**T.C.**
**İSTANBUL ÜNİVERSİTESİ-CERRAHPAŞA**
**LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ**

# YÜKSEK LİSANS TEZİ

# İSTANBUL SAATLİK TOPLU ULAŞIM VERİSİNE GÖRE GELECEĞE DÖNÜK GEÇİŞ SAYILARININ TAHMİNLENMESİ VE SEFER SAYILARININ OPTİMİZASYONU

**Emel GEÇİCİ**

**DANIŞMAN**
**Dr. Öğr. Üyesi Gülsüm Zeynep GÜRKAŞ AYDIN**

**Bilgisayar Mühendisliği Anabilim Dalı**

**Bilgisayar Mühendisliği, Tezli Yüksek Lisans Programı**

**Kasım, 2023**

# BÜTÇE DESTEKLERİ

## İSTANBUL SAATLİK TOPLU ULAŞIM VERİSİNE GÖRE GELECEĞE DÖNÜK GEÇİŞ SAYILARININ TAHMİNLENMESİ VE SEFER SAYILARININ OPTİMİZASYONU

Bu tez çalışması için herhangi bir kurumdan bütçe desteği alınmamıştır.

# TEŞEKKÜR

Ulaşımla ilgili yaşanan problemler günlük yaşantımızda sıkılıkla karşılaştığımız ve giderek artan bir hal almaya başlamıştır. Yoğunluğun olması, istenilen talebin karşılanamaması, sefer sayılarının yeterli olmamasına bağlı olarak yaşanan aksaklıklar ulaşımı etkilemekte özellikle toplu ulaşımı kullanan yolcular için zorlu bir hale gelmiştir.

Bu çalışmada ulaşımda kullanılan saatlik geçiş verilerinin geleceğe yönelik tahmin edilmesi üzerine çalışma yapılmıştır. Yapılan çalışma sonucu hareketliliğin fazla olduğu saatler tespit edilmeye çalışılmış ve yoğunluğun fazla olduğu saatler için o noktalarda sefer sayılarının arttırılması önerilmiştir.

Tez çalışmam sırasında değerli ve kıymetli bilgi, birikim ve tecrübeleri ile bana destek olan değerli danışman hocam Sayın Dr. Gülsüm Zeynep GÜRKAŞ AYDIN hocama sonsuz teşekkür ve saygılarımı sunarım. Yüksek Lisans Eğitimim ve Tez Dönemim boyunca bana destek olan, bilgi birikimi ve tecrübeleri ile yol gösteren kıymetli hocam Doç. Dr. Muhammed Ali AYDIN hocama sonsuz teşekkür ve saygılarımı sunarım.

Kıymetli ailem annem Şennur ÇORUH, ablam Ebru GEÇİCİ ve kardeşim Buğra GEÇİCİ ’ye her daim yanımda oldukları için çok teşekkür ediyorum.

Bizlere örnek olan yol göstericimiz, yolu daima yolumuz olacak büyük önderimiz başöğretmen Gazi Mustafa Kemal ATATÜRK’ e sonsuz saygı ve teşekkürü borç bilirim.

Kasım 2023 **Emel GEÇİCİ**

# İÇİNDEKİLER

**Sayfa No**

**BÜTÇE DESTEKLERİ** ..... **iii**

**TEŞEKKÜR** ..... **iv**

**İÇİNDEKİLER** ..... **v**

**ŞEKİL LİSTESİ** ..... **vii**

**TABLO LİSTESİ** ..... **viii**

**SİMGE VE KISALTMA LİSTESİ** ..... **ix**

**ÖZET** ..... **xi**

**ABSTRACT** ..... **xiii**

**1. GİRİŞ** ..... **15**

**2. KAVRAMSAL ÇERÇEVE** ..... **18**

2.1. GÖZETİMLİ ÖĞRENME ..... 18

2.1.1. DOĞRUSAL (LİNEER) REGRESYON ..... 20

2.1.2. RİDGE REGRESYON ..... 22

2.1.3. LASSO ..... 22

2.1.4. POLİNOM REGRESYON ..... 24

2.1.5. ELASTİC NET REGRESYON ..... 24

2.1.6. K-EN YAKIN KOMŞULAR (K-NN) ALGORİTMASI ..... 25

2.1.7. KARAR AĞAÇLARI ..... 26

2.1.8. RASTGELE ORMANLAR ALGORİTMASI ..... 28

2.1.9. DESTEK VEKTÖR MAKİNELERİ ..... 29

2.2. YAPAY SİNİR AĞLARI ..... 30

2.2.1. İleri Beslemeli Ağlar: ..... 31
2.2.2. Geri Beslemeli Ağlar: ..... 31
2.3. GÖZETİMLİ ÖĞRENME ALGORİTMANLARININ KARŞILAŞTIRILMASI ..... 32
2.4. GÖZETİMSİZ ÖĞRENME: KÜMELEME ..... 33
2.5. MAKSİMUM KÜME ÖRTME PROBLEMİ ..... 34
2.6. LİTERATÜR TARAMASI ..... 35
**3. YÖNTEM ..... 48**
3.1. UYGULAMA BÖLGESİ: İSTANBUL ..... 48
3.2. UYGULAMA VERİ SETİ ..... 52
3.3. GELECEK YOLCULUK SAYILARININ TAHMİN EDİLMESİ ..... 56
3.4. SEFER SAYISI ARTIRILACAK HATTI BELİRLENMESİ ..... 57
**4. BULGULAR ..... 58**
**5. TARTIŞMA ..... 60**
**6. SONUÇ ve ÖNERİLER ..... 62**
**KAYNAKLAR ..... 63**

# ŞEKİL LİSTESİ

**Sayfa No**

Şekil 1.1. İstanbul nüfusunun yıllara göre değişimi.....15

Şekil 2.1. Öğrenme Tekniklerinin Sınıflandırılması.....19

Şekil 2.2. Basit Doğrusal Regresyon.....21

Şekil 2.3. Lasso ve Ridgre Regresyonun Karşılaştırılması: (a) Lasso, (b) Ridge Regresyon (James, Witten, Hastie, & Tibshirani, 2013).....23

Şekil 2.4. Doğrusal Regresyon vs. Polinom Regresyon.....24

Şekil 2.5. (a) Lasso, (b) Ridge regresyon, (c) Elastic Net (Zou & Hastie, 2005).....25

Şekil 2.6. Karar Ağaçları Genel Yapısı (Java T Point, 2023).....27

Şekil 3.1. İstanbul'un Mahallelere göre Nüfus Yoğunluğu.....48

Şekil 3.2. Kümeler için Oluşturulan Dirsek Grafikleri: (a) Avrupa Yakası, (b) Anadolu Yakası, c) Tüm İlçeler.....51

# TABLO LİSTESİ

**Sayfa No**

Tablo 1.1. İstanbul'da kullanılan toplu taşıma türleri ..... 16

Tablo 2.1. Tahminleme Yöntemleri ..... 19

Tablo 2.2. Mesafelerin Hesaplanması için Kullanılan Uzaklık Formülleri ..... 26

Tablo 2.3. Regresyon Problemlerinde Algoritmaların Karşılaştırılması için Kullanılan Performans Göstergeleri ..... 32

Tablo 2.3. (Devamı) Regresyon Problemlerinde Algoritmaların Karşılaştırılması için Kullanılan Performans Göstergeleri ..... 33

Tablo 3.1. İstanbul'un İlçeleri ..... 49

Tablo 3.2. İstanbul'un Avrupa Yakasında Yer Alan İlçeleri ..... 49

Tablo 3.3. İstanbul'un Anadolu Yakasında Yer Alan İlçeleri ..... 49

Tablo 3.4. İstanbul'un İlçelerini Kümelemek için Kullanılacak olan Değişkenler Listesi ..... 50

Tablo 3.5. İstanbul'un İlçelerinin Kümelenmesi ..... 52

# SİMGE VE KISALTMA LİSTESİ

| Simgeler | Açıklama |
|---|---|
| **β** | : Beta |
| **ϵ** | : Epsilon |
| **x** | : Bağımsız değişken |
| **y** | : Bağımlı değişken |
| ∑ | : Toplama sembolü |
| **λ** | : Lambda |

| Kısaltmalar | Açıklama |
|---|---|
| **VISUM** | : Ağ ve talep modellemesi |
| **AA** | : Armoni Araştırması (Sezgisel Optimizasyon tekniği tabanlı model) |
| **LINGO** | : Optimizasyon yapan programlama dili |
| **ANFIS** | : Bulanık çıkarıma dayalı yapay sinir ağı yöntemi, Uyarlamalı ağ tabanlı çıkarım |
| **BCa** | : Ortalama yolcu talebi ve medyan yolcu talebi için boostrap güven aralıklarını hesaplayarak tahmini yolcu talebi |
| **EDM** | : Ekonometrik Dinamik Model |
| **Arellano Bond** | : Dinamik panel veri modellerini tahmin etmek için kullanılan genelleştirilmiş bir moment tahmin edici yöntemi |
| **AAH** | : Havayolu Havaalanı Merkezi konum problemi |
| **VMD** | : Varyasyonel Model Araştırma |
| **ADF** | : Bir zaman serisi içerisindeki durağanlığı test eder |
| **GVAR** | : Global Vector Auto Regression |
| **GAPAXDE** | : Avustralya'nın iç hat havayolu yolcu talebi için uçaklı yolcular tarafından ölçülen genetik algoritma optimizasyon modeli |
| **GARPKSDE** | : Gerçekleştirilen Avustralya iç hat havayolu geliri yolcu kilometresi için genetik algoritma optimizasyon modeli |
| **ABD** | : Amerika Birleşik Devletleri |
| **SSA** | : Singular Spectrum Analysis |

**SSA-WPDCNN-SVR** :Kısa vadede bir şehirlerarası yüksek hızlı demiryolunun günlük O-D yolcu akışını tahmin etmek için kullanılan metot

**RPK** : Revenue Passenger Kilometers-Ücretli yolcu kilometresi

**HKT** : Hata Kareler Toplamı

**KNN** : K-Nearest Neighbor, K-En yakın komşu

**SVM** : Support Vector Machines- Destek Vektör Makineleri

**YSA** : Yapay Sinir Ağlar

**Örn** : Örnek

# ÖZET

## YÜKSEK LİSANS TEZİ

## İSTANBUL SAATLİK TOPLU ULAŞIM VERİSİNE GÖRE GELECEĞE DÖNÜK GEÇİŞ SAYILARININ TAHMİNLENMESİ VE SEFER SAYILARININ OPTİMİZASYONU

**Emel GEÇİCİ**

**İstanbul Üniversitesi-Cerrahpaşa**

**Lisansüstü Eğitim Enstitüsü**

**Bilgisayar Mühendisliği Anabilim Dalı**

**Bilgisayar Mühendisliği, Tezli Yüksek Lisans Programı**

**Danışman : Dr. Öğr. Üyesi Gülsüm Zeynep GÜRKAŞ AYDIN**

Ulaşımla ilgili yaşanan problemler gelişmiş ve gelişmekte olan ülkelerde oldukça yoğun olarak görülen ve karşılaşılan bir problemdir. Çoğu ülkede gözle görülen ve oldukça önemli noktada yer alan ulaşım sorunları bizim ülkemizde de görülmektedir. Ulaşımla ilgili yaşanan problemler tek bir konu özelinde olmamakla beraber pek çok alanda karşımıza çıkmaktadır. Yaşanan problemler yolcu sayısının fazlalığı, sefer sayısının yeteri kadar olmaması sebebiyle yolcu talebine karşılık verememesi. Toplu ulaşımda yaşanan hat güzergâh düzenlemesinin yolcu isteğine göre karşılanmaması, eksik güzergâh olması ve sefer sıklığının yeterli olmaması sebebiyle yolcuların toplu ulaşım yerine şahsi araçlarını tercih etmesi**.** Sefer saatlerinin yolcuların olduğu saatlerde ihtiyaca cevap verememesi, ulaşımda toplu taşımanın yoğun olarak kullanıldığı saatlerde seferlerin yetersiz kalması bunlardan bazılarıdır. Belirtilen bu aksaklıklar ulaşımı olumsuz koşullarda etkilemekte, Yolcuların memnuniyetini azalmaktadır. Ulaşımda yaşanan aksaklıkların fazla olması beraberinde organizasyon açısından aksamalara sebep olmakta ve maliyet açısından fazla bir götürüye sebebiyet vermektedir. Bu çalışmada

süreci etkileyen tüm bu etmenler göz önünde bulundurulduğunda Saatlik toplu ulaşım veri setinde kullanılarak geçiş sayılarının geleceğe yönelik tahminlemesi farklı tahminleme yöntemleri ile tahmin edilmeye çalışılmış ve hata oranı en düşük sonuca ulaşılmaya çalışılmıştır. Elde edilen sonuçlar doğrultusunda yolcu talebine göre sefer sayılarının yetersiz kaldığı noktalar ve yoğun olarak kullanılan saatler göz önünde bulundurularak bu bölgelerde sefer sayılarının arttırılması üzerine öneriler sunulmuştur.

Kasım 2023 , 71 sayfa.

**Anahtar kelimeler:** Ulaşımda Yolcu Talebi, Ulaşım Yoğunluğu, Sefer Sıklığı, Yapay Öğrenme, Optimizasyon

# ABSTRACT

## M.Sc. THESIS

**FORECASTING THE NUMBER OF PASSES AND OPTIMIZATION OF THE NUMBER OF TRANSPORTATION ACCORDING TO ISTANBUL HOURLY PUBLIC TRANSPORTATION DATA**

**Emel GEÇİCİ**

**İstanbul University-Cerrahpaşa**

**Institute of Graduate Studies**

**Department of Computer Engineering**

**Computer Engineering**

**Supervisor : Assist. Prof. Dr. Gülsüm Zeynep GÜRKAŞ AYDIN**

[Problems related to transportation are a problem that is seen and encountered quite intensely in developed and developing countries. Transportation problems, which are visible and at a very important point in most countries, are also seen in our country. Although the problems related to transportation are not specific to a single issue, they appear in many areas. The problems experienced in transportation are the excess number of passengers and the inability to respond to the passenger demand due to the insufficient number of flights. Passengers prefer their personal vehicles instead of public transportation due to the fact that the route arrangement in public transportation is not met according to passenger requests, there are incomplete routes and the frequency of voyages is not sufficient. The inability of the departure times to meet the needs during the hours of the passengers, Inadequacy of the flights during the hours when public transportation is used intensively. These disruptions affect the transportation in negative conditions and decrease the satisfaction of the passenger. The high number of disruptions in transportation causes disruptions in terms of organization and causes a high lump sum in

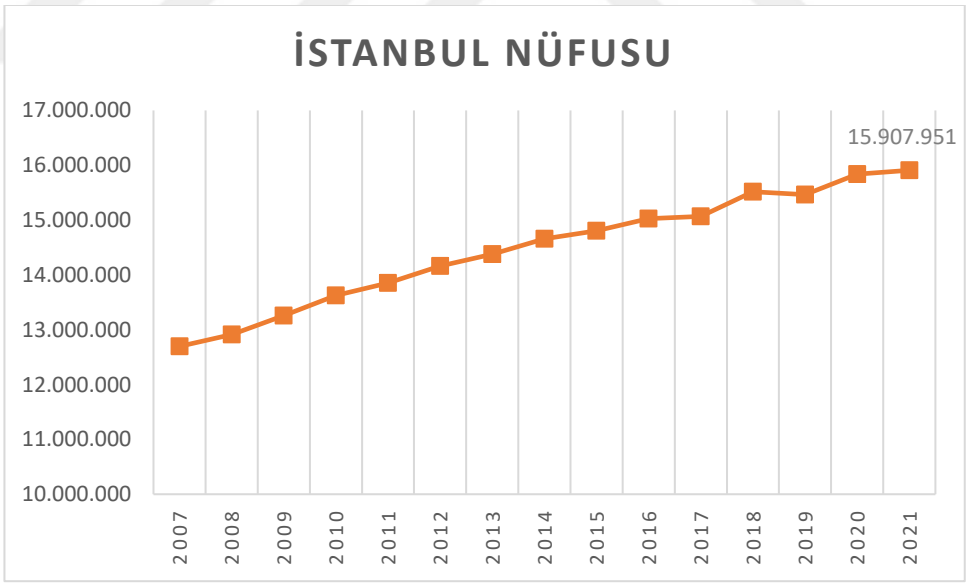
terms of costs. Considering all these factors affecting the process in this study, different estimation methods were tried to be estimated for the future estimation of the number of crossings using the Hourly public transportation dataset, and the lowest error rate was tried to be reached. In line with the results obtained, suggestions were made to increase the number of flights in these regions, taking into account the points where the number of flights is insufficient according to the passenger demand and the hours used intensively.

November 2023, 71 pages.

**Keywords:** Passenger Demand in Transportation, Transportation Intensity, Frequency of Flights, Machine Learning, Optimization

# 1. GİRİŞ

İstanbul, Türkiye'nin nüfusu en yüksek olan şehri olma özelliği göstermektedir. İstanbul'un nüfusu ülkedeki tüm nüfus ile karşılaştırıldığında ülke nüfusunun yaklaşık beşte birlik kısmı İstanbul'da yaşamaktadır. Şehirde sanayi alanlarına bağlı olarak iş imkânının fazla olması, eğitim anlamında ilkokul, ortaokul, lise ve en önemlisi üniversite olanaklarının fazla olması gibi imkânlar şehir nüfusunun gün geçtikçe artmasına neden olmaktadır. Şekil 1.1'de İstanbul'un 2007 yılından sonra yıllara göre nüfusunun değişimi gösterilmektedir. Grafikte de görüldüğü gibi son yıllarda nüfus artış hızında azalma görülse de nüfus sayısı artmaya devam etmektedir. Bu artış başta ulaşım olmak üzere bireysel olarak ihtiyaç duyulan gereksinimlere talebin artmasına neden olmaktadır. Bu da mevcut kaynakların yeterli olmayacağı ve bireylerin ihtiyaçlarının karşılanması için gerekli önlemlerin alınması gerektiği sonucunu ortaya çıkarmaktadır.




**Şekil 1.1.** İstanbul nüfusunun yıllara göre değişimi

Bireylerin hayatlarına devam etmesi ve kaliteli yaşama sahip olmaları için gerekli olan ihtiyaçlar Maslow'un temel ihtiyaçlar hiyerarşisi altında şu şekilde sınıflandırılmıştır: Fizyolojik ihtiyaçlar, güvenlik ihtiyacı, sosyal ihtiyaçlar, değer verilme/saygınlık ihtiyacı ve son olarak kendini gerçekleştirme ihtiyacı (Kavzoğlu ve Çölkesen, 2010). Bu

ihtiyaçlar göz önüne alındığında bireylerin ihtiyaç duyduğu temel şeyler fizyolojik ihtiyaçlar olarak adlandırılmakta ve bireylerin yaşaması için ihtiyaç duyulan gereksinimleri göstermektedir. Bireylerin ihtiyaç duydukları ihtiyaçlara erişmeleri için gerekli olan kaynaklardan biri de ulaşım alanında ortaya çıkmaktadır. Ulaşım, bireylerin ya da toplulukların bir yerden başka bir noktaya erişmesi şeklinde özetlenmektedir. Bu şekilde bir yer değiştirmenin sağlanabilmesi için farklı araçlar kullanılabilmektedir. Karayolu ulaşımı, denizyolu ulaşımı, havayolu ulaşımı ve demiryolu ulaşımı olmak üzere temelde dört ulaşım türü olarak sınıflandırılmaktadır. Şehir içi ulaşım incelendiği zaman bu ulaşım türlerinden daha çok karayolu ulaşımı tercih edilmekte ve şehir yoğunluğunun artmasına bağlı olarak talebin karşılanması için metro ve tramvay gibi demiryolu ulaşım türleri de geçmişten günümüze yaygınlaşmaya devam etmektedir. Şehrin bulunduğu coğrafi konuma bağlı olarak denizyolu ulaşım türleri de kullanılabilmektedir. Havayolu ulaşımı ise henüz bireysel olarak kullanımda teknolojinin yeterli seviyede gelişmemesi ve helikopter gibi mevcut araçların maliyetli olmasından dolayı daha çok şehirlerarası ve ülkeler arası toplu ulaşımların sağlanmasında ya da yük taşımacılığında tercih edilmektedir. İstanbul ilini ele aldığımızda ise şehir içi ulaşımda demiryolu, karayolu ve denizyolu ulaşım türlerinin altında yer alan ulaşım araçları kullanılmaktadır. Bu kapsamda otomobil gibi bireysel araçlar kullanılabildiği gibi toplu taşıma araçları da yaygın olarak ulaşımın sağlanmasında tercih edilmektedir. İstanbul'da toplu taşıma amacıyla kullanılan toplu taşıma araçları Tablo 1.1'de yer almaktadır.

**Tablo 1.1.** İstanbul'da kullanılan toplu taşıma türleri

| | | | | |
|---|---|---|---|---|
| Otobüs | Metrobüs | Füniküler | Vapur | Arabalı Vapur |
| Minibüs | Taksi | Marmaray | Feribot | |
| Metro | Tramvay | Teleferik | Dolmuş | |

Kullanılan toplu taşıma türleri incelendiğinde çok fazla alternatif gözükmesine rağmen nüfus yoğunluğunun fazlalığı hem bireysel araçlarla trafik yoğunluğunun artmasına hem de toplu taşıma araçlarının gerekli talebi karşılamada yetersiz kalmasına neden olmaktadır. Günümüzde özellikle bireysel araçlanmada araç fiyatları ve maliyetlerin artmasına bağlı olarak toplu taşımaya yönelim artmıştır. Çevresel olarak zararın

azaltılmasını sağlamak için toplu taşıma yöntemlerinin desteklenmesi de toplu taşıma kullanımının artmasına neden olan bir başka faktör olmuştur. Bu doğrultuda toplu taşıma talebi her geçen gün artmakta ve var olan kaynakların talebi karşılamada yetersiz kalacağı anlamına gelmektedir. Bu doğrultuda bireylerin ulaşım ihtiyacını karşılayarak talebin sağlanması için bu anlamda gerekli önlemlerin alınması gerekmektedir. Bu çalışmada öncelikli olarak İstanbul Büyükşehir Belediyesi tarafından saatlik olarak paylaşılan yolculuk sayısı (İBB Açık Veri Portalı, 2023) kullanılarak öncelikle gelecek dönemlerdeki toplu taşımayı kullanan yolcu sayısı ve yolculuk sayısı tahmin edilecektir. Bu işlem için makine öğrenmesi yöntemleri olan regresyon, karar ağaçları, rassal orman gibi yöntemler kullanılacaktır. Bu yöntemden elde edilen veriler sonrasında matematiksel modele girdi olarak sunulacaktır. Sonrasında ilçeler için var olan okul sayısı, halk ekmek büfe sayısı, araç sayısı, hane sayısı gibi faktörler ve yolculuk sayıları kullanılarak ilçelerin kümelenmesi ve talebin yoğun olduğu bölgelerin belirlenmesi sağlanacaktır. Sonrasında küme kapsama problemi kullanılarak gelecekteki talebin karşılanması için yoğun olan bölgelerin belirlenmesi sağlanacaktır.

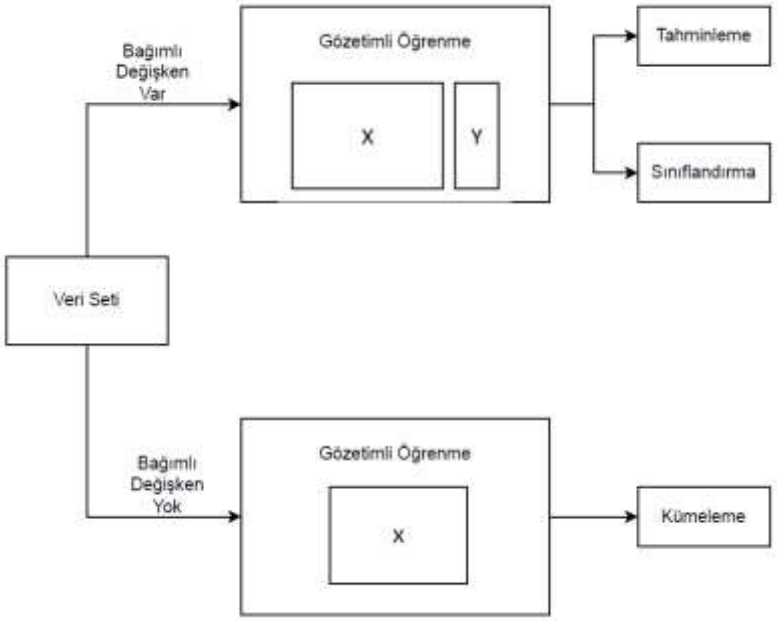
# 2. KAVRAMSAL ÇERÇEVE

Tez çalışmasının bu bölümünde analizler için kullanılacak olan yöntemlerden ve bu alanda yapılan literatür çalışmalarından bahsedilecektir.

Bu kapsamda öncelikle gelecek talebinin belirlenmesi için gerekli olan tahminleme yöntemlerinden bahsedilecektir. Makine öğrenmesi içinde regresyon, k-en yakın komşular, karar ağaçları ve rassal orman yöntemleri kullanılacaktır. Sonrasında ise hangi alanlara atama yapılmasını belirleyeceğimiz matematiksel model aşaması gelecektir. Bu doğrultuda literatürde küme örtme problemi olarak geçen doğrusal programlama yöntemi kullanılacaktır. Sonrasında analizlerde kullanılacak olan makine öğrenmesi ve optimizasyon tekniklerinin literatürde nasıl yer aldığına dair analizler verilecektir.

## 2.1. GÖZETİMLİ ÖĞRENME

Makine öğrenmesi, mevcut verileri kullanarak veriler için en uygun modelin belirlenmesi için kullanılan yöntemlerden oluşmaktadır (Yıldırım ve Demir, 2022). Makine öğrenmesi yöntemleri en temel anlamda ikiye ayrılmaktadır: Gözetimli öğrenme ve gözetimsiz öğrenme. Ortaya çıkan bu ayrım veri setinin yapısındaki farklılığa göre gözlemlenmektedir. Analizler için kullanılacak olan veri seti bağımsız değişken olarak adlandırılan değişkenler ve bağımlı değişkenden oluşuyorsa gözetimli öğrenme teknikleri kullanılmaktadır. Bu tür analizlerde bağımlı değişkenin sürekli ya da kesikli olmasına bağlı olarak sınıflandırma ya da tahminleme işlemleri yapılmaktadır. Ancak veri seti sadece bağımsız değişkenlerden oluşuyorsa ve bu değişkenlere göre gözlem değerleri hakkında bilgi edinilmeye çalışıyorsa bu durumda gözetimsiz öğrenme teknikleri kullanılmaktadır ve en bu alanda en çok bilinen yöntemler kümeleme olarak adlandırılmaktadır. Öğrenme tekniklerindeki bu ayrım Şekil 2.1'de gösterilmiştir. Bu tez çalışmasında gözetimli öğrenme teknikleri başlığı altında yer alan tahminleme yöntemleri kullanılacaktır. Literatürde yer alan temel tahminleme yöntemleri Tablo 2.1'de yer almaktadır.




**Şekil 2.1.** Öğrenme Tekniklerinin Sınıflandırılması

**Tablo 2.1.** Tahminleme Yöntemleri

| | |
|---|---|
| Basit Doğrusal Regresyon | Çoklu Doğrusal Regresyon |
| Polinom Regresyon | K-En Yakın Komşular |
| Ridge Regresyon | Karar Ağaçları |
| Lasso | Rassal Ormanlar |
| Elastic Net | Destek Vektör Makineleri |

Regresyon bağımlı ve bağımsız değişkenler kullanılarak bu değişkenler arasındaki ilişkiyi veya bağlantıyı belirlemede kullanılır. Bir ya da birden fazla bağımsız değişken ve bir bağımlı değişkenin yer aldığı veri setlerine uygulanmaktadır. Regresyon analizi, bağımsız değişkenlerden bir veya birkaçı değiştiğinde bağımlı değişkenin buna bağlı olarak nasıl değişkenlik gösterdiğini belirlemeye yardımcı olmaktadır. Kullanılan bu yöntem sayesinde değişkenler arasındaki ilişki ve neden sonuç bağlantısı bulunmaktadır. Regresyon uygulaması genel olarak bağımsız değişkenlerin sayısına ve bağımlı değişkenler ile arasında kurulan ilişkiye göre farklılık göstermektedir.

Tahminleme yapmak için farklı regresyon çeşitleri kullanılmaktadır. Literatürde kullanılan pek çok regresyon türü bulunmaktadır:

- Doğrusal Regresyon
    - Basit Doğrusal Regresyon
    - Çoklu Doğrusal Regresyon
- Ridge Regresyon
- Lasso Regresyon
- Polinom Regresyon
- Elastic-Net Regresyon

### 2.1.1. DOĞRUSAL (LİNEER) REGRESYON

Regresyonda en çok kullanılan tekniklerden bir tanesi de *doğrusal regresyon*dur. Doğrusal regresyon, seçilen konu özelinde var olan, bilinen değerlerin kullanılıp bilinmeyen değerlerin tahmin edilmesi üzerine yapılan çalışmadır. Bir çeşit veri analizi tekniğidir. Bilinmeyen ya da bağımlı olan değişken ile bilinen ya da bağımsız olan değişken ile işlem yaparak modelleme yapar. Doğrusal regresyon diğer modellere oranla uygulama açısından daha kolay bir yöntemdir. Kullanım açısından çok geniş bir alanı olduğundan makine öğrenmesi ve yapay zekâ gibi pek çok uygulamada kullanılmaktadır.

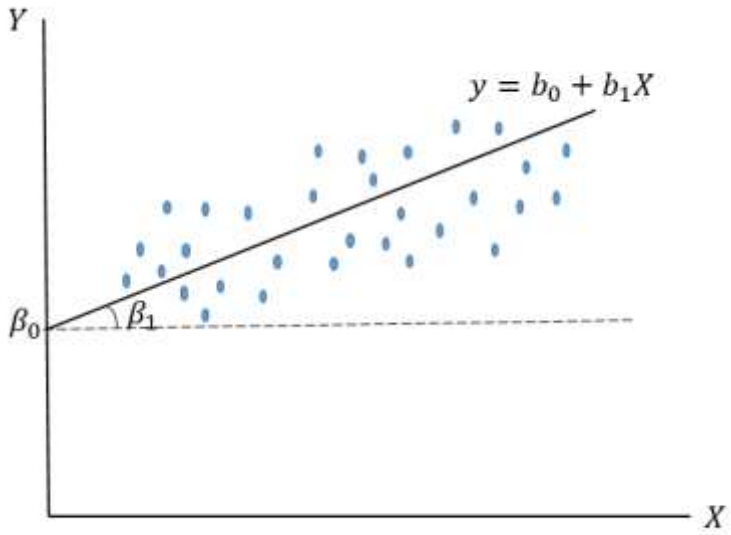
Doğrusal regresyon kendi içerisinde ikiye ayrılmaktadır. Doğrusal regresyon yöntemleri bağımsız değişken sayısına bağlı olarak temelde ikiye ayrılmaktadır: (i) Basit doğrusal regresyon ve (ii) Çoklu doğrusal regresyon. Bağımsız değişken sayısının bir olduğu yani veri setinin sadece bir bağımlı ve bir bağımsız değişkenden oluştuğu analizlerde *basit doğrusal regresyon* modeli kullanılmaktadır. Eğer bağımsız değişken sayısı birden fazla olursa bu durumda regresyon analizi *çoklu doğrusal regresyon* olarak adlandırılmaktadır. Her iki modelde de bağımlı değişken sürekli bir yapıya sahiptir. Örneğin ev satış fiyatının belirlenmesi ya da enerji fiyatlarının tahminlemesi gibi. Ancak her zaman bağımlı değişken sürekli formda olmak zorunda değildir ve kesikli bir yapıya sahip olabilir. Bu durumda regresyon tahminleme yerine sınıflandırma için kullanılacak olup yöntem de lojistik regresyon olarak adlandırılmaktadır.

#### *2.1.1.1. Basit Doğrusal Regresyon*

Bir önceki bölümde de bahsedildiği gibi eğer bağımlı değişkeni açıklamak için kullanılacak değişken sayısı bir ise kullanılan regresyon modeli basit doğrusal regresyon modeli olarak adlandırılmaktadır. Basit doğrusal regresyon modeli bağımsız değişken ile bağımlı değişken arasında bir ilişki varsa bağımsız değişken yardımıyla bağımlı değişkenin tahmin etmesi için kullanılan yöntemdir. Denklem 2.1'de temel basit doğrusal regresyon modeli yer almaktadır.

$$Y = \beta_0 + \beta_1 X + \epsilon$$ **Denklem 2.1**

Denklem 2.2'de yer alan $\beta_0$ denklemin sabitini, $\beta_1$ ise $X$ değişkenin modeli açıklamaktaki katsayısını göstermektedir. $\epsilon$ ise modelde değişkenler ile açıklayamadığımız hata değerini göstermektedir. Şekil 2.2'de yer alan basit doğrusal regresyon grafiği de incelendiğinde formülde yer alan $\beta_0$ kesim noktasını ve $\beta_1$ ise doğrunun eğimini göstermektedir.




**Şekil 2.2.** Basit Doğrusal Regresyon

### *2.1.1.2. Çoklu Doğrusal Regresyon*

Çoklu doğrusal regresyon ise basit doğrusal regresyon modelinde olduğu gibi bir bağımlı değişken içermektedir. Ancak bağımsız değişken birden fazladır. Denklem 2.2'de $p$ adet bağımsız değişkenin yer aldığı bir veri seti için oluşturulan çoklu doğrusal regresyon modeli yer almaktadır.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + .. \beta_p X_p + \epsilon$$ **Denklem 2.2**

Basit doğrusal regresyonda olduğu gibi $\beta_0$ modelde sabiti ifade etmektedir. Bağımsız değişkenlere karşılık gelen katsayısı ise değişken sayısı kadar değeri içeren bir verktörden oluşmaktadır ve bu katsayılar ilgili değişkenlerin bağımlı değişkeni açıklamaktaki etkisini göstermektedir. Özetle, çoklu doğrusal regresyon kullanılan değişken sayısı birden fazla olmakla beraber bunların sonuç üzerindeki etkisini modeller.

Basit doğrusal regresyon ve çoklu doğrusal regresyon analizlerinde $X$ değişkenlerine ait katsayıların bulunabilmesi için hata kareler toplamı (HKT) minimize edilmeye çalışılmaktadır. Kullanılan hata kareler toplamı ise Denklem 2.3'te yer almaktadır.

$$HKT = \sum_{i=1}^{n} \left( y_i - \beta_0 - \sum_{j=1}^{p} \beta_j X_{ij} \right)^2 \qquad \textbf{Denklem 2.3}$$

Çoklu doğrusal regresyon analizinde modele eklenen bağımsız değişken sayısı arttıkça modelin açıklama gücü olarak adlandırılan parametrenin arttığı görülmektedir. Ancak bu durumda modele katkısı çok olan değişkenlerin bazı durumlarda modele eklenmesi maliyet oluşturacağından modelde kullanılmamak istenmektedir. Bu durum regülarizasyon olarak adlandırılmakta ve temelde iki yöntem önerilmektedir: (i) ridge regresyon ve (ii) lasso.

### 2.1.2. RİDGE REGRESYON

Ridge regresyon modelinde çok değişkenli verileri analiz etmede kullanılır. Amaç hata kareler toplamını en aza indiren katsayıları kullanarak, bu katsayılara ceza uygulamak ve bunları bulmaktır. Bu modelde, veri setinde yer alan tüm bağımsız değişkenler modele dahil edilebilir ancak etkisi az olan bağımsız değişkenlerin katsayı değerleri sıfıra yaklaşmaktadır. Bu şekilde değişkenlerin katsayılarını elde etmek için kullanılan amaç değeri ise **Denklem 2.4**'teki gibidir.

$$\sum_{i=1}^{n} \left( y_i - \beta_0 - \sum_{j=1}^{p} \beta_j X_{ij} \right)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 = HKT + \lambda \sum_{j=1}^{p} \beta_j^2 \qquad \textbf{Denklem 2.4}$$

Burada yer alan $\lambda$ değeri hiperparametre olarak adlandırılmaktadır ve model çalıştırılmadan önce tanımlanması gerekmektedir. Farklı $\lambda$ değerleri için farklı sonuçlar elde edileceği ve bu nedenle en uygun modelin bulunması için farklı $\lambda$ değerlerinin denenmesi gerektiğini belirtilmelidir.

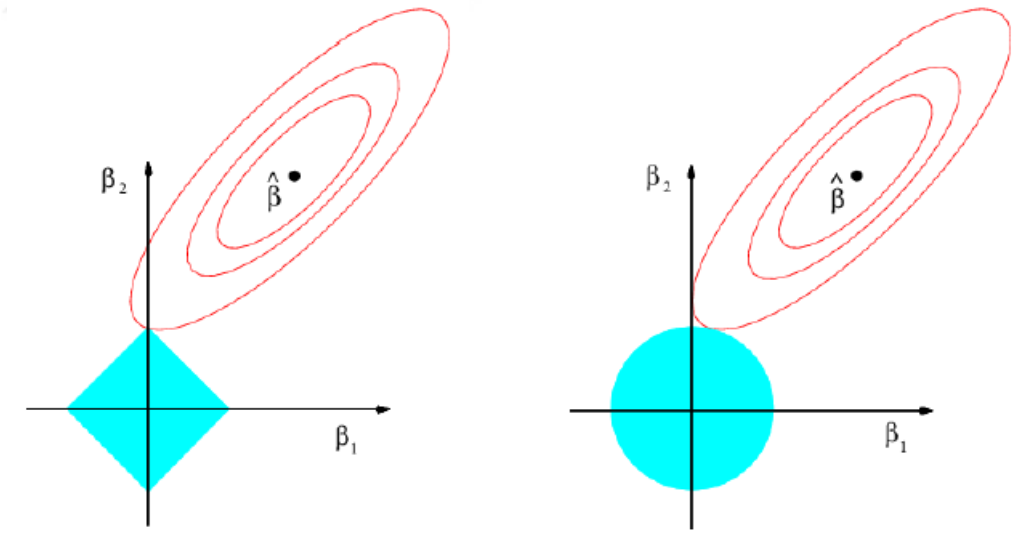
### 2.1.3. LASSO

Ürettiği modelin doğruluğunu arttırmaya dayalı olarak diğer regresyonlardan farkı değişken seçimi yapmasıdır. Rigde regresyonla birbirlerine benzemektedirler. Ancak ridge regresyonda

etkisi az olan değişkenlerin kat sayısı düşürülerek modelde tutulmaya devam edilirken lassoda katkısı az olan değişkenler modelden uzaklaştırılmaktadır. Başka bir deyişle katsayıları 0 yapılmakta ve bu da değişken seçimine olanak sağlamaktadır. Lasso için kullanılan amaç fonksiyonu Denklem 2.5'te yer almaktadır.

$$\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j X_{ij}\right)^2 + \lambda\sum_{j=1}^{p}|\beta_j| = HKT + \lambda\sum_{j=1}^{p}|\beta_j| \qquad \textbf{Denklem 2.5}$$

Ayrıca regülarizasyon için kullanılan ridge regresyon ve lassoda sırasıyla L2 düzenleme ($\lambda\sum_{j=1}^{p}\beta_j^2$) ve L1 düzenleme ($\lambda\sum_{j=1}^{p}|\beta_j|$) yapılmaktadır. Aralarındaki en büyük fark budur. L1 düzenlemesinde katsayıların mutlak değeri alınır. Aralarındaki bu farkı açıklamak için Şekil 2.3 kullanılmaktadır.




(a) (b)

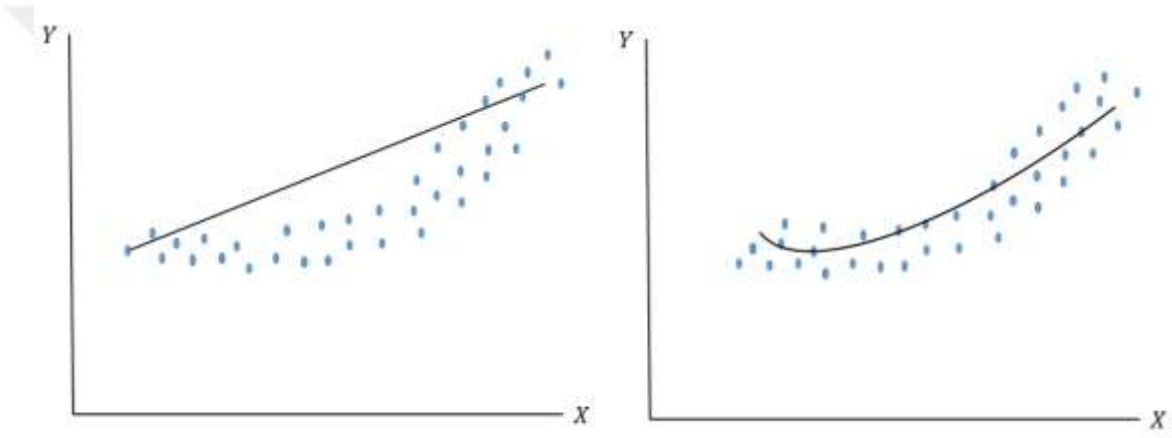
**Şekil 2.3.** Lasso ve Ridgre Regresyonun Karşılaştırılması: (a) Lasso, (b) Ridge Regresyon (James, Witten, Hastie, & Tibshirani, 2013)

### 2.1.4. POLİNOM REGRESYON

Doğrusal regresyonun yeterli olmadığı ya da değişkenler arasındaki ilişki doğrusal olmadığı durumlarda polinom regresyon kullanılır. Polinom regresyona ait Denklem 2.6'da yer almaktadır.

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + .. + \beta_p X^p \qquad \textbf{Denklem 2.6}$$

Doğrusal regresyon ile polinom regresyon arasındaki farkı göstermek için Şekil 2.4 kullanılabilir. Doğrusal yapıda olmayan bir veri seti için polinom regresyonun kullanılması tahminlerin doğru şekilde tahmin edilmesi değerini artıracaktır.

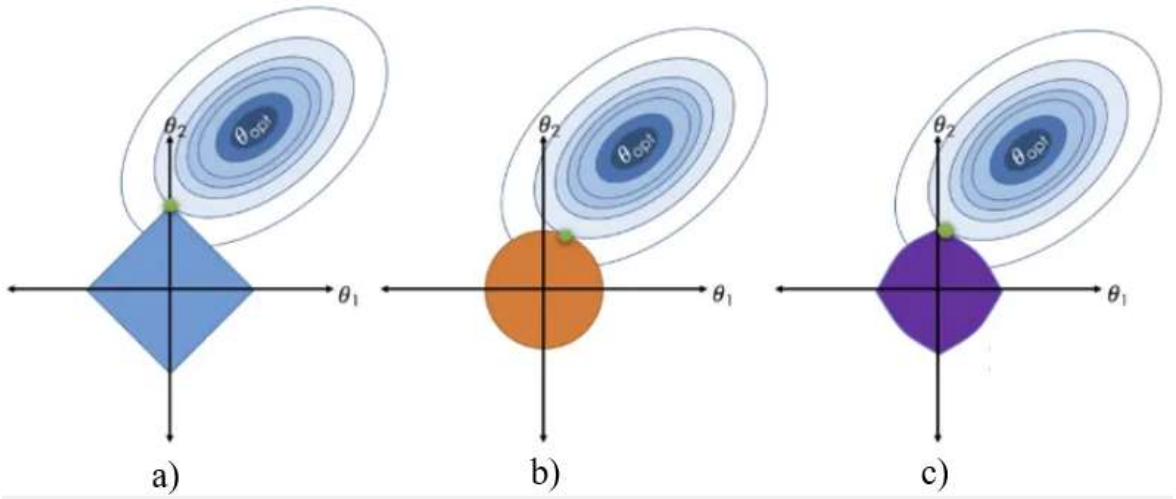



**Şekil 2.4.** Doğrusal Regresyon vs. Polinom Regresyon

### 2.1.5. ELASTİC NET REGRESYON

Elastic net regresyonu ridge ve lasso regresyonlarının birleşmiş yapısıdır. Elastic net regresyonunda $\alpha$ parametrelerine göre tahmin etme işlemi yapılmaktadır. Formülde kullanılan değer $\alpha = 0$ ise ridge regresyon, $\alpha = 1$ ise lasso regresyon kullanılmaktadır. Elastic net regresyonu için Denklem 2.7 kullanılmaktadır. Şekil 2.5 ise rigde regresyon, lasso ve elastic net arasındaki ilişkiyi açıklamak için kullanılabilir.

$$\sum_{i=1}^{n}\left(y_i-\beta_0-\sum_{j=1}^{p}\beta_j X_{ij}\right)^2 + \lambda\left((1-\alpha)\sum_{j=1}^{p}\beta_j^2+\alpha\sum_{j=1}^{p}|\beta_j|\right) \quad \textbf{Denklem 2.7}$$




**Şekil 2.5.** (a) Lasso, (b) Ridge regresyon, (c) Elastic Net (Zou & Hastie, 2005)

### 2.1.6. K-EN YAKIN KOMŞULAR (K-NN) ALGORİTMASI

K-en yakın komşular algoritması (K-nearest neighbor, KNN) en çok kullanılan sınıflandırma algoritmalarından biridir. Ancak sürekli değerlerin yer aldığı regresyon uygulamaları için de kullanılabilmektedir. Tembel bir öğrenme algoritmasıdır. Algoritmanın çalışmasında bir *k* değeri kullanılır.

KNN algoritması kullanılırken öncelikle k parametresi belirlenir. *k*, belirlenen noktaya en yakın komşuların sayısıdır. Belirlenen nokta için *k* kadar komşuya göre işlem yapılır. Var olan veri setine eklenecek olan yeni verilerin belirtilen noktalara göre uzaklığı hesaplanır. Hesaplama yapılırken KNN için Tablo 2.2'da yer alan mesafe hesaplama formüllerinden biri seçilir. En çok tercih edilen uzaklık hesaplama yöntemleri ise Manhattan ve Öklid hesaplamalarıdır. Belirlenen uzaklıklardan en yakın komşuya göre işlem yapılır. Elde edilen değerlere göre *k* komşu veya komşuların sınıfına atanır. Ataması yapılan sınıf istenilen gözlem değeri sınıfı olarak kabul edilir. Böylelikle veri etiketlenmiş olur.

**Tablo 2.2.** Mesafelerin Hesaplanması için Kullanılan Uzaklık Formülleri

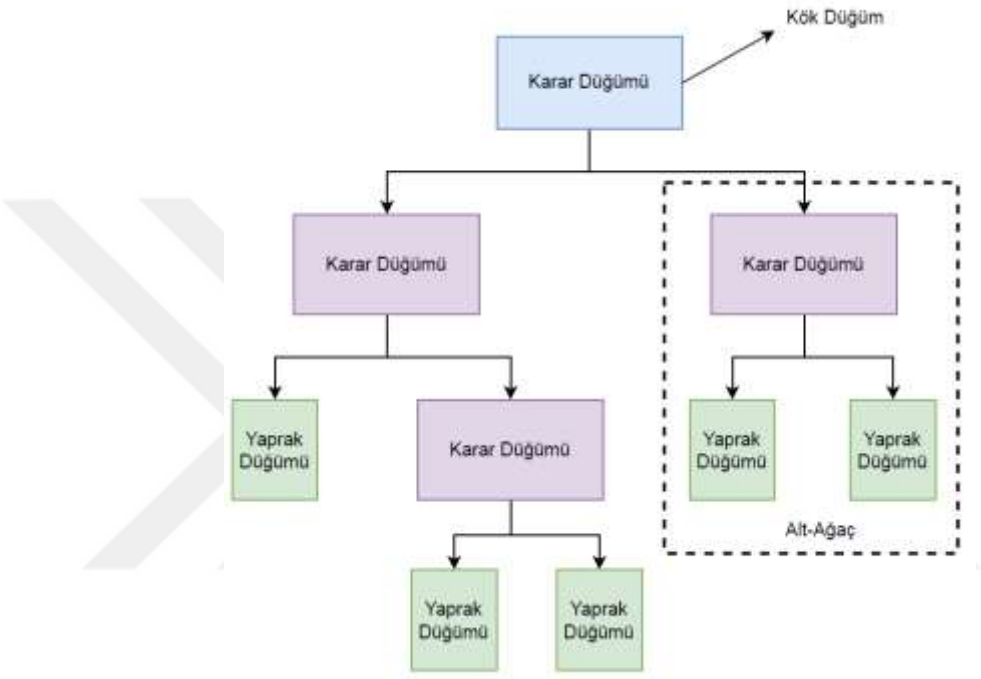
| **Yöntem** | **Formül** |
|---|---|
| Manhattan Uzaklığı | $d(x,y) = \sum_{i=1}^{n} \lvert x_i - y_i \rvert$ |
| Öklid Uzaklığı | $d(x,y) = \sum_{i=1}^{n} (x_i - y_i)^2$ |
| Minkowski Uzaklığı | $d(x,y) = \left( \sum_{i=1}^{n} \lvert x_i - y_i \rvert^p \right)^{1/p}$ |
| Chebyshev Uzaklığı | $d(x,y) = \max \lvert x_i - y_i \rvert$ |
| Mahalanobis Uzaklığı | Bir p noktası ile D dağılımı arasındaki mesafenin ölçülmesi için kullanılmaktadır. |
| Haversiene Uzaklığı | Enlem ve boylam değerlerini kullanarak bir küre üzerinde yer alan iki nokta arasındaki mesafenin belirlenmesi için kullanılmaktadır. |
| Sørensen-Dice Uzaklığı | Belirlenen iki nesnenin birbirine benzerliğini ölçmek için kullanılan hesaplama tekniğidir. |
| Jacckard Uzaklığı | Seçilen iki nokta/kümenin birbirine benzerliğini ölçmektedir. |

### 2.1.7. KARAR AĞAÇLARI

Regresyon ve sınıflandırma için birçok metot kullanılmaktadır ve bu yöntemlerden bir tanesi de karar ağacı algoritmasıdır. Karar ağacı olası sonuçları elde etmek için modelleme yaparken ağaç yapısına benzer bir şekil aldığı için bu şekilde adlandırılmaktadır. Karar ağaçları regresyon ve sınıflandırma da çok kullanılmasıyla birlikte yöneylem araştırmalarında da kullanılmaktadır (Kavzoğlu ve ark., 2012).

Karar ağaçları, parametrik olmayan gözetimli öğrenme algoritmasıdır (Kingsford ve Salzberg, 2008). Kök düğüm, dallar, iç düğümler ve yaprak düğümlerden oluşa bir yapıya sahiptir. Elde edilen şekildeki her dal modellemenin sonucunu temsil etmektedir. Böylece hiyerarşik yapıya sahip modeller oluşturulmaktadır. Ayrıca karar ağacı algoritması tipik bir tümevarım yaklaşımıdır. Karar ağacı algoritmaları başlangıç olarak kök olan ilk düğümden başlar ve Kök düğümden yola çıkılarak devam edilen aşamada gelen dallar bulunmaktadır. Buradaki dallar

daha sonraki adımlarda kullanılacak olan karar düğümleridir (Kingsford ve Salzberg, 2008). İç düğümlerden sonra gelen adımda ise yapraklar olarak da bilinen uç düğümler gelmektedir. Karar ağacının sınıflandırılmasında herhangi bir parametre düzenlemesi, alan bilgisi gerekmemektedir (Kingsford ve Salzberg, 2008). Ağaç yapısında en uç kısımda bulunan yaprak düğümler kullanılan veri setinin kullanılmasıyla beraber elde edilebilecek olası sonuçlara karşılık gelmektedir. ’da karar ağacında düğümlerin nasıl göründüğü örneklendirilmiştir.




**Şekil 2.6.** Karar Ağaçları Genel Yapısı (Java T Point, 2023)

Karar ağacı algoritmasında sonuca veri modellemeyi tamamlamak için böl ve fethet mantığı benimsenmektedir. Bu yöntemin kullanılmasının sebebi ise kök düğümden sonra devam edecek olan adımlarda en uygun dallanma ve bölünmeyi elde edebilmek içindir. Kullanılan veri setinin elde edilecek dallar özelinde homojenliğin elde edilip edilememesi durumu karar ağacının karmaşık olmasına bağlıdır. Veri setleri ile oluşturulacak ağaçlar küçük olur ise sonuca ulaşmak kolay olabilir. Karmaşık yapının artmasıyla elde edilen sonuçlarda sapmalar yaşansa da büyük veri setleri için de kullanılmaktadır. Ancak veri setinin büyümesi oluşturulacak ağaç modelinin dallanmasına ve karmaşıklığa ulaşmasına sebebiyet vereceği için sonuca ulaşmak daha zor bir hal alabilmektedir. Bu durumda ağaç karmaşıklaştıkça elde edilen sonuçlar istenildiği gibi olamadığı için uydurmayı önlemek adına *budama* yöntemi kullanılmaktadır. Bu işlem sayesinde önemsiz veya sonuca etkisi az olacak olan dallar ağaçtan çıkarılmaktadır. Bu yöntem için var olan avantaj ve dezavantajlar ise aşağıdaki gibi sıralanabilir. Avantajları;

- Karar ağacı yönteminde kurallar istenildiği şekilde üretilebilir.
- Çok fazla sınıflandırmaya gerek kalmadan sınıflandırma yapılabilmektedir.
- Tahmin ve sınıflandırma da önemli noktaları net bir şekilde belirleyebilir.
- Kullanım kolaylığı sunmaktadır. Kullanımı basit olmakla beraber oldukça geniş bir alanda kullanılabilmektedir.
- Veri setinin büyümesiyle beraber karar ağaçları karmaşıklaşmakla beraber bu tip veri setleri için kullanıma uygundur.
- Doğrusal olmayan ilişkilere sahip olan veri setleri için uygun bir seçim haline gelebilir.

Dezavantajları;

- Veri setlerinin büyütülmesi karar ağaçlarını karmaşıklaştıracağı için maliyeti yükseltebilmektedir.
- Veri setinin büyümesiyle beraber budama işlemlerinin de yapılması maliyeti arttırabilmektedir.
- Birçok sınıfın olması sebebiyle sınıflandırma işlemleri yapılırken hata oranında artma görülmesi yüksektir.

### 2.1.8. RASTGELE ORMANLAR ALGORİTMASI

Rastgele ormanlar algoritması birden fazla karar ağacının kullanımına dayalı bir algoritmadır. Karar ağaçlarında doğruluğun korunabilmesi adına rastgele ormanlar algoritması kullanılabilir. Özellikle ağaçların birbiriyle bağlantısı veya ilişkisi yok ise sonuçları daha doğru bir şekilde tahmin edebilmektedir. Birden fazla karar ağacı kullanılarak doğruluk oranı daha yüksek modeller üretilmeye çalışılır. Kullanılan bu modeller ile tahmin yapılması sağlanır. Ayrıca üretilen bu modeller sınıflandırma yapılmasını sağlamaktadır. Farklı modeller üretmesi sebebiyle kullanılan veri setlerini yeniden kullanıp farklı tahminler yapılmasını sağlamaktadır.

Karar ağaçlarının kullanılması sebebiyle kesikli bir yapıya sahiptir. Bu sebeple belirlenen aralıktaki istenen tahminler aynı sonucu üretmektedir. Rastgele ormanlar algoritması regresyon ve sınıflandırma problemlerinde kullanılmaktadır. Karar ağaçları ile rastgele ormanlar algoritması arasındaki en önemli fark rastgele ormanlar algoritmasındaki kök düğümü bulma ve düğümlerin bölünmesi işleminin rastgele olmasıdır.

Rastgele ormanlar algoritmasında öncelikli olarak veri seti hazırlanır. Problemin türüne göre veri setinde düzenleme yapılabilmektedir. Algoritmada veri setine bağlı olarak kullanılan her

bir yöntem için karar ağacı oluşturulur ve oluşturulan her bir karar ağacı için tahmin sonucu elde edilir. Elde edilen tahmin sonucunda oylama işlemi gerçekleştirilir. Yapılan oylama işlemi her değer için geçerlidir. Sınıflandırma problemi için mod, regresyon problemi için ortalama kullanılarak gerçekleşir. Sonuç olarak ise en çok oylanan değer üzerinden ilerleyerek tahmin etme işlemini gerçekleştirir ve sonuç oluşturur.

Yöntemin avantajları ve dezavantajları aşağıdaki gibi sıralanabilir (Arıkan ve Yıldız, 2023)

- Çok yönlü ve birden fazla modelleme ile tahminleme yapabildiği için sınıflandırma ve regresyon problemlerinde kullanılmaktadır.
- Çok fazla modelleme yapabilmesine karşın fazla olan ağaç sayısı sebebiyle yavaş sonuç verebilmektedir. Veri setinin yüklenmesi hızlı, tahmin süresi uzun sürebilmektedir.
- Önemli olan parametrelerin ve özelliklerin belirlenmesinde kolaylık sağlamaktadır.
- Büyük veri setleri için başarı oranı yüksek tahmin sonuçları elde edebilmektedir.
- Bilinmeyen verilerin olması durumunda eksik veri işlemede olumlu sonuçlar elde edebilmektedir.
- Kurallara bağlı kalarak yaklaşım önerdiği için verilerin normalleştirilmesine gerek olmamaktadır.
- Ancak veriler çok seyrek olduğunda iyi sonuçlar üretmemektedir.
- Tanımlamaktan ziyade tahmine dayalı bir modelleme aracıdır.
- Model yavaş olduğu için doğru bir tahmin sonucuna varılabilmesi için çok sayıda ağaca gerek duyulmaktadır.

### 2.1.9. DESTEK VEKTÖR MAKİNELERİ

Destek vektör makineleri (Support Vector Machines – SVM) genellikle sınıflandırma problemlerinde kullanılmaktadır. Sınıflandırma problemleriyle birlikte regresyon analizi için de kullanılmaktadır (Medium, 2023; Kavzoğlu ve ark., 2012). Kullanılan veri setleri için her zaman doğrusal modeller kullanılamamaktadır. Doğrusal modeller kullanılmadığı durumlarda destek vektör makinesi gibi vektör algoritmalarını kullanabilir (Demirezen ve Çetin, 2021). Doğrusal vektör makinelerinin ilk çıkış amacı sınıflandırmadır ve SVM olarak bilinmektedir. SVM algoritması, iki sınıfı birbirinden ayırır. Diğer algoritmalardan farklı olarak amaç fonksiyonu maksimizasyon olarak çalışmaktadır. Yapılan ayrımı ise sınırdaki elemanlara göre

yapmaktadır. Yapılan bu ayrımı eldeki verilerin paraleline en uzak olan çizgiyi çizerek yapmaktadır. Bu çizgi elde edildikten sonra eldeki veriler ve gelecek olan veriler için sınıflandırma yapabilir. Grafikte bulunan paralel aralığın kestiği noktalar destek noktalarıdır. Kullanılan veri setleri küçük ve karmaşık olabilir. Bu tipteki veri setlerini tahmin etmede kullanmak için uygun olan bir algoritmadır. Biyoloji, görüntü işleme gibi pek çok bilim alanında verilerin sınıflandırılması için kullanıldığı görülmektedir.

## 2.2. YAPAY SİNİR AĞLARI

Modelleme için kullanılan yöntemlerden bir tanesi de yapay sinir ağları (YSA)'dır. Yapay sinir ağları çalışma prensibi olarak insan beyninden yola çıkılarak geliştirildiği için bu şekilde anılmaktadır. YSA ile biyolojik sinir sistemi en basit haliyle taklit edilmektedir (Öztürk ve Şahin, 2018). Nöron hücrelerimizin ve bu hücrelerin kendileri arasında kurduğu sinaptik bir bağ vardır. YSA için bu sinaptik bağın modellenmesidir denilebilir. Nöronlar iletişim halinde bağlanarak ağlar oluştururlar. Bu ağlar öğrenme, hafıza ve veriler arasındaki ilişkiler gibi pek çok yeteneği ortaya çıkarma kapasitesine sahiptirler. YSA gözlem yapabilme, düşünme gibi insana ait yetenek gerektiren özelliklere sahip olması sebebiyle problemlere çözüm getirebilmektedir. YSA'da öğrenme işlemi yapılırken mevcutta bulunan örnekler kullanılmaktadır (Öztürk ve Şahin, 2018). Öğrenme sırasında giriş çıkış bilgilerinin durumuna göre kurallar belirlenir ve bu şekilde çalıştırılır. YSA, veri setine göre öğrenme işlemi gerçekleştirirken dışarıdan yardım almamaktadır (Ataseven, 2013). Makine bu işlemi kendi kendine öğrenerek gerçekleştirmektedir. YSA'da modelleme yapılırken giriş ve çıkış parametrelerinin arasındaki ilişkiyi ifade etmek için $y = f(x) + Hata$ formülü kullanılmaktadır. $x$ giriş, $y$ ise çıkış değişkelerini belirtmektedir. $f(x)$ ise çıkış değerleri üretmek için kullanılan fonksiyonu ifade etmektedir. Makine öğrenmesi algoritmalarında olduğu gibi amaç hatayı azaltmaktır ve bilgiler bu doğrultuda ayarlanmaktadır. Yapay sinir ağları kullanımı sırasında gizli düğüm ya da düğümler olabilir. Giriş çıkışların arasında saklı bulunan bir katman ya da modelle ilgisi bulunmayan katmanlar olabilir. Bu gizli katmanlar yapay sinir ağlarının anlaşılması için yardımcı olmaktadır.

Bilindiği gibi YSA'lar nöronların çalışma prensibinden yola çıkılarak elde edilmiş bir yöntemdir. Hataları düzeltmek ve gelişmek üzerine kurulu bir sistemdir. Yüz tanıma özet çıkarma gibi karmaşık problemleri çözüp hata oranı düşük olan şekilde doğru sonuca ulaşmaya çalışır (Öztürk ve Şahin, 2018). Doğrusal olmayan karmaşık verileri girdi ve çıktıları

doğrultusunda modelleyebilmektedir. YSA'larda aynı anlama gelen farklı ifadeleri birbirinden ayırt edebilirler ve aynı anlamda olduklarını belirtip çıkarım yapabilirler. Doğru sonuca ulaşma açısından düşük hata oranına sahip oldukları için pek çok alanda kullanılmaktadırlar. YSA sınıflandırma, analiz, tahmin etmenin yanı sıra filtreleme, teşhis etme, kontrol etme gibi alanlarda da kullanılırlar. Hangi ağın hangi problemde kullanılacağını belirlemek için giriş çıkış parametrelerine ve problemin neler istediğine bakmak ve istenilenleri karşılaştırmak gerekmektedir.

Genel olarak kullanım alanlarını aşağıdaki şekilde sıralayabiliriz:

- Tıbbi işlemler için, görüntü sınıflandırma, tıbbi tanılama,
- Veri analizi aracılığıyla pazarlama sektöründe,
- Verileri işleyerek finansal tahminleme işlemlerinde,
- Enerji sektörü ve talep tahmininde,
- Kalite kontrolünde,
- Kimya sektöründe.

Yapay Sinir Ağları ileri beslemeli ve Geri Beslemeli olmak üzere ikiye ayrılmaktadırlar.

### 2.2.1. İleri Beslemeli Ağlar:

İleri Beslemeli YSA'larda eldeki verileri doğrultusunda aktarılan bilgiler ileri taraflıdır (Ataseven, 2013). İleri doğru hareket etmektedir. Veri akışı girişten çıkışa doğrudur. Bir katmanda bulunan nöronlar sadece kendinden sonraki katmana gider. Sonunda ise sonuç-çıkış katmanına gider. Giriş katmanında bulunan bilgiler YSA tarafından herhangi bir değişikliğe uğramadan çıkış katmanına iletilir. Girişte kullanılan veri çıkış nöronuna yönlendirilir.

### 2.2.2. Geri Beslemeli Ağlar:

Geri Beslemeli YSA'larda geciktirme işlemi eleman üzerinden yapılmaktadır (Ataseven, 2013

). YSA'da en az bir hücrenin çıkışı diğer bir hücrenin girişi olacağı için gecikme bu sebeple elemanlar üzerinden yapılmaktadır. Geri besleme işlemi katmandaki hücreler arasında olamayabileceğinden doğrusal bir ilişkiden söz edilemez. Bu sebeple veri setinin farklılığına göre Geri Beslemeli YSA değişkenlik gösterebilir.

YSA çok katmanlı yapılara sahip olabilmektedir. Bu şekilde katmanlı olan YSA'larda momentum katsayısı bulunur. Momentum katsayısı ağların yerel sonuçlarda engele takılmaması için kullanılır. Bu katsayı hata oranını manipüle edip aşağı çekebilir. Ancak

bilgisayarın gücü de bunun için önemli bir etmendir. Momentum katsayısı ile öğrenme oranı arasında da bir ilişki bulunmaktadır. Çok katmanlı olan bir YSA modelinde öğrenme oranı ile momentum işlem süresini uzatabilir. Momentum değerinin yüksek ya da düşük olması işlemleri etkileyeceği için belirlenirken dikkat edilmelidir. YSA modelleme işlemi yapılırken fazla bir oran girilip öğrenme işlemi yapılmaya çalışılırsa model öğrenmekten ziyade ezberleme işlemi yapar. Bu da hatalı sonuç elde dilmesine sebep olur.

YSA'lara ait avantajlar aşağıdaki şekilde sıralanabilir (Öztürk ve Şahin, 2018):

- Eş zamanlı olarak çalışır ve birçok hücrenin birleşmesiyle meydana gelir ve karmaşık problemler için oldukça elverişlidir,
- Öğrenme yeteneği sayesinde farklı algoritmalar üzerinde de çalışma yapabilirler,
- Daha önce elde edilmemiş çıktılara ulaşabilir. Gözetimsiz öğretim yaparlar,
- Sınıflandırma yapabilirler,
- Eksilikleri tamamlayabilirler,
- Eksik veya belirsizi bir bilgi olması durumunda da çalışabilirler. Hata toleransına sahiptirler,
- Gerçek zamanlı bilgiyi işleyebilirler ve paralel çalışabilirler.

## 2.3. GÖZETİMLİ ÖĞRENME ALGORİTMANLARININ KARŞILAŞTIRILMASI

Uygulanan algoritmalardan hangisinin var olan veri setine uygun olduğunu belirlemek için algoritmaların karşılaştırılması gerekmektedir. Bunun için de performans göstergeleri kullanılmaktadır. Regresyon problemlerinde modellerin karşılaştırılması için performans göstergeleri Tablo 2.3'te yer almaktadır.

**Tablo 2.3.** Regresyon Problemlerinde Algoritmaların Karşılaştırılması için Kullanılan Performans Göstergeleri

| **Yöntem Adı** | **Formülasyon** |
|---|---|
| Hata Kareler Ortalaması ($MSE$) | $\frac{1}{n}\sum_{i}(y_i - \hat{y}_i)^2$ |
| Hata Kareler Ortalaması Karekökü ($RMSE$) | $\sqrt{\frac{\sum_i (y_i - \hat{y}_i)^2}{n}}$ |

**Tablo 2.4. (Devamı)** Regresyon Problemlerinde Algoritmaların Karşılaştırılması için Kullanılan Performans Göstergeleri

| | |
|---|---|
| Ortalama Mutlak Hata ($MAE$) | $\frac{1}{n}\sum_{i}\lvert y_i - \hat{y}_i\rvert$ |
| Ortalama Mutlak Yüzde Hatası ($MAPE$) | $\frac{1}{n}\sum_{i}\left\lvert\frac{y_i - \hat{y}_i}{y_i}\right\rvert$ |

Yukarıda belirtilen yöntemler kullanılarak gelecek yolculuk sayılarının tahminlemesi yapılacaktır. Buna ilaveten veri setinde yer alan hatlar arasından çalışılacak bölgelerin belirlenebilmesi için gözetimsiz öğrenme tekniklerinden biri olan kümeleme algoritması kullanılacaktır.

## 2.4. GÖZETİMSİZ ÖĞRENME: KÜMELEME

Makine öğrenmesi altında yapılan bir diğer çalışma türü ise gözetimsiz öğrenme teknikleridir. Gözetimsiz öğrenme teknikleri Şekil 2.1'de belirtildiği gibi veri setinde ulaşılmaya çalışılan bir bağımlı değişkenin yer almadığı veri setleri için kullanılmaktadır. Bu algoritmaların kullanılmasındaki temel amaç veri setinde yer alan gözlem değerleri arasındaki örüntüyü bulmak başka bir deyişle benzer olan gözlem değerlerini bir araya getirerek yorumları bu kümelere göre yapmayı sağlamaktadır. Literatürde çeşitli kümeleme algoritmaları yer alsa da en çok bilinen yöntemler k-ortalamalar ve hiyerarşik kümeleme yöntemleridir. İlk yöntem olan k-ortalamalar yönteminde belirlenen k sayıda kümenin belirlenmesi sağlanırken hiyerarşik kümelemede veri setinin ayrılabileceği küme sayısı belirlenmektedir. Bu doğrultuda birbirinden farklılaşmaktadırlar. Bu tez çalışmasında çalışılacak hatların belirlenmesi için seçilen kümeleme yöntemi k-ortalamalar yöntemidir. K-ortalamalar yöntemi için aşağıdaki adımlar takip edilerek kümelerin belirlenmesi gerçekleştirilmektedir (James, Witten, Hastie, & Tibshirani, 2013).

1. $C_1, C_2, \ldots, C_K$ ile her kümedeki gözlemlerin indekslerini içeren kümeler oluşturulur.

2. Her bir gözlem değeri en az bir kümede olacak şekilde $C_1 \cup C_2 \cup \ldots \cup C_K = \{1, \ldots, n\}$ olarak atanır. Burada n toplam gözlem sayısını ifade etmektedir.

3. Gözlem değerleri en fazla bir kümede yer alabilirler, $C_k \cap C_{k'} = \emptyset \; her \; k \neq k'$ için.

Yukarıdaki adımlar kullanılarak küme içindeki farklılıklar minimize edilecek şekilde kümeler oluşturulur. Bu minimizasyon için $\min_{C_1,C_2,\ldots,C_K} \left\{ \sum_{k=1}^{K} \frac{1}{|C_k|} \sum_{i,i' \in C_k} \sum_{j=1}^{p} \left(x_{ij} - x_{i'j}\right)^2 \right\}$ amaç fonksiyonu kullanılmaktadır. Belirtilen fonksiyon minimize edildikten sonra kümeler elde edilmektedir. Bu işlemlerin çalıştırılabilmesi için modele önceden k değerinin verilmesi gerekmektedir. Küme sayısının belirlenebilmesi için dirsek grafiği yöntemi kullanılmaktadır. Dirsek grafiği her bir küme sayısı için yukarıdaki amaç fonksiyonu değerini tutarak x-ekseninde küme sayısı y-ekseninde amaç değeri olacak şekilde oluşturulan grafiklerdir. Bu grafiklerde küme sayısının değiştiği noktalarda değişim oranının azalmaya başladığı kısım küme sayısının belirlenmesine olanak sağlamaktadır.

## 2.5. MAKSİMUM KÜME ÖRTME PROBLEMİ

Tez çalışmasında kullanılan son yöntem ise hatlar için artırılması gereken bölgelerin belirlenmesi amacıyla kullanılan maksimum küme örtme modelidir. Bu yöntemde en çok talebi kapsayacak şekilde yerlerin belirlenmesi işlemi gerçekleştirilmektedir (Current ve diğerleri, 2002). Bu yönteme ait model aşağıdaki gibidir.

$$max \sum h_i z_i \qquad \textbf{Denklem 2.8}$$

$$\sum_{j \in N_i} x_j - z_i \geq 0\,, \qquad i \in I \qquad \textbf{Denklem 2.9}$$

$$\sum_{j=1}^{n} x_j = p \qquad \textbf{Denklem 2.10}$$

$$z_i, x_j \in \{0,1\}, \qquad i \in I, j \in J \qquad \textbf{Denklem 2.11}$$

Bu modelde yer alan $h_i$ talebi ve $z_i$ ise hizmet alan talep noktasını $(I)$ göstermektedir. $x_j$ise talep noktalarına hizmet etmek için açılan tedarik noktalarını $(J)$ göstermektedir. Belirtilen $z_i$

ve $x_j$ modelin değişkenlerini oluşturup 0 ya da 1 değerlerini almaktadır. Bu kapsamda model incelendiği zaman 0-1 tam sayılı programlama olarak adlandırılmaktadır. Denklem 2.9 ile $z_i$'nin eğer $x_j$ tesisi açılırsa talebinin karşılanabileceğini göstermektedir. Denklem 2.10 ise kaç adet tesis açılması gerektiğini ifade etmektedir. 2.8 ile belirtilen amaç fonksiyonunda açılacak tesis yerlerinin belirlenmesinde ulaşılabilecek talep noktasının maksimize edilmesi sağlanmaktadır.

Tahminleme ve optimizasyon için kullanılan ve kullanılmaya devam edilen bu yöntemler araştırmalarda sonuca ulaşılmak için kullanılmaya devam etmektedir. Veri setine uygunluk ve elde edilmek istenilen sonuca bu yöntemler akademik çalışmalarda yer almaktadır.

Bir sonraki bölüm olan literatür taramasında başlığı altında ise yukarıda yöntemler de dahil olmak üzere makine öğrenmesi, optimizasyon teknikleri ve bunlara ek olarak önerilen yöntemlerin ulaşım alanında rotaların belirlenmesi gibi çalışmalardaki uygulamaları sunulacaktır.

## 2.6. LİTERATÜR TARAMASI

Ulaşım bir yolcunun bir yerden başka bir yere ulaşması için kullandığı bütün araçları içeren taşımacılık sistemlerini kapsamaktadır. Günümüzde artan trafik yoğunluğu ve ulaşım problemlerinden dolayı yolcular genellikle toplu taşıma araçlarına yönelmektedir. Ancak yoğunluğun fazla olmasından dolayı ulaşım kaynakları gerekli ihtiyacı karşılayamaz hale gelmektedir. Durum böyle olunca konu ile ilgili mevcut kaynakların en iyi şekilde kullanılması için çalışmalar yapılmakta ve gerekli olması durumunda ise geliştirme için öneriler oluşturulmaktadır. Araştırmacılar da bu konuda çalışmalar yapmakta ve literatürde bu konuyla ilgili çalışmalar yer almaktadır. Bu konuda yapılan çalışmalarda sefer sıklığı optimizasyonu, rota optimizasyonu, yeni rotaların tahminlemesi, sefer sıklığının artırılması, ulaşım tercihleri ile ilgili çalışmaların olduğu görülmektedir.

Yaylalı ve Dilek yapmış oldukları çalışmada hava yolu ulaşımının kullanımını incelemişlerdir (Yaylali & Dilek, 2009). Erzurum'da ulaşım için kullanılan hava yolu talebini incelemişler ve talep olması durumunda bu bölgeye yeni hatlar açmak isteyen havayolu şirketleri olabileceğinden bahsetmişlerdir. Ceylan ve Özcan ise kent içi otobüs ulaşımında sefer sıklığı için kullanılan algoritmalardan yöntemlerden en iyi sonucu alabilmek için iki seviyeli bir simülasyon modeli önermişlerdir. (Ceylan & Özcan, 2018). Önerdikleri modelde armoni

araştırması optimizasyon tekniğidir. Bu modelde aktarma sürelerinin kesin tespiti üzerine çalışmışlardır. Bunun için Logit tabanlı bir model ve VISUM yazılımını kullanmışlardır. Armoni araştırması (AA) tabanlı optimizasyon için problemin çözümünde başarılı sonuçlar elde edilebileceğini göstermişlerdir. Özcan ise kent içi toplu taşıma sistemlerinde sefer sıklığı optimizasyonu ile ilgili çalışma yapmıştır (Özcan, 2018). Yapmış olduğu bu çalışmada Ceylan ve Özcan (Ceylan & Özcan, 2018)'da olduğu gibi armoni optimizasyonu ile bir model geliştirmiştir. Performans değerlendirmesi için ise ağ tasarımı üzerinde yapılan çalışmalara uygun test ağı geliştirmiştir. Çorlu otobüs ağı üzerinde yapılan örnek bir çalışma ile modelin etkinliğini ortaya koymuştur. AA tabanlı modelin problem çözümünde etkin olduğunu ve kalkış zamanı öteleme parametresi ile modelin faydalarının artacağını belirtmiştir.

Otobüs ve sefer sıklığının optimize edilmesi ile ilgili çalışmalar ulaşım açısından önemli bir noktada bulunduğu için konu ile ilgili çalışmalar oldukça fazladır. Bununla ilgili çalışma yapan Demirkollu çalışma sefer sayılarının tespiti üzerine çalışmış ve sefer sıklıkları ile ilgili öneride bulunmuştur (Demirkollu, 2017). Bu çalışmada, Denizli ilinde bulunan otobüslerin sefer sıklığını incelemiştir. Gün içerisinde otobüs kullanımının zirve olduğu saatlerde seferlerin yolcuların ihtiyaçlarına göre hesaplanması gerektiğini belirtmiştir. Bu doğrultuda sefer sıklıklarını doğrusal en iyileme mantığına göre yeniden hesaplamıştır. Sadece Denizli değil farklı illerde de yolcu sayısı ve araç yetersizliği üzerine çalışmalar yapılmıştır. Deri ise çalışmasında akıllı kart verilerini kullanmıştır (Deri, 2012). Bu verileri kullanarak İzmir'de toplu ulaşımda yolculuk sefer sayıları üzerine çalışmıştır. Yaptığı çalışmaya göre sabah zirve saatlerinde kent merkezine gidiş yönünde talebe karşılık gelecek en uygun otobüs sefer sıklığını doğrusal hedef programlama yöntemiyle hesaplamıştır. Ulaşımda yolcu talebine göre sefer sıklığının düzenlenmesi yolcu memnuniyeti ve hizmet oranının yükseltilmesi oldukça önem arz etmektedir. Bu sebeple aynı ya da yakın bölgelerin seçilmesi ve araştırılması, çalışmalara konu olmaya devam etmiştir. İzmir'deki ulaşımla ilgili yapılan çalışmalarından bir diğeri ise Uludağ'a aittir (Uludağ, 2010). Uludağ, çalışmasında bulanık optimizasyon kullanarak otobüs hatlarının modellenmesi üzerine çalışmıştır. İncelemek için İzmir'in Lozan ve Montrö ilçeleri için çalışma yapmış ve iki durak belirlemiştir. Bu iki bölgenin aktif olarak kullanıldığını özellikle sabah saatlerinde yoğunluğun fazla olduğunu ifade etmiştir. Bu hatlara ait yolcu talepleri, yolculuk süreleri, araçların kapasiteleri gibi pek çok parametreyi göz önünde bulundurmuştur. Doğrusal programlama modelini kullanarak yolcuların araç içinde geçirdiği sürenin en aza indirilmesi için çalışma yapmıştır. Bu doğrultuda seçtiği bölgelerdeki otobüs ve sefer sıklıkları ile ilgili ulaşım hizmetlerinin artırılması gerektiğini belirtmiştir. Ancak sefer

sıklığının gerektiğinden fazla olması durumunda ise çevresel açıdan olumsuz sonuçlara yol açabileceğini belirtmiştir. Bulanık doğrusal programlama yöntemini kullanarak maliyet hesaplaması yaparak en uygun olan çözümü öneri olarak sunmuştur. Ulaşımda optimizasyon ve tahminleme pek çok ulaşım aracı üzerinde çalışmaya konu olmuştur. Gencer ve arkadaşları ise çalışmalarında Ankara metro hattı üzerinde bir çalışma yapmışlardır (Gencer, Alakaş, Tamer, & Hamurcu, 2018). Çalışmalarında ulaşımla ilgili problemlerden ve trafik sıkışıklığından bahsetmişlerdir. Sorunun giderilmesi için ulaşımla ilgili çalışma yapılması ve bu çalışmalar doğrultusunda bir planlamanın olması gerektiğini belirtmişlerdir. Yeni bir planlama yapabilmek için yolcu sayısı, sefer saatlerinin uzunluğu ve yolcu kapasitesi gibi değişkenlere ihtiyaç olduğunu belirtmişlerdir. Kızılay-Batıkent hattını incelemişler ve geçmişe dönük yolcu sayılarını değerlendirerek tahmin yapmışlardır. Yapılan tahminler doğrultusunda yolcu sayılarına göre tren sayısı ve sefer sıklığı ile ilgili yeni bir planlama üzerine çalışmışlardır.

Ulaşımda yapılan sefer sıklığının tahminleme çalışmaları dışında yolcuların kullanması için alternatif güzergâhların oluşturulabileceği çalışmalarda yapılmıştır. Engin, yapmış olduğu çalışmada zaman ve maliyet açısından değerleri en az değerde bulunan alternatifleri bulmak için genetik algoritmayı ve optimizasyon yöntemlerini kullanmıştır (Engin, 2013). Seçtiği yöntemleri Çanakkale'deki toplu ulaşım üzerinde uygulamış yeni modeller oluşturmuş ve bunları değerlendirmiştir. Şehir içi ulaşımda yapılan çalışmalar ve planlamalar nüfusun ve ulaşımda talebin artmasıyla giderek artar bir hale gelmiştir. Sağbaş ve Polat ise şehir içi toplu ulaşım modellemesi ile ilgili yapmış oldukları çalışmalarında doğrusal hedef programlama yöntemiyle yolcu taleplerini esas alarak şehir içi ulaşım için otobüs sefer çizelgeleme optimizasyonu yapmışlardır (Sağbaş & Polat). Sağbaş ve Polat'ın bir diğer çalışmasında ise ulaşım planlamasıyla ile ilgili başka bir çalışma daha yapmışlardır (Polat, Sağbaş, & Dermenci, 2022). Çalışmalarında Tekirdağ/Çorlu Belediyesi toplu taşıma sistemleri ile ilgili doğrusal hedef programlama modeli geliştirmişlerdir. Günlük sefer sayıları, sefer sıklıkları ve yolcu taşıma kapasitelerini değerlendirerek sefer sayısı optimizasyonu yapmışlardır. Kullandıkları model LINGO paket yardımıyla çözülmüş ve optimum sefer sayısını belirlemiştir. Ulaşımda sadece sefer sıklığı optimizasyonu için çalışmalar yapılmaktadır. Ulaşımda olması muhtemel her türlü potansiyelden faydalanmak ve maliyeti azaltmak için birçok çalışma ve araştırma yapılmış ve yapılmaya da devam etmektedir. Palamutçuoğlu yapmış olduğu çalışmasında öncelikli olarak biniş noktalarını temel alarak iniş noktalarını tahminlemeyi amaçlamıştır (Palamutçuoğlu, 2020). Bu doğrultuda gelecekteki talebin öngörülmesi ile sefer sıklığının belirlenmesi ve optimize edilmesi konuları üzerine durmuştur. Çalışmasında rastgele atama

tahmin yöntemi, yapay sinir ağları ve genetik algoritma yöntemini kullanmıştır. Ulaşım sorunu ile ilgili olan problem nüfus ve yoğunluğuna göre artarak devam etmektedir. Bu konuyla ilgili yapılan çalışmalarda buna paralel olarak artış göstermektedir. Nar ise yapmış olduğu çalışmada artan nüfus hızı ve yolcu taleplerini göz önünde bulundurarak Yenikapı M1 – Kirazlı M1 hattında yolcu talebiyle ilgili çalışma yapmıştır (Nar, 2022). Çalışmasında yolcuların bekleme süreleri ve yolcu memnuniyeti artıma amacıyla yapay sinir ağları, makine öğrenmesi ve yolcu talep tahmini yapmıştır. Çalışmasında elde ettiği sonuçlara göre istasyon bazlı yoğunluk ve talebe göre en doğru ve güvenilir sonuca ulaşmak için makine öğrenme yöntemlerinden biri olan karar ağacının daha başarılı sonuçlar verdiğini belirtmiştir. Çalışmasını iki aşamalı olarak gerçekleştirmiştir. İkinci aşamada ise raylı ulaşımda sefer sıklığı için çalışma yapmış ve sefer sıklığı belirlemede öncelikli referansların neler olduğunun bilinmesi gerektiğini belirtmiştir. Ulaşımda talebin karşılanması için yenilikler ve buna bağlı olarak sürekli olan değişimler gözlenmektedir. Kızıltaş ve Akkaya, talebin karşılanması ve arz talep ilişkisi ile ilgili bir çalışma yapmışlardır (Kızıltaş & Akkaya, 2022). Çalışmalarında ulaşımı etkileyen faktörlerin nüfusun fazla ve etkin olmasına bağlı olduğunu ve talebi etkileyebileceğini belirtmişlerdir. Öncelik olarak kent ve ulaşımın ilişkili olduğunu belirtmişler sonrasında ise sıklık konusuna değinmişlerdir. Bu doğrultuda sıklık ile ilgili optimizasyon çalışmalarına değinmişler bununla ilgili modelleme üzerine çalışıp önerilerini sunmuşlardır. Sefer sıklığı optimizasyonu farklı alanlarda çalışmalara konu olduğu gibi hat çizelgeleme ve belirlemede de çalışmalarda yer edinmiştir. Erpik ise yapmış olduğu çalışmasında kentleşmenin arttığını, artan nüfusa bağlı olarak toplu ulaşımda kullanılan modellerinde gelişmesi gerektiğini belirtmiştir (Erpik, 2019). Nüfusun artışı ile birlikte kişilerin toplu ulaşıma yöneldiğini ifade etmiştir. Yolcuların toplu ulaşımı etkin bir şekilde kullanabilmesi için ulaşımın nüfus artışına göre talebin karşılanması gerektiğini belirtmiştir. Bunun için etkili olabilecek planlamanın yapılması gerektiğini vurgulamıştır. Ulaşımda sefer sıklığının belirlenmesini etkileyen en önemli etmenlerden bir tanesi yolcu sayısı ve yolcuların talebine göre sefer sayılarının belirlenmesidir. Korkmaz ve arkadaşları şehir içinde kullanılan toplu ulaşım için kısa süre aralığına karşılık gelen yolcu tahminlemesi üzerine çalışmışlardır (Korkmaz, İlyas, & Efe). Çalışmalarını gerçekleştirirken python programlama dili ve spyder ara yüzü üzerine çalışmışlardır. Bu çalışmada yolcu sayısını tahmini yapmaya çalışmışlardır. Bursa Uludağ Üniversitesini varış yeri olarak belirleyerek kısa dönemli bir yolcu sayısı tahmini yapmışlardır. Ulaşımla ilgili pek çok çalışma yapılırken yolcu sayısı, sefer sıklığının yanı sıra hat belirleme ve var olan hatlara ek yapılması konusunda da öneriler sunulmaktadır. Şaşmaz, yapmış olduğu çalışmasında İstanbul'daki en önemli

sorunlardan birinin ulaşımda olduğunu nüfusun artması ve trafikteki araç sayısının artmasının da bu konuyu olumsuz etkilediğini belirtmiştir (Şaşmaz, 2019). Yolcu yoğunluğunun düşük olduğu bölgelerdeki taleplerin yoğun olduğu bölgelere aktarılarak karşılandığını belirtmiştir. Çalışması kapsamında ulaşımdaki kesişme noktalarını temel alarak İstanbul'un çeşitli noktalarına aktarma merkezleri önermiştir. Ulaşımdaki sefer sıklığı ve yolcu talebiyle ilgili yapılan çalışmalar karayolu ve demir yolu dışında hava ulaşımında önemli çalışmalara konu olmuştur. Havayolu ulaşımı günümüzde oldukça fazla kullanılır bir konuma gelmiştir. Havayolu ulaşımında Efendigil ve Eminler talep tahmini ile ilgili çalışma yapmışlardır (Efendigil & Eminler, 2017). Çalışmalarında havayolu ulaşımının hız, güvenlik ve rahatlık açısından kara ve demir yoluna göre daha fazla tercih edildiğini belirtmişlerdir. Çalışmada yolcu talebinin ciddi bir öneme sahip olduğunu belirtmişler ve yapay zekâ yöntemlerini kullanarak talep tahmini üzerine çalışmışlar ve en doğru sonucu veren yöntemi bulmak için sonuçları karşılaştırmışlardır. Elde ettikleri sonuca göre ANFIS ve yapay sinir ağları yöntemlerinden en güvenilir sonucun yapay sinir ağları ile elde edildiğini belirtmişlerdir. Havayolu ulaşımının yanı sıra demiryolu ulaşımının da planlanması ve yolcu talebine göre hareket edilmesi doğru kararların alınmasında önemli bir yer edinmektedir. Çakır, demiryolu ile ilgili yapmış olduğu çalışmasında yolcu taşıma talebinin belirlenmesi için dokuz bağımsız değişken yolcu sayısı için regresyon analizi ve yapay sinir ağları ile modelleme yapılmıştır (Çakır, 2020). Yapılan çalışma sonucunda yapay sinir ağlarının demiryolu ulaşımında yolcu sayısını belirlemede daha güvenilir sonuç verdiğini ve gerçeğe yakın değerler oluşturduğunu belirtmiştir. Şehir içi ulaşımda talebin artması gelişmekte olan ülkelerde artmakta ve trafikle ilgili ciddi sorunlara yol açabilmektedir. Yaşar, yapmış olduğu çalışmasında bu konuya değinmiştir (Yaşar, 2009). Toplu ulaşım sistemlerinde kalitenin arttırılabilmesi için bu talebe göre çalışılmasının gerekliliğini belirtmiştir. Yapmış olduğu çalışmada talep analizi ve matematiksel yeni metotların kullanılmasıyla talep tahmini yapmıştır. Kraft Talep modeli ile 4 farklı alternatif oluşturarak arz talep konusu için en iyi senaryonun belirlenmesini sağlamıştır. Ulaşımın en önemli etmenlerinden biri olan yolcu ve yolcu sayısı ile ilgili olan çalışmalar Türkiye'de olduğu gibi yurtdışında da çalışmalara konu olmuştur. Carmona-Benítez ve arkadaşları havalimanında yolcu taşıması ve yolcu sayısının tahmini ile ilgili çalışma yapmışlardır (Carmona-Benítez & Nieto-Delfín, 2015). Çalışmalarının amacı medyan yolcu sayısını belirlemek ve buna paralel olarak havaalanlarında mesafe bazlı olarak yolcu talebini belirlemek ve tahmin etmek için yeni yöntemler önermektir. Çalışmalarını aşamalara ayırmışlardır. İlk olarak mesafe bazlı ortalama yolcu ve medyan yolcu talebinin belirlenmesi

için önyükleme prosedürlerinin uygulanması, ikincisi hızlandırılmış yöntem olarak bilinen BCa metodunun kullanılması, üçüncüsü ise önyükleme prosedürlerinin performansının analiz edilmesi için Monte Carlo deneylerinin gerçekleştirilmesi. Yapılan bu üç aşamalı çalışma sonrasında Havayolu ulaşımında yolcu sayısının ve talebinin belirlenmesi için yolcu talebi medyanının önemli bir nokta olduğunu belirtmişlerdir. Carmona-Benítez ve arkadaşlarının başka bir çalışmasında ise havayolu ulaşımında yolcu sayısını tahmin etmek için Ekonometrik Dinamik Model (EDM) önermişlerdir (Carmona-Benítez, Nieto, & Miranda, 2017). Çalışmalarında Sargan testi ve Arellano Bond oto korelasyon kullanılmış ve doğrulanan EDM'nin test edilmesi için Arellano-Bover yöntemini uygulamışlardır. Bunlara bağlı olarak yolcu talebinin belirlenmesi için EDM'yi kullanan bir yaklaşım benimsemişlerdir. Yapmış oldukları çalışma Meksika hava taşımacılığı ile ilgilidir ve tahmin edilen her yıl için Havayolu Havaalanı Merkezi konum problemi (AAH) kullanılmaktadır. Ulaşımdaki yolcu sayısının belirlenmesi ve geleceğe yönelik sayının belirlenmesi için tahminleme çalışmalarının yapılması sunulan hizmetin artması için önemli bir etmendir. Kullanılan yöntemler ile tahminleme yapılması sonuca ulaşmak için yeterli olsa da bazı durumlarda karmaşık bir hal alabilmektedir. Jin ve arkadaşları ulaşımda yolcu tahminlemesi üzerine çalışırken bu karmaşıklıktan söz etmişlerdir (Jin, Li, Sun, & Li, 2020). Çalışmalarında bu karmaşık durumu kesin olarak belirleyebilmek için bazı yöntemler kullanmışlardır. Varyasyonel Mod ayrıştırma (Variational Mode Decomposition-VMD), Otoregresif hareketli ortalama (ARMA) ve VMD-ARMA/KELM-KELM önermişlerdir. Öncelikle karmaşıklığı azaltmak için VMD'yi benimsemişlerdir. Ardından kararsız serileri sınıflandırmak için kök test (ADF) kullanmışlardır. Sonuç olarak kullanılan tüm bu bileşenlerin tahmin sonuçlarını KELM yöntemiyle birleştirmişlerdir. Yaklaşımlarının geçerliliğini sağlamlaştırmak için Pekin, Guangzhou ve Pudong havaalanlarında çalışmalar yapmışlardır. Çalışmanın sonuçlarına göre kullandıkları yöntemin diğer modellere göre daha avantajlı olduğunu ve talep tahmini için ikna edici bir araç olduğunu ifade etmişlerdir. Yolcu sayısının belirlenmesi planlama ve sefer sayısının bu doğrultuda düzenlenmesi için önemli bir etmendir. Ulaşımın her alanında olduğu gibi havayolu ulaşımı için de aynı şekilde önemlidir. Kim, yapmış olduğu çalışmasında yolcu sayısının tahmin edilmesinin havalimanı operasyonun belirlenmesinde kritik bir yeri olduğunu ifade etmiştir (Kim, 2016). Havaalanı işletmelerinin uzun vadeli sonuçların beraberinde acil durumlar için kısa dönemli tahminlerinde etkili olduğunu ifade etmiştir. Ayrıca birçok havaalanının doğruluğa bağlı olarak kısa vadeli yolcu talebini tercih eder ve kullanır. Ek olarak yolcu talebindeki kısa vadeli dalgalanmalarında yansıtılması gerektiğini ifade etmiştir. Yapmış

olduğu çalışmasında büyük verileri kullanarak kısa vadeli yolcu talebinin tahmini için bir model geliştirmeyi amaçlamıştır. Önerilen model ortalama %5,3 'lük bir hata oranıyla beraber ileriye dönük sekiz aylık süreçte yolcu sayısının 195.000 civarında artacağını belirtmiştir. Ulaşımla ilgili yapılan yolcu talebine göre tahminleme çalışmaları yurtdışında ağırlıklı olarak havayolu ulaşımındaki yolcu sayısını tahminleme üzerine yapılmaktadır. Gunter ve çalışma arkadaşı yapmış oldukları çalışmada havayolu ulaşımı yolcu sayısının tahminlemesi üzerine çalışma yapmışlardır. Tahminleme için bir Global Vector Auto Regression (GVAR) modeli kullanmışlardır (Gunter & Zekan, 2021). Çalışmalarında tahminleme yapmak için dünyada en çok tercih edilen ve kullanılan ilk 20 havalimanı ile Asya-Pasifik ve Latin Amerika-Karayip bölgelerinin yolcu sayılarını kullanmışlardır. Buradaki yolcu sayılarının tahmin edilmesi için GVAR modelini kullanmışlardır. Yolcu sayıları başta olmak üzere birçok değişken kullanmışlardır. Çalışmalarının sonucuna göre GVAR modeli birkaç havaalanı için en iyi sonucu vermiştir. Ancak genel olarak değerlendirme yapılması gerektiğinde havaalanlarının bağlana bilirliğinin dikkate alınması gerektiğini belirtmişlerdir. Yolcu sayısının geleceğe yönelik tahminlemesi sefer sayılarının artırılması ve kapasitelerin artırılması açısından oldukça önemlidir. Suryani ve çalışma arkadaşları da yapmış oldukları çalışmalarında yolcu sayısının tahminlemesi ve talebe göre bazı çalışmaların yapılabileceğini, pist ve terminal kapasitelerinin bu doğrultuda düzenlenebileceğini belirtmişlerdir (Suryani, Chou, & Chen, 2010). Yapılabilecek olan tahminleme ile yolcu sayısının uçak biletine etkisini, hizmet düzeyi etkisini, günlük uçuş sayısının belirlenmesinin, havaalanında bekleme süresini etkilemesini, pist kullanımı ve terminal kapasitesi ile ilgili gerekli olan ve ihtiyaç duyulan toplam ek alanın belirlenmesinde önemli rol oynadığını belirtmişlerdir. Havayolu planlamasında yolcu sayısının tahmin edilmesi birçok ülkede çalışmalara konu olmuştur. Adeniran ve arkadaşları Nijerya'daki iç hat yolcu talebini tahmin etmek üzerine çalışma yapmışlardır (Adeniran, Kanyio, & Owoeye, 2018). Çalışmalarını 2010 ve 2017 yılı arasındaki iç hat yolcu talebi verilerine göre kullanmışlardır. Çalışmalarının amacı ise Nijeryada yer alan havaalanı, havayolu ve diğer aşamaların planlama sürecine fayda sağlamasıdır. Hava taşımacılığı talebinin hava taşımacılığı arzından fazla olması veya aşırı hava taşımacılığı arzının talep üzerinde olması sorunlarının önlenmesine yardımcı olacaktır. Yolcu tahmini ve talebin karşılanması ile ilgili çalışmaların yapılmasıyla beraber çeşitli yöntemler ve metotlarda kullanılmaktadır. Gelhausen ve çalışma arkadaşları, Alman Havacılık ve Uzay Merkezinde yapmış oldukları çalışmalarında Alman havaalanlarında yolcu ve uçuş hacmi tahmini için dört aşamalı bir model geliştirmişler ve uygulamışlardır (Gelhausen, Berster, & Wilken, 2018). Kullandıkları verilerde bazı

eksikliklerin olması sebebiyle verilerin doğrulanması zor bir hal almıştır. Bu sebeple yolcu sayısını ve uçuş kapasitesini tahmin etmek için eş bütünleme teorisine dayalı başka bir model geliştirmişlerdir. Kullandıkları bu yöntemi Brexit 2016-2018 yılları arası Alman Havaalanı trafik hacmi üzerindeki etkilerini tahmin etmek için kullanmışlardır. Almanya da olduğu gibi Avusturalya da iç hatlarda havayolu yolcu talebini modellemek için tahminleme çalışmaları yapılmıştır. Srisaeng ve çalışma arkadaşları genetik algoritma kullanmışlar ve deneysel olarak test etmişlerdir (Srisaeng, Baxter, Richardson, & Wild, 2015). Genetik algoritmayı test edebilmek için 74 adet veri seti kullanmışlardır. Ayrıca genetik algoritmanın doğruluğunu test etmek için örneklem dışı 13 veri seti kullanmışlardır. Çalışmanın sonuçlarına göre GAPAXDE (Genetic algorithm optimization model for Australia's domestic airline passenger demand, as measured by enplaned passengers) ve GARPKSDE (Genetic algorithm optimization model for Australia's domestic airline revenue passenger kilometres performed) modellerinin lineer modellere kıyasla daha doğru, güvenilir ve daha doğru tahmin etme özelliğine sahip olduğunu belirtmişlerdir. GAPAXDE ve GARPKSDE ikinci dereceden modeller için örnek dışı test veri setindeki ortalama mutlak yüzde hatası sırasıyla %2,55 ve %2,23'tür. Çoğu yerde yolcu talebinin tahmin edilmesi ile ilgili çalışmalar yapılmaktadır. Scarpel, yapmış olduğu çalışmasını Sao Paulo uluslararası havalimanındaki yolcu sayısını tahmin etmek üzerine yapmıştır (Scarpel, 2013). Çalışmasında yerel uzmanlar modeli karışımı kullanmış ve kullanılan yöntem sonucunda oluşturulan tahminlerin tatmin edici olduğunu belirtmiştir. Tahminlerdeki belirsizliğin belirlenmesinin yanı sıra farklı senaryoları göz önünde bulundurarak oluşturulan model yeni tahminleri yapar. Avrupa da hızlı büyüyen havayolu ulaşımı ve sayıların artması her ülkede olduğu gibi Türkiye'de de belirgin bir şekilde fark edilmektedir. Koç ve Arslan çalışmalarında hızla büyüyen havayolu ulaşımı ve rekabetçi ortamdan bahsetmişlerdir (Koç & Arslan, 2018). Bu rekabetçi ortamda ticari faaliyetler açısından başarılı tahmin yapmanın önemli bir yeri olduğunu belirtmişlerdir. Tahmin çalışmalarında doğrusal regresyon analizi, geri beslemeli sinir ağları, yerçekimi modelleri ve pek çok modelin kullanıldığını belirtmişlerdir. Yapmış oldukları çalışmada ise Türkiye'de iç hat hava taşımacılığı üzerine yapay sinir ağı modeli kullanmışlardır. 01.01.2007 - 01.11.2015 tarihleri arasındaki veri seti için AzureML, RScript, Matlab kullanmışlar ve başarılı sonuçlar elde etmişlerdir. Değerlendirme performansı açısından önerilen yapay sinir ağı modelinin kapasite planlaması, altyapı gibi pek çok alanda talep tahmininde kullanılabilir bir yöntem olduğunu belirtmişlerdir. Yolcu sayısının tahmin edilmesi sefer sıklığı ile ilişkili olduğu gibi yolculuk ücretlerinin düzenlenmesi, fiyat tutarlılığı ve zaman yönetimi ile ilgili önemli sonuçların da elde edilmesini sağlar. Hsiao ve Hansen yapmış

oldukları çalışmasında bu konu ile ilgili çalışma yapmışlardır (Hsiao & Hansen, 2011). Ampirik analiz ve beraberinde kullanılan alternatifler arasında korelasyon modelinin üç seviyeli iç içe logit modeli tanımlanabileceğini belirtmişlerdir. Havayolu için yolcu talep tahmini yoğun olarak yapılmakla beraber ulaşımda kullanılan bütün türler için çalışmalar yapılmaya devam etmektedir. Ulaşım talebini tahmin etmek, ulaşım sistemleri açısından oldukça önemlidir. Operasyonların planlanması ve çizelgelenmesinde ulaşım talebinin bilinmesi kritik bir öneme sahiptir. Anvari ve çalışma arkadaşları raylı sistemlerde ulaşım talebi ile ilgili çalışma yapmışlardır (Anvari, Tuna, Canci, & Turkay, 2016). Operasyonel ve planlama açısından talebe göre hareket etmenin önemini belirtmişlerdir. Çalışmalarında Box-Jenkins yöntemine dayanarak tahminleme çalışması yapmışlardır. Önerdikleri yöntemin kapsamlı bir şekilde kullanılması için matlab kullanmışlardır. Çalışmalarında veri seti olarak İstanbul metrosundan alınan gerçek yolcu bilgilerini kullanmışlardır. Sonuç olarak kullandıkları yöntemin literatürdeki birçok yöntemden daha yüksek doğruluğunun olduğunu belirtmişlerdir. Ulaşımda planlamanın yolcu talebine göre yapılması oldukça önemli bir noktada yer almaktadır. Önem arz eden ve aciliyetini bulunan durumlar için yolcu talebinin bilinmesi ve talebin karşılanması gerekmektedir. Jafari, yapmış olduğu çalışması Covid-19 etkisi ve beraberinde getirdiği aciliyet durumlarının olduğunu belirterek yolcu talebinin etkili bir şekilde tahmin edilmesinin öneminden bahsetmiştir (Jafari, 2022). Havayollarının buna uygun planları geliştirmesinin kurtarma faaliyetleri açısından herhangi bir olumsuz durumu engelleyebileceğini söylemiştir. Ulaşımda yolcu sayısının belirlenmesi ve tahmin edilmesi sefer sıklığı optimizasyonu dışında altyapının düzenlenmesi için de çalışmalara konu olmaktadır. Plakandaras ve arkadaşları Amerika Birleşik Devletleri (ABD) iç hatlarında hava, karayolu ve tren taşımacılığı tahmini üzerine çalışma yapmışlardır (Plakandaras, Papadimitriou, & Gogas, 2019). Çalışmalarında ulaşım talebinin bilinmesinin altyapı iyileştirmesi için alınan kararların doğru orantılı olduğunu belirtmişlerdir. Çalışmalarında 2000-2015 yılları arasını incelemişler ve destek vektörü regresyonunu kullanmışlardır. Planlama açısından yolcu talebinin belirlenmesi oldukça yüksek öneme sahip olduğu için pek çok çalışma ve araştırmaya konu olmaya devam etmektedir. Marie-Sainte ve arkadaşları yapmış oldukları çalışma bu konuya değinmişlerdir (Marie-Sainte, Saba, & Alotaibi, 2019). Yolcu talebinin tahmin edilebilmesi için doğrusal regresyon ve yapay sinir ağı gibi tekniklerin kullanılması gerektiğini belirtmişlerdir. Yöntemlerin doğruluğuyla beraber hata oranını en aza indirebilmek için bazı değişkenlere gereksinim olduğunu belirtmişlerdir. Buna en uygun yöntem olarak genetik algoritmayı araştırmışlardır. Doğrusal regresyonda en uygun sayıları bulmak için ateşböceği algoritması ve parçacık sürü

optimizasyonunu önermişlerdir. Sonuçlara bağlı olarak parçacık sürü optimizasyonunu düşük bir hata oranıyla en iyi tahmin sonuçlarını verdiğini belirtmişlerdir. Yolcu talebi için yapılan tahminleme sonuçları kullanılan metotlara göre farklılık göstermektedir. Sonuçlar metotların hata oranlarına göre farklılık göstermektedir. Aynı zamanda kullanılan yöntemlerin bazıları da gözlemleme açısından tahmin aşamasını zora sokmaktadır. Zhang ve Wang yapmış oldukları çalışmada bu konudan bahsetmişlerdir (Zhang, Zheng, & Wang, 2019). Model belirsizliği göz önünde bulundurularak stokastik sınır analizi modeli oluşturmuşlardır. Bu model için birçok alternatif belirlemişlerdir. Çalışmanın sonuçlarında ise önerdikleri yöntemin yolcu trafiğini tahmin etmesi açısından yaygın olarak kullanılan diğer yöntemlere göre daha iyi performans gösterdiğini belirtmişlerdir. Demiryolu için yapılan yolcu talep tahmini uzun vadeli ve kısa vadeli zaman aralıkları kullanılarak tahmin edilmeye çalışılmaktadır. Demiryolu planlaması ve işletimi için önemli bir etmendir. Zhao ve Mi yapmış oldukları çalışmalarında bu konuya dikkat çekmişlerdir (Zhao & Mi, 2019). Yüksek hızlı demiryolu ulaşımı için yeni bir hibrit model önermişlerdir. Çalışmalarının sonucunda Yüksek hızlı tren talep tahmini için önerdikleri SSA-WPDCNN-SVR modelinin kullanılmakta olan diğer modellere göre daha iyi performans gösterdiğini belirtmişlerdir. Demiryolu planlamasıyla beraber karayolu için kullanım da planlamanın yapılabilmesi için talep tahminin yapılması oldukça önemlidir. Cyril ve arkadaşları yapmış oldukları çalışmada bu konudan bahsetmişlerdir (Cyril, Mulangi, & George, 2018). Planlama yapılması ve otobüslerin yoğun olan ağa tahsis edilmesini ulaşımdaki sorunlardan bazılarına yardımcı olabileceğini belirtmişlerdir. Önerdikleri ARIMA (Trivandrum-Kerela arası 5 ilçe için ilçeler arası toplu ulaşım talep tahmin modeli) modelinin seçmiş oldukları bölgede kısa vadede talep tahmini için doğru sonuçlar verdiğini ifade etmişlerdir. Yolcu talep tahminleri kullanılan değişkenlere göre farklılık veya düzensizlikler gösterebilmektedir. Xiao ve arkadaşları çalışmalarında bu düzensizlikten bahsetmişlerdir (Xiao et al., 2014). Çalışmalarında SSA (Singular Spectrum Analysis), ANFIS ve geliştirilmiş parçacık sürü optimizasyonunu birleştiren bir yaklaşım önermişlerdir. Çalışmanın sonucunda önerdikleri modelin yüksek bir potansiyele sahip olduğunu ve kullanılabilecek bir potansiyeli olduğunu belirtmişlerdir. Yolcu talep tahmininde yapay sinir ağlarının birçok yöntemi kullanılmaktadır. Srisaeng ve arkadaşı yapmış oldukları çalışmalarında yapay sinir ağı örneklerinin sınırlı kullanılmasına karşın Avustralya için ilk kez bir yapay sinir ağını uygulamışlardır (Srisaeng, Baxter, & Wild, 2015). Avustralya'nın havayolu yolcu talebini PAX Modeli, yolcu-kilometre gelirini (RPKs modeli) tahmin etmek için iki yapay sinir ağı modeli oluşturmuşlardır. Hava ulaşımında kararları desteklemek için tahmin yöntemlerine güvenilmektedir. Ancak bazı

durumlarda bu tahminler bazı sorunlara sebep olmaktadır. Alekseev ve çalışma arkadaşı yapmış oldukları çalışmada bu probleme değinmişlerdir (Alekseev & Seixas, 2002). Problemin, Brezilya bazında problem teşkil ettiğini belirtip havayolu talep tahmini için yapay sinir ağlarına dayalı modeller geliştirmişlerdir. Nöron işlemenin daha iyi performans göstereceğini belirtip önceden işlenmiş verilerle tahmin hatasının %0,6'dan küçük olarak değerlendirileceğini ifade etmişlerdir. Tahmin etme yöntemi olarak pek çok farklı metot kullanılmaktadır. Srisaeng ve arkadaşları yapmış oldukları çalışmalarında düşük maliyetli taşıyıcıların talebini tahmin etmek için Genetik Algoritma modelleri önermişler ve test etmişlerdir (Srisaeng, Baxter, & Wild, 2015). GAPAXDE ve GAR-PKSDE modellerinin daha güvenilir ve yüksek tahmin yeteneğine sahip olduğunu belirtmişlerdir. Yolcu sayısını tahmin etmek ülke içinde yurtdışında tahmin edildiği gibi şehirlerarası ulaşımda da sıklıkla kullanılmaktadır. Terekhov ve arkadaşları bu konuya değinmiş ve bunun için ağırlıklı benzerlik tabanlı bir algoritma kullanıldığını ifade etmişlerdir (Terekhov, Evans, & Gollnick, 2016). Üç ayrı yıla ait modellenen ve bilinen verilerin karşılaştırmasıyla belirlemişlerdir. Sonuç olarak kaynak bir indeks belirlenmesi gerektiğini ifade etmişlerdir.

Yukarıda bahsedilen ve literatürde yer alan bu çalışmalar çalışmanın yapıldığı ülke, kullanılan yöntem ve hangi ulaşım türlerinin kapsadığını gösterecek şekilde Tablo 2.'de özetlenmiştir. Çalışmaların özeti incelendiğinde Türkiye'nin farklı illeri için çalışmaların yapıldığı görülmektedir. Ayrıca çalışmaların geneli incelendiğinde çoğunluğunun tahminleme için yapıldığı ve hem optimizasyon hem de tahminleme yöntemlerinin bir arada olduğu çalışmaların sayısının çok küçük bir kısımda yer aldığı görülmektedir. Yapılan çalışmalar ulaşım türleri özelinde incelendiğinde ise deniz ulaşımının modellere dâhil edilmediği (çalışmanın yapıldığı bölgenin özelliklerinden dolayı) ve çoğunlukla kara ya da hava ulaşımında çalışmaların yapıldığı görülmüştür. Bu doğrultuda literatürde yer alan çalışmalar ve özet tablo incelendiğinde yurt dışında kullanılmış olsa da İstanbul'a ait toplu ulaşım için yapılan çalışmalarda hem optimizasyon hem de tahminleme yöntemlerinin aynı anda kullanılmadığı gözlemlenmiştir. Bu da yapılan tez çalışmasının literatürdeki yenilikçi yönünü oluşturmaktadır.

**Tablo 2.4.** Literatür Özeti

| Yazar(lar) | Yıl | Ülke/Şehir | Yöntem | | Ulaşım Türü | | | |
|---|---|---|---|---|---|---|---|---|
| | | | Optimizasyon | Tahminleme | Hava | Kara | Deniz | Demir |
| Alekseev ve Seixas | 2002 | Brezilya | | X | X | | | |
| Yaşar | 2009 | Türkiye | | X | | X | | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Yaylalı ve Dilek | 2009 | Türkiye / Erzurum | | X | X | | | |

**Tablo 2.4 (Devamı).** Literatür Özeti

| **Yazar(lar)** | **Yıl** | **Ülke/Şehir** | **Yöntem** | | **Ulaşım Türü** | | | |
|---|---|---|---|---|---|---|---|---|
| | | | **Optimizasyon** | **Tahminleme** | **Hava** | **Kara** | **Deniz** | **Demir** |
| Suryani, Chou ve Chen | 2010 | | | X | X | | | |
| Uludağ | 2010 | Türkiye / İzmir | X | | | X | | |
| Hsiao ve Hansen | 2011 | | | X | X | | | |
| Deri | 2012 | Türkiye / İzmir | X | X | | X | | |
| Engin | 2013 | Türkiye / Çanakkale | X | | | X | | |
| Scarpel | 2013 | Brezilya / Sao Paulo | | X | X | | | |
| Yi, Liu, Hu, Wang, Lai ve Wang | 2014 | Hong Kong | | X | X | | | |
| Anvari, Tuna, Canci ve Türkay | 2015 | Türkiye / İstanbul | | X | | | | X |
| Carmona-Benítez ve Nieto-Delfín | 2015 | | | X | X | | | |
| Srisaeng, Baxter, Richardson ve Wild | 2015 | Avustralya | | X | X | | | |
| Srisaeng, Baxter ve Wild | 2015 | Avustralya | | X | X | | | |
| Kim ve Shin | 2016 | | | X | X | | | |
| Srisaeng, Richardson, Baxter ve Wild | 2016 | Avustralya | | X | | X | | |
| Terekhov, Evans ve Gollnick | 2016 | Global | | X | X | | | |
| Carmona-Benítez, Nieto ve Miranda | 2017 | Meksika | | X | X | | | |
| Gelhausen, Berster ve Wilken | 2018 | Almanya | | X | X | | | |
| Gencer, Alakaş, Eren ve Hamurcu | 2018 | Türkiye / Ankara | | X | | | | X |
| Koç ve Arslan | 2018 | Türkiye | | X | X | | | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Özcan | 2018 | Türkiye | X | | | X | | |

**Tablo 2.4 (Devamı).** Literatür Özeti

| Yazar(lar) | Yıl | Ülke/Şehir | Yöntem | | Ulaşım Türü | | | |
|---|---|---|---|---|---|---|---|---|
| | | | Optimizasyon | Tahminleme | Hava | Kara | Deniz | Demir |
| Erpik | 2019 | Türkiye / İstanbul | | X | | X | | |
| Larabi, Tanzila ve Sihaam | 2019 | | X | X | X | | | |
| Plakandaras, Papadimitriou ve Gogas | 2019 | Birleşik Amerika | | X | | X | | X |
| Şaşmaz | 2019 | Türkiye / İstanbul | X | | | X | | |
| Zhang, Zheng ve Wang | 2019 | | | X | X | | | |
| Zhao ve Mi | 2019 | | | X | | | | X |
| Çakır | 2020 | Türkiye | | X | | | | X |
| Jin, Li, Sun ve Li | 2020 | Çin | | X | X | | | |
| Palamutçuoğlu | 2020 | Türkiye | X | X | | X | | |
| Gunter ve Zekan | 2021 | Pasifik Asya-Amerika Karayipler | | X | X | | | |
| Jafari | 2022 | Birleşik Amerika | | X | X | | | |
| Kızıltaş ve Akkaya | 2022 | Türkiye | X | | | X | | |
| Korkmaz, Özer ve Efe | 2022 | Türkiye / Bursa | | X | | X | | |
| Nar | 2022 | Türkiye / İstanbul | | X | | | | X |
| Polat, Sağbaş ve Dermenci | 2022 | Türkiye / Tekirdağ | X | | | X | | |

# 3. YÖNTEM

Tez çalışmasının bu bölümü uygulamaların yapılması ve elde edilen sonuçların sunulmasından oluşmaktadır. Bu doğrultuda ülkemizin nüfus yoğunluğu olarak 5'te 1'lik kısmına denk gelen İstanbul ili uygulama bölgesi olarak seçilmiştir. Bu nedenle ilk önce İstanbul iline mevcut veriler hakkında bilgi verilerek sonrasında benzer ilçelerin bir arada yer almasını sağlayan kümeleme işlemi yapılacak ve yolculuk sayılarını kullanarak gelecekteki yolculuk talepleri belirlenecektir. Sonrasında her iki aşamada elde edilen bilgiler de kullanılarak nerelere kaynak aktarılması gerektiğine dair oluşturulan matematiksel model sonuçları verilecektir.

## 3.1. UYGULAMA BÖLGESİ: İSTANBUL

İstanbul ili Türkiye'nin en fazla nüfusunun yaşadığı ve ancak yüz ölçümünün fazla olmamasından dolayı özellikle şehir merkezinde nüfus yoğunluğunun fazla olduğu bir ildir. Şekil 1.1'de İstanbul'un mahallelerindeki nüfus bilgilerine göre oluşturulan nüfus yoğunluklarını göstermektedir (Atlas Big, 2023). Açık sarı renkten koyu kırmızı renge doğru gidildikçe nüfus yoğunluğu artmakta ve kalabalık yerlerin şehrin merkezine toplandığı görülmektedir. Bu da özellikle kırmızı bölgelerde özel araç ve toplu taşıma araçları olmak üzere trafik yoğunluğunun fazla olduğunu göstermektedir. Bu tez çalışmasının amacını da bu yoğunluk bölgelerinde talebin nasıl hareket edeceğinin incelenmesi ve geliştirmek için nasıl önerilerde bulunulması gerektiğini kapsamaktadır.



**Şekil 3.1.** İstanbul'un Mahallelere göre Nüfus Yoğunluğu

İstanbul ili toplamda 954 mahalle ve 39 ilçeden oluşmaktadır. Bu çalışmada ise talep noktaları olarak ilçeler kullanılacaktır. İstanbul'un ilçeleri Tablo 3.1'de yer almaktadır. İstanbul ili Avrupa ve Anadolu olmak üzere iki yakadan oluşmaktadır ve 39 ilçeden 14 tanesi Anadolu

yakasında ve kalan 25 tanesi ise Avrupa yakasında yer almaktadır. Tablo 3.2 ve Tablo 3.3'te ise sırasıyla Avrupa yakası ve Anadolu yakasında yer alan ilçeler verilmektedir.

**Tablo 3.1.** İstanbul'un İlçeleri

| | | |
|---|---|---|
| Adalar | Büyükçekmece | Maltepe |
| Arnavutköy | Çatalca | Pendik |
| Ataşehir | Çekmeköy | Sancaktepe |
| Avcılar | Esenler | Sarıyer |
| Bağcılar | Esenyurt | Silivri |
| Bahçelievler | Eyüpsultan | Sultanbeyli |
| Bakırköy | Fatih | Sultangazi |
| Başakşehir | Gaziosmanpaşa | Şile |
| Bayrampaşa | Güngören | Şişli |
| Beşiktaş | Kadıköy | Tuzla |
| Beykoz | Kâğıthane | Ümraniye |
| Beylikdüzü | Kartal | Üsküdar |
| Beyoğlu | Küçükçekmece | Zeytinburnu |

**Tablo 3.2.** İstanbul'un Avrupa Yakasında Yer Alan İlçeleri

| | | | | |
|---|---|---|---|---|
| Arnavutköy | Başakşehir | Büyükçekmece | Fatih | Sarıyer |
| Avcılar | Bayrampaşa | Çatalca | Gaziosmanpaşa | Silivri |
| Bağcılar | Beşiktaş | Esenler | Güngören | Sultangazi |
| Bahçelievler | Beylikdüzü | Esenyurt | Kâğıthane | Şişli |
| Bakırköy | Beyoğlu | Eyüpsultan | Küçükçekmece | Zeytinburnu |

**Tablo 3.3.** İstanbul'un Anadolu Yakasında Yer Alan İlçeleri

| | | | | |
|---|---|---|---|---|
| Adalar | Çekmeköy | Maltepe | Sultanbeyli | Ümraniye |
| Ataşehir | Kadıköy | Pendik | Şile | Üsküdar |
| Beykoz | Kartal | Sancaktepe | Tuzla | |

Öncelikle İstanbul'un ilçeleri belirlenen kriterlere göre kümeleme yapılarak benzer yapıya ya da karakteristiğe sahip olan ilçelerin bir arada olması sağlanacaktır. Burada bahsedilen kümeleme işlemi gözetimsiz öğrenme olarak adlandırılan makine öğrenmesinin bir alt

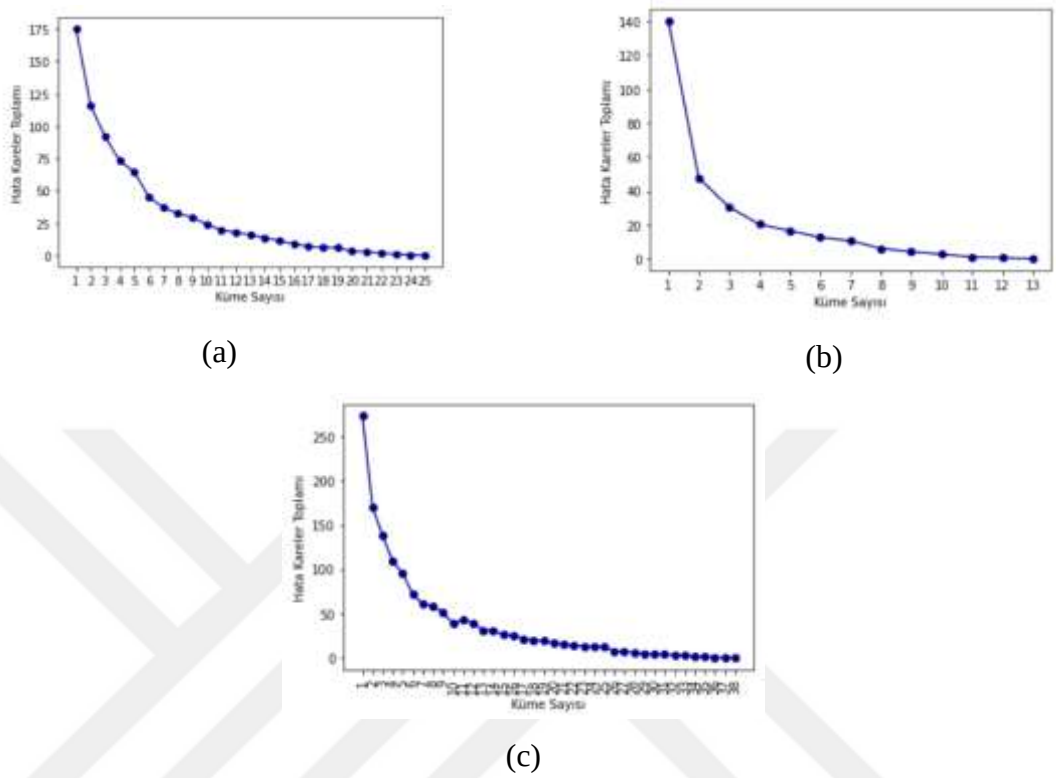
başlığıdır. Temelde bağımlı değişken olarak adlandırılan bir çıktı değerinin olmaması nedeniyle var olan bağımsız değişkenler kullanılarak ilgili değerlerin (bu tez çalışmasında kümeleme işleminde ilgili gözlem değeri olarak adlandırılan ilgili değerler ilçelerden oluşmaktadır) gruplandırılması işlemine dayanmaktadır. Kümeleme işleminin yapılabilmesi için İstanbul'un karakterini yansıttığını düşündüğümüz ve Tablo 3.4'de yer alan veriler kullanılmıştır.

**Tablo 3.4.** İstanbul'un İlçelerini Kümelemek için Kullanılacak olan Değişkenler Listesi

| | |
|---|---|
| Hastane Sayısı | İlçelerde Yer Alan Raylı İstasyon Sayısı |
| İlçelerdeki Konut Satış Oranı | Okul Sayısı |
| İlçelere göre Okuma Yazma Bilen Sayısı | Otomobil Sayısı |

Tablo 3.4'de yer alan veriler kullanılarak en uygun küme sayısının bulunabilmesi için öncelikli olarak dirsek grafiği oluşturulmuştur. Dirsek grafiği en az bir küme ve en çok var olan gözlem sayısı değerini içerecek şekilde, 1'den başlayıp gözlem sayısı değerine kadar tüm değerler için kümeleme algoritmasının oluşturulması ve elde edilen değişimlerin bir grafikte gösterilerek en uygun küme sayısının belirlenmesine olanak sağlamaktadır. Bu grafiklerin oluşturulması için python programlama dili ve sklearn paketinde yer alan K-Means algoritması kullanılmıştır. K-ortalamalar olarak Türkçe ismiyle bilinen algoritmanın çalıştırılması için başlangıç fonksiyonu random olarak adlandırılan rassal başlangıç noktasından başlanmıştır. Küme sayısı [1, n) olacak şekilde iterasyon şeklinde ifade edilmiştir (tanım aralığında yer alan ve üst değer olarak gösterilen n veri setimizde toplam ilçe sayısına denk gelmektedir). Algoritmanın farklı merkezler kullanılarak çalıştırılması için gerekli olan parametre 10 olarak belirlenmiştir. Algoritmanın tek bir koşturumunda yapılacak olan maksimum iterasyon sayısı ise 300 olarak belirlenmiştir. Burada belirtilen parametreler birden çalıştırmadan sonra karar verilmiştir. Sonrasında parametreler kullanılarak Avrupa yakasında yer alan ilçeler, Anadolu yakasında yer alan ilçeler ve tüm ilçeler olmak üzere üç farklı koşturma elde edilmiş olup her bir koşturma için elde edilen dirsek grafikleri sırasıyla Şekil 4.2 (a), Şekil 4.2 (b) ve Şekil 4.2 (c)'de yer almaktadır. Elde edilen grafikler incelendiğinde en büyük değişimin birinci ve ikinci küme oluşturulmasında elde edildiği görülmektedir. Grafikteki değişimlerin oranının azaldığı nokta ise Avrupa yakası için oluşturulan dirsek grafiğinde beş kümenin oluştuğu noktada, Anadolu yakası için üç kümenin oluştuğu noktada ve tüm ilçeleri ele aldığımızda ise dokuz kümenin yer aldığı noktada oluşmaktadır. dokuz kümenin seçilmesi mevcut gözlem değeri sayısının yaklaşık

dörtte birlik bir kısmına denk gelmektedir, bu nedenle orta noktanın elde edilebilmesi amacıyla Avrupa yakası grafiğinde elde edilen küme sayısı olan beş kullanılacaktır.




**Şekil 3.2.** Kümeler için Oluşturulan Dirsek Grafikleri: (a) Avrupa Yakası, (b) Anadolu Yakası, c) Tüm İlçeler

Şekil 3.2'de yer alan dirsek grafikleri kullanılarak küme sayısı belirlendikten sonra İstanbul'un ilçeleri Tablo 3.5'te yer aldığı şekilde kümelere ayrılmış olur. Tablo 3.5'te yer alan Küme başlığı hangi küme olduğunu gösteriyorken Kümede Yer Alan İlçeler başlığı altında ilgili kümenin içinde sınıflandırılan ilçeler sınıflandırılmaktadır. Burada kümelerin rassal olarak yerleştirildiğini de belirtmeliyiz. Şekil 1.3'te ise İstanbul haritasında ilçelerin gösterimi yer almaktadır.

**Tablo 3.5.** İstanbul'un İlçelerinin Kümelenmesi

| **Küme** | **Kümede Yer Alan İlçeler** |
|---|---|
| Küme 1 | Bakırköy, Bayrampaşa, Beşiktaş, Beyoğlu, Esenler, Fatih, Güngören, Şişli, Zeytinburnu |
| Küme 2 | Esenyurt, Küçükçekmece, Pendik, Ümraniye |
| Küme 3 | Bağcılar, Bahçelievler, Kadıköy, Maltepe, Üsküdar |
| Küme 4 | Adalar, Arnavutköy, Beykoz, Çatalca, Silivri, Sultanbeyli, Şile, Tuzla |
| Küme 5 | Ataşehir, Avcılar, Başakşehir, Beylikdüzü, Büyükçekmece, Çekmeköy, Eyüpsultan, Gaziosmanpaşa, Kâğıthane, Kartal, Sancaktepe, Sarıyer, Sultangazi |

## 3.2. UYGULAMA VERİ SETİ

Kümeleme işlemi tamamlandıktan sonra yapılan talep noktalarının gelecekteki taleplerinin belirlenmesi için tahminleme işlemi gerçekleştirilmektedir. Bunun için İstanbul Büyükşehir Belediyesi veri portalı tarafından yayınlanan veriler kullanılmaktadır.

Yukarıda da belirtildiği gibi uygulama bölgesi olarak İstanbul ili seçilmiştir. Bu kapsamda ulaşım verilerinin elde edilebilmesi için İstanbul Büyükşehir Belediyesi tarafından "***Açık Veri Portalı***"nda yer alan "Saatlik Toplu Ulaşım Veri Seti" (**İBB Açık Veri Portalı, 2023**) kullanılmıştır. Paylaşılan bu veri seti İstanbul'da saatlik olarak taşınan yolcu ve yolculuk sayısını göstermektedir ve bu veriler aylık verileri içerecek şekilde 2020 Ocak ayından beri paylaşılmaktadır. Veri seti, tarih, saat, ulaşım türü, rotası, yolculuk hattı, yolcuların transfer yolcusu olup olmadığı o hat için geçiş sayısını ve yolcu sayısı değişkenlerini içermektedir.

Veri seti "Saatlik Toplu Ulaşım Verisi" kullanılmaktadır (İBB Açık Veri Portalı, 2023). Paylaşılan veri seti 2020 Ocak ayı ile 2023 Nisan ayı arasındaki ayları kapsayarak aylık toplu ulaşım tarafından takip edilen rotalarda ne kadar yolculuk yapıldığını göstermekte ve Tablo 4.6'da yer alan değişkenlerden oluşmaktadır. "***Tarih***" değişkeni geçişin yapıldığı günü kapsıyorken, "***Saat***" değişkeni hangi saatte o rotadan geçiş yapıldığı ifade etmektedir. "***Ulaşım Tipi***" tabloda verildiği gibi üç farklı kategoride ifade edilmekle birlikte 1,2 ve 3 sayıları ile kategorik veri olarak tutulmaktadır. Bu değerlerin sözel olarak ifade edilmesi, başka bir deyişle Otoyol, Deniz ya da Raylı şeklinde ifade edilmesi "***Rota Adı***" değişkeni altında sunulmuştur.

"***Ulaşım Hattı***" ise hangi hattın kullanıldığını ifade etmektedir. "***Transfer Tipi***" ise o hattı hangi amaçla kullanıldığını göstermektedir. Örneğin, *Normal* ifadesi İETT kartının belirlenen periyotta ilk kez geçiş yapıldığını gösterirken *Aktarma* o kartın en az ikinci kez kullanıldığını ifade etmektedir. "***Yolculuk Sayısı***", verilen saatte, transfer tipi ve ulaşım tipine göre yolculuk sayısını içeriyorken; "***Yolcu Sayısı***", verilen saatte, transfer tipi ve ulaşım tipine göre yolculuk yapan tekil yolcu sayısını içerir.

**Tablo 3.6.** İstanbul Saatlik Toplu Ulaşım Verisi'nde Yer Alan Değişkenler

| | |
|---|---|
| Tarih | Rota Adı |
| Saat | Ulaşım Hattı |
| Ulaşım Tipi (1=Otoyol, 2=Raylı, 3=Deniz) | Transfer Tipi (Normal, Aktarma) |
| Yolculuk Sayısı | Yolcu Sayısı |

Veri seti yukarıda da belirtildiği gibi 2020-2023 yılları arasındaki verileri kapsamaktadır. Ancak 2020 ve 2021 yılları covid-19 olarak adlandırılan pandemi dönemine denk gelmektedir ve bu dönem yolculuk davranışlarını tam olarak yansıtmamaktadır. Başka bir deyişle beklenen talebi tam olarak göstermemektedir. Ayrıca 2023 yılında yer alan verilen tüm yılı kapsamamaktadır. Bu nedenle talep tahminin tam olarak gerçekleştirebilmek amacıyla 2022 yılında yer alan veriler kullanılmıştır. Bu veriler incelendiğinde 7.866.727 satır verinin yer aldığı görülmektedir. Bu da 2022 yılı içerisinde yapılan tüm yolcukları ifade etmektedir. Bu veri setine ait örnek gözlem değerleri Şekil 4.3'te yer almaktadır.

| | Date | Hour | TransportationType | Road | Line | TransferType | NumberofPasage | NumberofPassenger |
|---|---|---|---|---|---|---|---|---|
| **6247610** | 2022-10-19 | 19 | 1 | OTOYOL | ALTUNIZADE-SULTANBEYLI | Aktarma | 235 | 235 |
| **5569277** | 2022-09-19 | 5 | 1 | OTOYOL | ALTUNIZADE-SULTANBEYLI | Normal | 461 | 452 |
| **438501** | 2022-01-21 | 18 | 1 | OTOYOL | ALTUNIZADE-SULTANBEYLI | Normal | 1426 | 1419 |
| **7837173** | 2022-12-30 | 16 | 1 | OTOYOL | ALTUNIZADE-SULTANBEYLI | Normal | 2077 | 2036 |
| **6319010** | 2022-10-23 | 4 | 1 | OTOYOL | ALTUNIZADE-SULTANBEYLI | Normal | 26 | 26 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **771727** | 2022-02-07 | 9 | 1 | OTOYOL | ALTUNIZADE-SULTANBEYLI | Normal | 1349 | 1337 |
| **894539** | 2022-02-12 | 19 | 1 | OTOYOL | ALTUNIZADE-SULTANBEYLI | Normal | 1578 | 1575 |
| **832404** | 2022-02-09 | 22 | 1 | OTOYOL | ALTUNIZADE-SULTANBEYLI | Normal | 741 | 739 |
| **4447554** | 2022-07-28 | 21 | 1 | OTOYOL | ALTUNIZADE-SULTANBEYLI | Normal | 977 | 948 |
| **6199310** | 2022-10-17 | 18 | 1 | OTOYOL | ALTUNIZADE-SULTANBEYLI | Aktarma | 228 | 227 |

**Şekil 3.3.** Ham Veri Setinin Yapısı

Ulaşım Hattı başlığı altında yer alan veriler incelendiğinde ise yapılan tüm yolculukların toplamda 939 üzerinden gerçekleştirildiğini göstermektedir. Belirlenen küme içerisinde yer alan ve bu hatlar içerisinde en çok tekrarlayan üç hat bu tez çalışmasının analizlerinin yapıldığı hattı göstermektedir. Bu şekilde analizlere eklenmesinin sebebi ise yoğun kullanımın olduğunu ifade ettiğini belirtmesidir.

Veri setindeki gözlem değerleri Ulaşım Hattındaki özgün değerlere göre en çok tekrarlayan hat olarak sınıflandırıldığında başka bir deyişle modu alındığında "ALTUNIZADE-SULTANBEYLİ, USKUDAR-BEYKOZ, USKUDAR-CEKMEKOY" hatlarının en fazla kullanıldığı gözlenmektedir. Bu nedenle Altunizade-Sultanbeyli, Üsküdar-Beykoz ve Üsküdar-Çekmeköy hatları talep tahmininin yapılacağı bölgeler olarak seçilmiştir. Bu kapsamda öncelikle veri setinde filtreleme işlemi yapılarak ilgi hatların olduğu alt veri setleri elde edilmiştir. Sonrasında analizlerin yapılabilmesi için veri setinde uygun dönüşümler yapılmıştır. Buna göre alt veri kümeleri oluşturulduktan sonra elde edilen veri kümelerinden bir örnek Şekil 4.3'teki yapıya sahip iken metin formatında olan veriler kategorik veriye dönüştürülerek Şekil 4.4'teki düzenlenmiş veri elde edilmiştir.

| | Date | Hour | TransportationType | Road | TransferType | NumberofPasage | NumberofPassenger |
|---|---|---|---|---|---|---|---|
| **215** | 2022-01-01 | 0 | 1 | 3 | 1 | 60 | 60 |
| **254** | 2022-01-01 | 0 | 1 | 3 | 2 | 134 | 133 |
| **293** | 2022-01-01 | 1 | 1 | 3 | 1 | 74 | 74 |
| **338** | 2022-01-01 | 1 | 1 | 3 | 2 | 254 | 253 |
| **390** | 2022-01-01 | 2 | 1 | 3 | 2 | 101 | 101 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| **7865336** | 2022-12-31 | 21 | 1 | 3 | 2 | 745 | 706 |
| **7865587** | 2022-12-31 | 22 | 1 | 3 | 1 | 111 | 111 |
| **7865620** | 2022-12-31 | 22 | 1 | 3 | 2 | 584 | 557 |
| **7866372** | 2022-12-31 | 23 | 1 | 3 | 1 | 61 | 61 |
| **7866610** | 2022-12-31 | 23 | 1 | 3 | 2 | 358 | 331 |

**Şekil 3.4.** Analizler için Hazırlanan Verinin Son Hali (Altunizade-Sultanbeyli Hattına ait Saatlik Ulaşım Verisi)

Şekil 4.4'te bir kısmı yer alan verinin elde edilmesi için Rota Adı ve Transfer Tipi değişkenlerinde veri tipi dönüşümü yapılmıştır. Bu doğrultuda ilk önce Rota (Road) Adı başlığı altında yer alan "Deniz", "Raylı" ve "Otoyol" tanımları sırasıyla 1,2 ve 3 olacak şekilde değiştirilmiştir. Transfer (Transfer Type) adı altında yer alan "Aktarma" ve "Normal" değerleri ise sırasıyla 1 ve 2 olacak şekilde güncellenmiştir. Sonrasında yapılan filtreleme sonucunda veri seti hep aynı hattı kapsamasından dolayı "Ulaşım Hattı" değişkeni veri setinden uzaklaştırılmıştır. Böylece toplamda 7 değişken ve 16465 veri satırından oluşan veri seti

analizler için hazır hale getirilmiştir. Bu veri seti için açıklayıcı istatistikler ise Şekil 4.5'te yer almaktadır. Bu açıklayıcı veriler tahminleme işlemi yapılmadan önce veri seti hakkında ön bilgi vermekte olup normalizasyon gerektirip gerektirmediğini göstermektedir. Açıklayıcı bilgi ise her bir değişken için o değişkende kaç gözlem değeri yer aldığını, ortalamayı, standart sapmayı, en büyük ve en küçük değeri ve değişkenin çeyreklik bilgilerini göstermektedir. Bu bilgiler normalizasyon gibi ayrıca tahminleme işlemi öncesi ön-işlem gerekip gerekmediğini de gösterebilmektedir.

| | Hour | TransportationType | Road | TransferType | \ |
|---|---|---|---|---|---|
| count | 16465.000000 | 16465.0 | 16465.0 | 16465.000000 | |
| mean | 12.048709 | 1.0 | 3.0 | 1.532038 | |
| std | 6.778097 | 0.0 | 0.0 | 0.498988 | |
| min | 0.000000 | 1.0 | 3.0 | 1.000000 | |
| 25% | 7.000000 | 1.0 | 3.0 | 1.000000 | |
| 50% | 12.000000 | 1.0 | 3.0 | 2.000000 | |
| 75% | 18.000000 | 1.0 | 3.0 | 2.000000 | |
| max | 23.000000 | 1.0 | 3.0 | 2.000000 | |

| | NumberofPasage | NumberofPassenger |
|---|---|---|
| count | 16465.000000 | 16465.000000 |
| mean | 648.927786 | 635.243243 |
| std | 688.409986 | 673.303385 |
| min | 1.000000 | 1.000000 |
| 25% | 132.000000 | 131.000000 |
| 50% | 242.000000 | 241.000000 |
| 75% | 1208.000000 | 1172.000000 |
| max | 3312.000000 | 3274.000000 |

**Şekil 3.5.** Analizlerde Kullanılan Veriye ait Açıklayıcı İstatistikler

Veri setinden talep tahmini yapabilmek için öncelikli olarak değişkenler ve bağımlı değişken olmak üzere ikiye ayrılmalıdır. Bu tez çalışmasında bağımlı değişken "Yolcu Sayısı" iken diğer değişkenler bağımsız değişkenleri göstermektedir. Bu işlem tamamlandıktan sonra verilerin ölçeklendirilmesi gerekmektedir. Bu işlemin amacı değişkenlerin birbiri üzerinde baskın olmasını önlemektir. Bu işlem için "Min-max Scaler" kullanılmıştır. Bu ölçeklendirme işlemi her bir sütun için yapılacak olup o sütundaki en büyük ve en küçük değer belirlenmektedir. Sonrasında her bir gözlem değeri için ilgili sütunda en büyük ve en küçük değerler ile öncelikle standart değeri hesaplanmakta ($Gözlem_{standart} = \frac{Gözlem - Minimum}{Maksimum - Minimum}$)ve sonrasında ölçeklendirilmiş değer ($Gözlem_{ölçek} = Gözlem_{standart} * (Maksimum - Minimum) + Minimum$) elde edilmektedir. Bu işlem sonucunda değerler [0,1] aralığında olacak şekilde elde edilmektedir. Bu işlem sonrasında modellerin oluşturulması, eğitilmesi ve en uygun modelin seçilebilmesi için belirlenen performans göstergeleri açısından karşılaştırılması gerekmektedir. Bunun için ölçeklendirilerek standart hale getirilmiş veri seti eğitim ve test verisi olarak ikiye ayrılmaktadır. Bu işlem için literatüre paralel olarak veri setinin %80'i eğitim verisi (13172

gözlem değeri) ve %20'si test verisi (3293 gözlem değeri) olacak şekilde ayrım işlemi tamamlanır.

### 3.3. GELECEK YOLCULUK SAYILARININ TAHMİN EDİLMESİ

Veri seti test ve eğitim diye ayrıldıktan sonra eğitim veri seti kullanılarak bir önceki bölümde yer alan yöntemler ve algoritmalar kullanılarak talep tahmini için en uygun modelin belirlenmesi sağlanır. Bu doğrultuda doğrusal regresyon, ridge regresyon, lasso, elastic net, k-en yakın komşu regreysonu, karar ağacı resgresyonu ve rastgele orman regresyon mevcut eğitim verisi ile çalıştırılmıştır. Bilindiği gibi doğrusal regresyon herhangi bir hiper parametre içermediği için elde edilen sonuçlar tek bir modelin için elde edilmektedir. Ancak doğrusal regresyon dışında yer alan diğer algoritmalarda farklı hiperparametre değerleri yer aldığı için her bir parametrede farklı parametre değerleri elde edilerek bulunan o algoritma için bulunan en iyi sonuca göre Tablo 4.7 oluşturulmuştur. Ridge ve lasso regresyonda alfa değeri için sırasıyla 0,5 ve 10 için en iyi performans göstergesi değerleri elde edilmiştir. Elastic net ise daha önce bahsedildiği gibi lasso ve ridge regresyonun bir birleşimi olup en iyi sonuçlar kullanılan fonksiyon verilen değerlerinde en iyi değer elde edilmiştir. Polinom regresyonda ise değişken sayısına kadar derece artımı sağlanmıştır. Burada en iyi değerler altı derece dışındaki dereceler olarak elde edilmiştir. Bu nedenle karmaşık yapısından uzaklaşmak için en düşük polinom derecesine sahip polinom regresyon seçilmiştir.

**Tablo 3.7.** Oluşturulan Modeller için Hata Değerleri (Eğitim Verisi)

| **Model** | **R2** | **MSE** | **RMSE** | **MAE** | **MAPE** |
|---|---|---|---|---|---|
| Doğrusal Regresyon | 0,960 | 0,0298 | 0,1726 | 0,0396 | 0,1297 |
| Ridge ($\alpha = 0{,}5$) | 0,997 | 0,0298 | 0,1726 | 0,0396 | 3,391 |
| Lasso ($\alpha = 10$) | 0,930 | 0,0298 | 0,1726 | 0,0396 | 67,099 |
| Elastic Net | 0,434 | 257602 | 507,545 | 419,157 | 686,529 |
| Polinom Regresyon (derece = 2) | 0,939 | 631,2 | 25,123 | 122,978 | 72,856 |
| K-En Yakın Komşular | 0,975 | 11,188 | 3,344 | 1,349 | 0,239 |
| Karar Ağacı | 0,999 | 0,001 | 0,001 | 0,001 | 0,001 |
| Rastgele Orman | 0,942 | 26317 | 162,227 | 120,773 | 0,001 |
| Yapay Sinir Ağı | 0,872 | 12,712 | 3,565 | 390,417 | 589,167 |

Tablo 4.7'de yer alan eğitim verisine ait değerler incelendiği zaman en iyi performans göstergelerinin karar ağacı modelinde elde edildiği görülmektedir. Ancak sonuçlara bakıldığında verilerin aşırı öğrendiği ve bu durumunun istenmemesinden dolayı tahminleme yöntemi olarak karar ağacı seçilmemiştir. Sonrasında en iyi olan doğrusal regresyon yöntemi seçilmiştir. Bu modele ait test verisi oluşturulan model ile çalıştırıldığında ise elde edilen ilgili performans göstergesinin 0,871 olduğu görülmektedir. Bu da elde edilen modelin talep tahmin etmek için kullanılabilir olduğunu göstermektedir. Bilindiği gibi elde edilen modelin kullanılabilmesi için performans göstergesi R2 değerinin en az 0,7 değerinde olması gerekmektedir. Burada kullanılan fermonas göstergelerinden "MSE", "RMSE", "MAE" ve "MAPE" değerleri hatayı gösteriyorken "R2" değeri modelin tahminleme kabiliyetini göstermektedir. Bu nedenle hata değerlerinin olabildiğince küçük, R2 değerinin ise olabildiğince büyük olması modelin iyi bir performans göstermesi açısından önemlidir. Buna göre elde edilen model doğrultusunda sonraki beş yıl için yolcu sayısı tahmin edilmiştir. Elde edilen bu değerler bir sonraki bölümde anlatına matematiksel modelde talep değerleri olarak girdi verisi olacak şekilde kullanılmıştır.

## 3.4. SEFER SAYISI ARTIRILACAK HATTI BELİRLENMESİ

Bu bölümde artırılması gereken bölgelerin belirlenmesi için kullanılan maksimum küme örtme modeli ile Altunizade-Sultanbeyli hattı arasında yer alan mahallelerde hangi kısımlara sefer eklenmesi gerektiği talep değerlerine göre belirlenecektir. Bu değer bir önceki bölümde yer alan doğrusal regresyon modelinden elde edilmiştir. Zamanlara göre elde edilen talepler daha sonrasında Altunizade-Sultanbeyli arasında yer alan mahaller arasında nüfusa göre orantısal olarak paylaştırılmış ve böylece o ulaşım hattında yer alan mahallerin talepleri elde edilmiştir. Daha sonrasında ise Altunizade-Sultanbeyli arasında yer alan noktalar kullanılarak talebin karşılanacağı noktalar belirlenmiştir. Elde edilen sonuçlara göre Altunizade hattının yer aldığı Üsküdar ilçesinin o hat içinde yoğunluğa sahip olması sebebiyle hat artımı yapılması gerektiği sonucuna ulaşılmıştır.

# 4. BULGULAR

Türkiye'nin 5 te birlik kısmına denk gelen popülasyona sahip olan İstanbul'da yaklaşık 17 farklı ulaşım türü kullanılmaktadır.

Farklı ulaşım türlerine bağlı olarak İstanbul toplu ulaşım talebinin belirlenmesi için 2022 yılına saatlik ulaşım verisi kullanılmıştır. Bu veri kullanılarak doğrusal regresyon, ridge regresyon, lasso, polinom regresyon, elastik net, k-en yakın komşular, karar ağacı regresyonu ve rastgele orman regresyonu yöntemleri eğitim verileri kullanılarak model oluşturulmuştur. Bu işlem sonucunda eğitim verileri ile en iyi modelin lasso regresyonu ile elde edildiği görülmüştür. Test verileri kullanılarak modelin test edilmesi ile beş yıllık talebin belirlenmesi için doğrusal regresyon yönteminin kullanılmasına karar verilmiştir. Bu işlem saatlik toplu taşıma verisinde yer alan Altunizade-Sultanbeyli, Üsküdar-Beykoz ve Üsküdar-Çekmeköy hatları için gerçekleştirilmiştir. Bu 3 hat 954 farklı hat ve toplam 7866727 geçiş sayısı üzerinden elde edilen kümeleme ve sefer sayıları üzerine bulunmuştur. Belirlenen 3 hattın toplamda kapladığı yüzde 0.5852370369532335'tir. Sonrasında bu hatlar içinde yer alan bölgeler kullanılarak talep bölgeleri oluşturulmuş ve doğrusal regresyon modeli sonucunda elde edilen talep değerleri nüfusla orantılı olacak şekilde dağıtılarak maksimum küme örtme modeli çalıştırılmıştır. Elde edilen sonuçlara göre Üsküdar ilçesinin o hat içerisinde hat sayısının artırılması sonucuna ulaşılmıştır.

Yapılan çalışmalar verilen analizler dikkate alındığında aşağıdaki adımlar takip edilerek sonuçlar özetlenebilir:

1. Öncelikle uygulamaların yapılacağı bölge seçilmiştir. Bu aşamada nüfusun en yoğun olduğu ve ulaşım alanında en fazla ihtiyacın yer aldığı İstanbul ili uygulama bölgesi seçilmiştir.

2. İstanbul ili büyük bir bölge olması sebebiyle oluşturulacak modellerin boyutu dikkate alındığında İstanbul'un belli bir bölgesi dikkate alınmıştır. Bu bölgenin belirlenmesi için gözetimsiz öğrenme tekniklerinden biri olan k-means yöntemi seçilmiştir. Bu

yöntemde küme sayısının belirlenmesi için dirsek grafiği kullanılmış ve küme sayısı beş olarak belirlenmiştir. İstanbul ili beş bölgeye ayrıldıktan sonra yoğunluk olarak en fazla yolcuya ulaşan küme seçilmiştir.

3. Bölge seçildikten sonra o bölgede yer alan gelecekteki yolculuk sayısı tahmin edilerek talebin belirlenmesi gerekmektedir. İstanbul toplu ulaşım talebinin belirlenmesi için 2022 yılına saatlik ulaşım verisi kullanılmıştır. Bu veri İBB Açık Veri Portalı (2023) tarafından yayınlanan "Saatlik Toplu Ulaşım Veri Seti'nden elde edilmiştir. Bu veriden gelecekteki yolculuk sayılarının tahmin edilmesi için gözetimli öğrenme tekniklerinden doğrusal regresyon, ridge regresyon, lasso, elastic-net, polinom regresyon, k-en yakın komşular, karar ağaçları, rastgele ormanlar ve yapay sinir ağları yöntemleri kullanılmıştır. Oluşturulan modeller farklı performans göstergeleri kullanılarak karşılaştırılmış ve tahminleme için en uygun yöntem doğrusal regresyon olarak belirlenmiştir. Böylece ulaşım için gelecek beş yıla ait talep belirlenmiştir.

4. Bu işlem için talebin belirlendiği bölge için hangi noktalara verilen ulaşım hizmetinin artırılması gerektiğine dair belirleme işlemi kullanılan matematiksel model ile belirlenmiştir. Gelecekteki yolculuk sayısı belirlendikten sonra son işlem olarak ekleme yapılacak noktaların belirlenmesi için maksimum küme örtme problemi çalıştırılarak Altunizade-Sultanbeyli, Üsküdar-Beykoz ve Üsküdar-Çekmeköy arasında yer alan noktalar talebin karşılanacağı noktalar olarak belirlenmiştir. Elde edilen sonuçlar optimal olarak elde edilmiştir.

# 5. TARTIŞMA

İstanbul sahip olduğu nüfusu ile Türkiye'nin 5'te birlik kısmına denk gelen bir popülasyonu içermektedir. Ancak İstanbul'da yer alan nüfus dikkate alındığında ilin yüz ölçümü olarak yeterli kapasitede olmadığı bu nedenle özellikle şehir merkezinde nüfus yoğunluğunun fazla olduğu görülmektedir. Bu yoğunluğa bağlı olarak bireylerin hayatlarını idame ettirebilmeleri için gerekli olan ihtiyaçların ise karşılanması için iyi bir planlama yapılması gerekmektedir. Bu ihtiyaçlardan biri de bireylerin bir noktadan farklı bir noktaya ulaşımının sağlanmasıdır. Bu tez çalışmasında da öncelikli olarak ulaşım için gelecek beş yıla ait talep belirlenmiştir. Sonrasında ise talebin belirlendiği bölge için hangi noktalara verilen ulaşım hizmetinin artırılması gerektiğine dair belirleme işlemi kullanılan matematikse model ile belirlenmiştir.

Yapılan çalışma işlem adımları olarak takip edildiğinde aşağıdaki basamaklar genel olarak sınıflandırıldığı zaman aşağıdaki gibi iki temel adım takip edilmektedir:

1. Makine öğrenmesi yöntemleri ile tahminleme,
2. Elde edilen veriler ve optimizasyon tekniklerinden olan matematiksel model ile yer seçimi modeli.

Verilen bu adımlar ve mevcut veri seti dikkate alındığında gelecek çalışmalarda bu algoritmalara ek olarak zaman serileri yöntemleri de kullanılarak talep tahmininde kullanılan yöntemlere ait alternatifler geliştirilerek daha fazla model arasından en iyi yöntem belirlenebilir. Böylece daha geniş bir havuzdan model seçimi yapılarak uygun olan en iyi yöntem seçilebilir. Ayrıca zaman serileri ile birlikte mevsimsel etkileri dikkate alacak faktörler de göz önünde bulundurulmuş olur. Ayrıca yapılan çalışma beş yıl sonrasının talebini kapsadığı için bu süre geliştirilerek daha uzun dönemli bir planlama periyodu dikkate alınabilir ve çok periyotlu bir matematiksel model kullanılarak genişletilmesi gereken hatların tek seferde düzenlenmesi yerine kademeli olarak işleme alınması sağlanabilir. Son olarak seçilen

matematiksel modelde farklı bir yöntem kullanılarak maksimum kişiye ulaşmak yerine tüm yolculara ulaşılabilen bir model tercih edilebilir.

# 6. SONUÇ ve ÖNERİLER

Türkiye'nin en fazla nüfusuna sahip ili İstanbul'da ulaşım gibi temel ihtiyaçlar ön plana çıkmakta ve fazla yolcunun yer aldığı şehirde büyük bir önem taşımaktadır. Bu nedenle başta toplu ulaşım olmak üzere ulaşım ağının uygun şekilde kurularak yolcular tarafından oluşan talebin en uygun şekilde sağlanması gerekmektedir. Ulaşımda karşılaşılan problemler, yolcu sayısının fazlalığı, sefer sayısının yeteri kadar olmaması sebebiyle yolcu talebine karşılık verememesi gibi nedenlerden kaynaklanabilmektedir. Örneğin, sefer saatlerinin yolcuların olduğu saatlerde ihtiyaca cevap verilememesi, ulaşımda toplu taşımanın yoğun olarak kullanıldığı saatlerde seferlerin yetersiz kalması gibi durumlar bunlardan bazılarıdır. Belirtilen bu aksaklıklar ulaşımı olumsuz olarak etkilemekte, yolcuların memnuniyetini azalmaktadır. Ulaşımda yaşanan aksaklıkların fazla olması beraberinde organizasyon açısından aksamalara sebep olmakta ve maliyet açısından olumsuz etkileri olmaktadır.

Bu çalışmada süreci etkileyen tüm bu etmenler göz önünde bulundurulduğunda İstanbul Büyükşehir Belediye'si tarafından açık veri portalında paylaşılan "Saatlik Toplu Ulaşım Veri Seti" kullanılarak geçiş sayılarının geleceğe yönelik beş yıllık tahminlemesi yapılmıştır. Bu işlem için makine öğrenmesi başlığı altında yera alan doğrusal regresyon, ridge, lasso, elastic-net, karar ağaçlar, yapay sinir ağları gibi farklı yöntemler kullanılmış ve en iyi sonuçlara doğrusal regresyon ile ulaşılmıştır. Sonrasında elde edilen talep bilgileri maksimum küme örtme modeline talep değeri olarak girilerek Altunizade-Sultanbeyli, Üsküdar-Beykoz ve Üsküdar-Çekmeköy arasında yer alan talep noktalarından nerelere yeni seferler eklenmesi gerektiği belirlenmiştir. Tüm değerler dikkate alındığında Üsküdar noktasında ulaşımın düzenlenmesi ve seferlerin artırılması gerektiği sonucuna ulaşılmıştır. Yapılması muhtemel bu düzenleme ve seferlerin artırılması ile geleceğe yönelik yolcu memnuniyetinin artması ve oluşabilecek karmaşanın azalması öngörülmektedir.

# KAYNAKLAR


Adeniran, A., Kanyio, O., & Owoeye, A. S. (2018). Forecasting methods for domestic air passenger demand in Nigeria.

Alekseev, K., & Seixas, J. (2002). *Forecasting the air transport demand for passengers with neural modelling.* Paper presented at the VII Brazilian Symposium on Neural Networks, 2002. SBRN 2002. Proceedings.

Anvari, S., Tuna, S., Canci, M., & Turkay, M. (2016). Automated Box–Jenkins forecasting tool with an application for passenger demand in urban rail systems. *Journal of Advanced Transportation, 50*(1), 25-49.

Arıkan, D., & Yıldız, F. (2023). Sentinel-2 Uydu Görüntülerinde Destek Vektör Makinesi ve Rastgele Orman Algoritmaları Kullanılarak Piksel Tabanlı Arazi Sınıflandırması. Osmaniye Korkut Ata Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 6(2), 1243-1260.

Ataseven, B. (2013). Yapay sinir ağlari ile öngörü modellemesi. Öneri Dergisi, 10(39), 101-115.

Atlas Big, "Istanbul'un Mahalle Bazinda Nüfus Yogunlugu Haritasi", Erişim: 07.07.2023, https://www.atlasbig.com/tr/istanbul-mahalleleri-nufus-yogunlugu

Carmona-Benítez, R. B., & Nieto-Delfín, M. R. (2015). *Bootstrap estimation intervals using bias corrected accelerated method to forecast air passenger demand.* Paper presented at the Computational Logistics: 6th International Conference, ICCL 2015, Delft, The Netherlands, September 23-25, 2015, Proceedings 6.

Carmona-Benítez, R. B., Nieto, M. R., & Miranda, D. (2017). An Econometric Dynamic Model to estimate passenger demand for air transport industry. *Transportation Research Procedia, 25*, 17-29.

Ceylan, H., & Özcan, T. (2018). Otobüs ağlarındaki sefer sıklıklarının Armoni Araştırması Algoritması ile optimizasyonu: Mandl test ağı üzerine bir uygulama. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi, 24*(6), 1107-1116.

Cyril, A., Mulangi, R. H., & George, V. (2018). Modelling and Forecasting Bus Passenger Demand using Time Series Method. *2018 7th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, 460-466.

Çakır, F. (2020). *Demiryolu yolcu taşıma talebinin yapay sinir ağları ile tahmini.* Aksaray Üniversitesi Fen Bilimleri Enstitüsü.

Current, J., Daskin, M., & Schilling, D. (2002). Discrete network location models. Facility location: Applications and theory, 1, 81-118.

Demirkollu, M. (2017). *Hedef programlama yönetmi ile otobüs sefer sayılarının tespit edilmesi.* Pamukkale Üniversitesi Fen Bilimleri Enstitüsü.

Demirezen, S., & Çetin, M. (2021). Rassal Orman Regresyonu ve Destek Vektör Regresyonu İle Piyasa Takas Fiyatinin Tahmini. Nicel Bilimler Dergisi, 3(1), 1-15.

Deri, A. (2012). *Akıllı kart verileri kullanılarak toplu ulaşım yolculuk talebinin belirlenmesi ve sefer çizelgeleme optimizasyonu.* DEÜ Fen Bilimleri Enstitüsü.

Efendigil, T., & Eminler, Ö. E. (2017). Havacılık sektöründe talep tahminin önemi: Yolcu talebi üzerine bir tahmin modeli. *Yaşar Üniversitesi E-Dergisi, 12*, 14-30.

Engin, T. (2013). *Genetik algoritma ile toplu ulaşım sistemi hareket çizelgesi optimizasyonu: Çanakkale örneği.* Marmara Universitesi (Turkey).

Erpik, Z. (2019). *Hat çizelgeleme problemine genetik algoritma yaklaşımı: İETT örneği.* Sosyal Bilimler Enstitüsü.

Gelhausen, M. C., Berster, P., & Wilken, D. (2018). A new direct demand model of long-term forecasting air passengers and air transport movements at German airports. *Journal of Air Transport Management, 71*, 140-152.

Gencer, M. A., Alakaş, H. M., Tamer, E., & Hamurcu, M. (2018). Ankara Metrosu M1 hattı hareket saatlerinin çizelgelenmesi: Bir karar destek sistemi uygulaması. *Akıllı Ulaşım Sistemleri ve Uygulamaları Dergisi, 1*(2), 108-128.

Gunter, U., & Zekan, B. (2021). Forecasting air passenger numbers with a GVAR model. *Annals of Tourism Research, 89*, 103252.

Hsiao, C.-Y., & Hansen, M. (2011). A passenger demand model for air transportation in a hub-and-spoke network. *Transportation Research Part E: Logistics and Transportation Review, 47*(6), 1112-1125.

İBB Açık Veri Portalı, "Saatlik Toplu Ulaşım Veri Seti", Erişim: 07.07.2023, https://data.ibb.gov.tr/dataset/hourly-public-transport-data-set.

Jafari, N. (2022). The chaos on US domestic airline passenger demand forecasting caused by COVID-19. *International Journal of Business Forecasting and Marketing Intelligence, 7*(3), 241-258.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112): Springer.

Java T point, "Decision Tree Classification Algorithm", Erişim: 07.07.2023, https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm

Jin, F., Li, Y., Sun, S., & Li, H. (2020). Forecasting air passenger demand with a new hybrid ensemble approach. *Journal of Air Transport Management, 83*, 101744.

Kavzoğlu, T., & Çölkesen, İ. (2010). Destek vektör makineleri ile uydu görüntülerinin sınıflandırılmasında kernel fonksiyonlarının etkilerinin incelenmesi. Harita Dergisi, 144(7), 73-82.

Kavzoğlu, T., Şahin, E. K., & Çölkesen, İ. (2012). Heyelan duyarlılığının incelenmesinde regresyon ağaçlarının kullanımı: Trabzon örneği. Harita Dergisi, 147(3), 21-33.

Kim, S. (2016). Forecasting short-term air passenger demand using big data from search engine queries. *Automation in Construction, 70*, 98-108.

Kingsford, C., & Salzberg, S. L. (2008). What are decision trees?. Nature biotechnology, 26(9), 1011-1013.

Kızıltaş, M. Ç., & Akkaya, D. (2022). Yenilikçi Toplu Ulaştırma ve Bu Bağlamda Sıklık Hizmet Parametresinin İncelenmesi. *Avrupa Bilim ve Teknoloji Dergisi*(41), 315-323.

Koç, İ., & Arslan, E. (2018). *Demand Forecasting for Domestic Air Transportation in Turkey using Artificial Neural Networks.* Paper presented at the 2018 6th International Conference on Control Engineering & Information Technology (CEIT).

Korkmaz, E. S., İlyas, Ö., & Efe, S. B. Şehir içi otobüs taşımacılığında kısa dönemli yolcu sayısı tahmini. *Akıllı Ulaşım Sistemleri ve Uygulamaları Dergisi, 5*(2), 133-145.

Marie-Sainte, S. L., Saba, T., & Alotaibi, S. (2019). Air passenger demand forecasting using particle swarm optimization and firefly algorithm. *Journal of Computational and Theoretical Nanoscience, 16*(9), 3735-3743.

Nar, M. (2022). Yenikapı M1–Kirazlı M1 hattı için istasyon ve hat bazlı yolcu talep tahmini ve raylı ulaşım sistemlerinde sefer sıklığı belirlemede kritik başarı faktörlerinin çok kriterli karar verme yöntemleri ile önceliklendirilmesi= Station and line-based passenger demand forecasting for Yenikapi M1-Kirazli M1 line and prioritization of critical success factors in headway determination in railway systems with multi-criteria decision-making methods.

Özcan, T. (2018). *Kentiçi toplu taşıma sistemlerinde sefer sıklığı optimizasyonu.* Pamukkale Üniversitesi Fen Bilimleri Enstitüsü.

Öztürk, K., & Şahin, M. E. (2018). Yapay sinir ağları ve yapay zekâ'ya genel bir bakış. Takvim-i Vekayi, 6(2), 25-36.

Palamutçuoğlu, B. T. (2020). Şehir içi toplu taşıma planlamasında akıllı kart veri madenciliği ile sefer sıklığı optimizasyonu.

Plakandaras, V., Papadimitriou, T., & Gogas, P. (2019). Forecasting transportation demand for the US market. *Transportation Research Part A: Policy and Practice, 126*, 195-214.

Polat, U., Sağbaş, A., & Dermenci, M. S. (2022). Sürdürülebilir ulaşım planlaması için şehir içi otobüs hatlarında sefer çizelgeleme optimizasyonu.

Sağbaş, A., & Polat, U. Doğrusal Hedef Programlama Yaklaşımı İle Şehir İçi Toplu Ulaşım Sistemlerinin Modellenmesi. *European Journal of Engineering and Applied Sciences, 5*(2), 112-121.

Scarpel, R. A. (2013). Forecasting air passengers at São Paulo International Airport using a mixture of local experts model. *Journal of Air Transport Management, 26*, 35-39.

Srisaeng, P., Baxter, G., Richardson, S., & Wild, G. (2015). A forecasting tool for predicting Australia's domestic airline passenger demand using a genetic algorithm. *Journal of Aerospace Technology and Management, 7*, 476-489.

Srisaeng, P., Baxter, G., & Wild, G. (2015). Using an artificial neural network approach to forecast Australia's domestic passenger air travel demand. *World Review of Intermodal Transportation Research, 5*(3), 281-313.

Suryani, E., Chou, S.-Y., & Chen, C.-H. (2010). Air passenger demand forecasting and passenger terminal capacity expansion: A system dynamics framework. *Expert Systems with Applications, 37*(3), 2324-2339.

Şaşmaz, H. *Toplu ulaşımda ana besleme hat modellemesi: İstanbul Mahmutbey Aktarma Merkezi örneği.* Fen Bilimleri Enstitüsü.

Terekhov, I., Evans, A., & Gollnick, V. (2016). *Forecasting a global air passenger demand network using weighted similarity-based algorithms.* Paper presented at the Complex Networks VII: Proceedings of the 7th Workshop on Complex Networks CompleNet 2016.

Yildirim, O., & Demir, S. (2022, October). Mathematical modelling of traffic with regression analysis and machine learning. In Sixth International Conference on Analysis and Applied Mathematics (p. 24).

Uludağ, N. (2010). Bulanık optimizasyon ve doğrusal hedef programlama yaklaşımları ile otobüs hatlarının modellenmesi.

Xiao, Y., Liu, J. J., Hu, Y., Wang, Y., Lai, K. K., & Wang, S. (2014). A neuro-fuzzy combination model based on singular spectrum analysis for air transport demand forecasting. *Journal of Air Transport Management, 39*, 1-11.

Yaşar, A. B. (2009). *Kentiçi otobüs taşımacılığında talep tahmini.* Pamukkale Üniversitesi Fen Bilimleri Enstitüsü.

Yaylali, M., & Dilek, Ö. (2009). Erzurum'da Yolcularin Havayolu Ulaşim Tercihlerini Etkileyen Faktörlerin Tespiti. *Marmara Üniversitesi İktisadi ve İdari Bilimler Dergisi, 26*(1), 1-21.

Zhang, X., Zheng, Y., & Wang, S. (2019). A demand forecasting method based on stochastic frontier analysis and model average: An application in air travel demand forecasting. *Journal of Systems Science and Complexity*, *32*(2), 615-633.

Zhao, S., & Mi, X. (2019). A Novel Hybrid Model for Short-Term High-Speed Railway Passenger Demand Forecasting. *IEEE Access*, *7*, 175681-175692.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, *67*(2), 301-320.