sağlayarak arka plan bilgisi saldırısını engellemeyi amaçlamıştır. Diferansiyel mahremiyette temel amaç veri tabanına yapılacak sorguya döndürülecek gerçek cevaba belirli ölçülerde gürültü eklemektir. Eklenen gürültü ile veri tabanına gönderilen toplam kayıt sayısı, ortalama, toplam değer gibi istatistiksel sorgularda saldırgana yanıltıcı cevaplar verilmiş olacaktır.

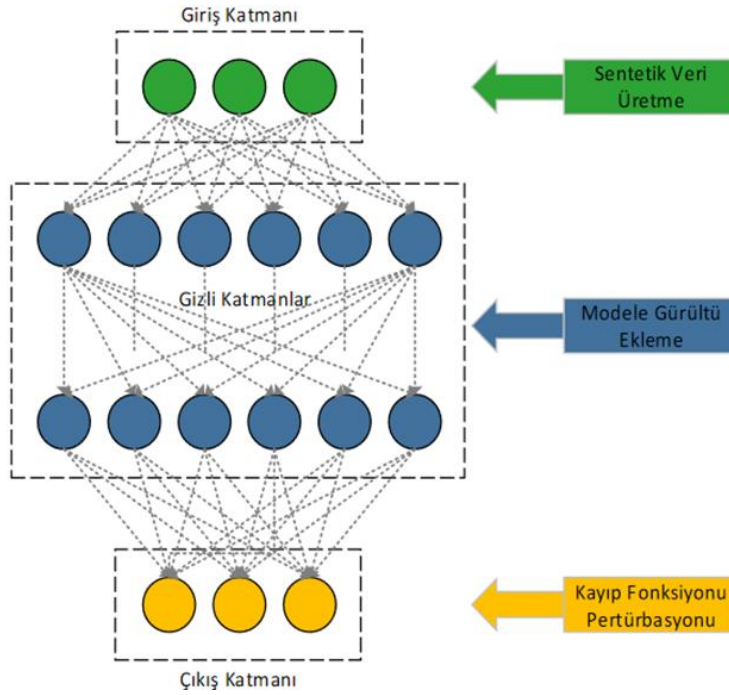
Diferansiyel mahremiyetin çalışma prensibi kısaca şu şekilde açıklanabilir. D_1 ve D_2 komşu veri kümeleri için bir f bir sorgu fonksiyonu ise; bu veri kümelerine uygulanacak herhangi bir f sorgusunun sonuçları arasındaki maksimum fark Δf olarak adlandırılmakta, yöntemde mahremiyet koruma D veri kümesine uygulanan rasgele bir R uzayında sağlanmaktadır. Rastgele uzay R , her S çıktı kümesi için herhangi D_1 ve D_2 komşu veri kümeleri için Eş 3.7 ile ifade edilen olasılık şartını sağlaması halinde ε (epsilon) aralığında ε -diferansiyel mahremiyet elde edilmektedir. Literatürde diferansiyel mahremiyeti sağlamak amacıyla kullanılan Laplace, Üssel ve Gauss gibi çeşitli mekanizmalar mevcuttur [106].

$$Olasılık[M(D_1) \in S] \leq \exp(\varepsilon) Olasılık[M(D_2) \in S] \quad (3.7)$$

Diferansiyel mahremiyet yöntemi son yıllarda sıklıkla veri mahremiyetini sağlamak için kullanılmakta, veri mahremiyeti konusu, mahremiyet korumalı veri yayınlama ve mahremiyet korumalı veri madenciliği olmak üzere iki ana başlık altında da ele alınmaktadır [107]. Mahremiyet korumalı veri yayınlama, veriden daha fazla değer üretilebilmesi, analiz edilebilmesi amacıyla üçüncü kişilerle paylaşılması iken, mahremiyet korumalı veri madenciliği ise veri madenciliği ve makine öğrenme yöntemlerin sürecinde kişisel ve tanımlayıcı verilere mahremiyet tekniklerinin uygulanmasıdır.

Şekil 3.4'te görülen makine öğrenme yöntemlerinden olan derin öğrenme yönteminde diferansiyel mahremiyet;

- Giriş katmanında sentetik veri üreterek,
- Eğitim verisi üzerinde ağırlık matrislerine gürültü eklenerek ve
- Çıktı katmanında kayıp fonksiyonunun pertübasyonu ile sağlanmaktadır.



Şekil 3.4. Derin öğrenmede kullanılan diferansiyel mahremiyet yöntemleri [107,108]

3.6. Hata Metrikleri

Trafik parametrelerinin tahmini için geliştirilen uygulamaların sonuçlarını değerlendirmek için geleneksel hata göstergeleri olan RMSE, MAPE ve MAE hata ölçüm değerleri kullanılmaktadır [109]. Hata oranları, n elemanlı bir dizi için her tahmin değerinden ($\bar{y}_i$) ile gerçek değer (y_i) arasındaki fark değerlerini dikkate alınarak hesaplanmaktadır. RMSE, MAPE ve MAE hesaplama eşitlikleri sırasıyla Eş. 3.8-Eş.3.10'da ifade edilmiştir. Mahremiyet korumalı tahmin modelinin değerlendirilmesi RMSE değerinin yanı sıra mekânsal verinin birbirinden ne kadar uzak olduğunun ölçüldüğü Öklid yöntemi kullanılmış [110,111] ve Eş. 3.11'de hesaplama eşitliği verilmiştir.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\bar{y}_i - y_i)^2} \quad (3.8)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|\bar{y}_i - y_i|}{y_i} \quad (3.9)$$

$$MAE = \sum_{i=1}^n \frac{|\bar{y}_i - y_i|}{y_i} \quad (3.10)$$

$$d(x, y) = \sqrt{\sum_{k=1}^n |x_k - y_k|^2} \quad (3.11)$$

3.7. Haversine Uzaklığı

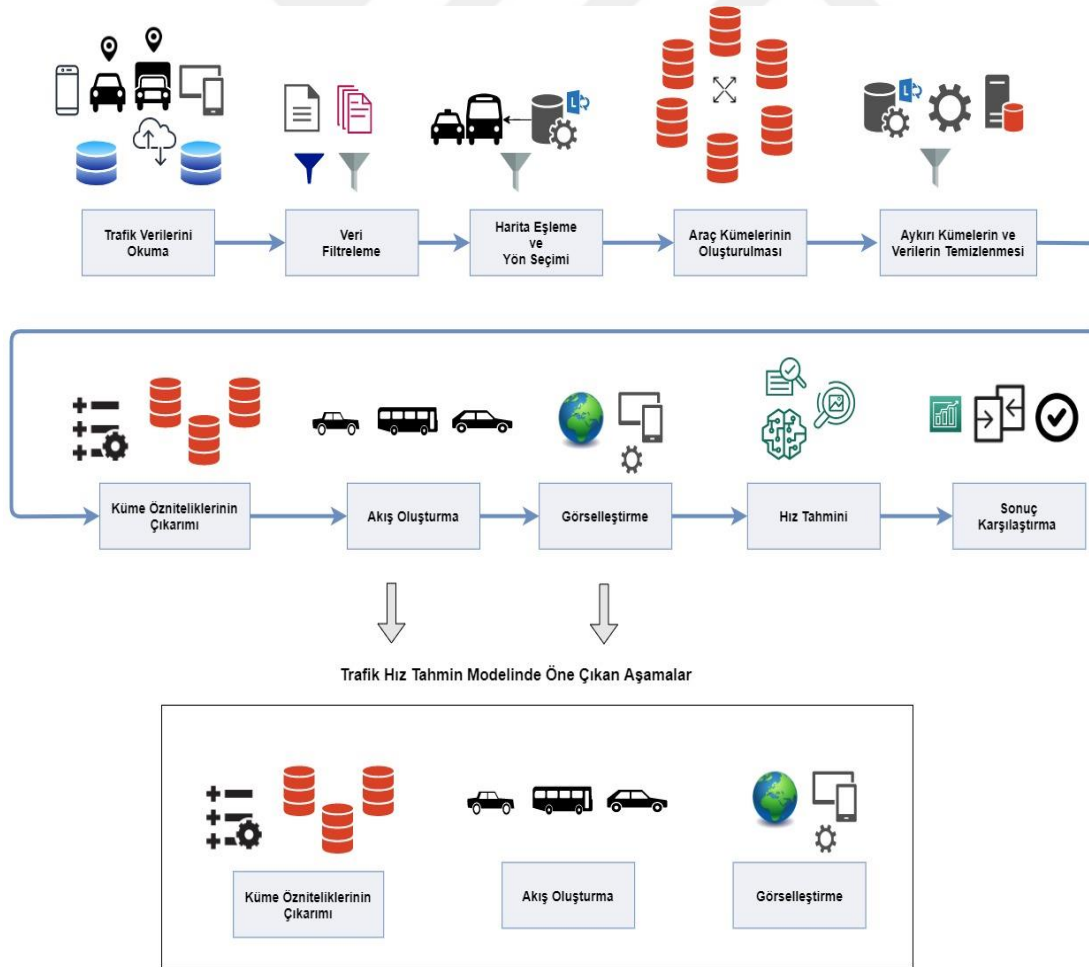
Haversine uzaklığı, küre bir düzlem üzerindeki iki nokta arasındaki en büyük daire mesafesini hesaplamak için kullanılmaktadır [112]. Dünyanın şekli kutuplardan basık bir küre olduğundan, GPS noktaları arasındaki mesafe Haversine uzaklığı ile hesaplanmaktadır ve bu hesaplama Eş. 3.12’de verilmiştir [113]. Formülde, uzaklık hesabı “ d ”, dünya yarıçapı “ r ”, enlem değerleri “ θ ”, boylam değerleri “ φ ” ile gösterilmiştir.

$$d = 2r \arcsin \left(\sqrt{\sin^2 \left(\frac{\theta_2 - \theta_1}{2} \right) + \cos(\varphi_1) \cos(\varphi_2) \sin^2 \left(\frac{\varphi_2 - \varphi_1}{2} \right)} \right) \quad (3.12)$$

Bu bölümde trafik hız tahmini ve mahremiyet korumalı modeller için kullanılan veri kümeleri ve özellikleri açıklanmış, harita üzerinde görselleştirilmiştir. Ayrıca önerilen modellerin aşamalarında kullanılan yöntemler, teknikler ve araçlar açıklanmıştır. Sonraki bölümlerde sırasıyla geliştirilen trafik hız tahmin modeli, mahremiyet korumalı tahmin modeli ve veri yayınlama için önerilen yörünge mahremiyet modeli detaylıca açıklanmış, test sonuçları sunulmuştur.

4. TRAFİK HIZ TAHMİN MODELİ

Bu bölümde, temel trafik parametrelerinden olan hız parametresi için yeni ve özgün bir tahmin modeli geliştirilmiş, gerçek araç verileri üzerinde test edilmiş ve sonuçları sunulmuştur. Geliştirilen tahmin modeli dokuz aşamadan ve birbirini tamamlayan sıralı metotlardan oluşmaktadır. Modelin veri hazırlama aşamaları filtreleme, harita eşleme, küme oluşturma ve aykırı veri temizleme gibi verilerin tahmin için hazırlıklarını içerirken sonraki aşamalar öznitelik üretimi, akış oluşturma, görselleştirme, hız tahmini ve doğrulama adımlarından oluşmaktadır. Model test ve değerlendirilmesi için Ankara ilinde bulunan iki ana güzergâh belirlenmiş, bu güzergâhlar çeşitli zamanlarda farklı tipte araç verilerinin sıklıkla gezinmesinden dolayı seçilmiştir. Test ve değerlendirme iki ana yol için yapılsa da geliştirilen model Türkiye’de herhangi bir şehirde ve yolda tahmin üretecek şekilde tasarlanmıştır. Geliştirilen trafik hız tahmin modelinin genel yapısı ve akışı Şekil 4.1’de, model sözde kodu ve matematiksel hesaplamaları Algoritma 4.1’de gösterilmiştir.



Şekil 4.1. Kısa zamanlı trafik hız tahmin model süreci ve öne çıkan aşamalar

Algoritma 4.1. Yoğunluk tabanlı kümelerle hız tahmin algoritması

Girdi: veri kümesi $V = \{ID, Zaman, Y, X, Hız, Açısıl Yön\}$, nx6 boyutunda matris

(n: veri kayıt sayısı olmak üzere)

Çıktı: hız tahmin değerleri $H_t = \{H_{t_1}, H_{t_2}, \dots, H_{t_m}\}$ m sayıda küme için tahmin edilen hız değerleri dizisi;

```

1  V ← Eskişehir veya İstanbul Yolu ham veri
2   $K_1, K_2$  ← Filtre konum değerleri
3   $Z_1, Z_2$  ← Filtre zaman değerleri
4   $V_{filtre} \leftarrow (K_1 < V \ \& \ V < K_2) \ \& \ (Z_1 < V \ \& \ V < Z_2)$ 
5  SHP ← Ankara iline ait geometrik şekil verisi
6  for each  $V_x, V_y \in V_{filtre}$  do
7       $V_{yol} \leftarrow$  Her konum verisini SHP verisi ile eşle ve yola ait konum verilerini elde et.
8  end for
9   $AY_1, AY_2$  ← Filtre açısıl yön bilgileri
10  $V_{yönlü} \leftarrow (AY_1 < V_{yol} \ \& \ V_{yol} < AY_2)$ 
11 for each  $i = 0, V_x, V_y \in V_{yönlü}$  do
12      $k_i = \{\}, p \leftarrow (V_x, V_y)$ 
13     for each  $p \in V_{yönlü}$  do
14         if  $N_\varepsilon(p) < MinEleman$  then (noktanın komşuluk değeri:  $N_\varepsilon(p)$ )
15             Tanımsız.
16         else
17              $k_i \leftarrow k_i \cup p$ 
18         end if
19     end for
20      $V_{küme} \leftarrow V_{küme} \cup k_i$ 
21 end for
22  $V_{küme} \leftarrow (k_i \in V_{küme})! = -1$ 
23 for each  $k_i \in V_{küme}$  do
24      $k_{stdhız}, k_{stdaçısıl} \leftarrow k_i$  için hız ve açısıl yön standart sapma değerleri hesapla.
25      $V_{küme} \leftarrow V_{küme} \cup k_{stdhız}, k_{stdaçısıl}$ 
26 end for
27  $V_{filtreküme} \leftarrow V_{küme}$  içerisinde  $V_{stdhız}, V_{stdaçısıl}$  değerleri ile filtrele.

```

Algoritma 4.1. (devam) Yoğunluk tabanlı kümelerle hız tahmin algoritması

```

28  $V_{k\ddot{o}znt} = \{ \}$ 
29 for each  $k_i \in V_{filtrek\ddot{u}me}$  do
30    $k_{ara\breve{s}sayı} \leftarrow k_i$  içerisindeki araç sayısı
31    $k_{hız} \leftarrow k_i$  içerisindeki araç hız ortalaması
32    $k_{önXY} \leftarrow k_i$  ön araç koordinatları
33    $k_{arkaXY} \leftarrow k_i$  arka araç koordinatları
34    $k_{uzunluk} \leftarrow Haversine(k_{önXY}, k_{arkaXY})$ 
35    $k_{yoğunluk} \leftarrow k_{ara\breve{s}sayı} / k_{uzunluk}$ 
36    $V_{k\ddot{u}m\ddot{o}zn} \leftarrow V_{k\ddot{u}m\ddot{o}zn} \cup k_{ara\breve{s}sayı}, k_{hız}, k_{önXY}, k_{arkaXY}, k_{uzunluk}, k_{yoğunluk}$ 
37 end for
38  $B_n \leftarrow$  Başlangıç Noktası seç.
39 for each  $k_i \in V_{k\ddot{u}m\ddot{o}zn}$  do
40    $V_{sıralı} \leftarrow Haversine(B_n, k_{arkaXY})$  hesapla ve yakınlığa göre sırala.
41 end for
42 for each  $k_i \in V_{sıralı}$  do
43    $k_{renk} \leftarrow k_i$  içerisindeki hız değerleri ile renk oluştur.
44    $H_{görsel} \leftarrow k_{renk}$  ile harita üzerinde renklendirme yap.
45 end for
46  $H_t = \{ \} : Tahmin kümesi$ 
47 for each  $k_i \in V_{sıralı}$  do
48    $k_{tahmin} \leftarrow k_i, k_{i+1}, k_{i+2}$  kullanarak 3 pencere aralığında RNN ile tahmin üret.
49    $H_t \leftarrow H_t \cup k_{tahmin}$ 
50 end for

```

Tahmin modelinin değerlendirilmesinde aynı güne ait iki farklı zaman diliminde yer alan veri kümeleri kullanılmış, kullanılan birinci ham veri kümesi (HV1) 27 Ocak 2020 14:59-15:59, ikinci ham veri kümesi (HV2) 09:59-10:59 arası bir saatlik araç verilerinden oluşmaktadır. Her iki veri kümesi zamansal filtre işlemlerinden geçirilmiş, 10 dakikalık zaman dilimi aralığına çekilmiş, birinci veri kümesi (VK1) ve ikinci veri kümesi (VK2) olarak ayrılmış, sonrasında VK1 ve VK2 mekânsal olarak Eskişehir ve İstanbul yollarını içeren bölgelere sınırlandırılmıştır. Filtrelenen ve sınırlandırılan her iki veri kümesi bütün aşamalardan geçirilmiş ve bu aşamaların bütün detayları alt başlıklarda açıklanmıştır.

Sürekli tekrar eden konum görüntüleri olmaması adına yoğun trafik verisi olan VK1'in bütün aşama sonuçları verilirken VK2 için sonuç ve karşılaştırma değerleri gösterilmiştir. Kullanılan veri kümelerinin detayları Çizelge 4.1'de verilmiştir.

Çizelge 4.1. Model için kullanılan veri kümelerinin özellikleri

Yoğunluk	Veri Kümesi	Tarih/Saat	Bölge/Yol	Yön	Kayıt Sayısı	Boyut
Yoğun	HV1	27 Ocak 2021 14:59-15:59	Ankara	Karışık	3.386.083	205 MB
	VK1	27 Ocak 2021 15:10-15:20	Eskişehir Yolu	K. Doğu/Doğu	5.781	348 KB
			İstanbul Yolu	Batı/K. Batı	7.169	440 KB
Orta Yoğun	HV2	27 Ocak 2021 09:59-10:59	Ankara	Karışık	2.006.421	122 MB
	VK2	27 Ocak 2021 10:30-10:40	Eskişehir Yolu	K. Doğu/Doğu	3.846	244 KB
			İstanbul Yolu	Batı/K. Batı	3.106	195 KB

4.1. Trafik Verilerini Okuma

Tahmin modeli için seçilen araç verileri aynı gün farklı saat dilimlerine ait olup veri boyutu dikkate alınarak orta yoğun ve yoğun trafik olarak ayrılmıştır. Yoğun veri kümesi olan HV1 ve orta yoğun veri kümesi HV2 olarak ilk aşamada okunan ve model testinde kullanılan veriler araçlara yerleştirilen mobil cihazlarından elde edilen verilerdir. Araç verileri, ID bilgisi, zaman damgası, enlem-boylam bilgisi, anlık hız değeri ve açılal yön bilgisinden oluşmaktadır ve ham veri Python-Pandas Kütüphanesiyle okunarak veri çerçevesine aktarılmıştır.

4.2. Veri Filtreleme

Elde edilen ve okunan veriyi istenilen yola yerleştirmek, sonrasında hız tahmini yapmak için trafik verisi zaman filtreleme ve bölge sınırlandırma işlemlerinden geçirilmiştir. Trafik hız tahmin modelinde kısa zamanlı bir tahmin yapılması amaçlandığından seçilen veri kümeleri zamansal olarak filtrelenmiş, yoğun trafik verisi HV1 15:10-15:20 zaman aralığına, orta yoğun trafik verisi HV2 10:30-10:40 zaman aralığına olacak şekilde 10'ar dakikalık zaman dilimine çekilmiştir. Zaman dilimi filtreleme işlemi sonrasında bölgesel sınırlandırma için

Ankara iline ait verilere hedef bölge olan İstanbul ve Eskişehir yol güzergâhlarının bulunduğu bölgeler için filtreleme işlemleri uygulanmıştır. Filtreleme işlemi uygulanırken iki farklı Enlem ve iki farklı Boylam noktası seçilmiş ve araç verileri bir dörtgen arasında kalmıştır.

Bu noktalar;

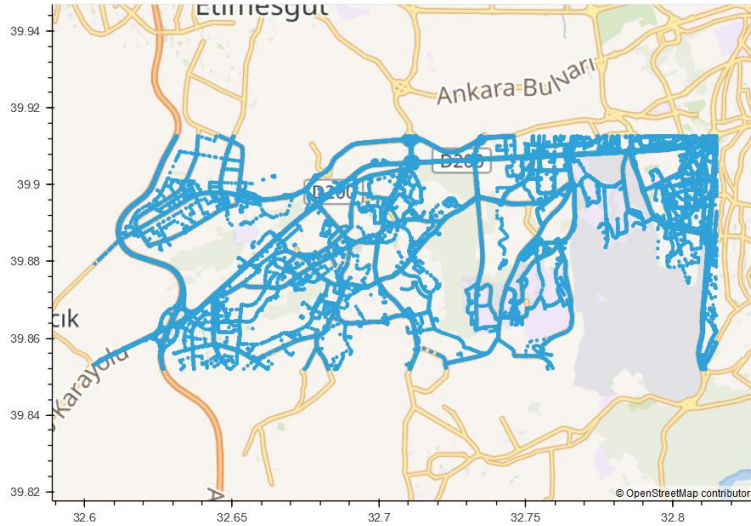
Eskişehir Yolu için;

- $32,815056^{\circ}$ K, $32,603345^{\circ}$ K Enlem noktaları ve $39,851896^{\circ}$ D, $39,912720^{\circ}$ D Boylam noktaları,

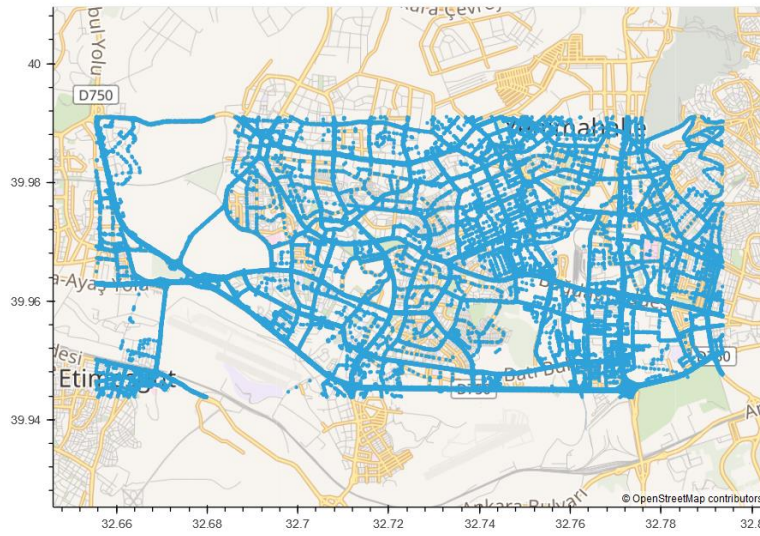
İstanbul Yolu için;

- $32,793608^{\circ}$ K, $32,655502^{\circ}$ K Enlem noktaları ve $39,991057^{\circ}$ D, $39,943762^{\circ}$ D Boylam noktaları,

olarak belirlenmiştir. Hem zamansal hem de mekânsal olarak yukarıda belirlenen filtreleme işlemleri uygulandığında Eskişehir ve İstanbul yolları üzerinde oluşan araç konum noktaları Şekil 4.2-4.3’de görülmektedir.



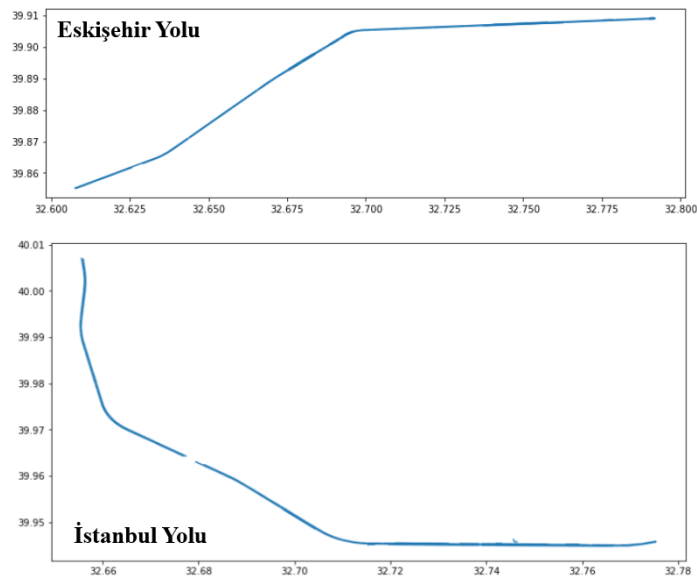
Şekil 4.2. VK1 - Eskişehir yolu filtrelenen araç konum noktaları



Şekil 4.3. VK1 - İstanbul yolu filtrelenen araç konum noktaları

4.3. Harita Eşleme ve Yön Seçimi

Hedef bölge ve tahmin aralığı oluşturulduktan sonra ilgili araç verileri için harita eşleme işlemi gerçekleştirilmiştir. Harita eşleme işlemi için geometrik şekillerden oluşan bir Yol Geometrik Dosyası (SHP) dosyası kullanılmış, her bir GPS noktası tampon oluşturularak seçilen SHP dosyası ile eşleştirme işlemi yapılmıştır. Eşleme işleminde SHP dosyalarının mekân indeksini kullanılmıştır. Tahmin modeli için seçilen yollara ait SHP bilgilerinin örnek gösterimi Çizelge 4.2’de, bu bilgilerle çizimleri Şekil 4.4’de verilmiştir.



Şekil 4.4. Eskişehir ve İstanbul yolları SHP çizimleri

Çizelge 4.2. Seçilen yollara ait SHP dosya bilgilerine ait örnek gösterim

ID	ADI	GEOMETRY
600**	Eskişehir Yolu	LINESTRING (32,78312 39,90847, 32,78326 39,90848)
603**	Eskişehir Yolu	LINESTRING (32,71574 39,90578, 32,72120 39,90599)
600**	Eskişehir Yolu	LINESTRING (32,62835 39,86292, 32,62147 39,86035)
600**	FSM Bulvarı	LINESTRING (32,73853 39,94533, 32,73841 39,94533)
600**	FSM Bulvarı	LINESTRING (32,73177 39,94503, 32,73558 39,94500)
600**	FSM Bulvarı	LINESTRING (32,70623 39,94771, 32,70576 39,94798)

Bu tez çalışmasıyla geliştirilen trafik hız tahmin modelinde, tahmin yapılacak yol için harita eşleme işlemi bir ön işlemdir. Bu işlem Ankara ilinde bulunan iki ana güzergâh Eskişehir ve İstanbul Yolları için yapılmıştır ve bu yollar iki yönlüdür. Yola eşlenen araç verileri iki yönde hareket etmesinden dolayı aynı yönde giden araçlar için açısız yön dikkate alınarak ekstra bir seçim işlemi gerçekleştirilmiştir. Kuzey yönü açısız yönü 0° olacak şekilde, Eskişehir yolu doğu ve kuzey-doğu yönünde 20-100°, İstanbul yolu için batı ve kuzey-batı yönünde 240-359° açısız yön aralığında veriler filtrelenmiştir. Bu değerler dikkate alınarak açısız yön filtreleme uygulandığında Eskişehir yolu trafiğin sağından, İstanbul yolu trafiğin solundan ilerleyen araçlar seçilmiştir.

4.4. Araç Kümelerinin Oluşturulması

Harita eşleme işlemi uygulandıktan ve hedef yol için yönlü araçlar seçildikten sonra enlem ve boylam bilgileri dikkate alınarak yoğunluk tabanlı kümeleme yöntemi OPTICS ile araç verileri kümelenmiştir. Kümeleme işlemi için minimum araç sayısı 20 olarak belirlenmiş ve bu sayı isteğe bağlı olarak değişmektedir. DBSCAN yerine OPTICS yönteminin kullanılması ile küme için sadece eleman sayısının verilmiş ve herhangi bir minimum mesafe belirlenmemiştir. Minimum mesafenin belirlenmemesi sayesinde araç verileri kendi değişkenliği içinde kümelenecek ve yol karakteristiğini oluşturacak araçlar belirlenecektir.

4.5. Aykırı Kümelerin ve Verilerin Temizlenmesi

Yoğunluk tabanlı kümeleme işlemi sonucunda kümelenmeyen GPS kayıtları -1 değeri olarak işaretlenmektedir ve kümelenmeyen bu kayıtlar veriden temizlenmiştir. Ayrıca kümeleme işlemi yapıldıktan sonra kavşaklarda farklı yönde/yolda hareket eden ve yol üzerinde duraklayan araçlar tespit edilmiş, aykırı veri temizliği için kullanılan [114] standart sapma

değeriyle açısıl yön ve hız değerleri dikkate alınarak aykırı araçlar kümelerden çıkarılmıştır. Küme içindeki araç sayısı “n”, açısıl yön ve hız değerleri “X” değerini göstermek üzere Standart Sapma ve ortalama hesabı Eş. 4.1, Eş. 4.2’de gösterilmiştir.

$$\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2} \quad (4.1)$$

$$\mu = \frac{1}{n} \sum_{i=1}^n X_i \quad (4.2)$$

4.6. Küme Özniteliklerinin Çıkarımı

Geliştirilen tahmin modelinde yol karakteristiği oluşturmak için her bir kümeye ait öznitelikleri üretilmiştir. Üretilen küme öznitelikleri Çizelge 4.3 ile listelenmiş ve hesaplama eşitlikleri gösterilmiştir.

Çizelge 4.3. Küme Öznitelikleri

Küme Öznitelikleri	Küme Öznitelik Çıkarımında Kullanılan Hesaplamalar
Küme Araç Sayısı (k_s)	Küme içerisinde her bir a aracının ürettiği sinyal toplamı ($\sum_{i=1}^n a_i$)
Küme Hızı (k_v)	Her aracın hızı v, toplam hızın küme araç sayısına oranı ($\frac{\sum_{i=1}^n v}{k_s}$)
Ön Araç X	Min($X_{Küme}$) Batıdan Doğuya hareket eden araçlardan minimum X (Boylam) değeri
Ön Araç Y	Min($X_{Küme}$) değerine sahip aracın Y (Enlem) değeri
Arka Araç X	Max($X_{Küme}$) Doğudan Batıya hareket eden araçlardan minimum X (Boylam) değeri
Arka Araç Y	Max($X_{Küme}$) değerine sahip aracın Y (Enlem) değeri
Küme Uzunluğu (k_u)	Ön araç ve Arka Araç koordinatları arası Haversine uzaklığı (metre)
Küme Yoğunluğu (k_y)	Küme uzunluğunun küme araç sayısına oranı (k_s/k_u)

4.7. Akış Oluşturma

Trafik hız tahmin modelimiz için akış, belirlenen bir başlangıç noktası ve küme merkezinin bu noktaya uzaklığı ile belirlenerek oluşturulmakta, başlangıç noktasından itibaren oluşan kümeler yakınlığına göre sıralanmaktadır. Eskişehir yolunda trafikte doğu yönünde

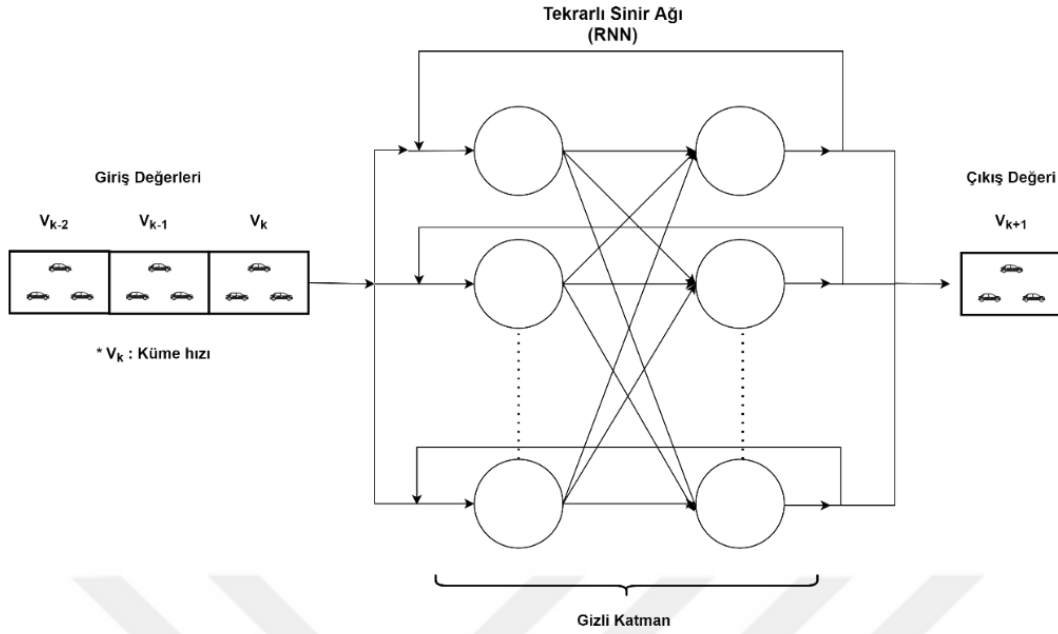
32,603345° K - 39,851896° D (Yaşamkent Bölgesi), İstanbul yolu batı yönünde 32,776352° K – 39,946020° D (Gimat Kavşağı) başlangıç noktaları seçilmiştir. Küme merkezlerinin başlangıç noktalarına uzaklığı Haversine uzaklığı ile hesaplanmıştır.

4.8. Görselleştirme

Akış oluşturulan ve başlangıç noktasına yakınlığına göre sıralanan kümelerin hızları, küme bölgelerinde trafiğin anlık hızını oluşturmuş, trafiğin anlık hız gösterimi için renklendirme yapılmış ve harita üzerinde görselleştirilmiştir. Anlık trafik hız gösterimi için farklı renkler kullanılmıştır. 60 km/s hızı ve üzerindeki değerler için trafiğin akıcı olduğunu belirtmek için mavi renk seçilmiştir. Daha yoğun trafiği göstermek için sırasıyla; 0-25 km/s kırmızı, 25-45 km/s kahverengi, 45-60 km/s turuncu renk ile gösterilmiştir.

4.9. Hız Tahmini ve Sonuçları Karşılaştırma

Geliştirilen tahmin modelinin trafik hız tahmini aşamasında kullanılan RNN varyantı olan LSTM ve GRU sinir ağı modelleri 2 katmandan oluşurken modellerde sırasıyla 128 ve 64 nöron kullanılmıştır. Modeller için kullanılan veri kümeleri %80 eğitim ve %20 test olarak ikiye ayrılmış, 200 adımda eğitilen modellerle tahmin yapılmıştır. Tahmin sonucu üretilen hız değerleri ve orijinal hız değerlerinin karşılaştırılması RMSE, MAE ve MAPE hata metrikleri kullanılarak yapılmıştır. Tahmin modeli için küme özniteliklerinden küme hızı girdi olarak kullanılmış ve çıktı olarak yine küme hızı elde edilmiştir. Tahmin modelinde uygulanan sinir ağının genel modeli Şekil 4.5’de gösterilmiştir.



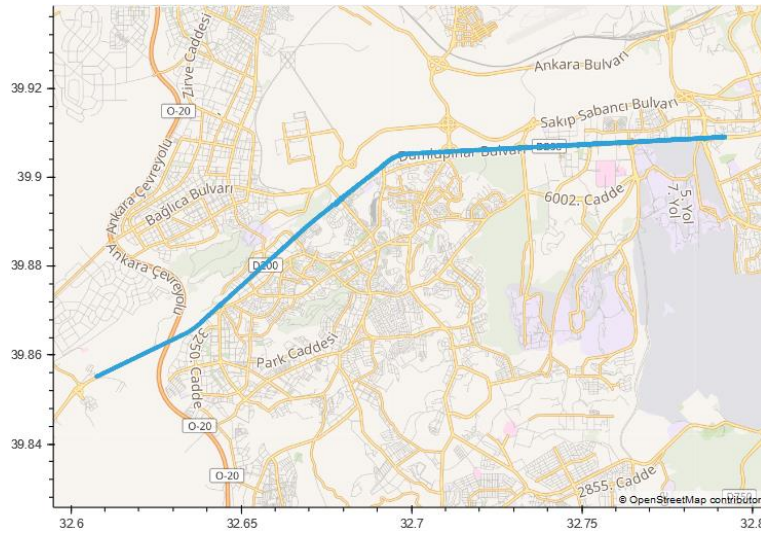
Şekil 4.5. Trafik hız tahmini için uygulanan sinir ağı modeli

4.10. Deneyisel Sonuçlar

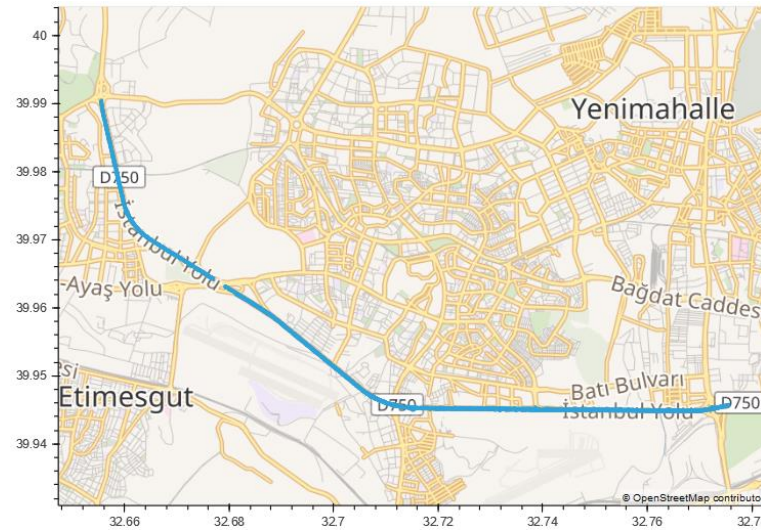
Şekil 4.1’de detayları açıklanan model gerçekleştirilmiş, modele ait test sonuç ve çıktıları bu bölümde açıklanmıştır. Modellerin test sonuçlarının doğrulaması için yoğun ve orta yoğun trafiğin olduğu iki zaman dilimine, Eskişehir ve İstanbul yolu olmak üzere iki ana arter yola ait veriler seçilmiştir. Seçilen veri kümeleri öncelikle, trafik hız tahmin modelimizin başlangıç aşamalarında ham veri okunmuş, belirlenen bölgeye filtrelenmiş ve yol ağına eşleştirilmiş, yoğunluk tabanlı kümeler elde edilerek veri temizleme işlemi yapılmıştır. Veri temizleme işlemi sonrasında elde edilen kümelerin özniteliklerinin üretilmiş ve bir başlangıç noktasına göre kümeler sıralanarak bir akış oluşturulmuştur. Bu işlemlerle yola ait karakteristik oluşturmuş ve sonrasında küme öznitelikleri ile önce görselleştirme yapılmış ve sonrasında küme hız değerleri LSTM ve GRU sinir ağıları için girdi sağlamış ve trafik hız tahmini yapılmıştır.

Tahmin modelini doğrulamak ve test etmek üzere Ankara iline ait 27 Ocak 2021 tarihine ait iki farklı zaman dilimi araç verisi Eskişehir ve İstanbul yollarını içeren 2 farklı bölge için filtrelenmiş ve literatürde de kısa zamanlı olarak ifade edilen 10 dakikalık zaman dilimine indirgenerek kısa zamanlı bir tahmin gerçekleştirilmiştir. Tahmin yapılacak yol için Eskişehir yolu için batı-doğu yönünde, İstanbul yolu için doğu-batı yönünde giden araçlar seçilmiştir. Filtreleme ve ön işlem yapılmayan Ankara verisi saat 09:59-10:59 arası 3

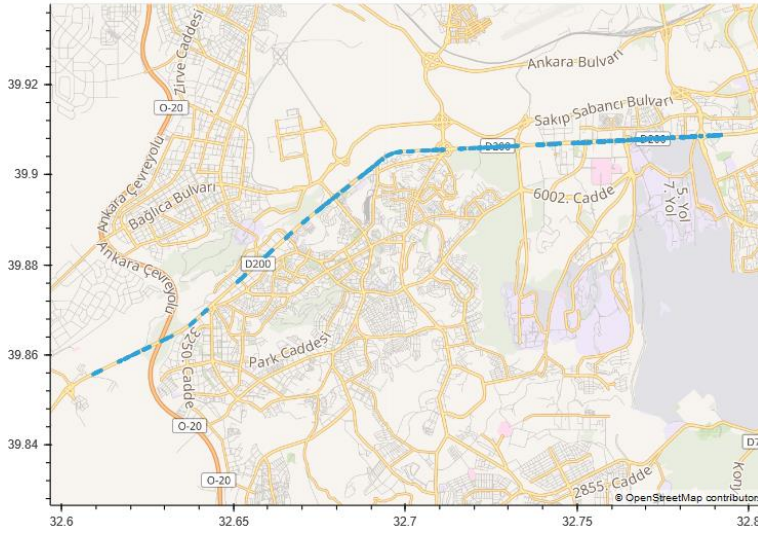
milyon, saat 14:59-15:59 arası yaklaşık 3,4 milyon GPS kaydını içermektedir. Filtreleme, zaman seçimi ve harita eşleme işlemleri sonucunda elde edilen VK2’de GPS kayıt sayısı Eskişehir Yolu 5,7 bin ve İstanbul Yolu 7,1 bin civarındadır. Detayları 4.2 ve 4.3 bölümünde anlatılan veri filtreleme ve harita eşleme işleminin sonuçları Şekil 4.6 ve 4.7’de harita üzerinde görselleştirilmiştir. Her iki yola yerleştirilen araç verileri yoğunluk tabanlı kümeleme yöntemi ile kümelenmiş ve kümeleme işlemi minimum araç sayısı olarak 20 araç seçilerek oluşan kümelerin harita üzerinde gösterimi Şekil 4.8 ve 4.9’da görülmektedir.



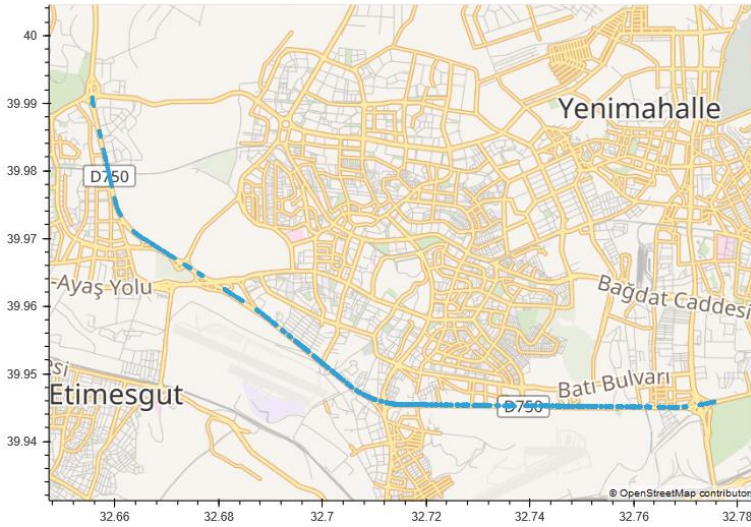
Şekil 4.6. Eskişehir yolu ağ bilgisi ile eşlenen noktaların görünümü



Şekil 4.7. İstanbul yolu ağ bilgisi ile eşlenen noktaların görünümü



Şekil 4.8. Eskişehir yolu - Elde Edilen Yoğunluk Tabanlı Kümeler



Şekil 4.9. İstanbul yolu - Elde Edilen Yoğunluk Tabanlı Kümeler

Kümeleme işlemi sonrasında yoğunluk tabanlı kümelerde standart sapma değerlerine göre veri temizleme işlemi gerçekleştirilmiştir. Veri temizleme işlemi küme içinde bulunan, duraklayan veya oldukça yavaşlayan böylelikle hız ortalamasını olağan dışı etkileyen ve ayrıca kavşaklarda dönüş yapan, yan yolda ilerleyen hedef yola ait olmayan araçların çıkarılması için yapılmıştır. Veri temizleme işlemi 3. Bölümde açıklanan standart sapma değeriyle gerçekleştirilmiş, bu değer kullanıcılara bağlı bir değer olup [114] çalışmalarda deneysel olarak belirlenen 1,2 değeri alınmıştır. Örneğin; Eskişehir yolu 6. Küme içerisindeki araçlardan hem yan yolda ilerleyen ve hız ortalamasını olağan dışı etkileyen araçlar mevcuttur ve bu araçlardan bazı örnekler Şekil 4.10 ve Çizelge 4.4'te gösterilmiştir.



Şekil 4.10. Eskişehir yolu güzergâh dışı örnek araçlar

Çizelge 4.4. Eskişehir yolu güzergâh dışı araçlardan örnek gösterim

ID	ZAMAN	X	Y	HIZ	A. YÖN	Küme No
c9f****	2020-01-27 15:14:49	32,679493	39,894852	7	68	6
fe9****	2020-01-27 15:13:00	32,679302	39,894840	10	96	6
f5d****	2020-01-27 15:11:24	32,680267	39,895489	73	52	6
01a****	2020-01-27 15:11:17	32,678818	39,894707	109	51	6
c6e****	2020-01-27 15:14:19	32,678902	39,894711	138	51	6

Aykırı ve güzergâh dışı veri temizleme işlemi gerçekleştirildikten sonra trafik yoğunluk gösterimi ve tahmin modeli için girdi hız değerleri oluşturulması ve görselleştirme yapabilmek için kümelere ait öznitelik üretim işlemi gerçekleştirilmiştir. Her küme için belirlenen ve üretilen öznitelikler; küme içindeki araç sayısı, küme araçlarının ortalama hızı, küme içerisindeki en ön ve arkadaki aracın enlem ve boylam bilgileri, metre cinsinden küme uzunluğu ve küme içerisindeki küme araç yoğunluğu olarak belirlenmiştir. Model değerlendirmesi için kullanılan VK1 ve VK2 veri kümelerinden VK1 için veri hazırlama aşamaları olan veri filtreleme, harita eşleme, yön seçimi, küme oluşturma ve aykırı verilerinin temizlenme aşamaları detaylıca açıklanmış, VK2 için tekrar aynı sonuçları vermemek adına küme öznitelik aşamasından sonraki süreçler anlatılmıştır. Bahsedilen süreçlerin ve aşamaların sonuçları takip eden alt başlıklarda verilmiştir.

4.10.1. VK1 için elde edilen sonuçlar

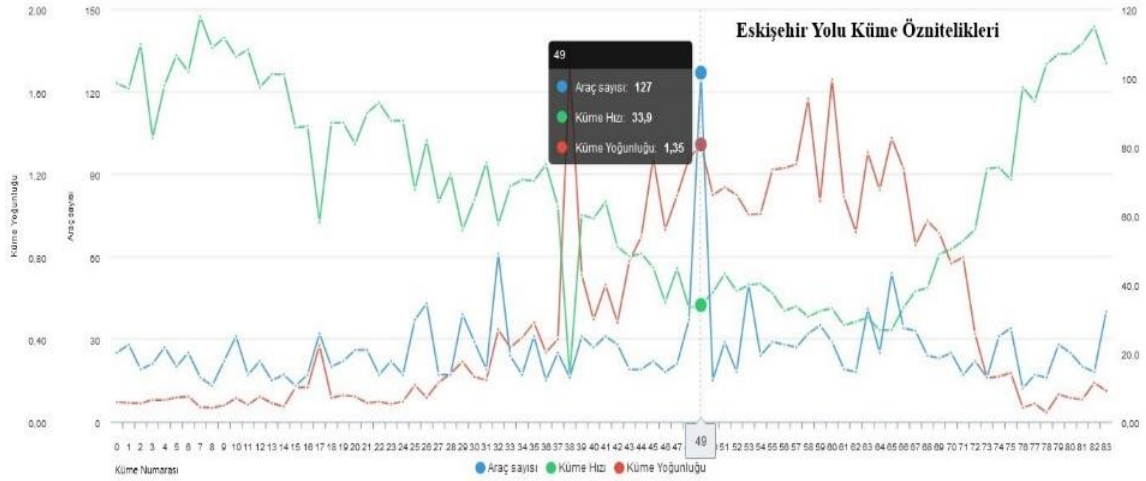
VK1'den Eskişehir yol güzergâhı için 84 ve İstanbul yol güzergâhı için 118 küme oluşmuş, Eskişehir yol güzergâhı kümeleri içerisinde en fazla araç sayısına 127 araç ile 49. Küme olmuştur. Bu sayı İstanbul Yol güzergâhı için 70 araç ile 27. kümedir. Küme içi araç yoğunluğunun en fazla olduğu küme Eskişehir yolu için 38. küme, İstanbul yol güzergâhı için 31. küme ve küme yoğunluğunun en fazla olduğu bu iki küme için hız ortalamaları sırasıyla 13,44 ve 28,27 km/h'dir. Çizelge 4.5-4.6'da Eskişehir ve İstanbul yolunda oluşan kümelere ait öznitelikler değerleri örnek olarak verilmiş olup hız, yoğunluk ve araç sayı grafikleri Şekil 4.11-4.12'de görülmektedir. Şekil 4.11 ve 4.12 incelendiğinde araç sayısının ve yoğunluğunun artması bize hız değerlerinin düştüğünü gösterirken, özellikle hız ve yoğunluk çizgi grafiğine dikkat edildiğinde 43. ve 70. kümeler için kesiştiği noktada trafik yoğunluğu görülmektedir. Çizgilerin ayrılmaya başladığı noktadan sonrasında aralarındaki ters orantı ilişkisi devam etmektedir.

Çizelge 4.5. VK1-Eskişehir yol güzergâhı için örnek küme öznitelikleri

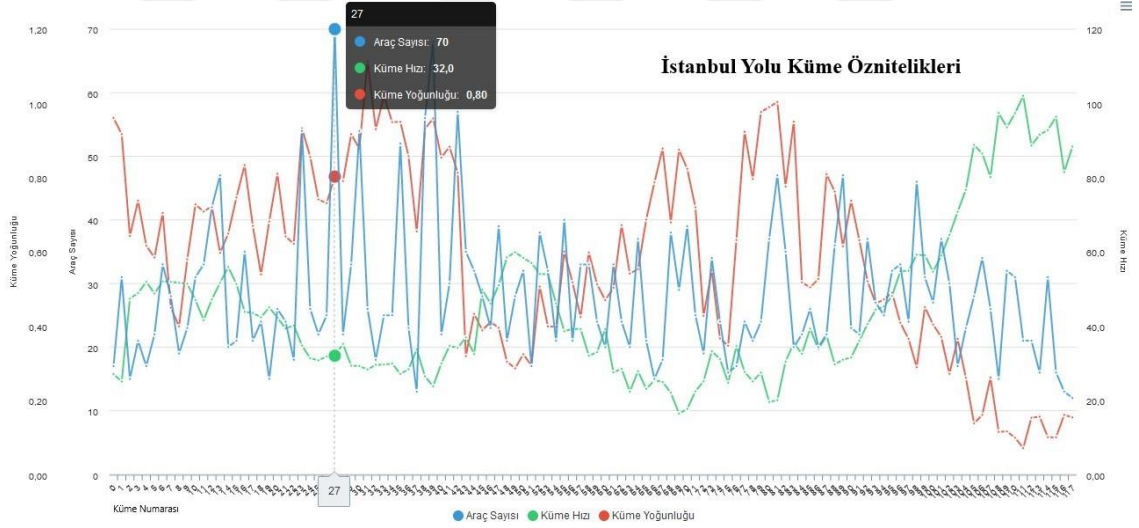
Küme No	Küme Araç Sayısı	Küme Hızı	Ön Araç X	Ön Araç Y	Arka Araç X	Arka Araç Y	Küme uzunluğu (m)	Küme Yoğunluğu
42	28	50,8929	32,7677	39,9079	32,7672	39,9078	58,2875	0,4804
60	29	32,9655	32,7779	39,9082	32,7778	39,9083	17,4601	1,6609
58	32	30,5312	32,7766	39,9082	32,7764	39,9082	20,4261	1,5666
38	16	13,4375	32,7637	39,9076	32,7636	39,9076	9,2291	1,7336
49	127	33,9055	32,7711	39,9080	32,7703	39,9079	94,4029	1,3453

Çizelge 4.6. VK1-İstanbul yol güzergâhı için örnek küme öznitelikleri

Küme No	Küme Araç Sayısı	Küme Hızı	Ön Araç X	Ön Araç Y	Arka Araç X	Arka Araç Y	Küme Uzunluğu (m)	Küme Yoğunluğu
52	38	54,0263	32,7291	39,9453	32,7298	39,9453	75,0777	0,5061
105	28	88,7857	32,6829	39,9615	32,6847	39,9608	204,1740	0,1371
103	17	70,7647	32,6883	39,9588	32,6887	39,9587	46,5699	0,3650
31	26	28,2692	32,7467	39,9452	32,7469	39,9453	23,3571	1,1131
27	70	31,9571	32,7490	39,9452	32,7498	39,9451	87,2221	0,8025



Şekil 4.11. VK1-Eskişehir yolunda oluşan kümelerin araç sayı, hız ve yoğunluk grafiği



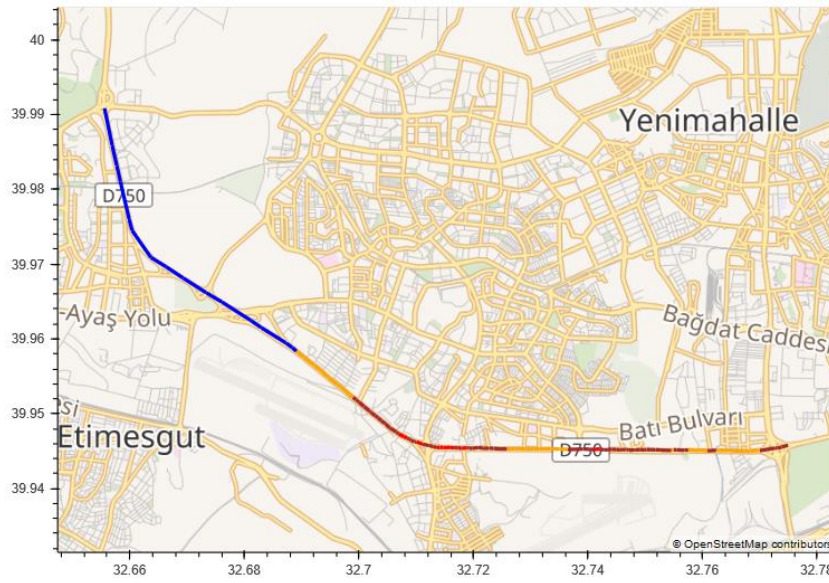
Şekil 4.12. VK1- İstanbul yolunda oluşan kümelerin araç sayı, hız ve yoğunluk grafiği

Küme öznitelikleri üretimi sonrasında belirlenen başlangıç noktasına göre Haversine uzaklığı ile kümeler sıralanmış ve akış oluşturulmuştur. Her iki yol için başlangıç noktasına en yakın 5 küme ve metre uzaklık değerleri Çizelge 4.7 ile gösterilmiştir.

Akış oluşturulan kümeler ve bu kümelere ait öznitelik değerleri kullanılarak harita üzerinde anlık trafik hız gösterimi yapılmıştır. Yoğunluk haritalarında mavi renk 60 km/s üzerindeki hızı gösterirken, koyu renkler daha düşük hızda yoğun trafiği göstermektedir. Harita üzerinde anlık trafik hızı kümelere ait ön ve arka araç değerleri kullanılarak çizdirilmiş, İstanbul yolu ve Eskişehir yolu güzergâhları için anlık hız trafiğinin gösterimi Şekil 4.13 ve 4.14'de gösterilmiştir.

Çizelge 4.7. VK1-Başlangıç noktalarına göre en yakın 5 küme ve uzaklık değerleri

Eskişehir Yol Güzergâhı		İstanbul Yol Güzergâhı	
Küme No	Başlangıç Noktasına Uzaklık (km)	Küme No	Başlangıç Noktasına Uzaklık (km)
83	0,330293	44	0,194372
82	1,097951	43	0,299727
81	1,348346	1	0,423741
80	1,625581	0	0,474760
79	2,000304	2	0,775014

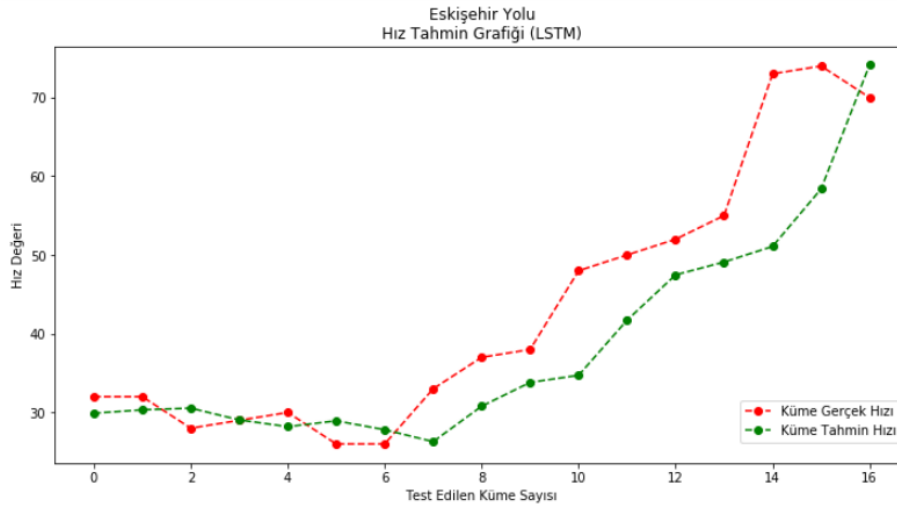


Şekil 4.13. VK1-İstanbul Yolu Anlık Trafik Hız Gösterimi

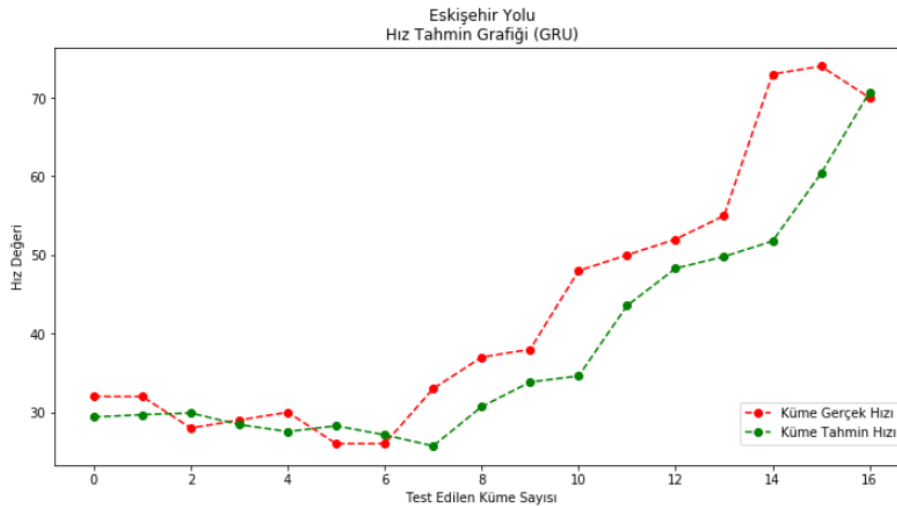


Şekil 4.14. VK1-Eskişehir Yolu Anlık Trafik Hız Gösterimi

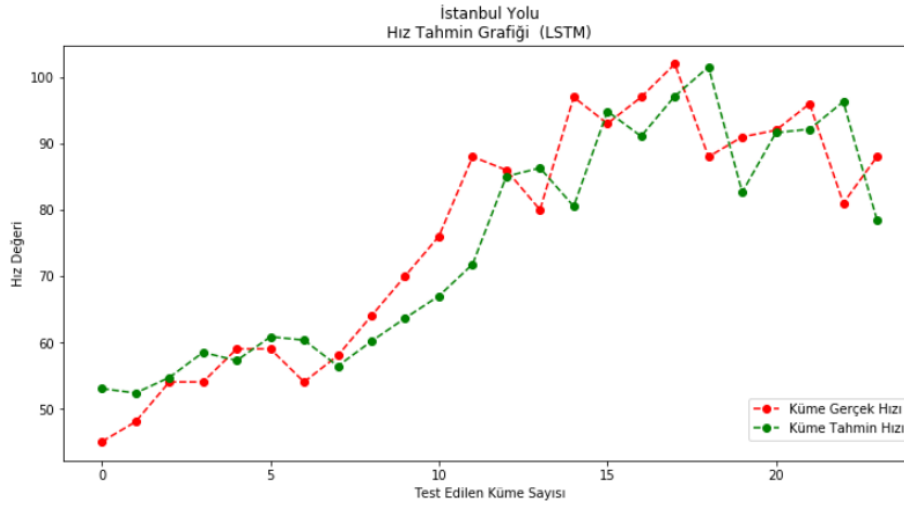
Trafik tahmin modeli Şekil 4.5’de verilen sinir ağı modeli temel alınarak testler yürütülmüştür. Kümelere ait küme hızı değerleri LSTM ve GRU sinir ağı modelleri için giriş değerleridir. Her iki yolda oluşan kümeler için %80 eğitim verisi seçilmiş, %20 küme sayısı ile modeller test edilmiştir. Farklı yollar ve veri kümeleri için bu orana bağlı olarak test ve eğitim küme sayıları değişkenlik göstermektedir. LSTM-GRU sinir ağları modelleri için 2 katman oluşturulmuş, birinci katmanda 128 nöron, ikinci katmanda 64 nöron kullanılmıştır. Bu ağlar, 200 turda eğitilmiş olup elde edilen gerçek ve tahmini hız değer grafikleri Şekil 4.15-4.18’de verilmiştir.



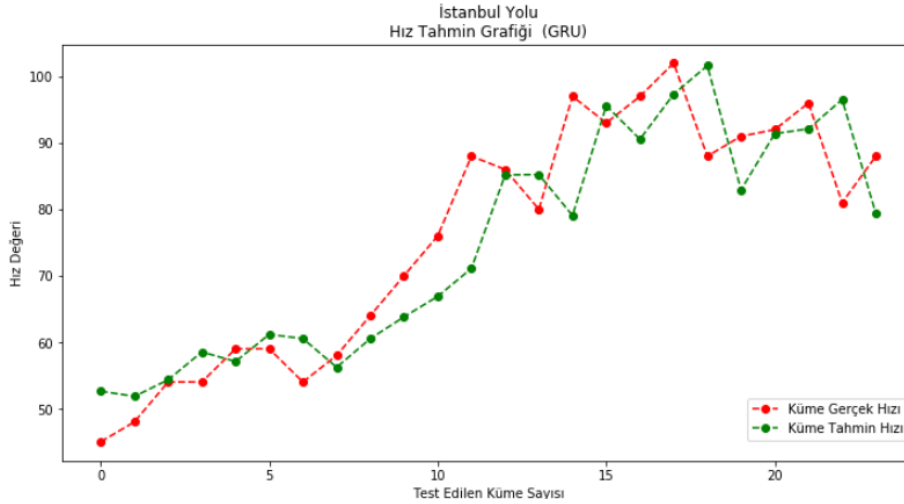
Şekil 4.15. VK1-Eskişehir Yolu LSTM hız tahmin sonuçları



Şekil 4.16. VK1-Eskişehir Yolu GRU hız tahmin sonuçları



Şekil 4.17. VK1-İstanbul Yolu LSTM hız tahmin sonuçları



Şekil 4.18. VK1-İstanbul Yolu GRU hız tahmin sonuçları

Grafiklerde görülen model tahmin sonuçlarının hata oranları RMSE, MAPE ve MAE hata metrikleriyle ölçülmüş ve hata değerleri Çizelge 4.8’de gösterilmiştir. Sinir ağıları modelleri için son üç küme hız değeri zaman penceresi olarak seçildiğinde tekrarlayan örüntülerle orijinal hız verilerine yakın değerler üretilmiştir ve birbirine yakın hata değerleri ölçülmüştür. Şekil 4.15 ve 4.16 incelendiğinde Eskişehir yolunda oluşan test kümelerinde ilk 7 kümede yoğun trafik oluşurken sonraki kümelerde trafik yoğunluğu azalmakta ve tahmin sonuçları her iki yoğunluktaki kümelerde paralel sonuçlar vermektedir. İstanbul yol güzergâhında bulunan test kümelerinin bulunduğu Şekil 4.17 ve 4.18’de trafik orta yoğunluktan daha akıcı olan trafiğe doğru yönelim göstermektedir. Test kümelerinden 11. küme sonrası hızlı değişimler oluşurken bu kümelerden 13 ile 14. ve 17 ile 18. kümelerde ani düşüş ve çıkış gözlemlenmektedir. Bu noktalarda ani değişimler her iki yol için kavşaklardaki

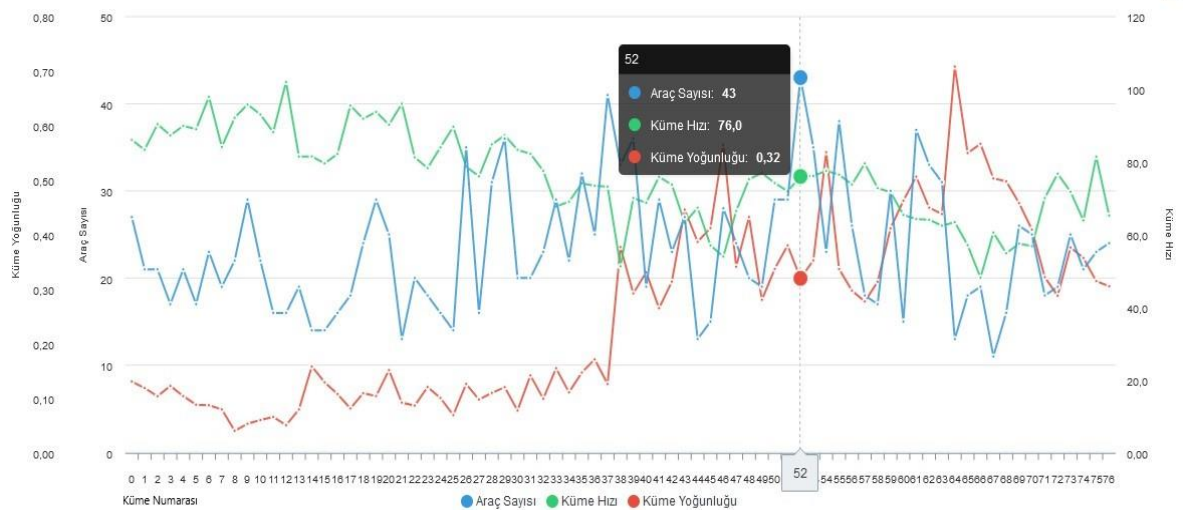
yavaşlamalardan kaynaklanmakta ve tahmin modelinin hata oranı attırmakta fakat diğer kümelerde yakın tahminler üretmektedir. Kavşaklardaki yavaşlamalar ve ani hız değişimleri hata oranı yükseltse de genel toplam hata oranı literatürdeki çalışmalar değerlendirildiğinde tatmin edicidir.

Çizelge 4.8. VK1-Trafik hız tahmini için hata değerleri

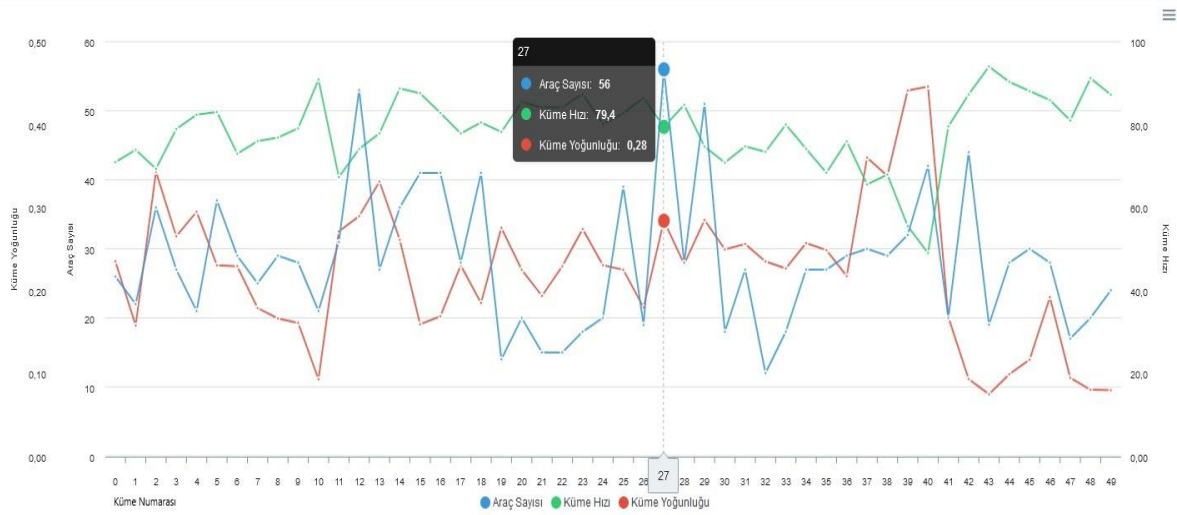
	Eskişehir Yolu Güzergâhı		İstanbul Yolu Güzergâhı	
	LSTM Modeli	GRU Modeli	LSTM Modeli	GRU Modeli
RMSE	8,302	7,814	7,958	8,084
MAPE (%)	12,597	11,842	8,327	8,312
MAE	6,098	5,606	6,323	6,351

4.10.2. VK2 için elde edilen sonuçlar

Orta yoğun araç verisi olan VK2 kullanarak yapılan tahmin sonuçlarında Eskişehir Yolu için 77 küme oluşmuş, oluşan kümeler içinde 52. kümede 43 araç ile en yüksek sayıda araç gruplanmıştır. En düşük hızı olan küme 66. küme olup küme hızı 48 km/s, küme yoğunluğu 0,5665'dur. İstanbul Yolunda elde edilen 50 küme içerisinde en fazla araç 56 araçla 27. kümede oluşmuş, 40. küme 49 km/s hız ortalamasıyla en düşük hız ortalamasına sahip küme olmuştur. Eskişehir ve İstanbul Yollarında oluşan örnek kümeler için grafikler Şekil 4.19-4.20'de, küme öznitelikleri Çizelge 4.9-4.10'da gösterilmiştir.



Şekil 4.19. VK2- Eskişehir Yolunda oluşan kümelerin araç sayısı, hız ve yoğunluk grafiği



Şekil 4.20. VK2- İstanbul Yolunda oluşan kümelerin araç sayısı, hız ve yoğunluk grafiği

Çizelge 4.9. VK2-Eskişehir yol güzergâhı için örnek küme öznitelikleri

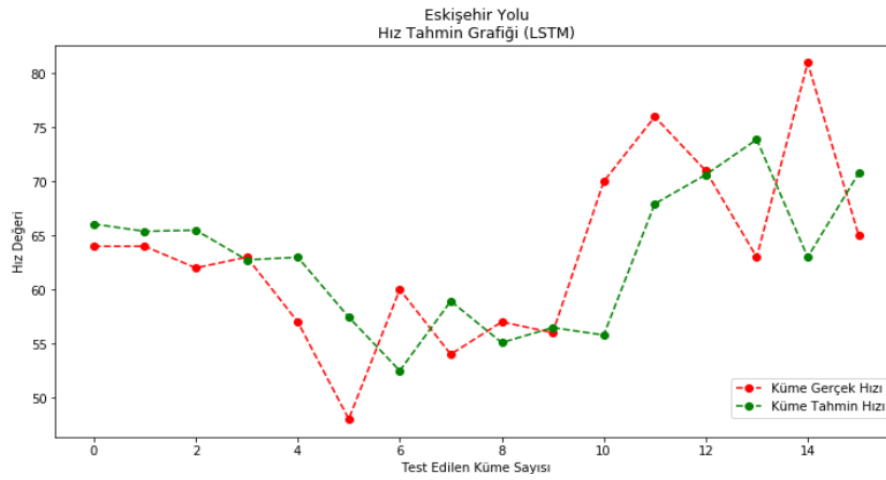
Küme No	Küme Araç Sayısı	Küme Hızı	Ön Araç X	Ön Araç Y	Arka Araç X	Arka Araç Y	Küme uzunluğu (m)	Küme Yoğunluğu
24	16	84	32,7075	39,9056	32,706	39,9056	160,353	0,0998
13	19	81,4737	32,6802	39,8957	32,6783	39,8945	241,546	0,0787
70	25	56,8	32,7857	39,9087	32,7852	39,9087	61,1377	0,4089
66	19	48,1053	32,7822	39,9087	32,7819	39,9086	33,5387	0,5665
52	43	75,9767	32,7698	39,9082	32,7686	39,9081	134,623	0,3194

Çizelge 4.10. VK2-İstanbul yol güzergâhı için örnek küme öznitelikleri

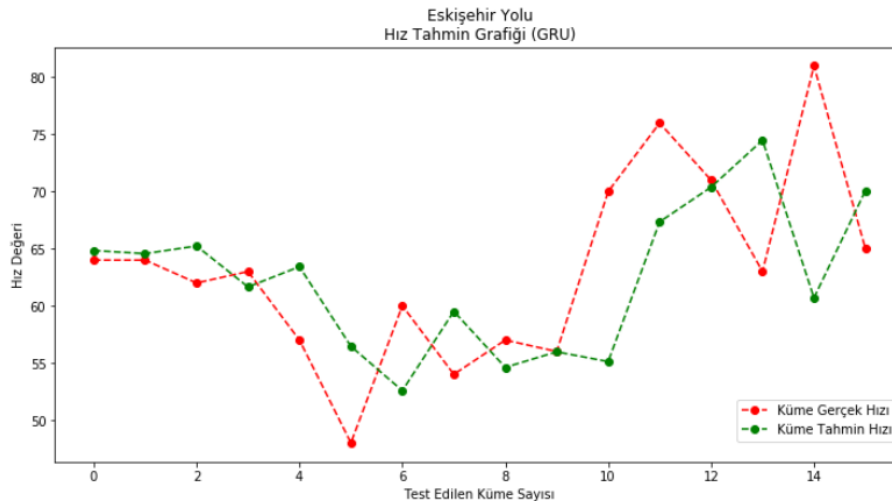
Küme No	Küme Araç Sayısı	Küme Hızı	Ön Araç X	Ön Araç Y	Arka Araç X	Arka Araç Y	Küme uzunluğu (m)	Küme Yoğunluğu
10	21	90,8571	32,6872	39,9594	32,689	39,9584	226,754	0,0926
13	27	77,8889	32,7185	39,9454	32,7192	39,9454	81,506	0,3312
39	32	55,50	32,7718	39,9453	32,7725	39,9454	72,5522	0,4410
40	42	49,01	32,7727	39,9453	32,7735	39,9455	94,2044	0,4458
27	56	79,4464	32,7514	39,9451	32,7531	39,9451	197,26	0,2838

VK2’de önceki bölümde VK1 için uygulanan sinir ağı tur ve nöron sayısı, test-eğitim küme oranları kullanılarak tahmin yapılmıştır. Elde edilen sonuçlar grafik olarak Şekil 4.21-4.24’de, hata değerleri Çizelge 4.11’de verilmiştir. Şekil 4.21 ve 4.22 incelendiğinde VK2 için Eskişehir Yolu ilk test kümelerinde akıcı olan trafik 5. kümede orta yoğun trafiğe dönüşürken sonrasında kademeli olarak açılan akıcı trafik haline gelmiştir. Eskişehir Yolunda ani değişimler 9. kümeden sonra oluşurken özellikle 13 ve 14. küme hız değişimi

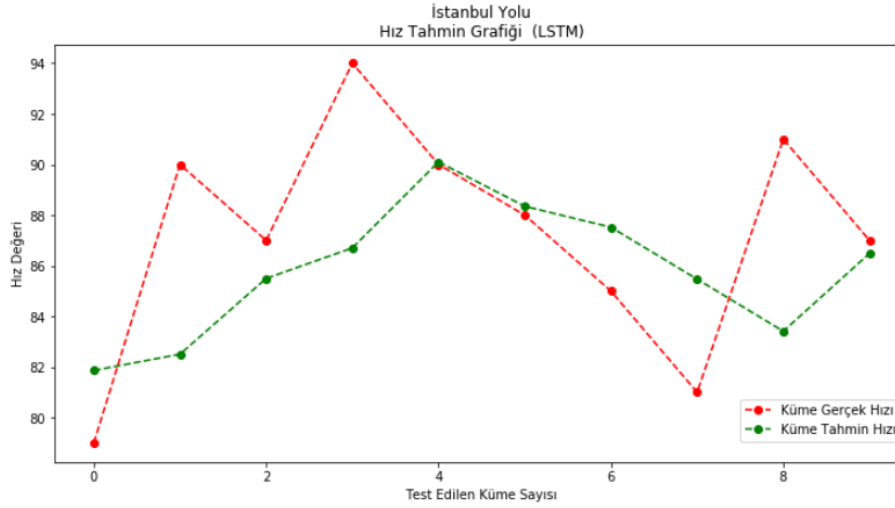
büyük miktardadır ve tahmin hata oranı bu noktada oldukça yüksektir. Bu değişimin sebebi VK1 deneyinde olduğu gibi bu yolda bulunan bağlantı ve kavşak noktasında oluşan yavaşlama ve sonrasında hareketlenen trafik örüntüsünden kaynaklanmaktadır. Şekil 4.23 ve 4.24 incelendiğinde İstanbul yolunda oluşan test küme sayısı diğer deney kümelerine nazaran daha az sayıda olup bu sayı yol üzerinde oluşan toplam küme sayısının az olması ve model eğitim ve test verilerinin %80-%20 oranında seçilmesinden dolayıdır. İstanbul yolu kümelerin grafik hareketleri hızlı değişim gösterse de hız değerlerine özellikle dikkat edilirse değişimin oldukça az olduğu görülecektir ve hata oranı Çizelge 4.11’de olduğu gibi oldukça düşüktür.



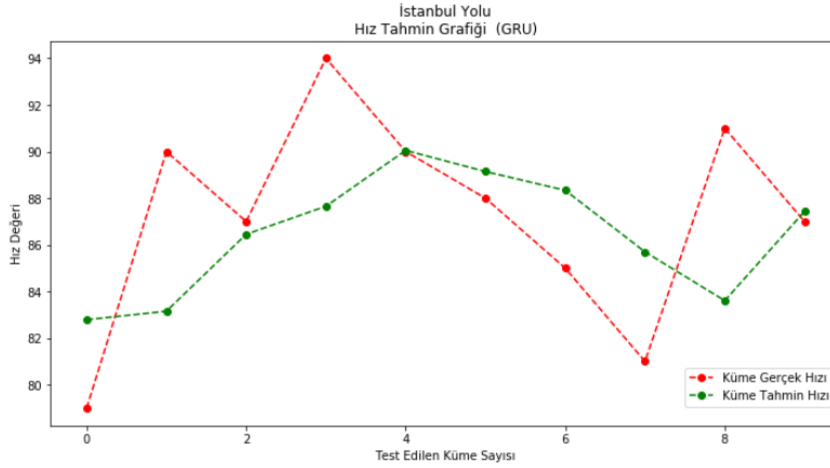
Şekil 4.21. VK2-Eskişehir Yolu LSTM hız tahmin sonuçları



Şekil 4.22. VK2-Eskişehir Yolu GRU hız tahmin sonuçları



Şekil 4.23. VK2-İstanbul Yolu LSTM hız tahmin sonuçları



Şekil 4.24. VK2-İstanbul Yolu GRU hız tahmin sonuçları

Çizelge 4.11. VK2 için Trafik hız tahmini için hata değerleri

	Eskişehir Yolu Güzergâhı		İstanbul Yolu Güzergâhı	
	LSTM Modeli	GRU Modeli	LSTM Modeli	GRU Modeli
RMSE	7,793	8,232	4,519	4,372
MAPE (%)	9,208	9,368	3,936	3,949
MAE	5,925	6,078	3,470	3,458

Bu bölümde, anlık trafik yoğunluğunu gösteren ve trafik parametrelerinden kısa zamanlı hız tahmini yapan yeni ve özgün bir model geliştirilmiş, test edilmiş ve başarı ile sunulmuştur. 9 aşamadan oluşan bu modelde, başlangıç aşamaları, anlık trafik hız gösterimi ve hız tahmini yapmak için mevcut trafik verisini hazırlama aşamalarını içerirken, son aşamalar hız tahmini, test ve karşılaştırma aşamalarını kapsamaktadır. Ankara iline ait iki ana yol olan

Eskişehir ve İstanbul Yollarında aynı gün farklı trafik yoğunluğunda test edilen modelde temel amaç yol için karakteristik oluşturmak olmuştur. Yol karakteristikleri elde edilen yollar için aynı zamanda harita üzerinde anlık trafik yoğunluk gösterimi ve renklendirmesi yapılmıştır. Geliştirilen modellere küme hız değerleri girdi olarak verilmiş ve modeller test edilmiş ve tahmini hız değerleri bulunmuştur. Trafikğin değişken yapısı için önerilen bu modelin, gerçek zaman uygulamaları için başarılı sonuçlar verdiği belirlenmiştir. Gelecek bölümde geliştirilen trafik hız tahmin modeli için mahremiyet dikkate alınarak geliştirilen ve önerilen yeni bir tahmin modeli anlatılmıştır.





5. MAHREMİYET KORUMALI TRAFİK HIZ TAHMİN MODELİ

Akıllı Ulaşım Sistemleri hareketliliği, konforu, güvenliği ve verimliliği artırmak adına algılama ve iletişim teknolojilerini seyahat altyapısına ve ulaşımına entegre etmeyi amaçlamaktadır. Akıllı ulaşım sistemlerini oluşturan akıllı araçlar, toplu taşıma sistemleri, araçlar arası ağlar gibi farklı bileşenlerden veriler toplanmakta, toplanan veriler kimlik ve konum bilgisi gibi tanımlayıcı bilgiler içermesinden dolayı mahremiyet konusunda endişeleri beraberinde getirmektedir [115]. Çünkü konum bilgisinden rota ve yörünge bilgisi oluşturulmakta veya bulundukları POI noktalarından kişiye ait çıkarımlar yapılabilmektedir [116]. Servis sağlayıcılar tarafından kamu veya üçüncü parti kurumlarla paylaşılabilen ve yayınlanan veriler, hassas veriler içerdiğinden, suiistimal edilerek kötü niyetli kullanıma imkân tanımakta, kullanıcılar servis kullanımından önce onayladıkları sözleşmelerden dolayı bu durumun farkında olmamaktadır [117].

Konum ve konumla birlikte oluşan yörünge verileri için uygun veri temizleme işlemi yapılmadan yayınlamak ve kullanmak, kullanıcıların gizliliğine yönelik ciddi tehditler oluşturmaktadır. Yasal düzenlemelerle korunmaya çalışılan konum verisi üzerine oldukça kapsamlı çalışmalar mevcuttur ve bu gibi problemlerin önüne geçmek için veri paylaşımında bazı koruyucu önlemler alınmaktadır. Mevcut çalışmalar, mahremiyeti yörünge veya konum birleştirme yöntemleriyle, girdi verisinden bazı rota/noktaların kaldırılarak temizlenmesi veya kimlik bilgilerinin genelleştirilmesi gibi yöntemlerle gerçekleştirmektedir [87]. Birleştirme, temizleme ve genelleştirme işlemleri verinin kullanılabilirliğini oldukça azaltmaktadır ve dolayısıyla veri mahremiyeti garanti edilirken kullanılabilirlikten feragat edilmektedir.

Akıllı ulaşım sistemlerinde kullanılan veri türleri hassas veriler içermekte, sistem için kullanılan bilgiler genellikle yol üzerinde geçen araç sayısı, hız veya yoğunluk gibi bilgiler olduğundan bütün bir veri kümesinin yayınlanması gerekmemektedir. Bunun yerine trafiği yönetecek veya trafik üzerine çalışma yapacak araştırmacılar için bölge, yol ve segment için karakteristik özelliklerin verilmesi yeterli olacaktır. Veri yayınlanması yerine istatistiki bilgilerinin paylaşılması veya veri tabanlarına elde edilecek sorgu sonuçlarına yönelik uygulanan bir mahremiyet yöntemi olan diferansiyel mahremiyet yöntemi literatürde yaygın olarak kullanılan bir yöntemdir [118] ve önerilen modelde bu yöntem kullanılmıştır.

Bu bölümde mahremiyet korumalı tahmin sonuçları üretecek ve trafik akış, yoğunluk ve hız bilgileri gibi istatistik bilgilerinin paylaşılması için mahremiyet sağlayacak yeni bir mahremiyet korumalı hız tahmin modeli önerilmiştir. Bölüm alt başlıklarında mahremiyet ihlalleri ve saldırılarına örnekler verilmiş, bu saldırılara karşı koruma yöntemleri açıklanmış, konum verisi kullanan sistemlerin algoritma düzeyinde korunması için kullanılacak diferansiyel mahremiyet gerekçeleri verilmiştir. Takip eden bölümlerde mahremiyet korumalı hız tahmin modeli aşamaları anlatılmış, yine 4. Bölümde sunulan model test sonuçları bu bölümde mahremiyet odaklı olarak uygulanmış, elde edilen sonuçlar sunulmuştur.

5.1. Veri Mahremiyet İhlalleri ve Saldırıları

Veri mahremiyeti problemi, veriden elde edilen fayda ile veriye sağlanan mahremiyet arasındaki dengeyi sağlama olarak ifade edilirken istatistiksel ifşa kontrolü, mahremiyet korumalı veri madenciliği ve mahremiyet korumalı veri yayınlama gibi farklı alanlarda araştırma konusudur [119,120]. Veri madenciliği finansal bankacılık, eğitim, sağlık ve trafik planlama gibi çeşitli alanlarda hizmet ve ürün geliştirme sağlamakta fakat verileri çıkarılan ve analiz edilen kişilerin mahremiyetine ilişkin endişeleri de arttırmaktadır [121].

Veri mahremiyetini ihlalleri ve saldırılarına örnek olarak;

- Narayan vd. [122] Netflix kullanıcılarının film tavsiye servislerini geliştirmek adına kullanıcı oylarını açık bir şekilde yayınladığı veri kümesini IMDB sitesinin veri tabanı ile eşleyerek kullanıcıları tanımlayarak açığa çıkarmıştır. Netflix yayınladığı veri kümesinde tanımlayıcı bilgileri veri tabanından çıkardığını ifade etse de araştırmacılar bağlantı atağı gerçekleştirerek kullanıcıları tanımlamayı başarmıştır.
- Amerika Ulusal Sağlık Enstitüsünün genetik kaynaklı hastalıkları tespitinde yeni araştırmaları destekleme amacıyla kişi kimlik bilgisi olmadan genetik işaretler ve hastalık bilgilerini yayınlamış, Wang vd. [123] X^2 istatistik yöntemiyle kişileri tanımlamıştır. Bu ifşa çalışmasının ardından enstitü veri yayınlamayı durdurmuştur.
- Amerika bulunan özel bir sigorta şirketi müşterilerine ait verileri isimleri ve sosyal sigorta numaraları olmadan açık bir şekilde yayınlamış, Sweeney [124] farklı bir kurumdan satın aldığı veri tabanı ile bu bilgileri eşleyerek kullanıcı bilgilerine ulaşmıştır.

Sweeney yaptığı çalışmada kişiye ait doğum günü, cinsiyet ve posta kodu bilgileriyle veri tabanındaki kişilerin %87'sini tanımlayabildiğini ifade etmiştir.

- Facebook kullanıcı bilgileri bir analiz şirketi ile kullanıcıların izni olmadan paylaşılmış politik eğilimler üzerinde kullanılmıştır. İzinsiz veri paylaşımının yanı sıra kullanıcıların arkadaşlarına ait bilgilerde farklı tekniklerle toplanmış, bağlantı temelli veri ihlali gerçekleşmiştir [125,126].

Yukarıda verilen ve literatürde yer alan veri mahremiyetini ihlal eden ve saldırılarını gerçekleştiren kişiler, özellikle farklı veri tabanları ile bağlantı kurarak bağlantı atakları gerçekleştirmekte veya arka plan bilgisini kullanarak çıkarım saldırıları yapmaktadır. Örneğin kimlik bilgisi olmadan sadece özel bir hastalık türü için pozitif ve negatif sonuç bilgilerinin paylaşıldığı bir hastane veri tabanında, 1000 hasta kaydında 990 pozitif 10 negatif sonucun kaydı istatistiki olarak paylaşılsa bile saldırgan kişi 999 kişinin arka plan bilgisini elde ettiğinde 1000. kişinin sonucunu ortaya çıkarılabilmektedir [127]. Yine hasta kayıtlarında olduğu gibi konum tabanlı servis kullanıcılarına ait bilgiler arka plan bilgisi ve farklı veri tabanı bağlantılarıyla kullanıcı kimlikleri de tespit edilebilmektedir [128].

Ayrıca konum bilgisi için bağlam saldırıları gerçekleşmekte, bağlam saldırıları için;

- Belirli bir zaman diliminde bir bölgede bulunan kullanıcı sayısının,
- Farklı kullanıcılar arası etkileşimlerin,
- Kullanıcının kimlik bilgisi ve konum bilgisinin ilişkisinin,
- Konumda bulunma kısıtlarının (Örneğin, POI noktası),
- Kullanıcıların bir bölgede bulunma olasılıklarının (Akşamları genelde evde bulunma gibi),

gibi bilgilerinin bilinmesi yeterli olacaktır [88].

5.2. Veri İhlal ve Saldırılarından Korunma Yöntemleri

Bir önceki bölümde örnekleri verilen mahremiyet ihlallerine yönelik birçok araştırma yapılmış, mahremiyeti koruma adına çeşitli yöntemler geliştirilmiştir. Bu yöntemlerden yaygın olarak uygulanan ve birçok çalışmada kullanılan yöntemler şu şekildedir;

- Sweeney [129] tarafından önerilen k-anonimlik yöntemi, genelleştirme ve baskılama teknikleri kullanarak eşlenik sınıflar oluşturmakta, böylelikle eşlenik sınıflar içinde bulunan bir kayıt k-1 tane veriden ayırt edilememektedir.
- Fakat eşlenik sınıfların homojen dağılması sonucu ve arka plan bilgileriyle tanımlayıcı bilgilerinin açığa çıkmasından dolayı Machanavajjhala vd. [130] tarafından önerilen l-çeşitlilik yöntemi her eşlenik sınıfta l-çeşitlilikte veri olmasını garanti etmektedir.
- Ninghu vd. [131] l-çeşitliliği geliştirerek veri dağılımını dikkate alan ve verinin ayrıntı düzeyini azaltan t-yakınlık yöntemini önermiş, önerilen modelde veri boyutu ve çeşitliliğinin artmasıyla tanımlayıcı bilginin açığa çıkma olasılığını arttığı da vurgulanmıştır [75].
- Nergiz vd. [132] k-anonimlik yönteminin yetersiz kaldığı durumları vurgulayarak bu durumlarda etkin olarak kullanılabilecek δ -Mevcudiyet yöntemini önermiştir. Yöntem hem veri kümesindeki bir kaydın varlığını hem genelleştirilen hem de örnek arka plan bilgisi veri tabanında bulunma olasılığını delta minimum ve maksimum aralığına çekmektedir.
- Dwork [105] iki veri tabanı arasında eşleşebilecek veri kaydının olasılığını azaltmak amacıyla çeşitli mekanizmaları kullanarak veriye gürültü ekleme prensibine dayanan diferansiyel mahremiyet yöntemini geliştirmiştir. Arka plan bilgisi, bağlantı eşleme ve olasılık bilgileri kullanarak yapılan saldırılara karşı oldukça güçlü olan yöntem saldırganların her türlü saldırısını minimize etmek amacıyla sıklıkla kullanılmaktadır.

Yukarıda bahsedilen yöntemler konum bilgilerinin mahremiyeti için anonimleştirme tekniği olarak kullanılırken, farklı karıştırma ve şaşırtma teknikleri de uygulanmaktadır. Bu uygulamalara kimlik bilgilerinin birbiriyle veya anonim bir kimlikle değiştirildiği karışık bölge [133], sahte konum [89] ve konum şaşırtma [134] çalışmaları örnek olarak verilebilir. Ayrıca konum bilgisinin coğrafi olarak ayırt edilememesini sağlama için Andres vd. [110], mekân bilgisinin yanı sıra zaman özelliklerini dikkate alan konum mahremiyeti çalışmasında Xiao vd. [135] diferansiyel mahremiyet yöntemini de kullanmıştır.

Önerilen mahremiyet korumalı tahmin modeli için 3. Bölümde detayları verilen ve Şekil 3.4'te görülen derin öğrenmede kullanılan diferansiyel mahremiyet yöntemlerinden, sinir ağı modeline “giriş aşamasında” sentetik veri yöntemi kullanılmıştır. 4. Bölümde önerilen trafik tahmin modelinin küme öznitelik oluşum aşamasında küme özniteliklerine diferansiyel mahremiyet uygulanmış, mahremiyet uygulanan öznitelik değerleri derin öğrenme

yönteminin girdi değerlerini oluşturmuştur. Önerilen mahremiyet korumalı modelin aşamaları ve detayları, test ve değerlendirme sonuçları bir alt başlıkta açıklanmıştır.

5.3. Model Aşamaları ve Değerlendirme

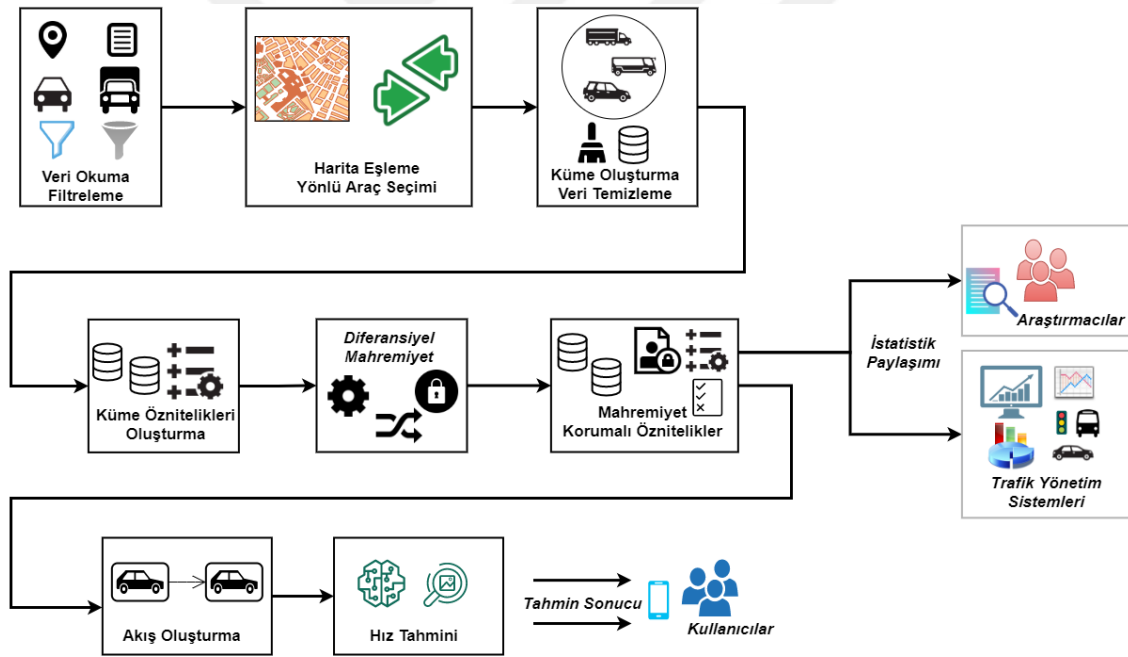
Mahremiyet korumalı hız tahmin modeli için Bölüm 4’de verilen ve Şekil 4.1’de görülen hız tahmin modeli referans alınmış olup bütün aşamalar aynı şekilde uygulanmıştır. Sadece mahremiyet korumalı tahmin modelinde tahmin modeline ek olarak küme öznitelik üretim aşamasında diferansiyel mahremiyet yöntemi uygulanmıştır. Diğer tüm aşamalarda 4. Bölümde açıklanan yöntemler ve parametreler uygulanmış, tekrar olmaması adına bu bölümde açıklanmamıştır. Mahremiyet korumalı modelin testlerinde yine aynı şekilde 4. Bölümde kullanılan yoğun ve orta yoğun trafik ham veri kümeleri olan HV1 ve HV2 kullanılmıştır. HV1 ve HV2 veri kümeleri bir saatlik araç verileri olup trafik tahmin modelinde uygulanan veri filtreleme, harita eşleme ve yönlü veri seçimi sonrasında aynı şekilde VK1 ve VK2 veri kümeleri elde edilmiştir. VK1 ve VK2 için yoğunluk tabanlı kümeler oluşturulup veri temizleme işlemi sonrasında küme hız değerleri ve arka araç koordinat bilgilerine küme öznitelik üretimi esnasında diferansiyel mahremiyet yöntemine tabi tutulmuştur. Diferansiyel mahremiyet yönteminde Laplace mekanizması [136] kullanılmıştır. Eş. 5.1’de görüldüğü üzere, bu mekanizmada rastgele bir K fonksiyonu D_1 ve D_2 veri kümelerinde yer alan bir elemanın farklılık gösterdiği olasılıkta ε -diferansiyel mahremiyet sağlamaktadır. Bu olasılıkta bir f sorgusu $f: D \rightarrow R^d$ olmak üzere tanımlanmış, bu sorgunun hassasiyeti Eş. 5.2 ile ifade edilmiştir.

$$Pr[K(D_1) \in S] \leq \exp(\varepsilon) \times Pr[K(D_2) \in S] \quad (5.1)$$

$$\Delta f = \max \|f(D_1) - f(D_2)\| \quad (5.2)$$

Model deneylerinde konum noktalarını yol üzerinden çıkarmayacak, küme hız değeri en alt ve üst sınırını aşmayacak şekilde Eş 5.1’de ifade edilen epsilon değeri 50 olarak alınmış, Eş 5.2’de tanımlanan hassasiyet değerleri hız özniteliği için küme içi minimum ve maksimum hız değerleri aralığında, arka araç koordinat bilgileri için -0,005 ve +0,005 aralığında gürültülü değerler elde edilmiştir. Elde edilen gürültülü veriler RNN varyantı LSTM ve GRU modelleri için giriş değerlerini oluşturmuş, trafik hız tahmin modelinde olduğu gibi Eskişehir ve İstanbul yollarında test edilerek değerlendirilmiştir.

Mahremiyet korumalı tahmin modelinin değerlendirilmesinde detayları yukarıda verilen araç veri kümeleri kullanılmış, veri kümeleri trafik tahmin modelinde olduğu gibi veri filtreleme, harita eşleme, küme oluşturma, temizleme işlemlerinden geçirilmiş, sonrasında küme öznitelikleri diferansiyel mahremiyet uygulanarak üretilmiştir. Mahremiyet korumalı öznitelikler üretildikten sonra literatürde kullanılan RMSE yöntemi [137] ve Öklid mesafesiyle [110] veri faydası ile ölçülmüştür. Yine hız tahmin modelinde olduğu gibi diferansiyel mahremiyet uygulaması sonrasında akış oluşturulmuş ve hız tahmini gerçekleştirilmiştir. Son olarak, mahremiyeti korunmuş küme öznitelikleri ile gerçekleşen hız tahmin değerleri orijinal küme hız değerleriyle karşılaştırılmıştır. Deney ve karşılaştırma sonuçları bir alt bölümde verilmiş olup 4. Bölümde önerilen trafik hız tahmin modelinin de özetlendiği mahremiyet korumalı hız tahmin model aşamaları ve süreci Şekil 5.1’de görülmektedir.



Şekil 5.1. Mahremiyet korumalı trafik hız tahmin modeli

5.3.1. VK1 için elde edilen deneysel sonuçlar

VK1 veri kümesinden Eskişehir ve İstanbul Yolları için sırasıyla elde edilen 84 ve 118 kümelerde oluşan orijinal ve diferansiyel mahremiyet uygulanmış küme hız değerleri, arka araç X-Y koordinat bilgileri için örnek küme gösterimleri Çizelge 5.1 ve Çizelge 5.2’de verilmiştir. Her iki çizelgede verilen değerler küme numarası (KN), orijinal küme hızı (OKH), diferansiyel mahremiyet uygulanmış küme hızı (DMKH), orijinal arka araç X

koordinatı (OAAX), diferansiyel mahremiyet uygulanmış arka araç X koordinatı (DMAAX), orijinal arka araç Y koordinatı (OAAY), diferansiyel mahremiyet uygulanmış arka araç X koordinatı (DMAAY) değerleridir. Eskişehir ve İstanbul güzergâhları için elde edilen, bütün kümelerle ait orijinal ve mahremiyet korumalı hız değerlerinin RMSE sonuçları Çizelge 5.3 sunulmuştur.

Çizelge 5.1. VK1-Eskişehir yolu örnek kümelerle ait orijinal ve mahremiyet uygulanan öznitelikler

KN	OKH	DMKH	OAAX	DMAAX	OAAY	DMAAY
83	102,372	102,9991	32,609226	32,609182	39,855606	39,855106
76	81,4545	81,6206	32,640976	32,640936	39,869045	39,868545
18	77,6552	78,0774	32,715504	32,715552	39,905788	39,905330
25	66,3721	57,9912	32,738346	32,738345	39,906708	39,906208
45	42,4286	41,7013	32,768394	32,768392	39,907822	39,907322

Çizelge 5.2. VK1- İstanbul yolu örnek kümelerle ait orijinal ve mahremiyet uygulanan öznitelikler

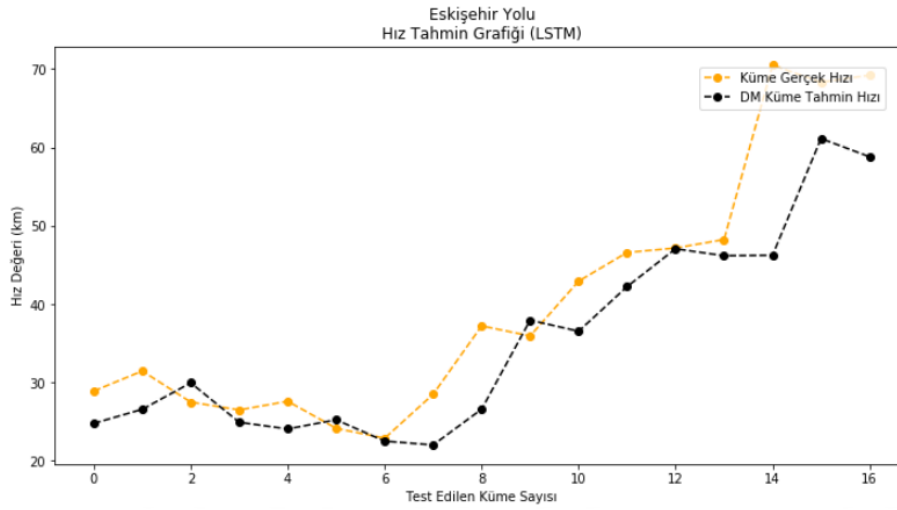
KN	OKH	DMKH	OAAX	DMAAX	OAAY	DMAAY
44	30,8537	30,8534	32,775609	32,775573	39,945847	39,945347
7	48,8649	50,1310	32,767239	32,767250	39,945118	39,944618
33	26,5862	26,5026	32,746067	32,746058	39,945187	39,944687
48	56,8148	56,1519	32,734081	32,734086	39,945251	39,944751
103	70,2857	71,5730	32,688736	32,688725	39,958664	39,958164

Çizelge 5.3. VK1 için oluşan bütün kümelerle ait hız ve arka araç X-Y koordinatı RMSE değerleri

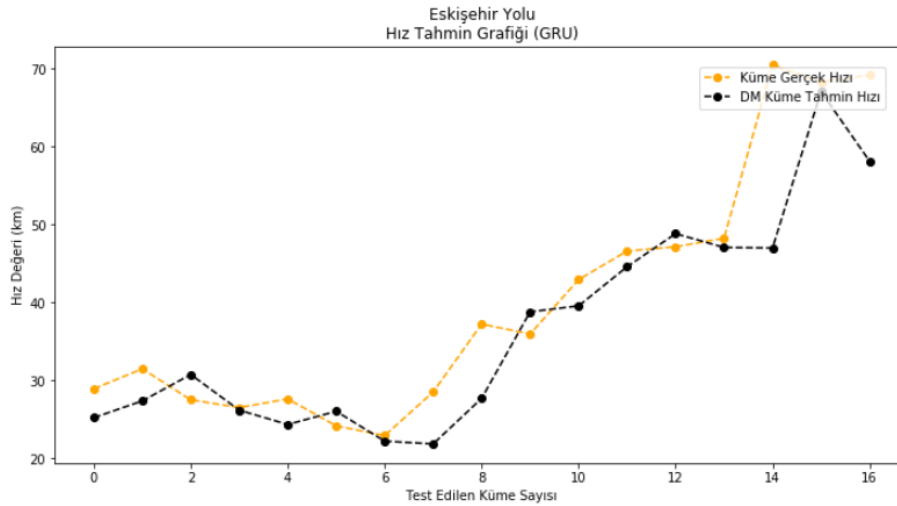
		RMSE	Öklid Mesafesi	
		Küme Hızı	Arka Araç X	Arka Araç Y
VK1	Eskişehir yolu	2,370	0,000259	0,004501
	İstanbul yolu	1,295	0,000295	0,005327

Çizelge 5.1 ve 5.2 incelendiğinde örnek küme özniteliklerinde değişimlerin az olduğu, orijinal ve mahremiyet uygulanmış küme değerleri birbirine yakın olduğu görülmektedir. Böylelikle veri kullanılabilirliği sağlanırken, istatistiki bilgi mahrem hale getirilmiştir. Genel hata oranlarını içeren Çizelge 5.3’de hem Eskişehir hem de İstanbul yolu hız RMSE hata değeri değişimin ortalama 1,5 km, konum bilgilerin değişiminin ölçüldüğü Öklid mesafe değişiminin binde 5 düzeyinde, hata oranlarının oldukça düşük ve veri faydasının yüksek olduğu görülmektedir.

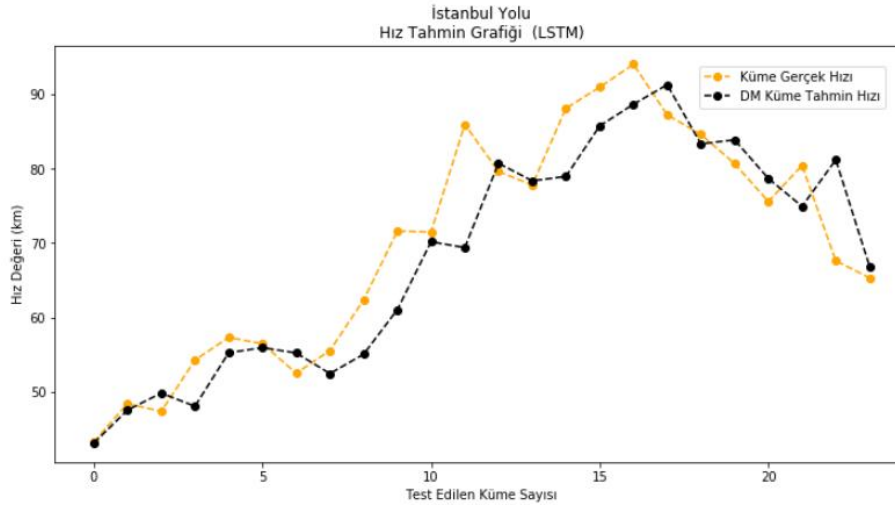
Küme özniteliklerine uygulanan mahremiyet işlemi sonrasında aynı trafik hız tahmin modelinde olduğu gibi öznitelik üretimi sonrasında başlangıç noktasına göre sıralan kümelerle bir akış oluşturulmuştur. Akış oluşturulan küme hızları tekrarlı sinir ağının iki varyantı LSTM ve GRU modellerine giriş değerlerini oluşturulmuş, mahremiyet korumalı hız tahminleri üretilmiş ve tahmin değerlerine ait grafikler Şekil 5.2–5.5’de verilmiştir. Üretilen mahremiyet korumalı tahmin değerlerinin orijinal hız değerleri dikkate alınarak hesaplanan RMSE değerleri Çizelge 5.4’de sunulmuştur.



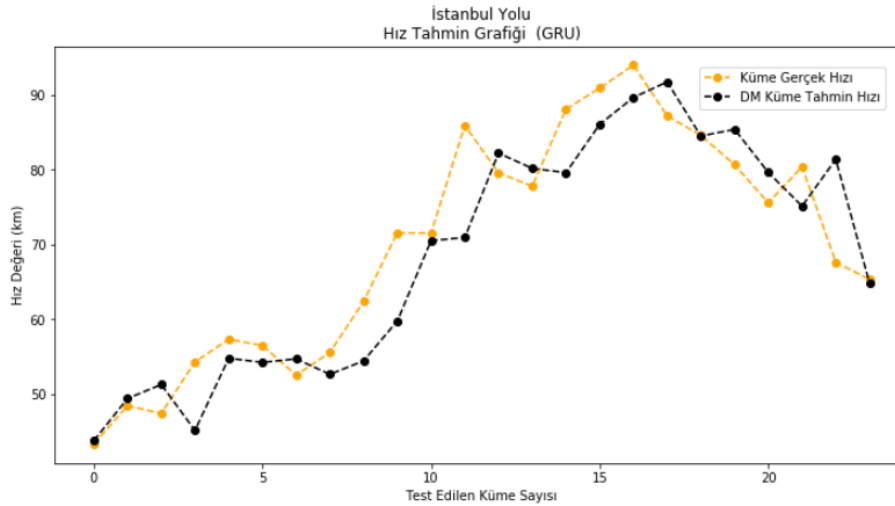
Şekil 5.2. VK1-Eskişehir yolu DM küme öznitelikleri ile LSTM model sonucu



Şekil 5.3. VK1-Eskişehir yolu DM küme öznitelikleri ile GRU model sonucu



Şekil 5.4. VK1-İstanbul yolu DM küme öznitelikleri ile LSTM model sonucu



Şekil 5.5. VK1-İstanbul yolu DM küme öznitelikleri ile GRU model sonucu

Çizelge 5.4. Mahremiyet korumalı tahmin modelinde VK1 için üretilen hız tahminleri hata değerleri

	Eskişehir Yolu Güzergâhı		İstanbul Yolu Güzergâhı	
	LSTM Modeli	GRU Modeli	LSTM Modeli	GRU Modeli
RMSE	7,793	7,250	6,127	6,318
MAPE (%)	12,296	11,047	6,204	6,845
MAE	5,402	4,723	4,472	4,824

Mahremiyet korumalı modelde üretilen tahmin sonuçlarının verildiği Çizelge 5.4 incelendiğinde Eskişehir yolu ve İstanbul yolu RMSE değerleri birbirine yakinken yüzdesel oran arasında ciddi fark oluşmuştur. Bu fark, Şekil 5.3 ve 5.4'te görüleceği üzere Eskişehir yolunda test edilen 7. Kümeden itibaren küme hızlarının artmaya başlaması, özellikle

örneğin 13. küme sonrasında çok hızlı bir artışın gerçekleşmesi ve buna bağlı hata oranının artmasından kaynaklanmaktadır. Şekil 5.5 ve 5.6’da incelendiğinde İstanbul yolundaki test küme hızlarında ani değişimler görülmemektedir. Sadece 6. Küme sonrası başlayan artış gözlemlenmiş, fakat sonraki küme hızları kademeli olarak arttığından ve hız değerlerinin artış veya azalışında yavaş değişimlerden dolayı daha düşük yüzdesel hata değerleri elde edilmiştir. Her iki yola ait tahmin sonuçlarının genel hata oranları dikkate alındığında tahmin sonuçları orijinal küme değerleri ile yapılan tahminlere yakındır ve tahminin yanı sıra mahremiyet de korunmuştur.

5.3.2. VK2 için elde edilen deneysel sonuçlar

VK2 veri kümesi kullanarak elde edilen yoğunluk tabanlı kümeler sırasıyla Eskişehir ve İstanbul yolları için sırasıyla 77 ve 50 küme elde edilmiştir. Kümeler için oluşan orijinal ve diferansiyel mahremiyet uygulanmış küme hız değerleri, arka araç X-Y koordinat bilgileri için örnek küme gösterimleri Çizelge 5.5 ve Çizelge 5.6’da gösterilmiştir.

Çizelge 5.5. VK2-Eskişehir yolu örnek kümelere ait orijinal ve mahremiyet uygulanan öznitelikler

KN	OKH	DMKH	OAAX	DMAAX	OAAAY	DMAAY
37	56,6719	57,6192	32,748699	32,748677	39,907269	39,906769
47	44,8077	48,1106	32,764591	32,764595	39,907974	39,907510
19	74,0909	72,1618	32,691216	32,691228	39,902348	39,901899
28	82,1026	79,5481	32,719368	32,719384	39,906021	39,905550
72	72,6087	71,9114	32,787228	32,787217	39,908783	39,908283

Çizelge 5.6. VK2-İstanbul yolu örnek kümelere ait orijinal ve mahremiyet uygulanan öznitelikler

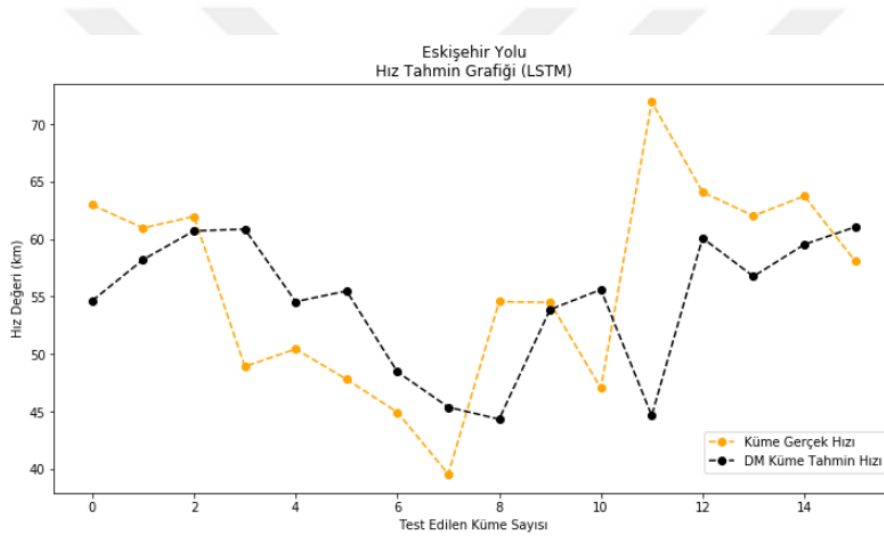
KN	OKH	DMKH	OAAX	DMAAX	OAAAY	DMAAY
27	73,6719	75,5232	32,753147	32,753142	39,945103	39,944603
40	49,4754	51,0468	32,773487	32,773478	39,945499	39,944999
15	85,4182	88,1557	32,725479	32,725474	39,945316	39,944816
20	50,3409	49,7746	32,739765	32,739782	39,945194	39,944702
47	64,6333	64,4797	32,658997	32,658972	39,979328	39,978830

VK2 için elde edilen küme orijinal hızları ve arka araç koordinatları DM değerler ile kıyaslandığında, Çizelge 5.7’de görülebileceği üzere oldukça düşüktür ve VK1 için elde edilen RMSE hata oranlarına yakın değerler elde edilmiştir. Bir önceki deneyde VK1 için

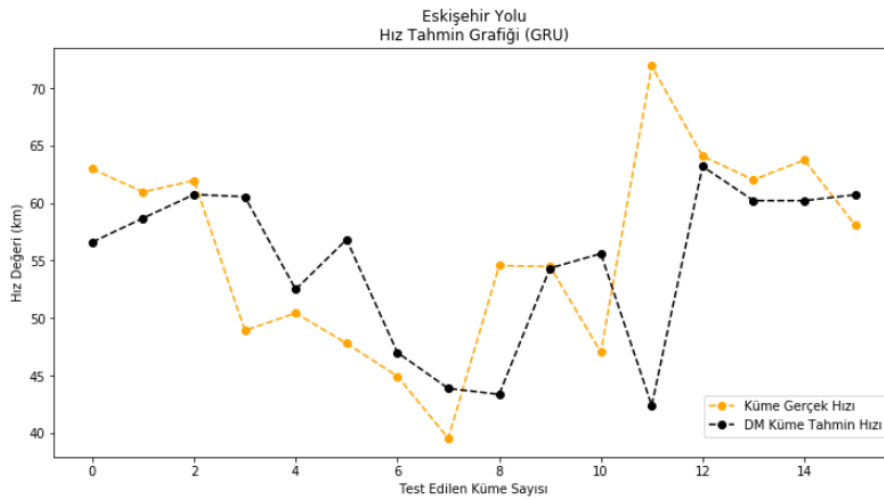
uygulanmış adımlar VK2 için de uygulanmış, DM uygulanmış küme hız değerleri yine tekrarlı sinir ağı modelleri olan LSTM ve GRU modellerine giriş olarak verilmiş, elde edilen sonuç grafikleri Şekil 5.6-5.9’da, RMSE değerleri Çizelge 5.8’de sunulmuştur.

Çizelge 5.7. VK2 için oluşan bütün kümeler için hız ve arka araç X-Y koordinatı RMSE değerleri

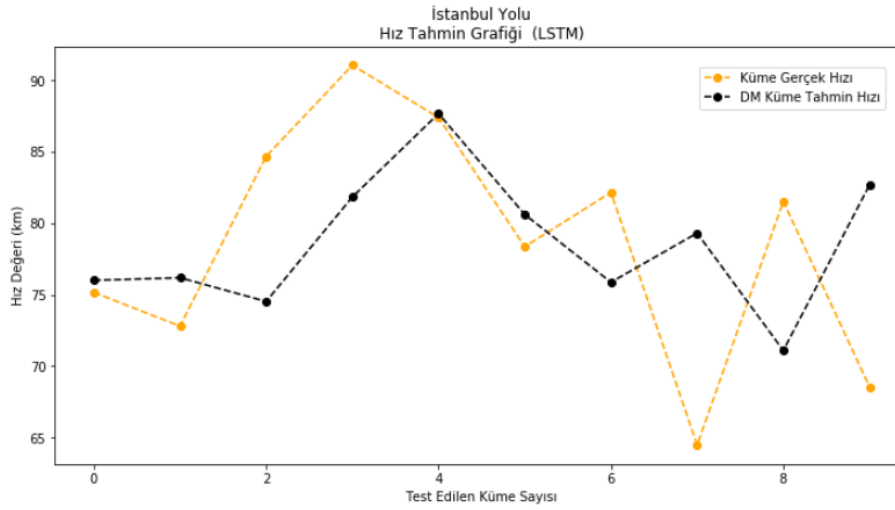
		RMSE	Öklid Mesafesi	
		Küme Hızı	Arka Araç X	Arka Araç Y
VK2	Eskişehir Yolu	2,397	0,000241	0,004288
	İstanbul Yolu	1,871	0,000164	0,003443



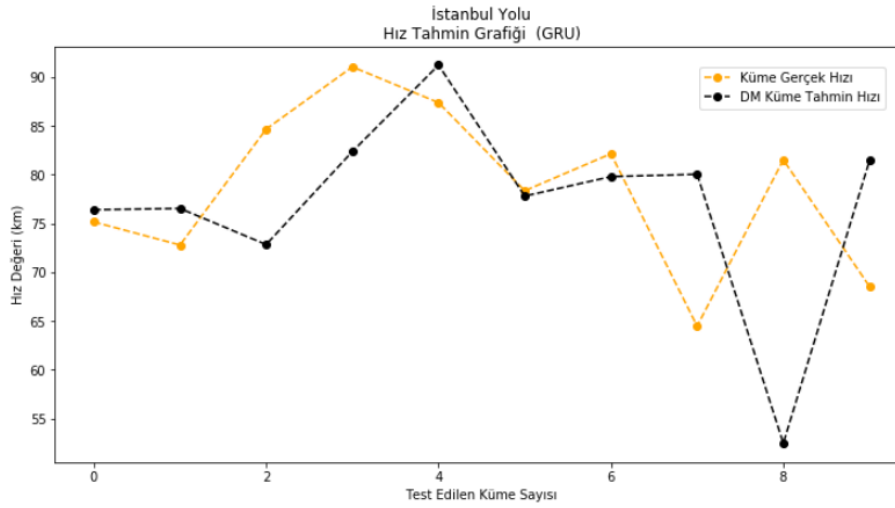
Şekil 5.6. VK2-Eskişehir yolu DM küme öznitelikleri ile LSTM model sonucu



Şekil 5.7. VK2-Eskişehir yolu DM küme öznitelikleri ile GRU model sonucu



Şekil 5.8. VK2-İstanbul yolu DM küme öznitelikleri ile LSTM model sonucu



Şekil 5.9. VK2-İstanbul yolu DM küme öznitelikleri ile GRU model sonucu

Çizelge 5.8. Mahremiyet korumalı tahmin modelinde VK2 için üretilen hız tahminleri hata değerleri

	Eskişehir Yolu Güzergâhı		İstanbul Yolu Güzergâhı	
	LSTM Modeli	GRU Modeli	LSTM Modeli	GRU Modeli
RMSE	9,135	9,306	8,789	12,256
MAPE (%)	12,090	10,786	9,509	11,680
MAE	6,783	6,080	7,176	8,971

VK2 için yapılan hız tahminleri için elde edilen RMSE değerlerinin verildiği Çizelge 5.8 incelendiğinde aynı VK1 deneyinde olduğu gibi başarılı sonuçlar almıştır. Her iki deney sonucu bizlere, DM uygulanan küme özniteliklerindeki değişim oranları tatmin edici

düzeyde olduğunu ve bir taraftan veri kullanılabilirliği diğer taraftan mahremiyet koruması da sağlandığını göstermiştir.

Bu bölümde 4. Bölümde önerilen trafik hız tahmin modeli referans alınarak mahremiyet korumalı trafik hız tahmin modeli önerilmiştir. Akıllı ulaşım sistemleri için kullanılan trafik verisinin yayınlanması yerine kurumlar ve araştırmacılar için akış, yoğunluk ve hız gibi temel parametrelerin istatistiki bilgisinin verilmesi adına önerilen model için trafik hız tahmin modelinde yol karakteristiğinin olduğu küme öznitelik üretimi aşamasında diferansiyel mahremiyet yöntemi uygulanmıştır. Mahrem hale gelen yol karakteristik bilgileri ile akış oluşturulan kümelere ait hız bilgileri tekrarlı sinir ağ yöntemiyle mahremiyet korumalı tahminler de üretilmiş, Ankara iline ait iki ana güzergâhta test edilmiş, başarılı sonuçlar alınmıştır.



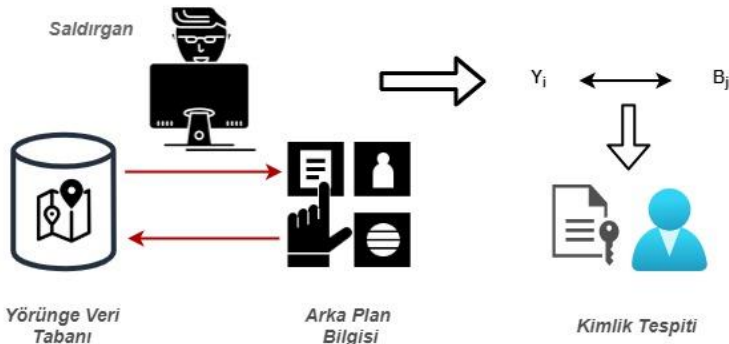
6. VERİ YAYINLAMA İÇİN ÖNERİLEN MAHREMİYET KORUMALI YÖRÜNGE MODELİ

Farklı kurum ve kuruluşlar tarafından veri paylaşımı ve yayınlama, araştırmacılar çalışmalarında kullanması açısından önemli olup aynı zamanda devlet kurumlarının şeffaflığı için bir gereklilik haline gelmektedir. Avustralya hükümeti özel hazırladığı bir internet sitesi aracılığıyla 2013 yılından beri 7000'den fazla veri kümesini kullanıcılarına açmıştır [138]. Ülkemizde İstanbul Büyükşehir Belediyesi kendi teknolojik araçlarıyla elde ettiği ekonomi, afet yönetimi, enerji, ulaşım ve trafik gibi çeşitli konularını içeren veri kümelerini kendi internet sitesinde yayınlamaktadır [139]. Fakat verinin paylaşılması ve yayınlanması hassas bilgilerin açığa çıkması gibi bazı riskleri de beraberinde getirmektedir. Bunun için verinin yayınlamadan önce tanımlayıcı bilginin çıkarılmasıyla mahremiyet sağlanmaya çalışılsa da saldırgan kişiler yarı tanımlayıcı veya arka plan bilgisi kullanarak kimlik tanımlı yapılabilmektedir. Kimlik tanımlı ve arka plan bilgisi hassas bilgileri ortaya çıkarmakta, kişilerin itibarını zedeleyerek finansal, fiziksel vb. konularda zararlar verebilmektedir.

Son yıllarda konum tabanlı servisleri kullanan navigasyon sistemleri ve akıllı telefonlar gibi mobil cihazların yaygınlaşmasıyla, servis sağlayıcıların konum bilgilerini toplaması ve yörünge verileri oluşturması daha kolay hale gelmiştir [87]. Kişiye ait hassas veri olan konum verileri ile birlikte zaman bilgileri de dijital uygulamalarda saklanmakta, mekân-zamansal bilgi olarak ifade edilen konum-zaman verileri zamana göre sıralandığında kişiye ait yörünge bilgileri elde edilebilmektedir. Yörünge verilerinden elde edilen davranış örüntülerinin analizi ve madenciliği, trafik tıkanıklık yönetimi, şehir sakinlerinin günlük yolculuklarını planlama gibi birçok konuda şehir planlama ve akıllı ulaşım sistemleri için karar mekanizmalarına destek olmaktadır [140]. Yörünge verilerinin kullanımının sağladığı ve getirdiği avantajlara rağmen kamuya ya da araştırmacılarla paylaşılması bazı riskler taşımakta, veriler kimliksizleştirilse bile farklı veri kaynaklarından elde edilen bilgiler ile eşleştirilmesi sonrası kimlik belirleme işlemi yapılabilmekte ve böylelikle kimlik ifşası gerçekleşebilmektedir. Örneğin, bir filo şirketine ait kimliksiz olarak yayınlanan yörünge veri kümesinden, kurban hakkında önceden edinilmiş bilgilerle eşleştirilmesiyle kurban hakkında sağlık, dini görüş, siyasi görüş, ev ve iş yeri bilgisi, araç kullanma hataları, sinirlilik durumu, sağlık durumu gibi pek çok hassas bilgiler elde edilebilmektedir [141]. Yörünge

verisi mahremiyeti, bu verilerden veri sahibine ulaşmayı mümkün olduğu kadar engelleyen kurallar bütünü olarak tanımlanmaktadır [142].

Bu bölümde, konum ve buna bağlı olarak elde edilebilen yörünge mahremiyeti için yeni bir mahremiyet modeli önerilmiştir. Mahremiyet modeli için ele alınan problem Şekil 6.1’de gösterilmektedir. Yörünge veri tabanı Y , kimlikli arka plan bilgi kümesi B , $Y=\{Y_1, \dots, Y_n\}$ ve $B=\{B_1, \dots, B_m\}$ olmak üzere, eğer bazı özellikler kullanılarak Y_i ile B_j eşleştirilirse bu durumda Y veri tabanındaki kurbanın kimliğini açığa çıkaran ifşa saldırısı gerçekleşmektedir.



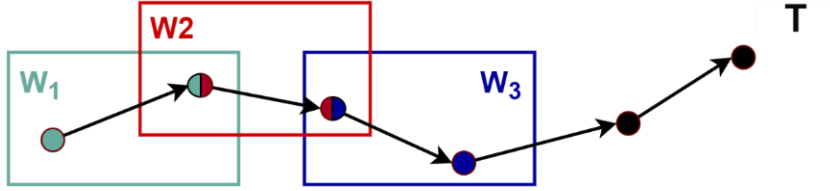
Şekil 6.1. Yörünge verilerinde ifşa saldırısı

Önerilen modelde Bölüm 3’de açıklanan diferansiyel mahremiyet yöntemi uygulanmış, veri faydası RMSE yöntemi ile ölçülmüştür. Veri kümesinde yer alan belirli bir zaman aralığındaki araç yörüngeleri ayrı ayrı ele alınmış, yörünge verisini oluşturan her bir konum verisine diferansiyel mahremiyet uygulanarak anonimleştirme yapılmıştır. Önerilen model, yörünge verilerinin pencere tabanlı gruplandırılması, grupların ortalamalarının alınması ve ortalamalara Laplace gürültüsü eklenmesi olmak üzere üç aşamadan oluşmakta olup bu aşamalar aşağıda açıklanmıştır.

6.1. Pencere Tabanlı Gruplama

Önerilen model yörünge verilerini pencere tabanlı olarak ele almaktadır. Pencere kavramı yörünge verilerinin belirli alanlarda gruplanmasıdır. Pencere Boyunun (PB) bir parametre olarak sisteme verilirken, ilgili pencere belirlenen bu boya göre her adımda bir sağa kaydırılarak yeni gruplar oluşturulur. Şekil 6.2’de örnek bir yörünge verisinde uygulanan pencere yaklaşımı gösterilmektedir. Verilen şekilde, W_i pencere yapısını, T ise ilgili yörünge

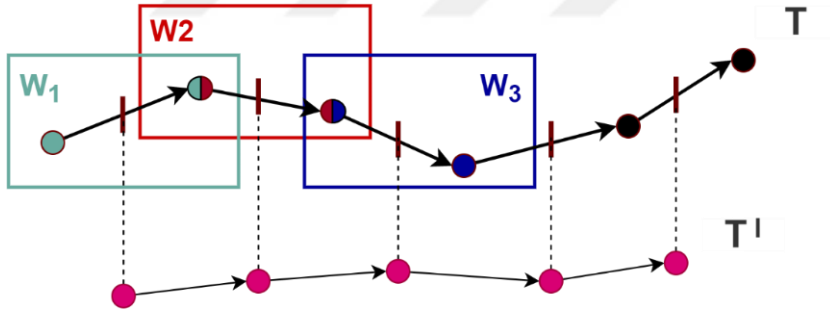
verisini temsil etmektedir. Bu grupta işleminin ardından, W_1 , W_2 ve W_3 pencereleri ve bu pencerelere düşen veri grupları oluşmaktadır.



Şekil 6.2. Yörünge verisine pencere yapısının uygulanması

6.2. Grupların Ortalamasını Alma

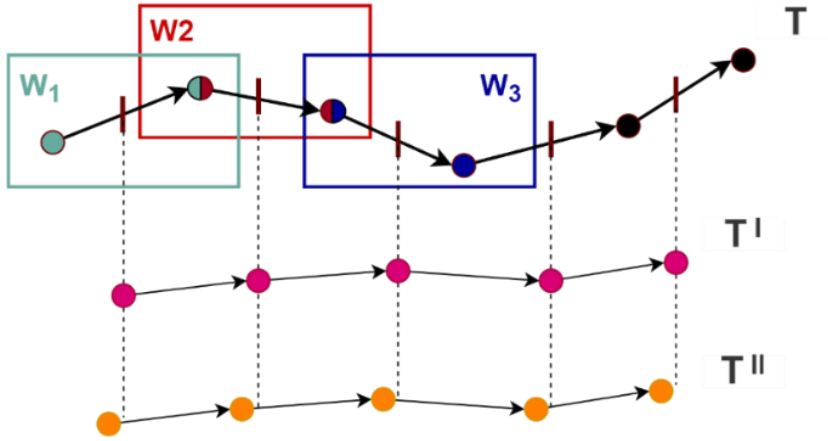
Pencere tabanlı grupta ile alt gruplara ayrılan veri kümesine diferansiyel mahremiyet yaklaşımını uygulayabilmek amacıyla, ilgili pencerelere düşen verilerin ortalamasının alınarak yeni bir yörünge verisinin (T') oluşturulduğu aşamadır. Şekil 6.2’de sunulan örnekteki pencereler içerisindeki verilerin ortalamaları alınarak yeni T' elde edilmiş ve bu Şekil 6.3’de gösterilmiştir.



Şekil 6.3. Pencere içindeki konum noktalarının ortalamasıyla T' yörünge verisi oluşturma

6.3. Ortalamalara Gürültü Ekleme

Bir önceki aşamada elde edilen T' bu aşamada Laplace gürültüsü eklenerek anonimleştirilmiş yörünge verisi (T'') oluşturulmakta ve bu süreç Şekil 6.4’te gösterilmektedir.



Şekil 6.4. T' yörünge verisine gürültü ekleyerek T'' anonim yörünge verisini oluşturma

Yörünge verileri genel olarak içerisinde zaman damgası ve konum bilgisi (koordinat) barındıran veriler olarak temsil edilir. x enlem, y boylam, Eş 6.1 ile görülen l yörünge bilgisi ve t ise zaman olmak üzere, T_i yörünge verisi Eş 6.2'deki gibi modellenebilir.

$$l_i = \{x_i, y_i\} \quad (6.1)$$

$$T_i = \{(t_1, l_1), (t_2, l_2), \dots, (t_n, l_n)\} \quad (6.2)$$

Yukarıda sunulan veri modelinde tek bir yörünge verisi modellenmiştir. Genel olarak veri kümeleri çok sayıda yörünge verisi içerebileceğinden dolayı böylesi bir T yörünge veri kümesi Eş 6.3'deki gibi modellenebilir.

$$T = \{T_1, T_2, \dots, T_m\} \quad (6.3)$$

Önerilen modelin sözde kodu Algoritma 6.1'de sunulmuştur. İlgili algoritmada ilk olarak, algoritmaya girdi olarak verilen yörünge veri kümesi pencere tabanlı olarak gruplanmakta, ikinci olarak, yörünge tabanlı elde edilen grupların her birinin kendi içerisinde ortalaması alınarak yeni yörünge verileri oluşturulmaktadır. Son olarak ise yeni yörünge verilerine gürültü eklenerek anonim yörünge verileri elde edilmektedir.

Algoritma 6.1. Önerilen anonimleştirme model algoritması

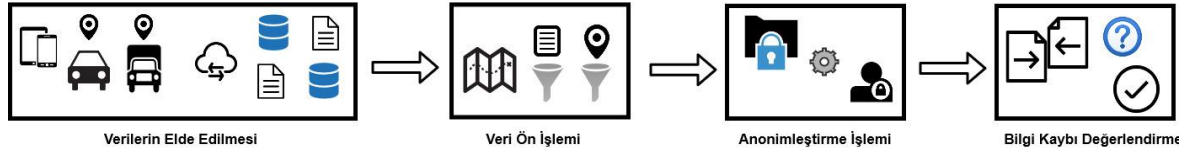
Girdi: veri kümesi $T = \{T_1, T_2, \dots, T_m\}$, pencere boyu s

Çıktı: anonim veri kümesi $T'' = \{T_1'', T_2'', \dots, T_m''\}$;

1	Her bir $T_i \in T$ için yinele:
2	Pencere tabanlı grupta:
3	$gruplar = \{g_1, g_2, \dots, g_m\} = w(T_i, s)$, $g_i = (t_x, l_x), (t_y, l_y), \dots, (t_s, l_s)$
4	Ortalama hesapla:
5	Her bir $g_i \in gruplar$ için yinele:
6	Ortalama hesapla; $g_j^{ortalama} = ortalama_hesapla(g_j)$
7	Yeni gruplar kümesine ekle; $gruplar_j^{ortalama} = g_j^{ortalama}$
8	Dönüştürülen yeni yörünge verisini oluştur; $T_i' = gruplar^{ortalama}$
9	Gürültü ekle:
10	Her bir $g_i \in gruplar$ için yinele:
11	Gürültü ekle; $g_j^{ortalama_laplace} = ortalama_laplace(g_j)$
12	Yeni gruplar kümesine ekle; $gruplar_j^{ortalama_laplace} = g_j^{ortalama_laplace}$
13	Dönüştürülen anonim yörünge verisini oluştur; $T_i'' = gruplar^{ortalama_laplace}$
14	Döndür: T'' anonim veri kümesi

Çalışma kapsamında yörünge verilerinin anonim hale getirilmesi için uygulanan genel iş akış süreci Şekil 6.5’de gösterilmektedir. Verilerin elde edilmesi ve veri ön işlemi 4. Bölümde açıklanan veri okuma, veri filtreleme, harita eşleme ve yön seçimi işlemlerini içermektedir. Verilerin elde edilme süreci olan ilk aşamada, araçlarda bulunan mobil cihaz verileri veri tabanına metin dosyası şeklinde kaydedilmektedir. Verilerin ön işlemi sürecinde araç verileri bir önceki bölümde belirtilen Eskişehir ve İstanbul yolu enlem-boylam noktaları arasında filtrelenmiş ve harita eşleme işlemi yapılmıştır. Harita eşleme işleminden sonra aynı yönde giden araçlar tespiti için yine bir önceki bölümde ayrıntıları verilen açısal yön değerleri dikkate alınmıştır. Veri filtreleme işleminde geniş zaman aralığında daha fazla araç yörünge verisi elde etmek için önceki bölümde uygulanan zaman dilimi arasına filtreleme işlemi yapılmamış, ham veri için geçerli olan bir saatlik zaman dilimi arasında araç verileri kullanılmıştır. Veri ön işlemi uygulanan araç verilerine üçüncü aşamada önerilen anonimleştirme işlemi uygulanmış ve son aşamada ise anonim verinin sunduğu veri faydası ölçülmüştür. Önerilen mahremiyet modeli için kullanılan verilerin detayları Çizelge 6.1’de

görülmektedir. Kullanılan veriler sadece modelin başarımını gösterim amaçlıdır. Model ülkemizde seçilen her bölge/yol üzerinde çalışabilecek şekilde geliştirilmiştir.



Şekil 6.5. Modelde uygulanan genel iş akış adımları

Çizelge 6.1. Model için kullanılan veri detayları

Zaman Aralığı	Bölge/Yol	Benzersiz Araç Sayısı	Toplam GPS Kayıt Sayısı	Veri Boyutu (MB)
27 Ocak 2020 14:59-15:59	Eskişehir Yolu	1379	26437	1,58
	İstanbul Yolu	1549	48443	2,95

Deneyisel çalışmalarda, farklı PB ve ε parametrelerine göre testler yapılmış ve sonuçlar RMSE metriği kullanılarak ölçülmüştür. PB değerleri sezgisel olarak seçilmiş ve sadece 3 farklı değere göre ölçümler yapılmıştır. Bu kapsamda, ilk deney için veri kümesinden en fazla GPS sinyali üreten araçlar seçilerek anonimleştirme işlemi uygulanmış ve ilgili test sonuçları alınarak grafiksel olarak gösterilmiştir. İkinci deneyde ise veri kümesindeki tüm araçlar için ilgili testler yapılmış ve elde edilen değerler sunulmuştur.

6.4. Örnek Araç için Test Sonuçları

Önerilen anonimleştirme modelinde kullanılan veri kümesi harita eşleme ve yön seçiminin ardından kimlik bilgisi ve zaman göre sıralanmıştır. Sıralamayla her aracın yörünge verisi oluşmuş, model değerlendirmesi için yörünge verisi oluşan araçlar içerisinde en fazla GPS sinyali üreten örnek bir araç her iki yolda ele alınmış, bu araçlarla ilgili testler yapılmış, yollara ait deney sonuçları aşağıda sunulmuştur.

6.4.1. İstanbul yolu test sonuçları

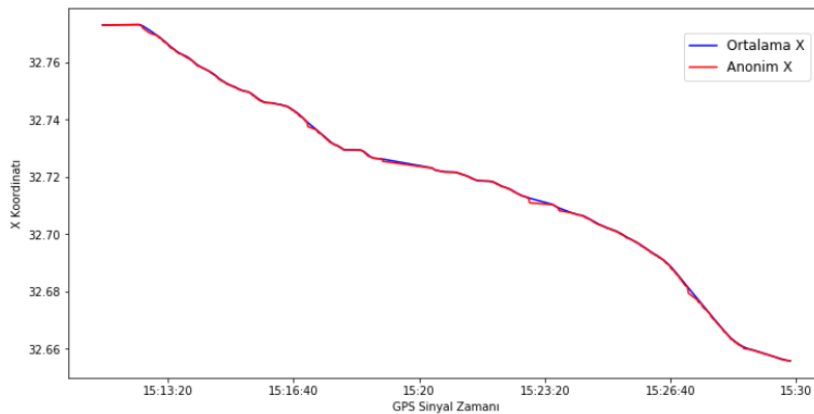
İstanbul yolunda seçilen örnek aracın yol boyunca ürettiği 402 GPS sinyali mevcuttur. Örnek araca ait sinyaller zamana göre sıralıdır ve sıralı sinyaller arası ortalama konum değeri elde edilmiş, ortalama değer sonrasında diferansiyel mahremiyet yöntemiyle yeni konum değerleri oluşturulmuştur. Örnek seçilen araca ait hata değerleri Çizelge 6.2’de verilmiştir.

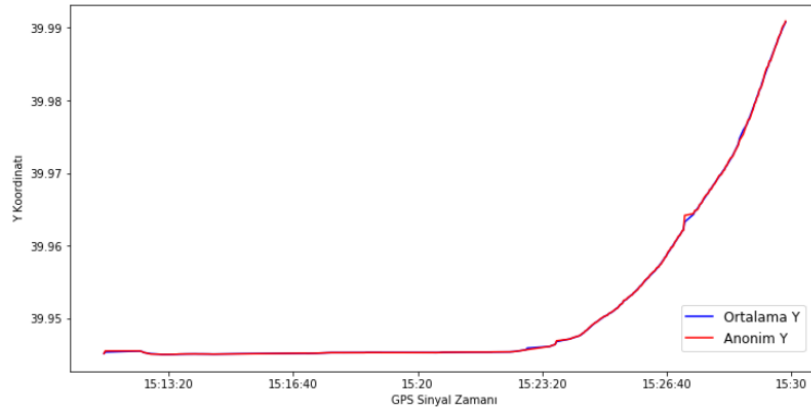
Çizelge 6.2. İstanbul yolu - Örnek araç için elde edilen sonuçlar

PB	ε	RMSE değeri	
		X	Y
2	0,5	0,00020804	0,00010104
	1,0	0,00017558	0,00009147
	1,5	0,00018551	0,00008538
	2,0	0,00011643	0,00008601
3	0,5	0,00038891	0,00019945
	1,0	0,00036539	0,00016429
	1,5	0,00028199	0,00015558
	2,0	0,00025050	0,00014493
4	0,5	0,00050431	0,00025283
	1,0	0,00044841	0,00025021
	1,5	0,00045279	0,00023383
	2,0	0,00044638	0,00022178

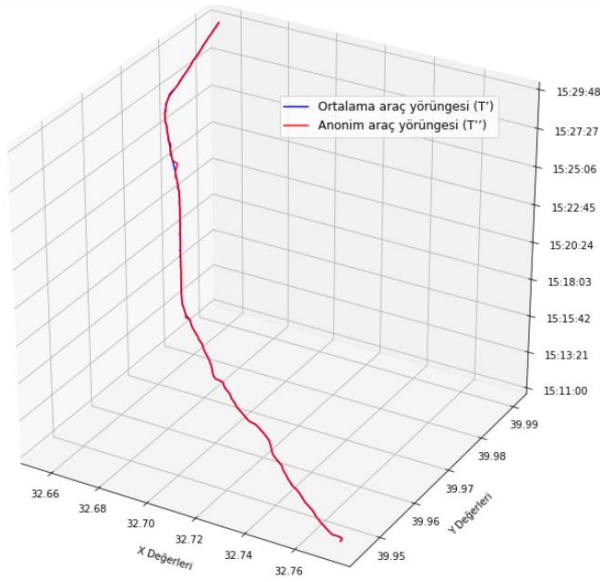
İstanbul yolu örnek aracın ortalama ile anonimleştirilmiş X ve Y değerlerinin grafikleri sırasıyla Şekil 6.6 ve Şekil 6.7’de çizdirilmiştir. Örnek araç İstanbul yolunda saat 15:11-15:30 arasında seyahat ederken, bu zaman diliminde anonimleştirilen konum verileri dikkate alındığında her iki grafikte de araç hareket grafiklerinin birbirine çok yakın olduğu görülmektedir. Şekil 6.8’de ise aynı örnek aracın X, Y ve GPS sinyal zamanı eksenlerinde hareketinin orijinali ile PB=2 ve $\varepsilon=2$ için anonimleştirilmiş yörüngesi gösterilmiştir.

Şekil 6.9’da örnek aracın orijinal yörünge verisi, Şekil 6.10’da ortalama alınmış yörünge verisi ve Şekil 6.11’de ise diferansiyel mahremiyet (DM) uygulanmış yörünge verisi harita üzerinde görselleştirilmiştir.

Şekil 6.6. İstanbul yolu - Örnek aracın ortalama ve anonimleştirilmiş X değerleri ($\varepsilon=2$)



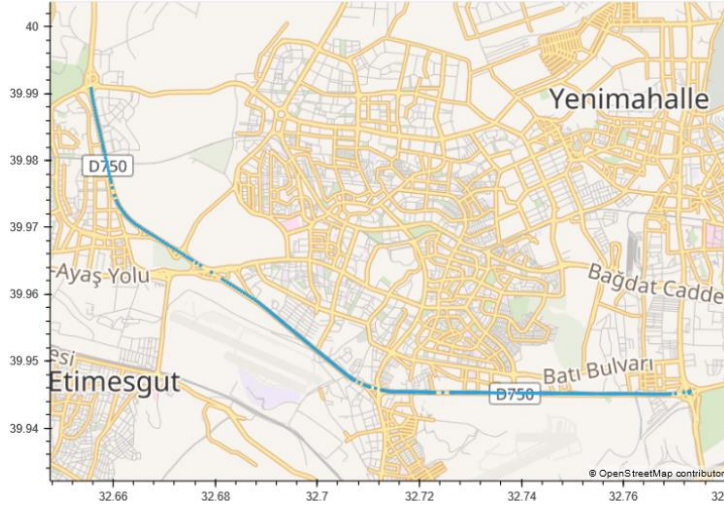
Şekil 6.7. İstanbul yolu - Örnek aracın ortalama ve anonimleştirilmiş Y değerleri ($\varepsilon=2$)



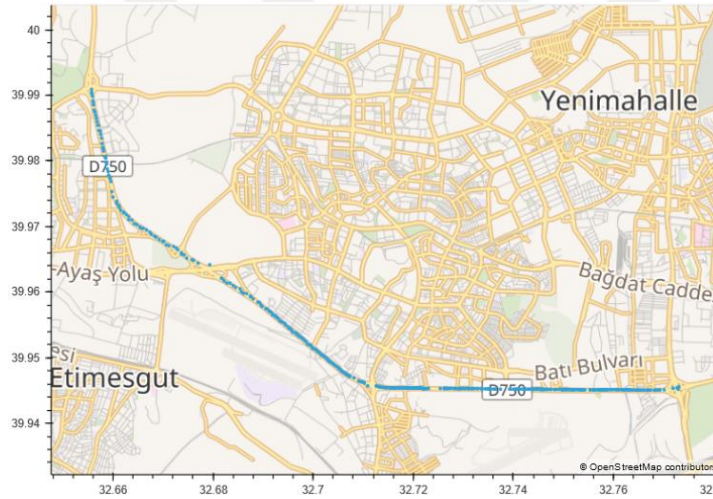
Şekil 6.8. İstanbul yolu - Örnek aracın X, Y ve zaman boyutunda GPS sinyal grafiği



Şekil 6.9. İstanbul yolu - Örnek aracın orijinal yörünge verisi ($T_{örnek}$)



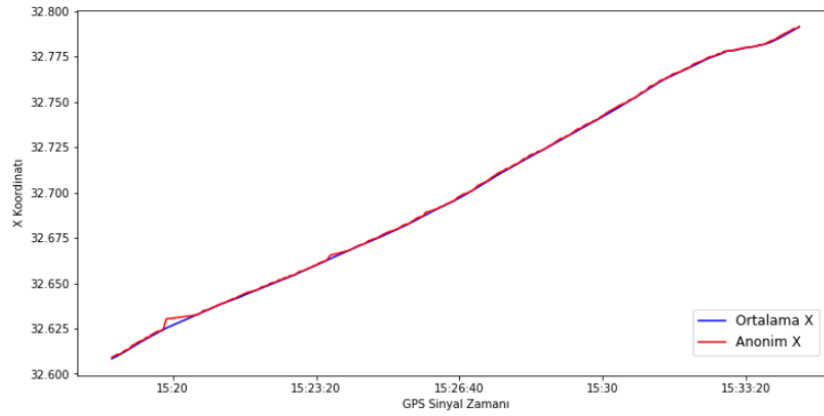
Şekil 6.10. İstanbul yolu - Örnek aracın ortalama alınmış yörünge verisi (T' örnek)



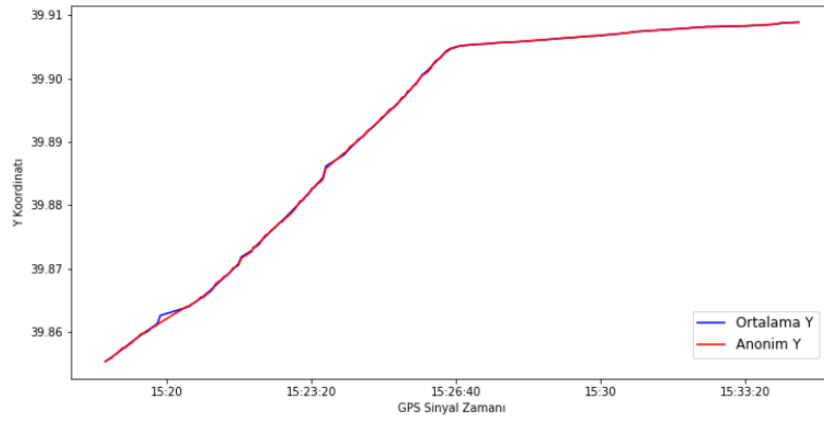
Şekil 6.11. İstanbul yolu örnek aracın DM uygulanmış yörünge verisi (T'' örnek)

6.4.2. Eskişehir yolu test sonuçları

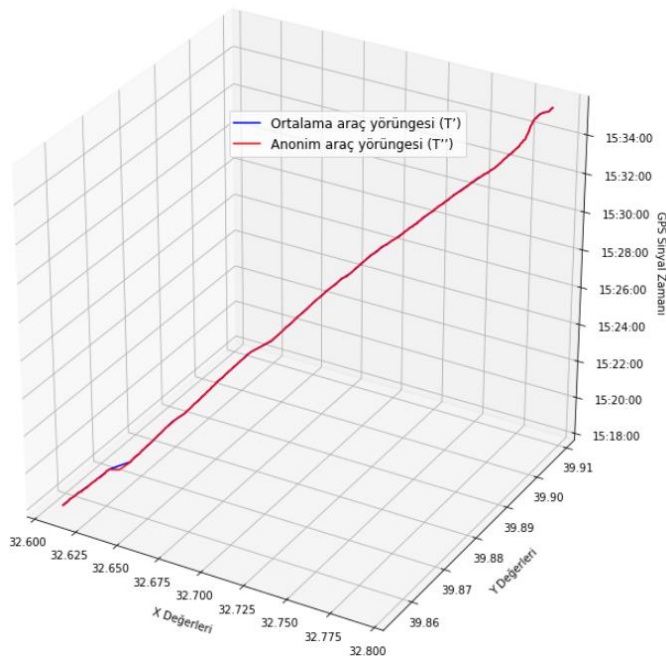
Eskişehir yolunda seçilen örnek aracın yol boyunca ürettiği 221 GPS sinyali mevcuttur. Üretilen sinyaller 15:18-15:34 arası olup, belirlenen güzergâhta gezinen araca ait sinyallere ortalama değerler ve diferansiyel mahremiyet yöntemi uygulandığında elde edilen hata değerleri Çizelge 6.3'de verilmiş, hareket eksenindeki X ve Y grafikleri Şekil 6.12 ve 6.13'de gösterilmiştir. Şekil 6.14'de ise aynı örnek aracın X, Y ve GPS sinyal zamanı eksenlerinde hareketinin orijinali ile $PB=2$ ve $\varepsilon=2$ için anonimleştirilmiş yörüngesi gösterilmiştir.



Şekil 6.12. Eskişehir yolu - Örnek aracın ortalama ve anonimleştirilmiş X değerleri ($\epsilon=2$)



Şekil 6.13. Eskişehir yolu - Örnek aracın ortalama ve anonimleştirilmiş Y değerleri ($\epsilon=2$)

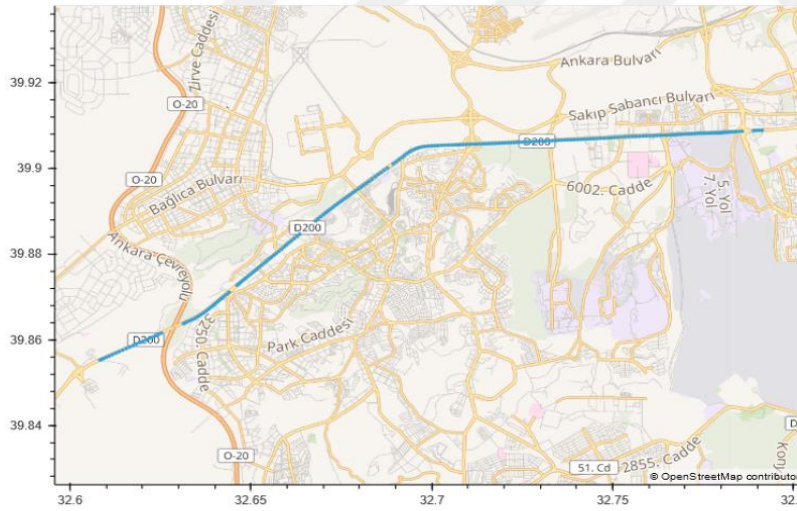


Şekil 6.14. Eskişehir yolu - Örnek aracın X, Y ve zaman boyutunda GPS sinyal grafiği

Çizelge 6.3. Eskişehir yolu - Örnek araç için elde edilen sonuçlar

PB	ε	RMSE değeri	
		X	Y
2	0,5	0,00045455	0,00019457
	1,0	0,00042366	0,00018799
	1,5	0,00037263	0,00017087
	2,0	0,00031973	0,00013009
3	0,5	0,00083346	0,00033209
	1,0	0,00083099	0,00032129
	1,5	0,00083819	0,00030099
	2,0	0,00063932	0,00026780
4	0,5	0,00122111	0,00049616
	1,0	0,00112780	0,00045153
	1,5	0,00102180	0,00041840
	2,0	0,00098913	0,00036975

Eskişehir yolu örnek aracın orijinal yörünge verisi Şekil 6.15’de, ortalama alınmış yörünge verisi Şekil 6.16’da ve diferansiyel mahremiyet uygulanmış yörünge verisi Şekil 6.17’de harita üzerinde görselleştirilmiştir.

Şekil 6.15. Eskişehir yolu - Örnek aracın orijinal yörünge verisi (*Törnek*)



Şekil 6.16. Eskişehir yolu - Örnek aracın ortalama alınmış yörünge verisi (T' örnek)



Şekil 6.17. Eskişehir yolu örnek aracın DM uygulanmış yörünge verisi (T'' örnek)

Her iki yol için seçilen örnek araçlara ait çizelgeler ve hata değerleri incelendiğinde, PB arttıkça X ve Y değerlerinde elde edilen RMSE değerlerinin de arttığı görülmüştür. Ayrıca her bir PB içerisinde ε değeri arttıkça RMSE değerinde genel olarak azalma meydana geldiği görülmektedir. Çizelge 6.2 ve 6.3’de sunulan sonuçlardan görülebileceği gibi, bu deneyde en az hata değeri PB=2 ve $\varepsilon=2$ iken elde edilmiş, bu parametre değerleriyle elde edilen örnek araçların ortalama değer ve diferansiyel mahremiyet sağlanmış yörünge verileri harita üzerinde Şekil 6.10-6.11 ve Şekil 6.16-6.17’de verildiği gibi görselleştirilmiştir. Örnek araçların orijinal yörünge verilerinin yer aldığı Şekil 6.9 ve Şekil 6.15 ile bu görseller kıyaslandığında modelin başarımı açıkça görülmektedir.

6.5. Veri Kümesindeki Tüm Araçlar için Elde Edilen Sonuçlar

Eskişehir ve İstanbul yollarına ait belirlenen zaman dilimi içinde oluşan tüm araçların yörüngelerinde uygulanan model sonuçları sırasıyla Çizelge 6.4 ve Çizelge 6.5’de sunulmuştur.

Çizelge 6.4. İstanbul yolu - Tüm araçlar için elde edilen ortalama RMSE değerleri

PB	ε	Ortalama RMSE değeri	
		X	Y
2	0,5	0,00112493	0,00049403
	1,0	0,00107773	0,00045832
	1,5	0,00100598	0,00043765
	2,0	0,00091260	0,00040587
3	0,5	0,00188254	0,00080501
	1,0	0,00168387	0,00075162
	1,5	0,00159350	0,00069840
	2,0	0,00147486	0,00063494
4	0,5	0,00252490	0,00110866
	1,0	0,00233460	0,00102218
	1,5	0,00212026	0,00093239
	2,0	0,00190452	0,00084913

Çizelge 6.5. Eskişehir yolu - Tüm araçlar için elde edilen ortalama RMSE değerleri

PB	ε	Ortalama RMSE değeri	
		X	Y
2	0,5	0,00149060	0,00059659
	1,0	0,00152422	0,00056827
	1,5	0,00139392	0,00054783
	2,0	0,00132915	0,00048220
3	0,5	0,00266316	0,00108066
	1,0	0,00248575	0,00099631
	1,5	0,00239207	0,00089190
	2,0	0,00198576	0,00086618
4	0,5	0,00361615	0,00146933
	1,0	0,00342388	0,00134813
	1,5	0,00314098	0,00124091
	2,0	0,00285329	0,00114256

Her iki çizelge incelendiğinde, PB arttıkça X ve Y değerlerinde elde edilen RMSE değerlerinin arttığı, her bir PB özelinde ε değeri arttıkça RMSE değerinde genel olarak azalma meydana geldiği görülmektedir. Elde edilen bu sonuçlara göre, bu deneyde de en az hata değerini PB=2 ve $\varepsilon=2$ iken elde edildiği gözlemlenmiştir. Hem örnek araç deneyinde hem de tüm araçlar deneyinde, hata değerlerin en küçük PB ve en büyük ε değerinde alındığı görülmektedir. PB=4 ve $\varepsilon=0,5$ değeri en büyük hata oranına sahip olsa da konum verileri kabul edilebilir ve kullanılabilir seviyededir.

Bu bölümde veri yayınlamada konum mahremiyeti ve konum mahremiyet ile birlikte yörünge mahremiyeti için yeni ve özgün bir model önerilmiş, Ankara iline ait olan iki ana güzergâh olan Eskişehir ve İstanbul yollarında deneyler gerçekleştirilmiştir. Her iki yolda yer alan tüm araçlara ait konum noktaları zamana göre sıralandığında yörünge verileri oluşmuştur. Oluşan yörünge verilerine uygulanan deney sonuçlarının hata değerleri ve veri kümesinden GPS sinyal sayısı en fazla olan birer araç seçilerek grafik ve harita görselleri verilmiştir. GPS sinyali en fazla olan araç İstanbul yolunda 402 sinyal üretmiş model uygulama sonrasında X-Y koordinatları RMSE değerleri sırasıyla 0,00011643 ve 0,00008601 olarak elde edilirken, 221 sinyal üreten Eskişehir yolundaki örnek aracın RMSE hata oranları 0,00031973 ve 0,00013009 olarak tespit edilmiştir. Bu hata oranları deneysel sonuçlardaki en düşük hata oranları olup her iki yol testinde de epsilon ve pencere boyutu “2” olduğunda elde edilmiştir. Farklı epsilon değerleri ve pencere boyutunda yapılan deneysel sonuçlara da bakıldığında veri yayınlamada yeni ve özgün olan bu modelin sonuç değerleri, başarımın ve veri faydasının yüksek olduğunu göstermiştir.

7. DEĞERLENDİRME VE SONUÇLAR

Trafik davranışlarını ve karakteristiğini anlamak ve bir trafik ağını tanımlamak için temel trafik parametreleri kullanılmaktadır. Bu parametreler akış, yoğunluk ve hız olarak ifade edilmekte, anlık ölçümler ve ileriye yönelik tahminler trafiği yöneten birim ve kurumların yönetim kabiliyetini arttırmakta, bireysel kullanıcıların günlük ve planlı işlerinde daha büyük kolaylıklar sağlamaktadır. Bu tez kapsamında trafik parametrelerinin tahmini ve mahremiyet odaklı çözümler için üç yeni model önerilmiş, elde edilen sonuçlar sunulmuş ve aşağıda başlıklarda katkılar özetlenmiştir.

- i. Yeni ve özgün bir trafik hız tahmin modeli geliştirilmiş, veri okuma, filtreleme, harita eşleme, küme oluşturma, veri temizleme, öznitelik çıkarımı, akış oluşturma, görselleştirme ve tahmin olmak üzere dokuz adet sıralı aşamadan oluşan tahmin modelinin testleri “gerçek araç” verileri kullanılarak Ankara iline ait iki ana yolda gerçekleştirilmiş ve deneysel sonuçlar sunulmuştur.
- Tahmin üzerine yapılan çalışmalarda kullanılan veri kaynakları incelendiğinde, tahmin yapılacak yollara yerleştirilen sabit veri kaynakları hem maliyet getirmekte hem de bu kaynaklarından elde edilen veriler ile sadece o yola ait tahminler üretilebilmektedir. Önerilen model, mobil cihazlardan elde edilen dağıtık GPS verisi kullanılarak istenilen yol için büyük bir maliyet getirmeden tahmin yapabilmektedir.
- Trafik hız tahmini, segment uzunluğuna bağımlı olmadan istenilen yol bölümünde, istenilen başlangıç-bitiş noktaları arasında yapılabilmektedir.
- Literatür çalışmalarına bakıldığında, aynı segment yola ait farklı günlerde geçmiş trafik verisini dikkate alınarak tahmin yapılmaktadır. Fakat trafik tahmini, hava durumu, olaylar, yol kazaları gibi birçok dış sorundan etkilenen oldukça zorlu süreçleri içerisinde barındırırken, trafiğin doğrusal olmayan yapısından dolayı tahmin yapmayı zorlaştırmaktadır. Geliştirilen modelle birlikte istenilen/belirlenen hedef yol üzerinde yoğunluk tabanlı kümeler oluşturularak yola ait karakteristikler oluşturulmuş ve trafiğin kendi değişken yapısında tahminler üretilmiştir.
- Geliştirilen model farklı 2 yolda ve 2 farklı durum için test edilmiş olsa da tüm ülkede çalışacak şekilde bir model olarak tasarlanmıştır.

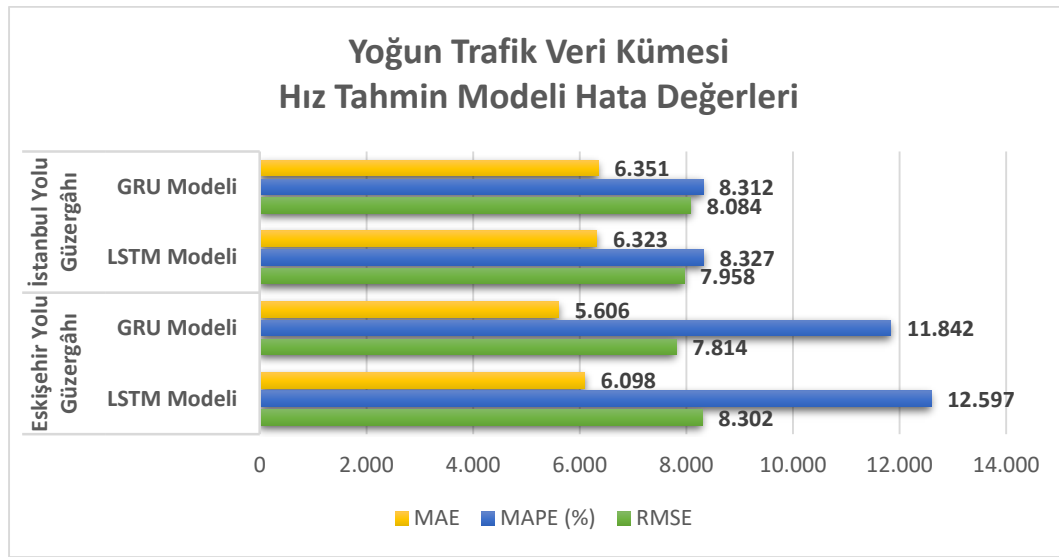
- Model testleri için deneylerde farklı çalışmalarla kıyaslamak ve zamansal performans elde etmek için DBSCAN ve OPTICS kümeleme algoritması kullanılmış, RNN sinir ağını iki farklı varyantı olan LSTM ve GRU modelleri ile tahminler başarı ile yapılmıştır.
 - Benzerlik gösteren 2 ticari uygulama kendi aralarında karşılaştırılmış, aynı zaman dilimi ve güzergâhta bu iki ticari çözüm farklı sonuçlar vermiştir. Ticari ürünlerin çalışma mekanizması ve veri kümesi paylaşılabılır olmadığından önerilen model bu çalışma kapsamında mukayese yapılamamış olsa da 2 ticari ürün arasındaki farkın kullanılan veri türü ve algoritma farklılığından olduğu, kabul edilebilir sonuçlar verdiği, bundan dolayı da piyasa da tercih edildiği görülmüştür. Sonuç olarak, geliştirilen modelin de kabul edilebilir sonuçlar verdiği değerlendirilse de TÜBİTAK kapsamında geliştirilen tahmin modeli ürünleştirildiğinde, canlı verilerle farklı ticari ürünlerle kıyaslanabilir bir yapı ancak o zaman oluşturulabilecektir.
 - Literatürde karşılaştırılabilir nitelikte veri paylaşılması, paylaşılan veri linklerinin bilim dili olan İngilizceden farklı dillerde paylaşılmasından dolayı herhangi bir kıyaslama yapılamamıştır. Ayrıca, yapılan çalışmalarda da sonuçların herhangi bir çalışma ile kıyaslanmadığı da tespit edilmiştir.
 - Önerilen tahmin modeli TÜBİTAK projesi kapsamında gerçek bir ticari ürün geliştirilmesinde kullanılacaktır.
 - Elde edilen sonuçlar, hipotezi doğrular niteliktedir.
- ii. Geliştirilen 2. modelde; trafik hız tahmini için geliştirilen model sonrasında konum mahremiyetini göz önünde bulundurarak model içerisinde diferansiyel mahremiyet yöntemi uygulanmış, tahmin modelinde olduğu gibi önerilen mahremiyet korumalı model 2 farklı durum ve 2 farklı veri seti ile test edilmiş, oluşturulan yol karakteristik bilgilerinin paylaşımı güvenli hale getirilmiş ve sonuç olarak tahmin modeli mahremiyet aşaması ile güçlendirilmiş, mahremiyet korumalı tahmin modeli geliştirilmiş ve başarı ile test edilmiştir. Mahremiyet korumalı hız tahmin modelinde elde edilen sonuçlar önerilen hipotezi doğrular niteliktedir.
- iii. Önerilen 3. model; daha genel kullanım için önerilmiş bir mahremiyet korumalı veri yayınlama modelidir. Önerilen bu model;

- Mobil cihazlardan elde edilen verilerin kullanımı ve analizleriyle faydalı sonuçlar alınsa da kişiyi tanımlayacak ve yörünge oluşturacak veriler içerdiğinden veri yayınlamada mahremiyeti sağlayacak sistemlerin geliştirilmesine katkılar sağlayacaktır.
- Literatüre katkı sunan bir model olup gerçek ortam verilerini anonimleştirirken, literatür çalışmaları incelendiğinde gerçek olmayan sentetik verilerin üretilmesi için modeller önerilmiş veya aynı güzergâh üzerinde olmayan farklı bölgeler için anonimleştirme işlemleri yapılmaktadır.
- Belirlenen yol üzerinde aynı yönde giden gerçek araç verileri için yörünge mahremiyetini sağlayacak yeni bir model olup, tahmin modelinde olduğu gibi Ankara iline ait iki ana yol üzerinde model test edilmiş ve başarılı sonuçlar alınmıştır.
- Veri yayınlama modelinin farklı uygulama alanlarında kullanılabilecek şekilde mahremiyet ihlalini önleme ve veri faydasını sağlamada katkı verebilecektir.
- Hipotezi doğrular niteliktedir.

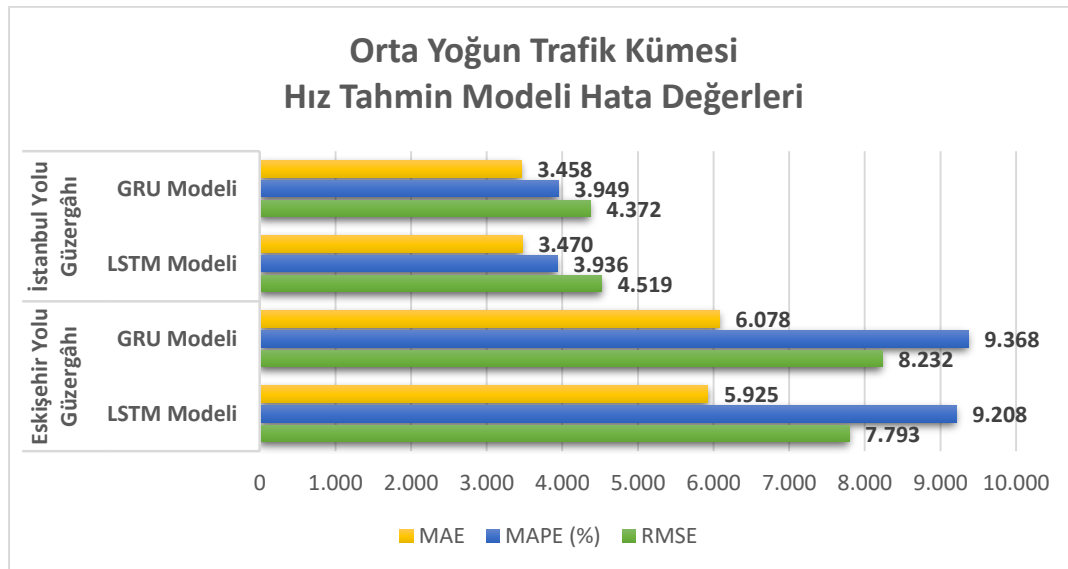
Geliştirilen modellerden elde edilen sonuçlar değerlendirilmiş ve aşağıda verilmiştir.

- Şekil 7.1 ve Şekil 7.2’de verilen tahminleme modelleri sonuçları incelendiğinde;
 - Eskişehir yolu yoğun trafik testlerinde, en az hata GRU modeliyle alınmış, 4. bölümde verilen Şekil 4.16 grafiğinde de görülebileceği üzere ilk 6 kümede yavaş akan trafik gözlemlenmiş, orijinal küme hız değerlerine yakın tahmin sonuçları üretilmiştir. 6. küme sonrasında trafik yoğunluğunun azalmasıyla beraber hız farkları oluşsa da başarılı tahmin sonuçları elde edilmiş, sadece 13 ve 14. Küme hız değişiminin fazla olmasından dolayı sadece o kümede dolayı hata oranı yüksek çıkmıştır. Genel hata oranına bakıldığında en az hata alınan GRU modeli için mutlak hata değeri 5,606 iken küme hızlarının tahminlerinde genel hata %11,842 olarak elde edilmiş, Şekil 7.1’de gösterilmiştir.
 - İstanbul yolu yoğun trafik testlerinde, LSTM modeli mutlak hata değeri 6,323 ile en az hata oranı verirken 4. Bölümde yer alan Şekil 4.17 incelendiğinde 23 adet test kümesinde ani değişimler görülmemekte sadece 7. küme hızından sonra kademeli bir artış gözlemlenmektedir. İstanbul yolu yüzdesel genel hata değeri %8,327 olarak elde edilirken ani hız değişimleri olmamasından dolayı Eskişehir yolu hata oranlarına nazaran İstanbul yolundan elde edilen hata oranları oldukça düşüktür ve iki test yolundaki hata değerlerinin farkları Şekil 7.1’de görülebilmektedir.

- c. Orta yoğun veri kümesinde ise en az mutlak değer hata sonucu Eskişehir yolunda 5,925 ile LSTM modelinden, İstanbul yolunda 3,458 ile GRU modelinden elde edilmiş, ilgili değerlerin grafikleri 4. Bölümde Şekil 4.22 ve 4.24’de gösterilmiş, bu yollar için yüzdesel hata oranı sırasıyla %9,208 ve 3,949 olmuştur.
- d. Kavşaklardaki yavaşlamalardan ve ani hız değişimlerinden dolayı yüzdesel hata oranları artsa da her iki veri kümesinden elde edilen ortalama yüzdesel hata oranları yoğun trafik için %9, orta yoğun trafik için %6 civarındadır ve bu ortalama hata oranları tatmin edici düzeydedir.

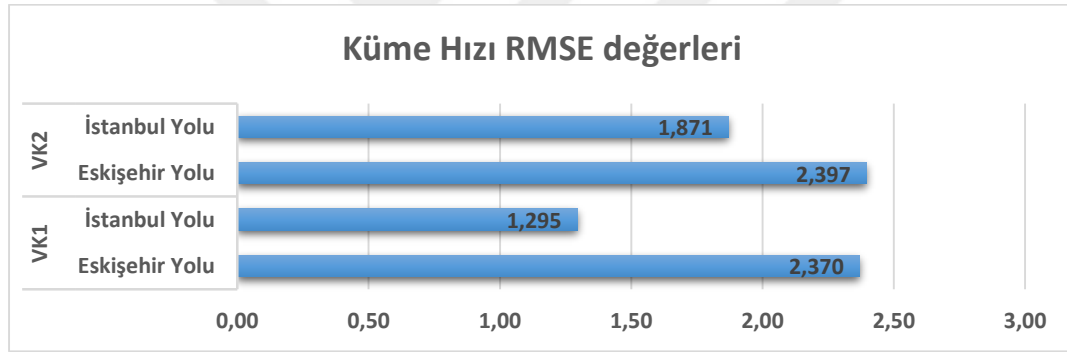


Şekil 7.1 Yoğun trafik veri kümesi için hız tahmin hata değerleri

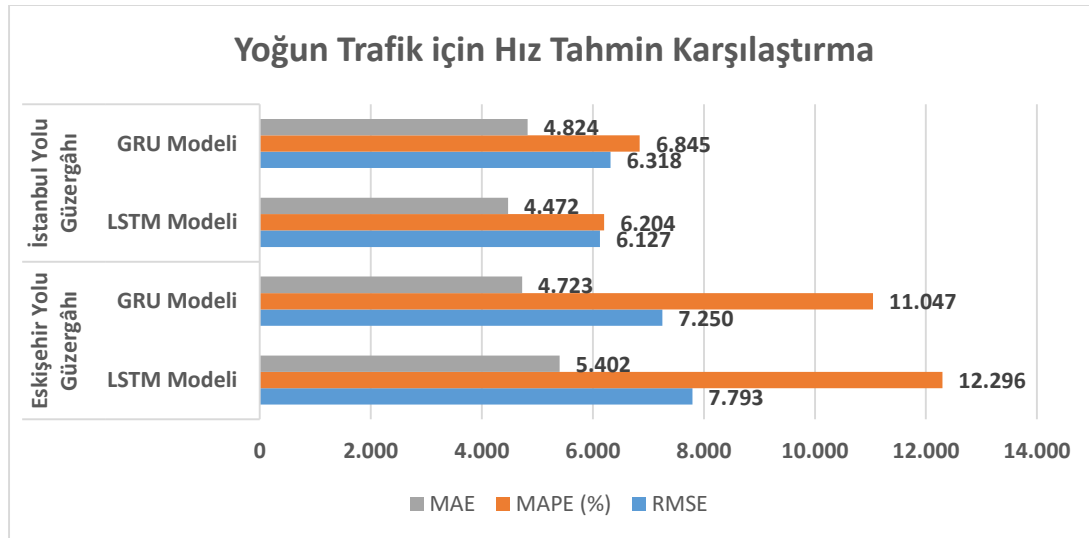


Şekil 7.2 Orta yoğun trafik veri kümesi için hız tahmin hata değerleri

- ii. Şekil 7.3'te dördüncü bölümde oluşan yoğunluk tabanlı kümelerin özniteliklerinden hız bilgileriyle, DM yöntemi uygulanmış küme özniteliklerinin kıyaslandığı RMSE değerleri verilmiştir. Şekil 7.3'teki hata değerleri arasındaki değişimin az olmasıyla Şekil 7.4'te görülebildiği üzere DM uygulanmış ve uygulanmamış özniteliklerle yapılan tahminler arasındaki hata oranları mahremiyet garanti edilirken benzer sonuçlar elde edilebileceğini göstermiştir ve bu sonuçlar 5. Bölümde verilen Şekil 5.3-5.6 grafiklerinde ve hata değerleri Çizelge 5.4'de de görülebilmektedir. Önerilen mahremiyet korumalı model ile elde edilen DM uygulanan küme öznitelikleri trafik yönetim sistemleri ve araştırmacılarla paylaşılabilir, ayrıca mahrem hale gelen öznitelik bilgileriyle trafik tahmini yapılabilir. Geliştirilen mahremiyet korumalı tahmin modeli literatürde önerilen ve bu alanda ilk olarak önerilen mahremiyet korumalı trafik hız tahmin modelidir.

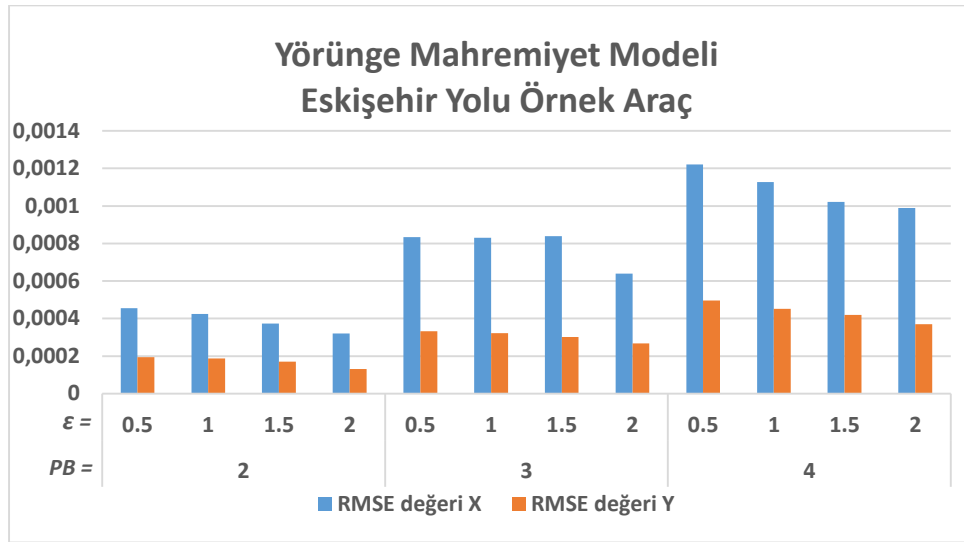


Şekil 7.3 Orijinal küme ve DM uygulanmış küme hızları arasında RMSE değerleri

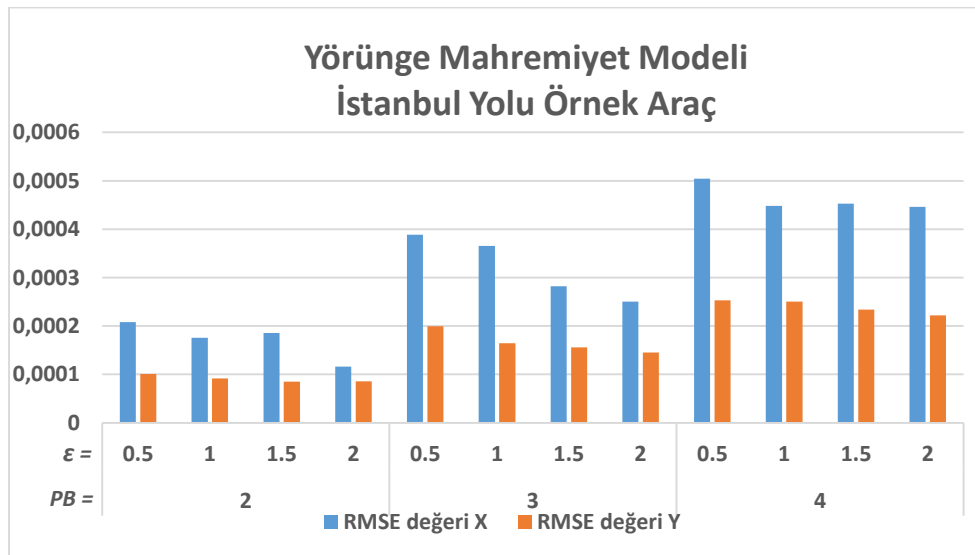


Şekil 7.4 Yoğun trafik veri kümesi için orijinal küme ve DM uygulanmış küme özniteliklerinden üretilen tahmin RMSE değerleri

- iii. Şekil 7.5 ve Şekil 7.6’da verilen yörünge mahremiyet modelleri sonuçları ve 6. Bölümde yer alan Çizelge 6.2 ve 6.3 incelenirse; X ve Y koordinat değerlerinin en düşük RMSE oranları $PB=2$ ve $\varepsilon=2$ değerleri ile önerilen modelde alınmıştır. Bu değerler oldukça küçük hata değerleri olup literatürde konum bilgilerini sıralı şekilde ele alıp mahremiyet işlemine tabi tutarak yörünge mahremiyeti sağlama adına elde edilen ve sunulan ilk sonuçlardır.



Şekil 7.5 Yörünge mahremiyet modeli örnek araç X-Y koordinatı RMSE değerleri



Şekil 7.6 Yörünge mahremiyet modeli örnek araç X-Y koordinatı RMSE değerleri

Sonuç olarak, tez kapsamında geliştirilen modellerin literatüre sunduğu katkılar aşağıda özetlenmiştir. Bunlar;

- i. Tez kapsamında 3 hipotez sunulmuş ve bu hipotezlerin doğrulaması yapılmıştır.
- **Birinci hipotez** için filtrelenen araç verileri haritada eşlenerek OPTICS algoritmasıyla yoğunluk tabanlı kümeler oluşturulmuş, oluşturulan kümelerin öznitelikleriyle yol karakteristikleri belirlenmiş, Eskişehir ve İstanbul yol güzergâhlarında bir başlangıç noktası seçilerek akış oluşturulmuş ve tekrarlayan örüntüler dikkate alınarak LSTM ve GRU modelleriyle hız tahmini yapılmış, tahmin sonuçları kıyaslanmış, elde edilen sonuçlara göre hipotez ispatlanmıştır.
 - **İkinci hipotezi** doğrulamak için, ilk hipotez için geliştirilen modelde yol karakteristiğini belirleyen küme öznitelik oluşumu esnasında diferansiyel mahremiyet yöntemi uygulanmış, birinci modelde olduğu gibi Eskişehir ve İstanbul yollarında test edilerek mahrem hale gelen küme öznitelik bilgileriyle mahrem tahmin sonuçları üretilmiş ve tahmin sonuçları kıyaslanmış, elde edilen sonuçlara göre hipotez ispatlanmıştır.
 - **Üçüncü hipotez** için, zamana göre sıralanan GPS verisi ile elde edilen araç yörünge bilgileri ve bu bilgileri oluşturan konum bilgileri yine sıralı ve ardışık gruplanmasıyla önerilen pencere yöntemi ve diferansiyel mahremiyet uygulanmış, yörünge bilgileri veri yayınlama için mahrem hale getirilmiş, iki ana güzergâhta testleri yapılarak veri faydası ölçülmüş olup elde edilen sonuçlara göre hipotez ispatlanmıştır.
- ii. Geliştirilen hız tahmin modeli yeni ve özgün yapısıyla tarihsel veriye veya hava durumu, kaza gibi dış etken verisine ihtiyaç duymadan trafiğin kendi değişkenliğinde tahmin yapma kabiliyetine sahiptir. Trafiğin sabit veri kaynaklarından zorunlu olarak üretilen tahmin yerine dağıtık GPS verisi kullanarak istenilen segmentte tahmin yapılabileceği gibi segment içinde başlangıç ve bitiş noktaları belirlenerek segment bağımsız istenilen yol bölümünde anlık trafik gösterimi ve tahmini yapılabilmektedir.
- iii. Trafik hız tahmini temel bir parametre olup, kümeler arası mesafe ölçülüp araç sayısına oranlanarak trafik yoğunluk gösterimi ve tahminini de yapılabilir. Farklı temel trafik parametrelerin belirlenmesi ve tahmini trafik rotalama için bir ön işlemdir ve bu parametrelerin tespiti ve tahmini sonucunda trafik yoğunluk durumuna göre kullanıcılara alternatif rotalar sunulabilir, trafik seyahat süreleri yüksek doğrulukla hesaplanabilir.
- iv. Tahmin modelinin aşamaların arasında yer alan küme özniteliklerinin üretimi esnasında uygulanan diferansiyel mahremiyet yöntemiyle de, trafik bilgisine ihtiyacı olan kurumlarla araç verileri yerine kümelere ait öznitelik bilgileri istatistiki olarak

paylaşılabilecektir. Böylelikle hem mahremiyet odaklı anlık gösterimler ve tahminler yürütülebileceği gibi trafiği yöneten veya kurumlara ihtiyacı olan istatistiki bilgiler paylaşılmış olacak, verilerden değer veya fayda elde edilmesinin önü de açılmış olacaktır.

- v. Veri yayınlamada önerilen konum mahremiyet modeli sayesinde; PB ve ϵ parametrelerinin değişmesine rağmen elde edilen hata oranının verilerin kullanılabilirliğini makul seviyelerde değiştirdiğini, böylelikle önerilen modelin yörünge verilerinin mümkün olduğu kadar anlamsal bütünlüğünü koruyarak mahremiyetini sağladığını ortaya koymuştur. Böylelikle veri mahremiyeti sağlanan araç verileri farklı platformlarda güvenlik endişesi olmadan paylaşılabilir, ihtiyaç duyan kurum veya kişiler yol üzerindeki yoğunluk bilgilerinden çıkarımlar yapabilecektir. Bu tez kapsamında önerilen, mahremiyet korumalı trafik verilerinin gelecekte farklı uygulamaların geliştirilmesine, çalışmaların yapılmasına katkı sağlayacağı değerlendirilmektedir.

Tez kapsamında çalışılmış olan konularla ilgili olarak gelecekte yapılabilecekler aşağıda maddeler halinde verilmiştir.

- Bu çalışma kapsamında önerilen modellerin performanslarının geliştirilmesi, gerçek ürüne dönüştürüldüğünde mevcut ürünlerde testleri yapılabilecektir.
- Bu tez kapsamında geliştirilen modellerin TÜBİTAK projesi kapsamında Başarsoft ürününe veya uygulamasına dönüştürülmesi hedeflenmektedir.
- Mahremiyet korumalı yeni modellerin ve projelerin geliştirilmesi planlanmaktadır.

KAYNAKLAR

1. İnternet: TUIK. Motorlu Kara Taşıtları, Ekim 2020. URL: <https://web.archive.org/web/20210316171357/https://tuikweb.tuik.gov.tr/PreHaberBultenleri.do?id=33658>, Son Erişim Tarihi: 16 Aralık 2021.
2. Djahel, S., Doolan, R., Muntean, G. M. and Murphy, J. (2014). A communications-oriented perspective on traffic management systems for smart cities: Challenges and innovative approaches. *IEEE Communications Surveys & Tutorials*, 17(1), 125-151.
3. İnternet: Fundamental parameters of traffic flow. URL: <http://web.archive.org/web/20210406205237/https://nptel.ac.in/content/storage2/courses/105101087/downloads/Lec-30.pdf>, Son Erişim Tarihi: 6 Nisan 2021.
4. El Sherief, M.M., Ramadan, I.M. and Ibrahim, A.M. (2016). Development of traffic stream characteristics models for intercity roads in Egypt. *Alexandria Engineering Journal*, 55(3), 2765-2770.
5. Kovačić, K., Ivanjko, E. and Jelušić, N. (2015). Measurement of road traffic parameters based on multi-vehicle tracking. *arXiv preprint arXiv:1510.04860*.
6. Fung, B.C., Wang, K., Fu, A.W.-C. and Philip, S.Y. (2010). *Introduction to privacy-Preserving Data Publishing: Concepts and Techniques*, Florida: CRC Press, 4-10.
7. Nagy, A.M. and Simon, V. (2018). Survey on traffic prediction in smart cities. *Pervasive and Mobile Computing*, 50, 148-163.
8. İnternet: Intelligent transport systems. URL: <http://web.archive.org/web/20200510203325/http://www.greendigitalcharter.eu/wp-content/uploads/2012/11/2010-European-Commission-Report-on-Intelligent-Transport-Systems.pdf> Son Erişim Tarihi: 10 Mayıs 2020.
9. İnternet: ISBAK-Architect of Smart city. URL: <http://web.archive.org/web/20200511161256/https://www.isbak.istanbul/wp-content/uploads/2018/01/a-guide-to-intelligent-transportation-systems-and-best-practices-a-guide-to-intelligent-transportation-systems-and-best-practices.pdf> Son Erişim Tarihi: 12 Mayıs 2020.
10. Sumalee, A. and Ho, H.W. (2018). Smarter and more connected: Future intelligent transportation system. *IATSS Research*, 42(2), 67-71.
11. İnternet: İBB, UYM Trafik Bilgilendirme. URL: <https://web.archive.org/web/20190502214712/https://uym.ibt.gov.tr/hizmetler/trafik-bilgilendirme>, Son Erişim Tarihi: 20 Nisan 2021.
12. Ersoy, P. and Börühan, G. (2015). *Intelligent transportation systems and their applications in road transportation industry in Turkey*. The International Conference on Logistics & Sustainable Transport, Celje, Slovenia, 11-13.

13. Zhang, J., Wang, F.-Y., Wang, K., Lin, W.-H., Xu, X. and Chen, C. (2011). Data-driven intelligent transportation systems: A survey. *IEEE Transactions on Intelligent Transportation Systems*, 12(4), 1624-1639.
14. Zhang, J., Liao, F., Arentze, T. and Timmermans, H. (2011). A multimodal transport network model for advanced traveler information systems. *Procedia-Social and Behavioral Sciences*, 20, 313-322.
15. Singh, B. and Gupta, A. (2015). Recent trends in intelligent transportation systems: a review. *Journal of Transport Literature*, 9, 30-34.
16. Jin, Q., Wu, G., Boriboonsomsin, K. and Barth, M. (2012). *Advanced intersection management for connected vehicles using a multi-agent systems approach*. 2012 IEEE Intelligent Vehicles Symposium, Madrid, Spain, 932-937 .
17. Bifulco, G.N., Cantarella, G.E., Simonelli, F. and Velonà, P. (2016). Advanced traveller information systems under recurrent traffic conditions: Network equilibrium and stability. *Transportation Research Part B: Methodological*, 92, 73-87.
18. Ding, Z., Chen, B., Zhang, L., Jiang, R., Wu, Y. and Ding, J. (2019). Segment travel time route guidance strategy in advanced traveler information systems. *Physica A: Statistical Mechanics and its Applications*, 534, 120432.
19. Kumar, B.A., Vanajakshi, L. and Subramanian, S.C. (2017). Bus travel time prediction using a time-space discretization approach. *Transportation Research Part C: Emerging Technologies*, 79, 308-332.
20. Jun, C. and Dongyuan, Y. (2013). Estimating smart card commuters origin-destination distribution based on APTS data. *Journal of Transportation Systems Engineering and Information Technology*, 13(4), 47-53.
21. Saharan, S., Bawa, S. and Kumar, N. (2020). Dynamic pricing techniques for Intelligent Transportation System in smart cities: A systematic review. *Computer Communications*, 150, 603-625.
22. Mandžuka, S. (2015). Intelligent transport system: Selected Lectures. *Fakultet Prometnih Znanosti, Zagreb*, 6-7.
23. Lingani, G.M., Rawat, D.B. and Garuba, M. (2019). *Smart traffic management system using deep learning for smart city applications*. 2019 IEEE 9th Annual Computing and Communication Workshop And Conference (CCWC), Las Vegas, NV, USA, 101-106.
24. Lanke, N. and Koul, S. (2013). Smart traffic management system. *International Journal of Computer Applications*, 75(7), 19-22.
25. Dimitrakopoulos, G. and Demestichas, P. (2010). Intelligent transportation systems. *IEEE Vehicular Technology Magazine*, 5(1), 77-84.

26. Mahdavinejad, M.S., Rezvan, M., Barekatain, M., Adibi, P., Barnaghi, P. and Sheth, A.P. (2018). Machine learning for Internet of Things data analysis: A survey. *Digital Communications and Networks*, 4(3), 161-175.
27. İnternet: Transportation, F.D.o. Managing Traffic. URL: https://web.archive.org/web/20210327004744if_/https://i4beyond.com/drive-safely/managing-traffic/#its, Son Erişim Tarihi: 20 Nisan 2021.
28. İnternet: IEEE. Internet of Things Technology will connect highways. URL: https://web.archive.org/web/20210420201632if_/https://spectrum.ieee.org/the-institute/ieee-products-services/internet-of-things-technology-will-connect-highways-street-lights-and-vehicles, Son Erişim Tarihi: 20 Nisan 2021.
29. İnternet: NYC toll plan can help keep Long Island moving. URL: <https://web.archive.org/web/20201108003535/https://www.newsday.com/opinion/editorial/mta-congestion-pricing-tolls-lirr-1.27934240> Son Erişim Tarihi: 20 Nisan 2021.
30. Abbas, Z., Al-Shishtawy, A., Girdzijauskas, S. and Vlassov, V. (2018). *Short-term traffic prediction using long short-term memory neural networks*. 2018 IEEE International Congress on Big Data (BigData Congress), San Francisco, CA, USA, 57-65.
31. Zhu, L., Yu, F.R., Wang, Y., Ning, B. and Tang, T. (2018). Big data analytics in intelligent transportation systems: A survey. *IEEE Transactions on Intelligent Transportation Systems*, 20(1), 383-398.
32. Jiang, H., Zou, Y., Zhang, S., Tang, J. and Wang, Y. (2016). Short-term speed prediction using remote microwave sensor data: machine learning versus statistical model. *Mathematical Problems in Engineering*, 2016(1), 1-13.
33. Gutierrez-Osorio, C. and Pedraza, C. (2020). Modern data sources and techniques for analysis and forecast of road accidents: a review. *Journal of Traffic and Transportation Engineering (English edition)*, 7(4), 432-446.
34. Hua, S. and Anastasiu, D.C. (2019). *Effective vehicle tracking algorithm for smart traffic networks*. 2019 IEEE International Conference on Service-Oriented System Engineering (SOSE), San Francisco, CA, USA, 67-76.
35. Jia, S., Hao, S., Gu, X. and Zhang, L. (2014). *Analyzing and relieving the impact of FCD traffic in LTE-VANET heterogeneous network*. 2014 21st International Conference on Telecommunications (ICT), Lisbon, Portugal, 88-92.
36. Ashwini, B. and Sumathi, R. (2020). *Data Sources for Urban Traffic Prediction: A Review on Classification, Comparison and Technologies*. 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), Thoothukudi, India, 628-635.
37. Heydecker, B. and Addison, J.D. (2011). Analysis and modelling of traffic flow under variable speed limits. *Transportation Research Part C: Emerging Technologies*, 19(2), 206-217.

38. Wardrop, J.G. (1952). Road paper. some theoretical aspects of road traffic research. *Proceedings of the Institution of Civil Engineers*, 1(3), 325-362.
39. İnternet: Mishra, S. Traffic Stream Model Relationship. URL: <https://web.archive.org/web/20170829011757/http://www.ce.memphis.edu/4162/QKV-v2.pdf>, Son Erişim Tarihi: 23 Nisan 2021.
40. Vlahogianni, E.I., Karlaftis, M.G. and Golias, J.C. (2014). Short-term traffic forecasting: Where we are and where we're going. *Transportation Research Part C: Emerging Technologies*, 43, 3-19.
41. Barros, J., Araujo, M. and Rossetti, R.J. (2015). *Short-term real-time traffic prediction methods: A survey*. 2015 International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS), Budapest, Hungary, 132-139.
42. Van Lint, J. and Van Hinsbergen, C. (2012). Short-term traffic and travel time prediction models. *Artificial Intelligence Applications to Critical Transportation Issues*, 22(1), 22-41.
43. Chen, X., Chen, H., Yang, Y., Wu, H., Zhang, W., Zhao, J. and Xiong, Y. (2021). Traffic flow prediction by an ensemble framework with data denoising and deep learning model. *Physica A: Statistical Mechanics and Its Applications*, 565, 125574.
44. Guo, F., Krishnan, R. and Polak, J. (2013). A computationally efficient two-stage method for short-term traffic prediction on urban roads. *Transportation Planning and Technology*, 36(1), 62-75.
45. Yuan, Z., Zhang, W. and Yang, M. (2016). *A Short-term Traffic Flow Prediction Approach of Neural Network Based on Cluster Analysis*. 2016 3rd International Conference on Engineering Technology and Application (ICETA 2016), Sarajevo, Bosnia and Herzegovina, 129-134.
46. Zhao, S., Zhao, Q., Bai, Y. and Li, S. (2019). A traffic flow prediction method based on road crossing vector coding and a bidirectional recursive neural network. *Electronics*, 8(9), 1006.
47. Do, L.N., Vu, H.L., Vo, B.Q., Liu, Z. and Phung, D. (2019). An effective spatial-temporal attention based neural network for traffic flow prediction. *Transportation Research Part C: Emerging Technologies*, 108, 12-28.
48. Duan, H., Xiao, X. and Xiao, Q. (2020). An inertia grey discrete model and its application in short-term traffic flow prediction and state determination. *Neural Computing and Applications*, 32(12), 8617-8633.
49. Polson, N.G. and Sokolov, V.O. (2017). Deep learning for short-term traffic flow prediction. *Transportation Research Part C: Emerging Technologies*, 79, 1-17.

50. Zhao, Z., Chen, W., Wu, X., Chen, P.C. and Liu, J. (2017). LSTM network: a deep learning approach for short-term traffic forecast. *IET Intelligent Transport Systems*, 11(2), 68-75.
51. Qu, W., Li, J., Yang, L., Li, D., Liu, S., Zhao, Q. and Qi, Y. (2020). Short-term intersection Traffic flow forecasting. *Sustainability*, 12(19), 8158.
52. Habtemichael, F.G. and Cetin, M. (2016). Short-term traffic flow rate forecasting based on identifying similar traffic patterns. *Transportation Research Part C: Emerging Technologies*, 66, 61-78.
53. Li, L., He, S., Zhang, J. and Ran, B. (2016). Short-term highway traffic flow prediction based on a hybrid strategy considering temporal-spatial information. *Journal of Advanced Transportation*, 50(8), 2029-2040.
54. Nam, D., Lavanya, R., Jayakrishnan, R., Yang, I. and Jeon, W.H. (2020). A deep learning approach for estimating traffic density using data obtained from connected and autonomous probes. *Sensors*, 20(17), 4824.
55. Raj, J., Bahuleyan, H. and Vanajakshi, L.D. (2016). Application of data mining techniques for traffic density estimation and prediction. *Transportation Research Procedia*, 17, 321-330.
56. Da Silva, T.L.C., Neto, A.C.A., Magalhaes, R.P., de Farias, V.A., de Macêdo, J.A.F. and Machado, J.C. (2014). *Efficient and Distributed DBScan Algorithm Using MapReduce to Detect Density Areas on Traffic Data*. ICEIS 2014 - Proceedings of the 16th International Conference on Enterprise Information Systems, Lisbon, Portugal, 52-59.
57. Aljamal, M.A., Abdelghaffar, H.M. and Rakha, H.A. (2020). Estimation of traffic stream density using connected vehicle data: Linear and nonlinear filtering approaches. *Sensors*, 20(15), 4066.
58. Zeroual, A., Harrou, F. and Sun, Y. (2019). Road traffic density estimation and congestion detection with a hybrid observer-based strategy. *Sustainable Cities and society*, 46, 101411.
59. Alkouz, B. and Al Aghbari, Z. (2020). SNSJam: Road traffic analysis and prediction by fusing data from multiple social networks. *Information Processing & Management*, 57(1), 102139.
60. Hu, H., Gao, Z., Sheng, Y., Zhang, C. and Zheng, R. (2019). Traffic density recognition based on image global texture feature. *International Journal of Intelligent Transportation Systems Research*, 17(3), 171-180.
61. Chung, J. and Sohn, K. (2017). Image-based learning to measure traffic density using a deep convolutional neural network. *IEEE Transactions on Intelligent Transportation Systems*, 19(5), 1670-1675.

62. Goves, C., North, R., Johnston, R. and Fletcher, G. (2016). Short term traffic prediction on the UK motorway network using neural networks. *Transportation Research Procedia*, 13, 184-195.
63. Cui, Z., Ke, R., Pu, Z., Ma, X. and Wang, Y. (2020). Learning traffic as a graph: A gated graph wavelet recurrent neural network for network-scale traffic prediction. *Transportation Research Part C: Emerging Technologies*, 115, 102620.
64. Niu, K., Zhang, H., Zhou, T., Cheng, C. and Wang, C. (2019). A novel spatio-temporal model for city-scale traffic speed prediction. *IEEE Access*, 7, 30050-30057.
65. Le, T.V., Oentaryo, R., Liu, S. and Lau, H.C. (2016). Local Gaussian processes for efficient fine-grained traffic speed prediction. *IEEE Transactions on Big Data*, 3(2), 194-207.
66. Ma, X., Tao, Z., Wang, Y., Yu, H. and Wang, Y. (2015). Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. *Transportation Research Part C: Emerging Technologies*, 54, 187-197.
67. Li, W., Wang, J., Fan, R., Zhang, Y., Guo, Q., Siddique, C. and Ban, X.J. (2020). Short-term traffic state prediction from latent structures: Accuracy vs. efficiency. *Transportation Research Part C: Emerging Technologies*, 111, 72-90.
68. Cui, Z., Ke, R., Pu, Z. and Wang, Y. (2020). Stacked bidirectional and unidirectional LSTM recurrent neural network for forecasting network-wide traffic state with missing values. *Transportation Research Part C: Emerging Technologies*, 118, 102674.
69. Zhao, J., Gao, Y., Yang, Z., Li, J., Feng, Y., Qin, Z. and Bai, Z. (2019). Truck traffic speed prediction under non-recurrent congestion: Based on optimized deep learning algorithms and GPS data. *IEEE Access*, 7, 9116-9127.
70. Jiang, B. and Fei, Y. (2016). Vehicle speed prediction by two-level data driven models in vehicular networks. *IEEE Transactions on Intelligent Transportation Systems*, 18(7), 1793-1801.
71. Kim, Y., Wang, P. and Mihaylova, L. (2019). Scalable learning with a structural recurrent neural network for short-term traffic prediction. *IEEE Sensors Journal*, 19(23), 11359-11366.
72. Jeon, S. and Hong, B. (2016). Monte Carlo simulation-based traffic speed forecasting using historical big data. *Future Generation Computer Systems*, 65, 182-195.
73. Ru, X., Xu, X., Zhou, Y. and Yang, C. (2020). Critical Segments Identification for Link Travel Speed Prediction in Urban Road Network. *Journal of Advanced Transportation*, 2020(1), 1-10.
74. Brandeis, L. and Warren, S. (1890). The right to privacy. *Harvard law review*, 4(5), 193-220.

75. Jain, P., Gyanchandani, M. and Khare, N. (2016). Big data privacy: a technological perspective and review. *Journal of Big Data*, 3(1), 1-25.
76. İnternet: Banisar, D. Privacy and Human Rights. URL: <https://web.archive.org/web/20210325082412/http://gilc.org/privacy/survey/intro.html>, Son Erişim Tarihi: 25 Mart 2021.
77. Binjubeir, M., Ahmed, A.A., Ismail, M.A.B., Sadiq, A.S. and Khan, M.K. (2019). Comprehensive survey on big data privacy protection. *IEEE Access*, 8, 20067-20079.
78. Patil, H.K. and Seshadri, R. (2014). *Big data security and privacy issues in healthcare*. 2014 IEEE International Congress On Big Data, Anchorage, AK, USA, 762-765.
79. İnternet: 2018 Verizon Data Breach Investigations Report. URL: https://web.archive.org/web/20210225190408/https://www.researchgate.net/publication/324455350_2018_Verizon_Data_Breach_Investigations_Report, Son Erişim Tarihi: 25 Şubat 2021.
80. De Capitani Di Vimercati, S., Foresti, S., Livraga, G. and Samarati, P. (2012). Data privacy: definitions and techniques. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 20(06), 793-817.
81. Beresford, A.R. and Stajano, F. (2003). Location privacy in pervasive computing. *IEEE Pervasive Computing*, 2(1), 46-55.
82. İnternet: KVKK, Kamuoyu Duyurusu. URL: <https://web.archive.org/web/20210214035733/https://www.kvkk.gov.tr/Icerik/6726/COVID-19-ILE-MUCADELEDE-KONUM-VERISININ-ISLENMESI-VE-KISILERIN-HAREKETLILIKLERININ-IZLENMESI-HAKKINDA-BILINMESI-GEREKENLER-2->, Son Erişim Tarihi: 14 Şubat 2021.
83. İnternet: General Data Protection Regulation (GDPR) and location data. URL: <https://web.archive.org/web/20210516224656/https://joinup.ec.europa.eu/collection/elise-european-location-interoperability-solutions-e-government/news/gdpr-and-location-data> Son Erişim Tarihi: 16 Mayıs 2021.
84. Krumm, J. (2009). *Realistic driving trips for location privacy*. International Conference on Pervasive Computing, Nara, Japan, 25-41.
85. Zhang, S., Li, X., Tan, Z., Peng, T. and Wang, G. (2019). A caching and spatial K-anonymity driven privacy enhancement scheme in continuous location-based services. *Future Generation Computer Systems*, 94, 40-50.
86. Sui, X., Chen, Z., Guo, L., Wu, K., Ma, J. and Wang, G. (2017). Social media as sensor in real world: movement trajectory detection with microblog. *Soft Computing*, 21(3), 765-779.
87. Li, M., Zhu, L., Zhang, Z. and Xu, R. (2017). Achieving differential privacy of trajectory data publishing in participatory sensing. *Information Sciences*, 400, 1-13.

88. Liu, B., Zhou, W., Zhu, T., Gao, L. and Xiang, Y. (2018). Location privacy and its applications: A systematic study. *IEEE Access*, 6, 17606-17624.
89. Hara, T., Suzuki, A., Iwata, M., Arase, Y. and Xie, X. (2016). Dummy-based user location anonymization under real-world constraints. *IEEE Access*, 4, 673-687.
90. Abul, O., Bonchi, F. and Nanni, M. (2008). *Never walk alone: Uncertainty for anonymity in moving objects databases*. 2008 IEEE 24th International Conference On Data Engineering, Cancun, Mexico, 376-385.
91. Huo, Z., Meng, X., Hu, H. and Huang, Y. (2012). *You can walk alone: trajectory privacy-preserving through significant stays protection*. International Conference on Database Systems for Advanced Applications, Busan, Korea, 351-366.
92. Zhang, S., Mao, X., Choo, K.-K.R., Peng, T. and Wang, G. (2020). A trajectory privacy-preserving scheme based on a dual-K mechanism for continuous location-based services. *Information Sciences*, 527, 406-419.
93. Ren, W. and Tang, S. (2020). EGeoIndis: An effective and efficient location privacy protection framework in traffic density detection. *Vehicular Communications*, 21, 100187.
94. Qi, Y., Hossain, M.S., Nie, J. and Li, X. (2021). Privacy-preserving blockchain-based federated learning for traffic flow prediction. *Future Generation Computer Systems*, 117, 328-337.
95. İnternet: Firma Profili | Başarsoft Bilgi Teknolojileri A.Ş. URL: <https://web.archive.org/web/20201201084806/https://www.basarsoft.com.tr/firma-profil/>, Son Erişim Tarihi: 1 Aralık 2020.
96. İnternet: Installation — HoloViews 1.14.3 documentation. URL: <https://web.archive.org/web/20210413122239/http://holoviews.org/>, Son Erişim Tarihi: 13 Nisan 2021.
97. İnternet: Installation — GeoViews 1.9.1 documentation. URL: <https://web.archive.org/web/20201031021639/https://geoviews.org/>, Son Erişim Tarihi: 31 Ekim 2020.
98. İnternet: OpenStreetMap. URL: <https://web.archive.org/web/20210518002120/https://www.openstreetmap.org/> Son Erişim Tarihi: 18 Mayıs 2021.
99. Ankerst, M., Breunig, M.M., Kriegel, H. P. and Sander, J. (1999). OPTICS: Ordering points to identify the clustering structure. *ACM Sigmod Record*, 28(2), 49-60.
100. İnternet: 2.3. Clustering — scikit-learn 0.20.3 documentation. URL: <https://web.archive.org/web/20210413235850/https://scikit-learn.org/stable/modules/clustering.html>, Son Erişim Tarihi: 13 Nisan 2021.

101. Ester, M., Kriegel, H. P., Sander, J. and Xu, X. (1996). *A density-based algorithm for discovering clusters in large spatial databases with noise*. KDD-96, Portland, USA, 226-231.
102. Cui, Z., Ke, R., Pu, Z. and Wang, Y. (2018). Deep bidirectional and unidirectional LSTM recurrent neural network for network-wide traffic speed prediction. *arXiv preprint arXiv:1801.02143*.
103. Wang, Y., Zhang, D., Liu, Y., Dai, B. and Lee, L.H. (2019). Enhancing transportation systems via deep learning: A survey. *Transportation Research Part C: Emerging Technologies*, 99, 144-163.
104. Işık, G. and Artuner, H. (2020). Turkish dialect recognition in terms of prosodic by long short-term memory neural networks. *Journal of the Faculty of Engineering and Architecture of Gazi University*, 35(1), 213-224.
105. Dwork, C. (2006). *Differential privacy*. International Colloquium on Automata, Languages, and Programming, Venice, Italy, 1-12.
106. Zhu, T., Li, G., Zhou, W. and Philip, S.Y. (2017). Differentially private data publishing and analysis: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 29(8), 1619-1638.
107. Canbay, Y. and Sağiroğlu, Ş. (2020). Derin Öğrenmede Diferansiyel Mahremiyet. *Uluslararası Bilgi Güvenliği Mühendisliği Dergisi*, 6(1), 1-16.
108. Zhao, J., Chen, Y. and Zhang, W. (2019). Differential privacy preservation in deep learning: Challenges, opportunities and solutions. *IEEE Access*, 7, 48901-48911.
109. Fusco, G., Colombaroni, C. and Isaenko, N. (2016). Short-term speed predictions exploiting big data on large urban road networks. *Transportation Research Part C: Emerging Technologies*, 73, 183-201.
110. Andrés, M.E., Bordenabe, N.E., Chatzikokolakis, K. and Palamidessi, C. (2013). *Geo-indistinguishability: Differential privacy for location-based systems*. Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security, Berlin, Germany, 901-914.
111. Chatzikokolakis, K., Andrés, M.E., Bordenabe, N.E. and Palamidessi, C. (2013). *Broadening the scope of differential privacy using metrics*. International Symposium on Privacy Enhancing Technologies Symposium, Bloomington, USA, 82-102.
112. Winarno, E., Hadikurniawati, W. and Rosso, R.N. (2017). *Location based service for presence system using haversine method*. 2017 international conference on innovative and creative information technology (ICITech), Salatiga, Indonesia, 1-4.
113. Chopde, N.R. and Nichat, M. (2013). Landmark based shortest path detection by using A* and Haversine formula. *International Journal of Innovative Research in Computer and Communication Engineering*, 1(2), 298-302.

114. Yang, J., Rahardja, S. and Fränti, P. (2019). *Outlier detection: how to threshold outlier scores?* Proceedings of the international conference on artificial intelligence, information processing and cloud computing.
115. Hahn, D.A., Munir, A. and Behzadan, V. (2021). Security and Privacy Issues in Intelligent Transportation Systems: Classification and Challenges. *IEEE Intelligent Transportation Systems Magazine*, 13(1), 181-196.
116. Hasan, A., Qu, Q., Li, C., Chen, L. and Jiang, Q. (2018). An effective privacy architecture to preserve user trajectories in reward-based LBS applications. *ISPRS International Journal of Geo-Information*, 7(2), 53.
117. Ometov, A., Bezzateev, S., Davydov, V., Shchesniak, A., Masek, P., Lohan, E.S. and Koucheryavy, Y. (2019). Positioning information privacy in intelligent transportation systems: An overview and future perspective. *Sensors*, 19(7), 1603.
118. Dwork, C. and Smith, A. (2010). Differential privacy for statistics: What we know and what we want to learn. *Journal of Privacy and Confidentiality*, 1(2), 135-154.
119. Canbay, Y. (2019). *Aykırı Veri Yönelimli Fayda Temelli Büyük Veri Anonimleştirme Modeli*, Doktora Tezi, Gazi Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
120. Canbay, Y., Vural, Y. ve Sağiroğlu, Ş. (2020). Mahremiyet korumalı büyük veri yayınlama için kavramsal model önerileri. *Politeknik Dergisi*, 23(3), 785-798.
121. Zorarpacı, E. and Özel, S.A. (2021). Privacy preserving classification over differentially private data. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 11(3), 1399.
122. Narayanan, A. and Shmatikov, V. (2008). *Robust de-anonymization of large sparse datasets*. 2008 IEEE Symposium on Security and Privacy (sp 2008), Oakland, CA, USA, 111-125.
123. Wang, R., Li, Y.F., Wang, X., Tang, H. and Zhou, X. (2009). *Learning your identity and disease from research papers: information leaks in genome wide association study*. Proceedings of the 16th ACM Conference on Computer and Communications security, Illinois, USA, 534-544.
124. Sweeney, L. (2000). Simple demographics often identify people uniquely. *Health (San Francisco)*, 671(2000), 1-34.
125. Xiong, X., Liu, S., Li, D., Cai, Z. and Niu, X. (2020). A comprehensive survey on local differential privacy. *Security and Communication Networks*, 2020(1), 1-29.
126. İnternet: Facebook–Cambridge Analytica data scandal. URL: https://web.archive.org/web/20201226151447/https://en.wikipedia.org/wiki/Facebook%E2%80%93Cambridge_Analytica_data_scandal, Son Erişim Tarihi: 26 Aralık 2020.

127. İnternet: Diferansiyel Mahremiyet ve Gizlilik. URL: <https://web.archive.org/web/20201026132715/http://ozdenercin.com/2018/09/19/differential-privacy-diferansiyel-gizlilik-ve-mahremiyet/>, Son Erişim Tarihi: 26 Ekim 2020.
128. Nussbaum, D., Omran, M.T. and Sack, J.-R. (2012). *Techniques to protect privacy against inference attacks in location based services*. Proceedings of the 3rd ACM SIGSPATIAL International Workshop on GeoStreaming, California, USA, 58-67.
129. Sweeney, L. (2002). k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(05), 557-570.
130. Machanavajjhala, A., Kifer, D., Gehrke, J. and Venkatasubramanian, M. (2007). l-diversity: Privacy beyond k-anonymity. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 1(1), 1-52.
131. Li, N., Li, T. and Venkatasubramanian, S. (2007). *t-closeness: Privacy beyond k-anonymity and l-diversity*. 2007 IEEE 23rd International Conference on Data Engineering, Istanbul, Turkey, 106-115.
132. Nergiz, M.E., Atzori, M. and Clifton, C. (2007). *Hiding the presence of individuals from shared databases*. Proceedings of the 2007 ACM SIGMOD international conference on Management of data, Beijing, China, 665-676.
133. Palanisamy, B. and Liu, L. (2015). Attack-Resilient Mix-zones over Road Networks: Architecture and Algorithms. *IEEE Transactions on Mobile Computing*, 14(3), 495-508.
134. Ardagna, C.A., Cremonini, M., Damiani, E., Di Vimercati, S.D.C. and Samarati, P. (2007). *Location privacy protection through obfuscation-based techniques*. IFIP Annual Conference on Data and Applications Security and Privacy, Redondo Beach, CA, USA, 47-60.
135. Xiao, Y. and Xiong, L. (2015). *Protecting locations with differential privacy under temporal correlations*. Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, Denver, USA, 1298–1309.
136. Dwork, C. (2008). *Differential privacy: A survey of results*. International conference on theory and applications of models of computation, Xi'an, China, 1-19.
137. Xie, H., Kulik, L. and Tanin, E. (2011). Privacy-aware collection of aggregate spatial data. *Data & Knowledge Engineering*, 70(6), 576-595.
138. Shaham, S., Ding, M., Liu, B., Dang, S., Lin, Z. and Li, J. (2020). Privacy preserving location data publishing: A machine learning approach. *IEEE Transactions on Knowledge and Data Engineering*, 33(9), 3270-3283.
139. İnternet: IBB-Açık Veri. URL: <https://web.archive.org/web/20201118091456/https://data.ibb.gov.tr/>, Son Erişim Tarihi: 18 Kasım 2020.

140. Liu, X. and Zhu, Y. (2020). Privacy and Utility Preserving Trajectory Data Publishing for Intelligent Transportation Systems. *IEEE Access*, 8, 176454-176466.
141. Hoh, B., Gruteser, M., Xiong, H. and Alrabady, A. (2007). *Preserving privacy in GPS traces via uncertainty-aware path cloaking*. Proceedings of the 14th ACM conference on Computer and communications security, Virginia, USA, 161-171.
142. Wang, T. and Liu, L. (2009). From data privacy to location privacy. *Machine Learning in Cyber Trust*, 217-246.





GAZİ GELECEKTİR..