



Teknoloji Fakültesi

## BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ

# Makine Öğrenimi ile Uçuş İptal Gecikme Tahmini ve Akıllı Uçak Bilet Sistemi

**BİTİRME PROJESİ**

Bilgisayar Mühendisliği Bölümü

**DANIŞMAN**

Dr. Öğr. Üyesi EYÜP EMRE ÜLKÜ

İSTANBUL, 2025



Teknoloji Fakültesi

## BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ

# Makine Öğrenimi ile Uçuş İptal Gecikme Tahmini ve Akıllı Uçak Bilet Sistemi

**BİTİRME PROJESİ**

Bilgisayar Mühendisliği Bölümü

**DANIŞMAN**

Dr. Öğr. Üyesi EYÜP EMRE ÜLKÜ

İSTANBUL, 2025

**MARMARA ÜNİVERSİTESİ**  
**TEKNOLOJİ FAKÜLTESİ**  
**BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ**

Marmara Üniversitesi Teknoloji Fakültesi Bilgisayar Mühendisliği Öğrencileri Nuri Can Birdemir ve Eda Nur Mutlu tarafından “Makine Öğrenimi ile Uçuş İptal Gecikme Tahmini ve Akıllı Uçak Bilet Sistemi” başlıklı proje çalışması, 19.03.2025 tarihinde savunulmuş ve jüri üyeleri tarafından başarılı bulunmuştur.

**Jüri Üyeleri**

Dr. Öğr. Üyesi xxx xxx  
Marmara Üniversitesi  
Prof. Dr. Xxx xxx  
Marmara Üniversitesi  
Prof. Dr. Xxx xxx  
Marmara Üniversitesi

**(Danışman)**

(Üye)

(Üye)

(İMZA).....

(İMZA).....

(İMZA).....

## **ÖNSÖZ**

Proje çalışmamız süresince karşılaştığımız bütün problemlerde, sabırla yardım ve bilgilerini esirgemeyen, tüm desteğini sonuna kadar yanımda hissettiğimiz sayın Dr. Öğr. Üyesi Eyüp Emre Ülkü' ye en içten teşekkürlerimizi sunarız.

# İÇİNDEKİLER

<b>1. GİRİŞ</b>	<b>1</b>
<b>1.1. Proje Çalışmasının Amacı ve Önemi</b>	<b>2</b>
<b>1.2. Literatür Taraması</b>	<b>3</b>
<b>2. MATERYAL VE YÖNTEM</b>	<b>7</b>
<b>2.1. Veri Seti</b>	<b>7</b>
<b>2.2. Veri Ön İşleme</b>	<b>12</b>
<b>2.3. Denetimli Öğrenme Modelleri</b>	<b>16</b>
<b>3. BULGULAR VE TARTIŞMA</b>	<b>21</b>
<b>3.1. Projede Şu Ana Kadar Yapılanlar</b>	<b>21</b>
<b>3.2. Projede Bundan Sonra Yapılacaklar</b>	<b>22</b>
<b>4. SONUÇLAR</b>	<b>22</b>

## ÖZET

Bu çalışma, uçuş iptali ve gecikmelerini tahmin edebilen bir makine öğrenimi modeli geliştirmeyi amaçlamaktadır. Hava yolu taşımacılığında, beklenmedik iptaller ve gecikmeler hem yolcular hem de havayolu şirketleri için önemli sorunlara yol açmaktadır. Bu bağlamda, geçmiş uçuş verileri, hava durumu bilgileri ve uçuş detayları kullanılarak tahminleme modelleri oluşturulmuştur.

Çalışmada öncelikle, Amerika Birleşik Devletleri Ulaştırma Bakanlığı (DOT) tarafından sağlanan açık uçuş verileri incelenmiş ve uygun veri setleri bir araya getirilmiştir. Daha sonra veri temizleme, eksik verileri giderme, zaman formatlarını düzenleme ve özellik mühendisliği gibi veri ön işleme teknikleri uygulanmıştır.

Makine öğrenimi aşamasında XGBoost, Random Forest, Decision Tree, KNN ve Gradient Boosting algoritmaları kullanılarak tahmin modelleri geliştirilmiştir. Modelin performansını artırmak amacıyla SMOTE yöntemi ile veri dengelenmiş ve modellerin doğruluk, precision, recall, f1-score ve ROC-AUC gibi metrikleri hesaplanmıştır. Son olarak, geliştirilen modelin uçuş iptali ve gecikme tahmininde ne kadar başarılı olduğu değerlendirilmiştir.

Çalışmanın sonucunda, uçuş iptali ve gecikmeleri tahmin edebilen başarılı bir model oluşturulmuş ve bu modelin uçak bileti satın alacak yolculara yardımcı olabilecek bir web platformuna entegre edilmesi planlanmıştır.

**Mart, 2025**

**Öğrenciler**

**Nuri Can Birdemir**

**Eda Nur Mutlu**

## **ABSTRACT**

This study aims to develop a machine learning model capable of predicting flight cancellations and delays. In air transportation, unexpected cancellations and delays cause significant problems for both passengers and airline companies. In this context, prediction models have been created using historical flight data, weather information, and flight details.

Initially, open flight data provided by the U.S. Department of Transportation (DOT) was examined, and relevant datasets were compiled. Various data preprocessing techniques were applied, including data cleaning, handling missing values, adjusting time formats, and feature engineering.

During the machine learning phase, XGBoost, Random Forest, Decision Tree, KNN, and Gradient Boosting algorithms were used to develop prediction models. To improve model performance, the SMOTE method was applied to balance the dataset, and evaluation metrics such as accuracy, precision, recall, f1-score, and ROC-AUC were calculated. Finally, the effectiveness of the developed model in predicting flight cancellations and delays was assessed.

As a result of this study, a successful model capable of predicting flight cancellations and delays was created. Future work includes integrating this model into a web platform to assist passengers in making informed flight booking decisions.

**March, 2025**

**Students**

**Nuri Can Birdemir**

**Eda Nur Mutlu**

## ŞEKİL LİSTESİ

<b>Şekil 1</b> Gecikme Durumu Grafiği	6
<b>Şekil 2</b> Hava Yolu Şirketlerine Göre Gecikme Durumu Grafiği	7
<b>Şekil 3</b> Gecikme Faktörü Dağılımı	7
<b>Şekil 4</b> Gecikme Faktörlerinin Isı Haritası	8
<b>Şekil 5</b> Uçuş İptal Tahmin Modeli Özellik Önem Derecesi	15
<b>Şekil 6</b> Uçuş İptal Kodu Tahmin Modeli Özellik Önem Derecesi	15
<b>Şekil 7</b> Uçuş Gecikme Tahmin Modeli Özellik Önem Derecesi	16



## TABLO LİSTESİ

<b>Tablo 1</b> Orijinal Veri Setinin Yapısı	5
<b>Tablo 2</b> Model Eğitim Veri Setinin Yapısı	9

## 1. GİRİŞ

Havayolu taşımacılığı, küresel ölçekte hızla büyüyen bir sektör olmasına rağmen, uçuş gecikmeleri endüstrinin karşılaştığı en önemli sorunlardan biri olmaya devam etmektedir. Uçuş gecikmeleri, yolcu memnuniyetini doğrudan etkilerken, operasyonel aksaklıklar nedeniyle havayolu şirketlerini ve havalimanlarını ekonomik kayıplarla karşı karşıya bırakmaktadır. Bununla birlikte, gecikmeler havalimanı trafiğini yoğunlaştırarak zincirleme gecikmelere yol açmakta ve hava sahası yönetimini karmaşık hale getirmektedir. Pandemi döneminde, havayolu taşımacılığına olan talepteki dalgalanmalar, uçuş gecikmelerine ve iptallere yönelik analizlerin önemini bir kez daha gözler önüne sermiştir. Bu tür sorunlara çözüm üretmek, hem havayolu şirketlerinin hem de yolcuların operasyonel ve bireysel ihtiyaçlarına cevap verebilecek yenilikçi yaklaşımları gerekli kılmaktadır.

Bu proje, uçuş gecikmelerini ve iptallerini tahmin etmek için makine öğrenimi ve yapay zeka yöntemlerini kullanarak bir model geliştirmeyi hedeflemektedir. Projenin ayırt edici yönlerinden biri, uçuş bilgileriyle birlikte Meteostat API gibi meteorolojik veri kaynaklarının entegrasyonunun sağlanmasıdır. Hava durumu, uçuş gecikmelerinin en belirgin nedenlerinden biri olduğu için, bu verilerin kullanımı modeli daha doğru ve güvenilir hale getirecektir. Literatürde sıklıkla yalnızca uçuş geçmişi veya operasyonel veriler kullanılırken, bu çalışma, hava durumu verilerini detaylı bir şekilde entegre ederek yenilikçi bir yaklaşım sunmaktadır.

Geliştirilen model, uçuş gecikme tahminlerini doğrudan bilet satış platformuna entegre ederek kullanıcı deneyimini iyileştirmeyi amaçlamaktadır. Yolcular, uçuş bilgilerini girdiklerinde, uçuşlarının gecikme riski hakkında tahminler alabilecek ve bu doğrultuda seyahat planlamalarını optimize edebileceklerdir. Alternatif uçuş önerileri, bilet fiyat analizi ve gecikme riskine göre daha iyi planlama seçenekleri sunularak yolculara değer katılacaktır. Aynı zamanda, havayolu şirketleri, operasyonel süreçlerini daha verimli hale getirebilecek, gecikme kaynaklarını daha iyi yönetebilecek ve gelir kayıplarını minimize edebileceklerdir.

Makine öğrenimi yöntemleri, büyük ve karmaşık veri setlerinden anlamlı bilgiler çıkarma konusunda güçlü bir araçtır. Bu proje kapsamında, uçuş gecikme ve iptallerini tahmin etmek için XGBClassifier algoritması tercih edilmiştir. XGBClassifier, yüksek doğruluk

oranı ve genellenebilirlik özellikleriyle öne çıkmaktadır. Bunun yanı sıra, bu algoritmanın esnek yapısı, modelin sürekli olarak yeni verilerle güncellenebilmesini sağlayarak gelecekteki performansını artırmakta ve dinamik bir çözüm sunmaktadır. Geleneksel istatistiksel yöntemlere kıyasla, XGBClassifier daha karmaşık ilişkileri modelleyebilme kabiliyeti sayesinde, uçuş gecikmeleri ve iptalleri gibi çok faktörlü problemlerde daha üstün bir performans sergilemektedir.

Çalışma, üç ana hedef kitleye değer yaratmayı amaçlamaktadır:

1. Yolcular: Gecikme riskine dair erken uyarılar alarak seyahatlerini daha bilinçli planlayabilecek ve alternatif uçuş seçeneklerine erişim sağlayabileceklerdir.
2. Havayolu Şirketleri: Gecikme kaynaklarını daha iyi analiz ederek operasyonel süreçlerini optimize edebilecek ve müşteri memnuniyetini artıracaklardır.
3. Havalimanı Yönetimi ve Sivil Havacılık Otoriteleri: Gecikme tahminlerine dayalı olarak uçuş planlamasını daha verimli hale getirebileceklerdir.

Sonuç olarak, bu proje, uçuş gecikmelerine yönelik çözüm arayışında makine öğrenimi teknolojilerinin sağladığı yenilikçi imkanlardan faydalanmaktadır. Projenin küresel havayolu taşımacılığına katkı sağlayacak bir model sunması ve özellikle pandemi döneminden sonraki süreçte karşılaşılan operasyonel zorluklara cevap verebilmesi hedeflenmektedir.

### **1.1. Proje Çalışmasının Amacı ve Önemi**

Bitirme projesinin temel amacı, havayolu uçuş gecikmelerini, iptallerini ve iptal uçuşların nedenini tahmin eden, makine öğrenimi tabanlı bir model geliştirmektir. Bu doğrultuda, uçuş gecikmelerini etkileyen faktörler kapsamlı bir şekilde analiz edilerek, faktörler arasındaki karmaşık ilişkiler ortaya çıkarılacaktır. Çalışma, özellikle meteorolojik koşulların gecikmelere olan etkisini derinlemesine incelemeyi hedeflemektedir. Bu kapsamda, Meteostat API'si aracılığıyla elde edilen hava durumu verileri, uçuş verileriyle entegre edilerek tahmin modelinin doğruluğu artırılacaktır.

Projenin teknik boyutunda, farklı makine öğrenimi algoritmalarının performansları karşılaştırmalı olarak değerlendirilecektir. Bu kapsamda XGBClassifier, Random Forest Classifier, Decision Tree Classifier, KNN Classifier ve Gradient Boosting Classifier algoritmaları kullanılacaktır. Algoritmaların performans karşılaştırması sonucunda en

yüksek başarıyı gösteren model seçilerek, bu model üzerinde optimizasyon çalışmaları gerçekleştirilecektir. Ön değerlendirmelerde XGBClassifier algoritmasının karmaşık veri setlerindeki örüntüleri etkili bir şekilde öğrenebilme ve yeni verilere hızla adapte olabilme özelliğiyle öne çıkması beklenmektedir. Geliştirilen model, çeşitli performans metrikleriyle değerlendirilerek, tahmin doğruluğunun sürekli iyileştirilmesi sağlanacaktır.

Çalışmanın pratik uygulama boyutunda, geliştirilen tahmin modelinin bilet satış platformlarına entegrasyonu hedeflenmektedir. Bu entegrasyon sayesinde, yolculara uçuş seçimi aşamasında gecikme risk değerlendirmesi sunulacak, alternatif uçuş rotaları önerilecek ve gecikme riskine dayalı dinamik fiyatlandırma stratejileri geliştirilecektir. Bu yaklaşım, yolcuların seyahat planlamalarını daha bilinçli yapabilmelerini sağlarken, havayolu şirketlerinin müşteri memnuniyetini artırmasına da katkıda bulunacaktır.

Havayolu şirketleri açısından proje, operasyonel verimliliği artırmaya yönelik bir karar destek sistemi oluşturmayı amaçlamaktadır. Bu sistem, gecikmelerin kök nedenlerini analiz ederek, operasyonel maliyetlerin optimize edilmesine ve gecikmelerin önlenmesine yönelik erken müdahalelerin yapılmasına olanak sağlayacaktır. Böylece havayolu şirketleri, veri odaklı kararlar alarak operasyonel süreçlerini iyileştirebilecek ve rekabet avantajı elde edebilecektir.

Proje, hem akademik literatüre özgün bir katkı sağlamayı hem de havacılık sektöründeki paydaşlar için pratik değer yaratmayı hedeflemektedir. Geliştirilen model ve metodoloji, gelecekteki çalışmalar için bir referans noktası oluşturarak, havacılık sektöründeki gecikme tahmin çalışmalarına yeni bir perspektif kazandıracaktır.

## **1.2. Literatür Taraması**

Günümüzde havayolu taşımacılığına olan talebin hızla artması, uçuş gecikmelerinin analizini önemli bir araştırma alanı haline getirmiştir. Araştırmacılar, uçuş gecikmelerini tahmin etmek için makine öğrenimi ve veri madenciliği tekniklerinden yaygın olarak yararlanmaktadır. Literatürdeki çalışmalar genellikle havalimanı tesislerinin konumu, hava durumu ve havalimanı kapasitesi gibi faktörlere odaklanmıştır. Makine öğrenimi teknikleri, büyük ölçekli veri setlerinin depolanmasını ve işlenmesini mümkün kılarak bu alanda önemli katkılar sunmaktadır [1]. Bununla birlikte, mevcut araştırmaların çoğu

belirli bir coğrafi bölgeye veya sınırlı sayıda faktöre odaklanmış olup, gecikme tahmininde bütüncül yaklaşımların geliştirilmesi konusunda eksik kalmıştır.

Bu çalışma, uçuş gecikmeleri için tahmin modellerine yeni bir perspektif kazandırmayı hedeflemektedir. Literatürdeki boşluğu doldurmak amacıyla, uçuş verilerinin yanı sıra meteorolojik veriler gibi çeşitli faktörlerin entegre edildiği kapsamlı bir veri seti kullanılmaktadır. Özellikle Meteostat gibi geniş kapsamlı meteorolojik veri kaynaklarının kullanımı, hava durumu etkilerinin daha hassas bir şekilde modellenmesine olanak tanımaktadır. Delahaye ve Puechmorel'in (2020) çalışmasında vurgulandığı gibi, literatürde genellikle ihmal edilen veya yüzeysel olarak ele alınan hava durumu faktörleri, bu çalışma kapsamında hava durumu faktörlerinin uçuş gecikmeleri üzerindeki etkisi derinlemesine analiz edilerek modelin doğruluğu ve güvenilirliği artırılmaktadır. [2].

Metodolojik açıdan, çalışmada Random Forest ve XGBoost gibi gelişmiş algoritmaların yanı sıra veri dengesizliği sorununu çözmek için SMOTE yöntemi uygulanmaktadır. Chawla ve diğerlerinin (2002) ortaya koyduğu gibi, veri dengesizliği problemi, gecikmelerin ve iptallerin nadir gözlemler içermesi nedeniyle literatürde önemli bir sorun olarak öne çıkmaktadır [3]. Bu çalışma, veri dengesizliği problemini sistematik bir şekilde ele alarak daha dengeli, genellenebilir ve yüksek doğrulukta tahmin modelleri geliştirmeyi amaçlamaktadır.

Khaksar ve Sheikholeslami yaptıkları çalışmada, hava yolu gecikmelerini tahmin etmek için çeşitli makine öğrenme algoritmaları uygulamaktadır [4]. ABD ve İran uçuş ağlarından elde edilen veriler üzerinde Decision Trees, Random Forest, kümeleme ve Bayesian sınıflandırma gibi yöntemler denenmiş ve gecikmelerin nedenlerini daha doğru tahmin etme üzerine yoğunlaşmıştır. Özellikle ABD uçuş ağında görüş mesafesi ve rüzgar hızının gecikmeleri önemli ölçüde etkilediği, İran uçuş ağında ise filo yaşı ve uçak tipi gibi faktörlerin öne çıktığı belirtilmiştir. Elde edilen tahminlerde karar ağaçları ve kümeleme yöntemleri ile hibrit bir sınıflandırma kullanılarak yaklaşık olarak %70 doğruluk oranına ulaşılmıştır.

Bojia Ye ve arkadaşları yaptıkları çalışmada, havaalanlarındaki uçuş gecikmelerini tahmin etmek için denetimli öğrenme yöntemlerini kullanan bir metodoloji önermektedir [5]. Çalışmada, Nanjing Lukou Uluslararası Havaalanı'ndan elde edilen operasyonel uçuş verileri ve hava durumu bilgileri işlenmiş ve tahmin modelleri için dört tür havaalanı ile

ilgili özellik oluşturulmuştur. Çalışma, özellikle 1 saatlik gecikme tahminlerinde yüksek doğruluk sağlamış ve LightGBM modeli %86,55 doğruluk oranıyla en iyi sonuçları vermiştir. Bu modelin sonuçları, operasyonel ve hava koşullarına göre tahmin yapılmasının önemini vurgulamaktadır. Benzer bir çalışmada Atlıoğlu, Türkiye'nin önde gelen bir havayolu şirketinden elde edilen operasyonel veri setini kullanarak, 11 farklı makine öğrenimi modeliyle değerlendirme yapmıştır [6]. Her model için çeşitli performans ölçütlerini karşılaştırarak, en yüksek doğruluğa ulaşmak için veri setindeki en uygun özellikleri belirlemeye çalışmıştır.

Shahinaz M. Al-Tabbakh ve arkadaşları, Mısır Hava Yolları'nın uçuş gecikmesi verilerini analiz etmek için çeşitli makine öğrenmesi tekniklerini uygulamışlardır [7]. Çalışmanın temel amacı, uçuş gecikmelerinin tahmin edilebilmesi için en uygun sınıflandırma algoritmasını belirlemektir. Araştırmacılar, veri hazırlama, sınıflandırma algoritmaları kullanma ve model performansını değerlendirme adımlarını içeren bir metodoloji izlemişlerdir. Sekiz farklı sınıflandırma algoritması (Karar Ağacı, Random Forest, REPTree, PART, Karar Tablosu, OneR, JRip) WEKA veri madenciliği aracında uygulanarak karşılaştırılmıştır. Sınıflandırma modellerinin performansı doğruluk, hassasiyet, özgülük, F1-skoru ve ROC alanı metrikleri kullanılarak değerlendirilmiştir. Analiz sonuçlarına göre, PART algoritması %83,1 doğruluk oranıyla en yüksek performansa sahip olmuştur. Diğer bir yandan, REPTree algoritması da %80,3 doğruluk oranı ve en hızlı çalışma süresiyle öne çıkan ağaç tabanlı sınıflandırıcı olarak belirlenmiştir. Çalışma, uçuş gecikmesi tahmininde makine öğrenmesi tekniklerinin etkinliğini göstermiş ve Mısır Hava Yolları'na uçuş operasyonlarını iyileştirmek için önemli bilgiler sağlamıştır. Araştırmacılar, gelecekte daha büyük veri kümeleri kullanarak büyük veri madenciliği teknolojilerini uygulamayı planlamaktadırlar.

Kurt'un çalışmasında, ABD iç hat uçuş verilerini kullanarak uçuş gecikmelerinin öngörülmesi amaçlanmıştır [8]. Bu doğrultuda, karar ağaçları, rastgele orman, torbalama, ekstra ağaçlar, Gradient Boosting ve XGBoost sınıflandırıcı gibi çeşitli makine öğrenmesi yöntemleri denenmiştir. Bu modellerin doğruluk, F1 skoru ve geri çağırma gibi başarı ölçütleriyle değerlendirildiği çalışmada, en yüksek doğruluk oranı %71,72 ile Gradient Boosting algoritması tarafından sağlanmıştır. Çalışma, farklı veri özelliklerini modele dahil ederek tahmin performansının iyileştirilebileceğini önermektedir. Tang'ın araştırmasında ise New York JFK Havalimanı'ndan kalkan uçuşların bir yıllık

verileri kullanılarak uçuş gecikmelerini tahmin etmek amacıyla yedi farklı sınıflandırma algoritması değerlendirilmiştir [9]. Bu algoritmalar arasında karar ağacı, %97,78 doğruluk oranı ile en yüksek başarıyı sağlamıştır. Özellikle ağaç tabanlı sınıflandırıcılar olan rastgele orman ve gradient boosting yöntemlerinin, diğer temel sınıflandırıcılara göre daha yüksek performans sergilediği tespit edilmiştir. Çalışmada verilerin dengesiz dağılımının ağırlıklı doğruluk gibi ölçütlerle giderildiği belirtilmiştir.

Meteorolojik verilerin entegrasyonu, gelişmiş makine öğrenimi algoritmalarının kullanımı ve veri dengesizliği probleminin sistematik şekilde ele alınmasını içeren bu çok yönlü yaklaşımla, havayolu sektöründeki operasyonel süreçleri iyileştirecek daha hassas ve genellenebilir tahmin modellerinin geliştirilmesi hedeflenmektedir. Literatürdeki mevcut çalışmaların sınırlamaları göz önüne alındığında, “Makine Öğrenimi Tabanlı Uçuş Gecikme Risk Analizi Öngörülü Bilet Satış Platformu” isimli bitirme tezi kapsamında hem teorik hem de pratik açıdan önemli katkılar sağlayacağı öngörülmektedir. Özellikle, meteorolojik verilerin kapsamlı entegrasyonu ve modern makine öğrenimi tekniklerinin kullanımı, gelecekteki araştırmalar için yeni bir çerçeve sunma potansiyeli taşımaktadır.

Bu araştırma kapsamında, 2018-2023 yılları arasında yayınlanan ve uçuş gecikmeleri üzerine makine öğrenimi yaklaşımlarını inceleyen çalışmalar sistematik olarak analiz edilmiştir. İncelenen çalışmaların çoğunluğunun tek bir havayolu veya havalimanına odaklandığı, meteorolojik verileri sınırlı düzeyde kullandığı ve veri dengesizliği problemini yeterince ele almadığı tespit edilmiştir. Çoğu mevcut çalışma, yalnızca belirli coğrafi bölgelerle sınırlı verileri veya tek bir faktör grubunu ele almakta ve geniş veri entegrasyonunu yeterince dikkate almamaktadır. Ayrıca, veri dengesizliği sorunu literatürde sıkça karşılaşılan bir engel olmakla birlikte, modern tekniklerin bu bağlamda yeterince uygulanmadığı gözlemlenmiştir.

Bu çalışmada, diğer çalışmalardan farklı olarak uçuş gecikmesi tahmininde kullanılan parametreler çeşitlendirilmiş ve genişletilmiştir. Tahmin modeli geliştirilirken uçuş kodu, hava şartları (örneğin, rüzgar hızı, sıcaklık, yağış durumu), önceki uçuşların iptal/gecikme oranları, uçak tipi ve uçuş yoğunluğu gibi parametreler dikkate alınmıştır. Bununla birlikte, uçuş gecikmelerini sadece "var" ya da "yok" şeklinde sınıflandırmak yerine, belirli dakika aralıklarında tahmin yapılması sağlanmıştır. Örneğin, bir uçuşun 15-30

dakika veya 30 dakika üzeri gecikme yaşama ihtimali, modelin güven oranına dayalı olarak tahmin edilebilmektedir. Bu yaklaşım, yalnızca uçuş iptali üzerine yoğunlaşan önceki çalışmalardan farklılaşarak, gecikme tahmini konusunda daha detaylı ve eyleme geçirilebilir çıktılar sunmaktadır. Ayrıca, bu çalışmada geniş veri entegrasyonu sağlanarak uçuş verileri ve Meteostat gibi kapsamlı meteorolojik veri kaynakları bir araya getirilmiştir. Bu yaklaşım, hava durumu faktörlerinin uçuş gecikmeleri üzerindeki etkisini daha hassas bir şekilde modellemeye olanak tanımaktadır. Veri dengesizliği sorunu ise SMOTE gibi modern veri işleme teknikleri kullanılarak ele alınmış ve modelin performansı artırılmıştır. Sonuç olarak, bu çalışma literatürdeki boşluğu doldurarak uçuş gecikmelerini tahmin etmek için dengeli, yüksek doğruluklu ve güvenilir bir model sunmayı hedeflemektedir. Bunun yanı sıra, önerilen modelin detaylı tahmin kabiliyetleri ve geniş veri entegrasyonu sayesinde gelecekteki araştırmalar için önemli bir temel oluşturacağı öngörülmektedir.

## **2. MATERYAL VE YÖNTEM**

Bu bölümde, çalışmanın gerçekleştirilmesinde kullanılan veri seti, veri ön işleme aşamaları ve denetimli öğrenme modelleri hakkında detaylı bilgi verilecektir. İlk olarak, uçuş ve hava durumu verilerinden oluşan veri seti tanıtılacak ve bu verilerin nasıl entegre edildiği açıklanacaktır. Ardından, veri ön işleme süreci ele alınacak; eksik veri, aykırı değerler ve dengesiz veri gibi sorunlara yönelik uygulanan adımlar anlatılacaktır. Son olarak, tahmin modeli oluşturulmasında kullanılan denetimli öğrenme algoritmalarına (örneğin, XGBClassifier, Random Forest ve XGBoost) odaklanılacak ve bu modellerin nasıl eğitildiği ve değerlendirildiği açıklanacaktır. Bu sayede, çalışmanın metodolojik yaklaşımı ve kullanılan teknikler kapsamlı bir şekilde sunulacaktır.

### **2.1. Veri Seti**

Bu çalışmada kullanılan veri seti, Amerika Birleşik Devletleri Ulaştırma Bakanlığı (DOT) ve Ulaştırma İstatistikleri Bürosu (Bureau of Transportation Statistics) tarafından sağlanan ve 2016 ile 2024 yılları arasındaki uçuş gecikme ve iptal verilerini içermektedir. Veri seti, DOT'nun "On-Time: Reporting Carrier On-Time Performance" (1987-günümüz) veritabanından alınmış olup, uçuş güzergahları (kalkış ve varış noktaları), olay



zaman aralıkları (dakika, yerel saat), gecikme ve iptal nedenleri gibi değişkenleri içermektedir. Bu veriler 3 açık kaynak verileri birleştirilerek oluşturulmuştur. İlk olarak <https://www.kaggle.com/datasets/patrickzel/flight-delay-and-cancellation-dataset-2019-2023/code> adresindeki açık verilerden iptal olan ve iptal olmayan uçuş verileri elde edilmiştir. Veri seti, 1.360.878 satır ve 32 sütun uçuş bilgisi içermektedir. Diğer bir veri seti <https://www.kaggle.com/datasets/threnjen/2019-airline-delays-and-cancellations/data> adresindeki açık verilerden sadece iptal edilmiş uçuşlar olan 64.097 satır verinin uçuş bilgileri alınmıştır. Son veri seti ise <https://www.kaggle.com/datasets/shubhamsingh42/flight-delay-dataset-2018-2024> adresinden açık verilerden sadece iptal edilmiş uçuşlar olan 21.824 satır verinin uçuş bilgileri alınmıştır. Toplam son veri setindeki satır sayısı 1.446.799 olmaktadır. Bu 3 benzer veri seti özelliklerinin ayrıntılı bir açıklaması Tablo 1'de sunulmuştur.

Çalışmada, bu veri seti ile birlikte, Metostat kütüphanesinden elde edilen hava durumu verileri de birleştirilmiştir. Metostat kütüphanesi, her bir havalimanı için coğrafi koordinatlar kullanılarak günlük bazda toplanan meteorolojik veriler ile uçuş gecikme ve iptalleri üzerinde hava koşullarının etkisini incelenmesine olanak tanımaktadır. Bu iki veri setinin entegrasyonu, uçuş performansını daha kapsamlı bir şekilde analiz edebilmek için önemli bir adım olmuştur.

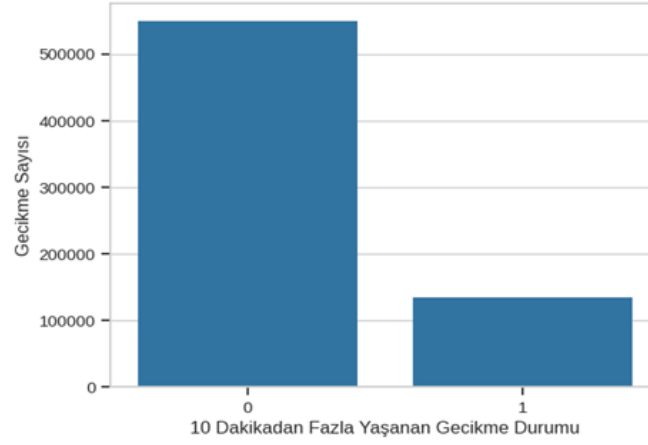
Çalışmada kullanılan veri kaynağı olan Meteostat kütüphanesi, çeşitli meteorolojik istasyonlardan elde edilen hava durumu verilerini sağlamaktadır. Bu veriler, günlük minimum ve maksimum sıcaklık (°C), toplam yağış miktarı (mm), kar yağışı miktarı (mm), rüzgar yönü (0-360°), ortalama ve en yüksek rüzgar hızı (km/saat), ortalama deniz seviyesi basıncı (hPa) ve toplam günlük güneşlenme süresi (dakika) gibi havacılık operasyonlarını doğrudan etkileyebilecek temel meteorolojik parametreleri içermektedir. Her bir havalimanı için coğrafi koordinatlar kullanılarak günlük bazda toplanan bu veriler, uçuş gecikmeleri ve iptalleri üzerinde hava koşullarının etkisini analiz etmek için ana veri setiyle entegre edilmiştir. Bu entegrasyon, uçuş operasyonlarının performansını etkileyen faktörlerin daha kapsamlı bir şekilde incelenmesine olanak tanımaktadır.

Tablo 1 Orijinal Veri Setinin Yapısı

Güncellenmiş Başlık	Kaynak Başlık	Veri Tipi	Açıklama
0 FL_DATE	FlightDate	object	Uçuş tarihi

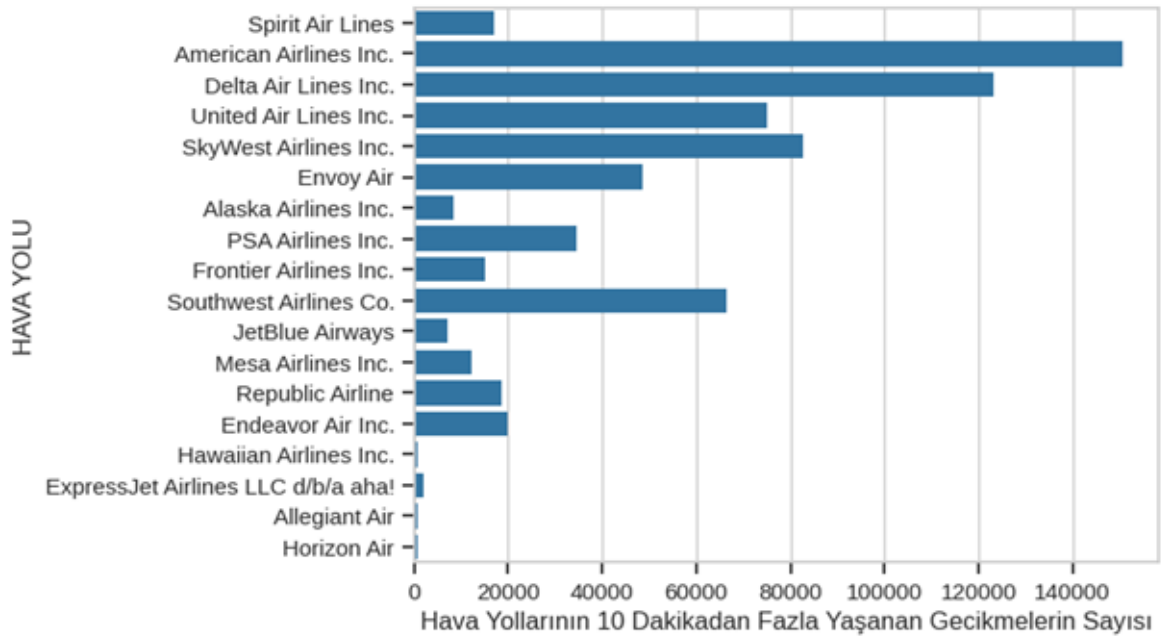
1	AIRLINE	Airline	object	Havayolu şirketinin adı
2	AIRLINE_DOT	AirlineDot	object	Havayolu için DOT tanımlayıcısı
3	AIRLINE_CODE	Reporting_Airline	object	Havayolu şirketinin adı
4	DOT_CODE	DOT_ID_Reporting_Airline	int64	Havayolu için DOT tanımlayıcısı
5	FL_NUMBER	Flight_Number_Reporting_Airline	int64	Uçuş numarası
6	ORIGIN	Origin	object	Çıkış havalimanı kodu
7	ORIGIN_CITY	OriginCityName	object	Çıkış havalimanı şehri
8	DEST	Dest	object	Hedef havaalanı kodu
9	DEST_CITY	DestCityName	object	Varış havalimanının bulunduğu şehir
10	CRS_DEP_TIME	CRSDepTime	int64	Planlanan kalkış saati
11	DEP_TIME	DepTime	float64	Gerçek kalkış saati
12	DEP_DELAY	DepDelay	float64	Kalkış gecikmesi
13	TAXI_OUT	TaxiOut	float64	Taksi yaparken harcanan zaman
14	WHEELS_OFF	WheelsOff	float64	Uçağın tekerleklerinin yerden ayrıldığı zaman
15	WHEELS_ON	WheelsOn	float64	Uçağın tekerleklerinin yere değdiği zaman
16	TAXI_IN	TaxiIn	float64	Taksi yaparken harcanan zaman
17	CRS_ARR_TIME	CRSArrTime	int64	Planlanan varış saati
18	ARR_TIME	ArrTime	float64	Gerçek varış zamanı
19	ARR_DELAY	ArrDelay	float64	Varış gecikmesi
20	CANCELLED	Cancelled	float64	Uçuşun iptal edilip edilmediğinin göstergesi (iptal için 1, iptal değil için 0)
21	CANCELLATION_CODE	CancellationCode	object	İptal nedeni (varsa)
22	DIVERTED	Diverted	float64	Uçuşun yönlendirilip yönlendirilmediğinin göstergesi (yönlendirildi için 1, yönlendirilmedi için 0)
23	CRS_ELAPSED_TIME	CRSElapsedTime	float64	Planlanan geçen süre
24	ELAPSED_TIME	ActualElapsedTime	float64	Gerçek geçen süre
25	AIR_TIME	AirTime	float64	Havada geçirilen zaman
26	DISTANCE	Distance	float64	Katedilen mesafe
27	DELAY_DUE_CARRIER	CarrierDelay	float64	Taşıyıcı nedeniyle gecikme
28	DELAY_DUE_WEATHER	WeatherDelay	float64	Hava koşulları nedeniyle gecikme
29	DELAY_DUE_NAS	NASDelay	float64	Ulusal Hava Sahası Sistemi (NAS) nedeniyle gecikme
30	DELAY_DUE_SECURITY	SecurityDelay	float64	Güvenlik nedeniyle gecikme
31	DELAY_DUE_LATE_AIRCRAFT	LateAircraftDelay	float64	Uçağın geç varması nedeniyle gecikme

Orijinal veri seti üzerinde yapılan analizler sonucunda Şekil 1’de de görüldüğü gibi 10 dakikadan fazla gecikme yaşayanlar “1”, yaşamayanlar ise “0” olarak nitelendirilerek gecikme durumunu ele alan bir grafiğe yer verilmektedir.



Şekil 1 Gecikme Durumu Grafiği

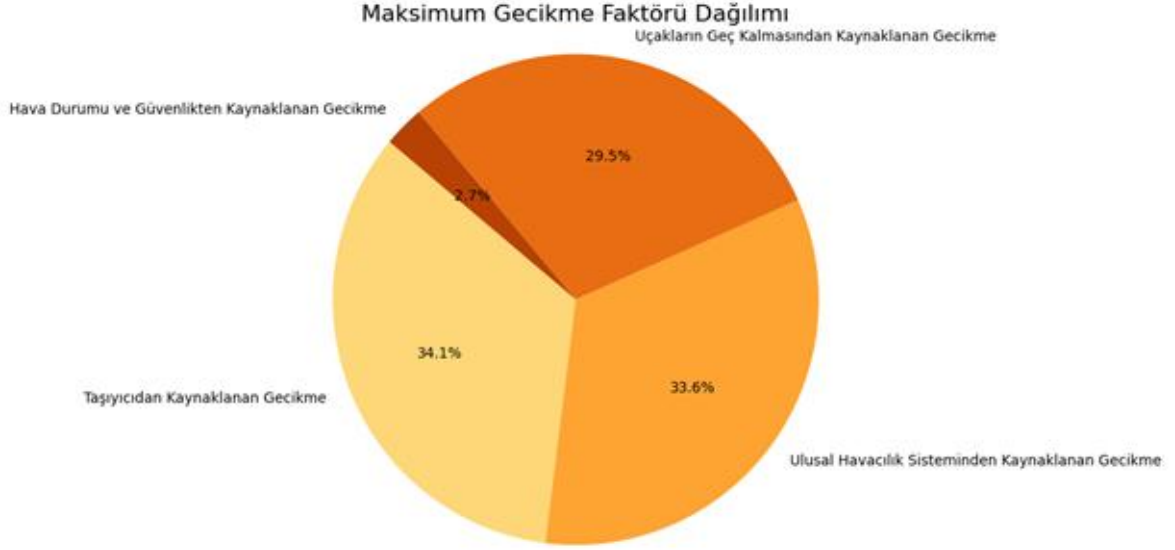
Şekil 2’de farklı hava yolu şirketlerine ait 10 dakikadan fazla gecikme yaşayan uçuş sayılarının gösterildiği bir bar grafiği yer almaktadır.



Şekil 2 Hava Yolu Şirketlerine Göre Gecikme Durumu Grafiği

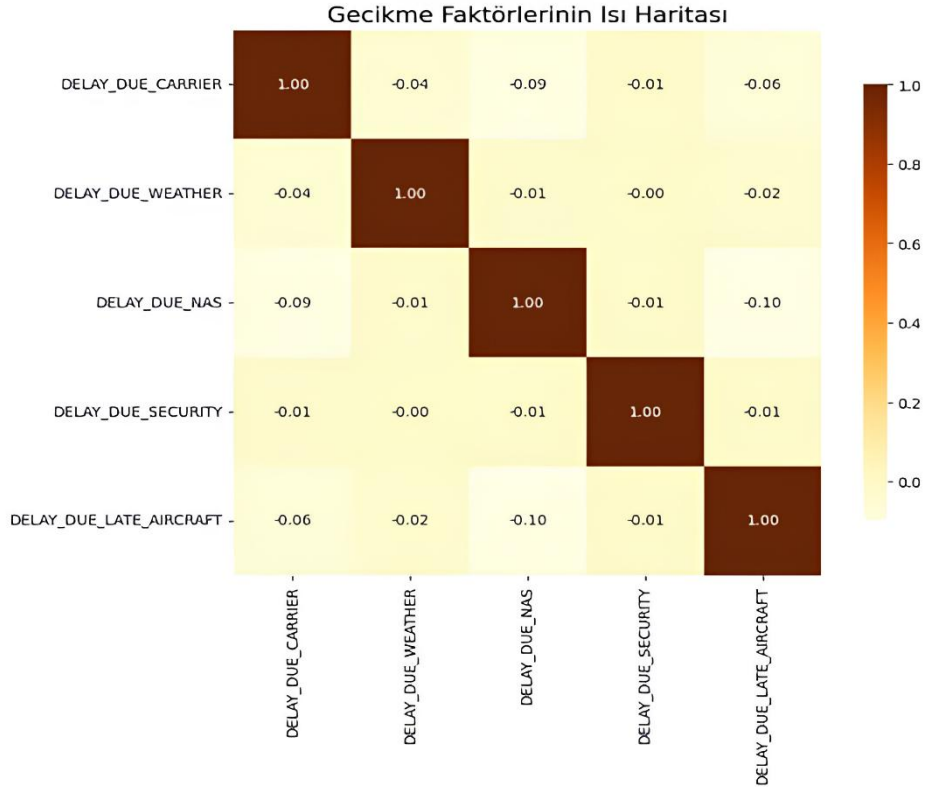
Elimizdeki veri setine göre Şekil 3’teki gibi maksimum gecikme faktörü dağılımı incelenmiştir. Bunun sonucunda, %34.1 taşıyıcıdan kaynaklanan gecikme, %33.6 ulusal

havacılık sisteminden kaynaklanan gecikme, %29.5 uçakların geç kalkmasından kaynaklanan gecikme, %2.7 hava durumu ve güvenlikten kaynaklanan gecikmelerin olduğu tespit edilmiştir.



Şekil 3 Gecikme Faktörü Dağılımı

Şekil 3'teki gecikme faktörü dağılımı, makine öğrenimi modellerinin performansını etkileyecek önemli içgörüler sağlamaktadır. Taşıyıcıdan kaynaklanan %34.1'lik gecikme oranı, model eğitiminde hassas bir özellik çıkarımı gerektirmektedir. Ulusal havacılık sisteminden kaynaklanan %33.6'lık gecikme, modellerin karmaşık sistemsel ilişkileri yakalama kapasitesini test etmektedir. Uçak gecikmelerinin %29.5'i, zaman serisi ve ardışık bağımlılık modellemelerinin önemini vurgulamaktadır. Düşük oranlı hava durumu ve güvenlik gecikmeleri (%2.7) ise sınıf dengesizliği sorunlarına işaret eder, bu da SMOTE gibi veri dengeleme tekniklerinin gerekliliğini ortaya koymaktadır.



Şekil 4 Gecikme Faktörlerinin Isı Haritası

Şekil 4'teki gecikme faktörlerinin ısı haritası, uçuş gecikmelerinin birbirleriyle olan korelasyonlarını görsel olarak temsil etmektedir. Renkli hücrelerin yoğunluğu, faktörler arasındaki ilişkilerin şiddetini göstermekte olup, makine öğrenimi modellerinin özellik seçiminde ve çoklu değişken etkileşimlerinin anlaşılmasında kritik bir rol oynamaktadır.

## 2.2. Veri Ön İşleme

Makine öğrenimi modellerinin başarısı, büyük ölçüde veri ön işleme aşamasının etkinliğine bağlıdır. Bu çalışmada, uçuş gecikme tahminlerinin doğruluğunu artırmak amacıyla kapsamlı bir veri ön işleme stratejisi uygulanmıştır. Özellikle, meteorolojik verilerin entegrasyonu ve veri kalitesinin iyileştirilmesi konularına odaklanılmıştır.

Veri ön işlemenin ilk aşamasında, uçuş verilerine meteorolojik verilerin entegrasyonu gerçekleştirilmiştir. Bu süreçte Meteostat kütüphanesi kullanılarak her bir uçuş noktası için detaylı hava durumu bilgileri elde edilmiştir. Meteorolojik veri entegrasyonu için öncelikle Nominatim API aracılığıyla havalimanlarının coğrafi koordinatları belirlenmiştir. Elde edilen koordinat bilgileri, Meteostat'ın Point sınıfı kullanılarak

meteorolojik veri noktalarına dönüştürülmüştür. Her uçuş tarihi için minimum sıcaklık, maksimum sıcaklık, yağış miktarı, kar yağışı, rüzgar yönü, rüzgar hızı, rüzgar hamlesi, atmosfer basıncı ve güneşlenme süresi gibi kritik meteorolojik parametreler toplanmıştır.

Meteorolojik verilerin uçuş verileriyle birleştirilmesi sürecinde, veri setinin büyüklüğü göz önünde bulundurularak parçalı (chunk) işleme stratejisi benimsenmiştir. Bu yaklaşım, bellek kullanımını optimize ederken veri işleme sürecinin kesintisiz devam etmesini sağlamıştır. Veriler, 'FL\_DATE' ve 'ORIGIN\_CITY' sütunları üzerinden eşleştirilmiş, eşleşmeyen kayıtlar ise veri kalitesi takibi için ayrı bir dosyada saklanmıştır.

Meteostat kütüphanesi kullanılarak her bir uçuş noktası için detaylı hava durumu bilgilerinin veri setine eklenmesi ile birlikte elde edilen veri setinin son hali Tablo 2'de gösterilmiştir.

Tablo 2 Model Eğitim Veri Setinin Yapısı

Güncellenmiş Başlık	Kaynak Başlık	Veri Tipi	Açıklama
1 FL_DATE	FlightDate	object	Uçuş tarihi
2 AIRLINE_CODE	Reporting_Airline	object	Havayolu şirketinin adı
3 ORIGIN_CITY	OriginCityName	object	Çıkış havalimanı şehri
4 DEST_CITY	DestCityName	object	Varış havalimanının bulunduğu şehir
5 CRS_DEP_TIME	CRSDepTime	int64	Planlanan kalkış saati
6 DEP_TIME	DepTime	float64	Gerçek kalkış saati
7 CRS_ARR_TIME	CRSArrTime	int64	Planlanan varış saati
8 CANCELLED	Cancelled	float64	Uçuşun iptal edilip edilmediğinin göstergesi (iptal için 1, iptal değil için 0)
9 CANCELLATION_CODE	CancellationCode	object	İptal nedeni (varsa)
10 DISTANCE	Distance	float64	Katedilen mesafe
11 TMIN	tmin	float64	Minimum sıcaklık (Celsius)
12 TMAX	tmax	float64	Maksimum sıcaklık (Celsius)
13 PRCP	prcp	float64	Yağış miktarı (mm)
14 SNOW	snow	float64	Kar yağışı miktarı (mm)
15 WDIR	wdir	float64	Rüzgar yönü (derece)
16 WSPD	wspd	float64	Rüzgar hızı (m/s)
17 WPGT	wpgt	float64	Rüzgar hamlesi (m/s)
18 PRES	pres	float64	Atmosfer basıncı (hPa)
19 TSUN	tsun	float64	Toplam güneşlenme süresi (saat)

Veri setindeki eksik değerler için kapsamlı bir analiz ve temizleme süreci uygulanmıştır. Eksikliklerin rastgele veya sistematik olup olmadığını belirlemek amacıyla eksik veri

dağılımları ve paternleri incelenmiştir. Sayısal değişkenlerdeki eksik değerler medyan ile doldurulurken, kategorik değişkenlerde mod kullanılmıştır. %30'dan fazla eksik veri içeren özellikler veri setinden çıkarılmış, meteorolojik verilerdeki eksiklikler ise zaman serisi karakterine uygun olarak en yakın zaman damgasındaki değerlerle doldurulmuştur.

Aykırı değerlerin tespiti ve işlenmesi için çok yönlü bir yaklaşım benimsenmiştir. Z-skoru ve IQR metodu gibi istatistiksel yöntemler, box plot ve scatter plot gibi görsel analiz teknikleri ve domain bilgisi bazlı kontroller bir arada kullanılmıştır. Gecikme süreleri için 24 saatten fazla olan değerler ve meteorolojik verilerdeki aykırı değerler detaylı bir şekilde incelenmiştir. Analiz sonucunda, veri setine anlamlı katkı sağlayan ve istatistiksel olarak mantıklı kabul edilen aykırı değerler korunmuş, hatalı veri girişlerinden kaynaklanan anormallikler ise uygun yöntemlerle düzeltilmiştir.

Model performansını artırmak amacıyla kapsamlı bir veri analizi ve tahmin için açıklayıcı değişkenlerin türetilmesi ve işlenmesi süreci uygulanmıştır. Zaman bazlı özellikler arasında uçuş saatinin günün hangi dilimine denk geldiği, haftanın günü, tatil günü bilgisi ve sezon bilgisi yer almaktadır. Meteorolojik özellikler, hava durumu verilerinden türetilen kategorik değişkenler, rüzgar şiddeti kategorileri ve görüş mesafesi sınıflandırması gibi yeni özellikleri içermektedir. Operasyonel özellikler kapsamında ise havalimanı yoğunluk göstergeleri, önceki uçuşların gecikme durumları ve rotaya özgü geçmiş performans metrikleri oluşturulmuştur.

Veri standardizasyonu ve normalizasyon aşamasında, sayısal özelliklerin ölçek farklılıklarından kaynaklanan sorunlar ele alınmıştır. Sürekli değişkenler için StandardScaler kullanılarak standardizasyon uygulanırken, sınırlı aralıktaki değişkenler MinMaxScaler ile [0,1] aralığına normalize edilmiştir. Kategorik değişkenler ise One-Hot Encoding yöntemi ile sayısallaştırılmıştır.

Gecikme sınıfları arasındaki dengesizlik problemi ele alınmıştır. SMOTE algoritması kullanılarak azınlık sınıfı örnekleri sentetik olarak artırılmış ve sınıf ağırlıkları parametresi ayarlanarak dengeli bir veri seti elde edilmiştir. Bu kapsamlı ön işleme adımları, modelin eğitim verilerini daha etkin kullanmasını ve gerçek dünya uygulamalarında daha güvenilir tahminler yapmasını sağlamıştır.

İptal ve gecikme nedeni tahmin modeli için veri ön işleme sürecinde, öncelikle zamansal verilerin ayrıştırılması gerçekleştirilmiştir. 'FL\_DATE' sütunu datetime formatına dönüştürülerek yıl, ay ve gün bilgileri ayrı özellikler olarak çıkarılmıştır. Kategorik değişkenler için LabelEncoder kullanılarak havayolu kodu (AIRLINE\_CODE), kalkış şehri (ORIGIN\_CITY) ve varış şehri (DEST\_CITY) sayısal değerlere dönüştürülmüştür. İptal kodu (CANCELLATION\_CODE) sütunundaki eksik değerler 'N' (iptal yok) değeri ile doldurularak encode edilmiştir.

Veri setindeki dengesizlik problemi, SMOTE (Synthetic Minority Over-sampling Technique) algoritması kullanılarak ele alınmıştır. İptal durumu için sampling\_strategy=0.5 parametresi ile azınlık sınıfı örnekleri sentetik olarak artırılırken, iptal nedeni sınıflandırması için 'auto' stratejisi kullanılmıştır. Eksik verilerin doldurulması için SimpleImputer ile medyan stratejisi uygulanmış, ardından StandardScaler ile özellikler ölçeklendirilmiştir.

Gecikme tahmin modeli için veri ön işleme sürecinde, öncelikle iptal edilmiş uçuşlar veri setinden çıkarılmıştır. Zamansal veriler benzer şekilde ayrıştırılmış ve kategorik değişkenler LabelEncoder ile dönüştürülmüştür. Gecikme süreleri dört farklı sınıfa ayrılmıştır:

- Sınıf 0: Zamanında veya erken ( $\leq 0$  dakika)
- Sınıf 1: Hafif gecikme (1-15 dakika)
- Sınıf 2: Orta gecikme (16-30 dakika)
- Sınıf 3: Ciddi gecikme ( $>30$  dakika)

Eksik değerler SimpleImputer kullanılarak medyan değerleri ile doldurulmuş ve StandardScaler ile özellikler normalize edilmiştir. Sınıf dengesizliği problemi, k\_neighbors=5 parametresi ile SMOTE algoritması kullanılarak çözülmüştür. Bu süreçte, her bir gecikme sınıfı için eşit sayıda örnek oluşturularak dengeli bir veri seti elde edilmiştir.

Her iki model için de özellik önem dereceleri analiz edilmiş ve görselleştirilmiştir. Bu analiz, modellerin tahmin performansını etkileyen en önemli faktörlerin belirlenmesine



ve özellik seçimi stratejilerinin değerlendirilmesine olanak sağlamıştır. Veri ön işleme adımları, scikit-learn kütüphanesinin Pipeline yapısı kullanılarak sistematik ve tekrarlanabilir bir şekilde uygulanmıştır.

Bu detaylı veri ön işleme süreci, modellerin eğitim verilerini daha etkin kullanmasını sağlamış ve tahmin performanslarını önemli ölçüde artırmıştır. Özellikle sınıf dengesizliği probleminin çözülmesi ve özellik mühendisliği adımları, modellerin gerçek dünya uygulamalarında daha güvenilir tahminler yapmasına katkıda bulunmuştur.

### **2.3. Denetimli Öğrenme Modelleri**

Makine öğrenimi, bilgisayarların veriyi analiz etmesini, olası desenleri keşfetmesini ve bu desenleri kullanarak tahminlerde bulunmasını sağlayan algoritmaların genel adıdır. Farklı ortamlarda öğrenmenin göreceli zorluğu hakkında bilgi verebilen bu algoritmalar çeşitli kategorilere ayrılmaktadır [10]. Makine öğrenimi algoritmalarının en yaygın iki türü ise denetimli ve denetimsiz öğrenmedir. Denetimli öğrenme algoritmaları, girdileri istenen çıktılara dönüştüren bir işlev oluşturur. Denetimli öğrenmenin temel türleri arasında regresyon ve sınıflandırma yer alır. Denetimsiz öğrenme ise etiketlenmiş örnekler olmadan bir girdi kümesini modellemeye odaklanır.

Bu çalışmada makine öğrenmesi, uçuş iptalleri ve gecikmelerinin tahmin edilmesi amacıyla kullanılmıştır. Havayolu operasyonlarındaki karmaşık ilişkileri modellemek ve çok sayıda değişkenin etkisini analiz etmek için makine öğrenmesi algoritmaları ideal araçlardır [11].

Karar ağacı algoritması, veriyi hiyerarşik bir yapıda bölerek sınıflandırma ve regresyon problemlerini çözen bir makine öğrenimi yaklaşımıdır. Her bir düğüm, bir özellik üzerinde karar verme sürecini temsil eder ve ağacın yapısı, en ayırt edici özelliklere dayalı olarak oluşturulur. Uçuş gecikmesi tahmininde, farklı faktörlerin etkisini net bir şekilde görselleştirme imkânı sağlar [12].

LightGBM, gradient boosting çerçevesinde çalışan, yüksek performanslı ve verimli bir makine öğrenimi algoritmasıdır. Büyük veri setlerinde hızlı eğitim yapabilme ve bellek kullanımını optimize etme özellikleriyle öne çıkar. Uçuş gecikmesi gibi karmaşık ve yüksek boyutlu veri setlerinde, diğer algoritmalara kıyasla daha yüksek doğruluk oranları elde edebilir [13].

Random Forest, birden fazla karar ağacını birleştiren bir ensemble öğrenme metodudur. Her bir ağaç, bootstrap örnekleme ve rastgele özellik seçimi ile eğitilir, bu sayede aşırı uyum (overfitting) riskini azaltır. Uçuş gecikmesi tahmininde, farklı alt veri setleri üzerinden tahminler yaparak daha güvenilir sonuçlar elde edilmesini sağlar [11].

Kümeleme algoritmaları, benzer özelliklere sahip veri noktalarını gruplandırarak, gizli desenler keşfetmeye olanak sağlar. Bayesian sınıflandırma ise olasılıksal bir yaklaşımla, öncül bilgileri ve gözlemsel verileri birleştirerek sınıflandırma yapar. Bu yöntemler, uçuş gecikmelerindeki karmaşık ilişkilerin anlaşılmasında tamamlayıcı rol oynar [12].

XGBoost, makine öğrenimi alanında oldukça popüler ve güçlü bir algoritmadır, özellikle büyük veri setleri üzerinde yüksek doğrulukla tahmin yapma yeteneği ile tanınır. Gradient boosting tekniklerinin geliştirilmiş bir versiyonu olan XGBoost, her bir iterasyonda modelin hatalarını düzeltmek için yeni ağaçlar ekleyerek modelin genel doğruluğunu artırır. Bu özelliği, özellikle karmaşık ilişkiler ve çoklu faktörlerin etkisi altında olan uçuş tahminleri gibi problemler için oldukça faydalıdır. Uçuş iptali ve uçuş gecikmesi gibi tahminler söz konusu olduğunda, XGBoost algoritması, verinin çok sayıda faktörden etkilenen karmaşık yapısını modellemek için etkili bir araçtır [10].

Uçuş iptali ve uçuş gecikmesi tahminlerinde doğru sonuçlar alabilmek için, algoritma seçiminde dikkatli olunması gerekti. Burada kullanılan XGBoost algoritması, özellikle iki önemli özelliği ile tercih edildi:

Sınıf dengesizliğiyle baş etme: Uçuş iptalleri ve gecikmeleri gibi olaylar genellikle dengesiz veri setleriyle ilişkilidir. Örneğin, iptal edilen uçuşların sayısı, gerçekleştirilen uçuşlara göre çok daha az olabilir. XGBoost, bu tür dengesizlikleri etkili bir şekilde yönetebilir ve azınlık sınıfını daha doğru şekilde öğrenebilir.

Yüksek performans ve esneklik: XGBoost, büyük veri setlerinde hızlı ve doğru tahminler yapabilme yeteneğine sahip bir algoritmadır. Özellikle uçuş gibi çok sayıda faktöre bağlı olan veri setlerinde, XGBoost'un güçlü özellik mühendisliği ve model optimizasyonu yetenekleri oldukça faydalıdır.

Bu nedenle, XGBoost, hem dengesiz veri setlerinde yüksek doğruluk sağlamak hem de karmaşık ilişkileri modellemek için ideal bir seçenek olarak tercih edilmiştir.

Bu bölümde, uçuş iptali ve iptal kodu tahmini için uyguladığımız veri işleme ve modelleme sürecini ayrıntılı olarak ele alacağız. İlk olarak, uçuş verilerindeki tarih

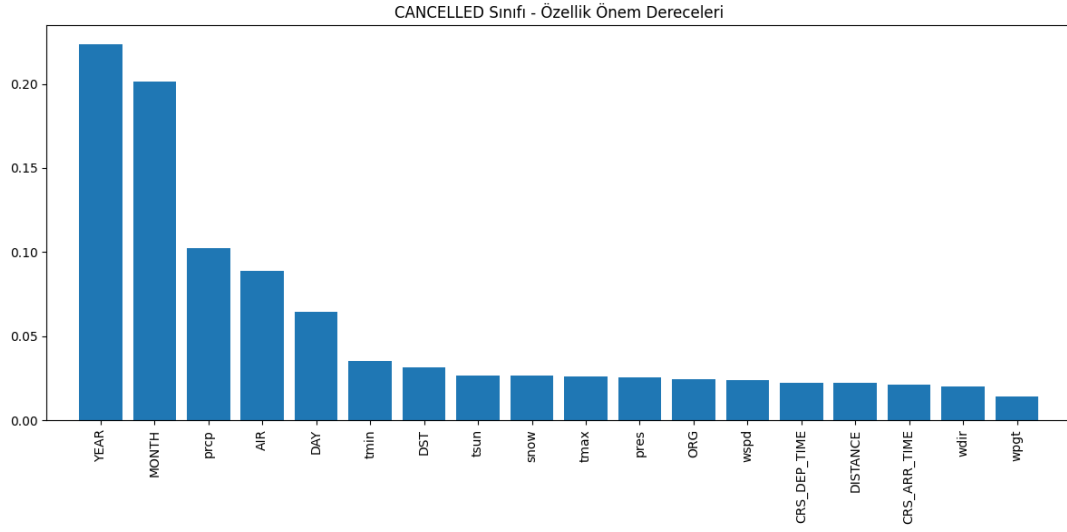
sütunları, yıl, ay ve gün olarak ayrılarak, zamanla ilişkili özellikler çıkarıldı. Kategorik veriler, LabelEncoder kullanılarak sayısal verilere dönüştürüldü. Bu işlem, modelin veriyi daha rahat anlayabilmesini sağladı.

Veri setindeki gereksiz ve modelin eğitimi için gerekli olmayan sütunlar, örneğin havayolu kodu, kalkış ve varış şehirleri gibi bilgiler çıkarıldı. Hedef değişkenler olarak uçuşun iptal durumu ve iptal kodu belirlendi. Veri seti daha sonra eğitim ve test setlerine ayrıldı ve eksik değerler, median değeriyle doldurularak verinin tutarlılığı sağlandı. Eğitim verileri, StandardScaler ile ölçeklendirilerek modelin daha düzgün bir şekilde eğitim alması sağlandı.

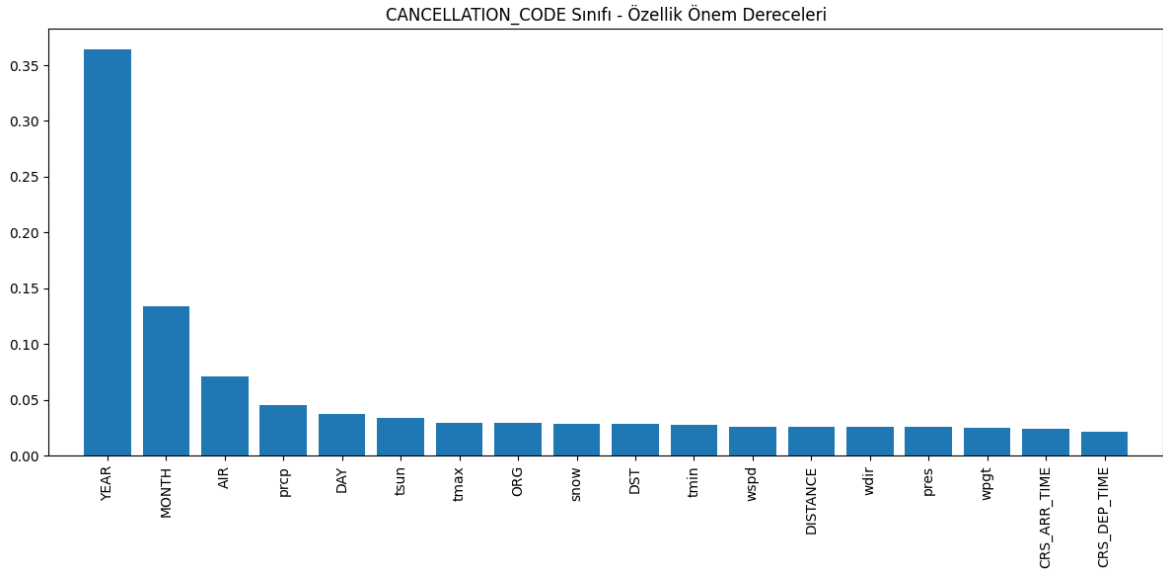
Modelin dengeli bir şekilde eğitilmesi için SMOTE (Synthetic Minority Over-sampling Technique) yöntemi kullanılarak yeniden örnekleme yapıldı. Bu adım, sınıf dengesizliğini gidermek ve modelin her sınıfı eşit derecede öğrenmesini sağlamak amacıyla gerçekleştirildi. XGBoost algoritması, sınıf dengesizliğiyle başa çıkabilen ve büyük veri setlerinde yüksek başarı sağlayan bir algoritma olarak tercih edildi.

Eğitilen modellerin doğruluk oranları, test seti üzerinde değerlendirildi. Bu değerlendirme sırasında, iptal durumu ve iptal kodu tahminleri için doğruluk oranları ayrı ayrı ölçüldü. Sonuçlar, her iki modelin de yüksek doğruluk oranlarına sahip olduğunu gösterdi. Ayrıca, model performansı classification report ve ROC-AUC skorları ile derinlemesine incelendi.

Modelin özellik önem dereceleri görselleştirilerek, hangi özelliklerin modelin tahminlerinde daha fazla etkili olduğu görsel olarak sunuldu. Bu görselleştirme, modelin nasıl çalıştığını daha iyi anlamamıza yardımcı oldu. Yeni uçuş verileri için tahminler yapılabilmesi adına, iptal durumu ve iptal kodu tahminleri gerçekleştiren bir fonksiyon geliştirildi. Bu fonksiyon, gerçek zamanlı tahminler yapılabilmesini sağladı.



Şekil 5 Uçuş İptal Tahmin Modeli Özellik Önem Derecesi



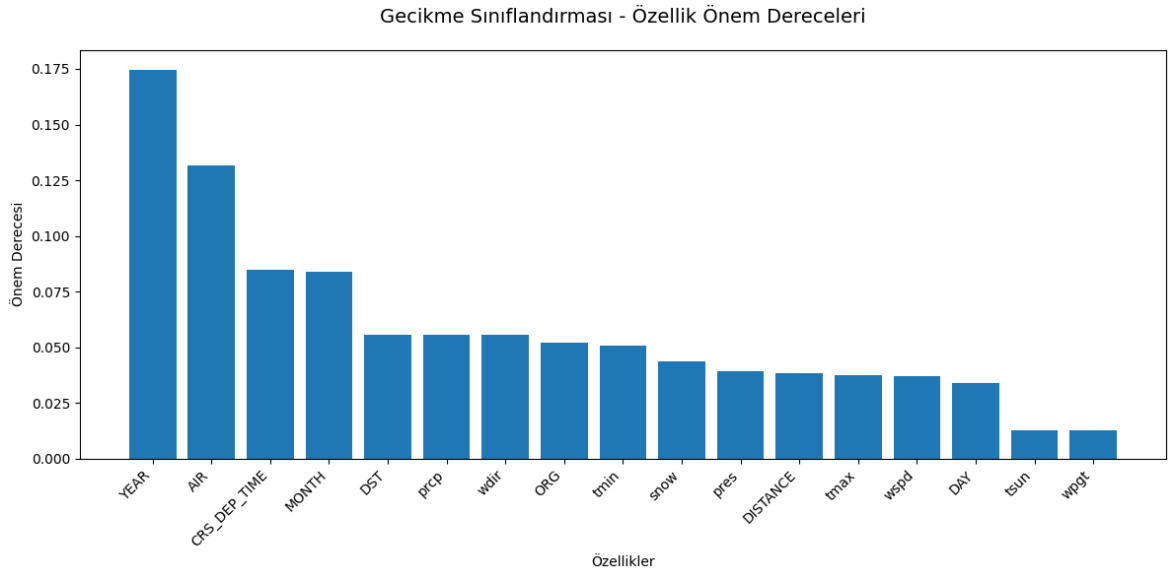
Şekil 6 Uçuş İptal Kodu Tahmin Modeli Özellik Önem Derecesi

Uçuş gecikmesi tahmini için uyguladığımız modelleme süreci de benzer adımları içermektedir. Veriler üzerinde gerekli tarih dönüşümleri yapıldı, uçuşların iptal edilip edilmediği kontrol edildi ve iptal edilmiş uçuşlar veri setinden çıkarıldı. Tarih sütunu üzerinden yıl, ay ve gün gibi zamanla ilgili özellikler oluşturuldu, kategorik veriler ise sayısal verilere dönüştürüldü. Eksik veriler, SimpleImputer kullanılarak median stratejisiyle dolduruldu ve veri setindeki özellikler StandardScaler ile ölçeklendirildi.

Uçuş gecikme süreleri, hedef değişken olarak sınıflara ayrıldı. Sınıflama, uçuşların gecikme süresine göre zamanında, hafif gecikme, orta gecikme ve ciddi gecikme olarak

yapılmıştır. Veri setindeki sınıf dağılımı gözden geçirilmiş ve sınıf dengesizliği tespit edilmiştir. Modelin dengeli bir şekilde eğitilmesi için SMOTE kullanılarak eğitim seti yeniden örneklenmiştir.

Model eğitimi aşamasında XGBoost sınıflandırıcı modeli seçilerek, modelin parametreleri üzerinde ince ayar yapılmıştır. Eğitim sürecinin ardından model, test seti üzerinde değerlendirilmiş ve doğru sınıflandırmalar elde edilmiştir. Modelin doğruluk oranı ve performansı, confusion matrix, classification report ve accuracy score ile analiz edilmiştir. Ayrıca, modelin hangi özelliklerinin daha önemli olduğunu belirlemek amacıyla, özelliklerin önem dereceleri görselleştirilmiş ve bu özelliklerin tahminlerdeki etkisi şekil 7’ de sunulmuştur.



Şekil 7 Uçuş Gecikme Tahmin Modeli Özellik Önem Derecesi

Son olarak, yeni uçuş verisiyle tahminler yapılabilmesi için bir örnek veri seti oluşturulmuş ve modelin tahminleri gerçekleştirilmiştir. Bu tahminler, uçuş gecikmesinin olası sınıfını ve her bir sınıfa ait olasılıkları içermektedir. Modelin doğruluğu ve güvenilirliği, bu tahminler üzerinden kontrol edilmiştir.

Algoritmaların seçilmesinde XGBoost’un tercih edilmesinin temel nedeni, özellikle büyük veri setlerinde yüksek başarı sağlaması ve sınıf dengesizliğini etkin bir şekilde yönetebilmesidir. Ayrıca, XGBoost’un sağladığı özellik önem dereceleri görselleştirmeleri, modelin karar verme süreçlerini daha anlaşılır kılmakta ve uçuş tahminleri gibi karmaşık problemlerde başarıyla uygulanmasını sağlamaktadır.

### 3. BULGULAR VE TARTIŞMA

Bu bölümde, proje başlangıcından itibaren tamamlanan aşamalar özetlenecek, ardından projenin kalan kısımlarına dair planlamalar paylaşılacaktır.

#### 3.1. Projede Şu Ana Kadar Yapılanlar

Proje Konu Belirleme ve Onayı:

- Uçuş iptalleri ve gecikmelerinin tahmin edilmesine yönelik bir makine öğrenimi projesi fikri belirlendi.
- Projenin uygulanabilirliği, literatürdeki benzer çalışmalar ve veri erişimi açısından değerlendirildi.
- Çalışmanın kapsamı ve hedefleri netleştirilerek proje onayı alındı.

Proje Konusuna Dair Açık Kaynak Veri Seti Araştırması ve İncelenmesi:

- Uçuş iptalleri ve gecikmelerine dair kamuya açık veri setleri araştırıldı.
- DOT (Ulaştırma Bakanlığı), Kaggle ve diğer açık kaynak platformlardan elde edilen veri setleri analiz edildi.
- Uygun veri setleri birleştirilerek proje için kullanılacak veri havuzu oluşturuldu.

Veri Ön İşleme ve Veri Seti Düzenleme İşlemleri:

- Eksik ve hatalı veriler tespit edilerek uygun yöntemlerle dolduruldu veya çıkarıldı.
- Zaman bilgileri standart hale getirilerek analiz edilebilir forma getirildi.
- Uçuş iptali ve gecikmelerini etkileyen faktörler belirlendi.
- Hava durumu, uçuş mesafesi ve hava yolu şirketi gibi ek değişkenler oluşturuldu.

Makine Öğrenimi Modellerinin Eğitilmesi:

- Uçuş iptali, iptal kodu ve uçuş gecikmesi tahmini için çeşitli modeller denendi.
- XGBoost, Random Forest, Decision Tree, KNN ve Gradient Boosting algoritmaları ile karşılaştırmalı performans analizleri yapıldı.
- SMOTE yöntemi kullanılarak, dengesiz veri setlerinde sınıflar arasındaki dağılım eşitlendi ve modelin doğruluğu artırıldı.
- Model doğruluk oranlarını artırmak için GridSearchCV ve RandomizedSearchCV gibi teknikler denendi.

Sonuçların Değerlendirilmesi:

- Modellerin doğruluk (accuracy), precision, recall, f1-score ve ROC-AUC gibi temel performans metrikleri hesaplandı.
- Elde edilen sonuçlar tablolar halinde düzenlendi ve yorumlandı.
- Literatürdeki çalışmalarla karşılaştırmalar yapılarak, modelin başarısı değerlendirildi.

### 3.2. Projede Bundan Sonra Yapılacaklar

Bu kısımda, proje kapsamında tamamlanacak aşamalar hakkında bilgi verilecektir.

Proje Raporu Düzenlenmesi:

- Proje süreci, kullanılan yöntemler ve elde edilen sonuçların raporlanmasının tamamlanması,
- Literatür çalışmaları ile proje sonuçlarının karşılaştırılması hedeflenmektedir.

Uçuk Bilet Web Sitesi Geliştirilmesi:

- Uçuş iptali ve gecikme tahminlerini gerçek zamanlı olarak kullanıcılara sunacak bir web platformu geliştirilmesi,
- Frontend: Kullanıcıların uçuşlarını sorgulayabileceği bir arayüz tasarımı (React, Next.js vb.),
- Backend: Model tahminlerini yapacak ve veritabanı yönetimini sağlayacak bir sunucu (Node.js, FastAPI vb.),
- API Entegrasyonu: Gerçek zamanlı uçuş verileri ve hava durumu bilgilerini çekmek için FlightAware, OpenWeather veya başka API'lerin entegrasyonu,
- Son kullanıcı testleri yapılarak, web sitesinin fonksiyonelliği ve doğruluğunun değerlendirilmesi hedeflenmektedir.

## 4. SONUÇLAR

Bu çalışma, havayolu sektöründe uçuş iptali, iptal nedenleri ve uçuş gecikmelerinin tahmini için gelişmiş makine öğrenmesi ve derin öğrenme yaklaşımları sunmaktadır. Elde edilen yüksek başarı oranları, geliştirilen modellerin havacılık sektöründeki tahmin problemlerinin çözümünde etkili bir şekilde kullanılabileceğini göstermektedir.

Çalışmamızın sunduğu yenilikçi metodolojiler ve teknikler, hem akademik literatüre katkı sağlamakta hem de havayolu şirketleri, havaalanı yönetimleri ve yolcular için pratik uygulamalar sunmaktadır.

Modelleme sürecinde XGBoost, Random Forest, Decision Tree, K-Nearest Neighbors (KNN) ve Gradient Boosting gibi algoritmalar karşılaştırılmış ve en iyi performans gösteren modeller belirlenmiştir. Elde edilen sonuçlar göstermiştir ki:

Uçuş iptal tahmini modeli, doğruluk ve diğer metrikler açısından başarılı sonuçlar üretmiş ve XGBoost algoritması %97 doğruluk oranı ile en iyi performansı sergilemiştir.

İptal kodu tahmini modeli, iptal edilen uçuşların nedenlerini belirleme konusunda tatmin edici sonuçlar vermiş olup, XGBoost %74 doğruluk oranı ile diğer algoritmalara kıyasla daha yüksek performans göstermiştir.

Uçuş gecikmesi tahmini modeli, uçuş gecikmelerini belirleme konusunda makul bir doğruluk oranı sunmuş, ancak hava durumu, hava trafiği ve operasyonel faktörler gibi değişkenlerin daha detaylı incelenmesi gerektiğini ortaya koymuştur.

Bu çalışmanın sağladığı yenilikler arasında, farklı veri kaynaklarından alınan açık verilerin birleştirilerek daha kapsamlı bir veri seti oluşturulması, farklı gecikme ve iptal nedenlerinin modellenmesi, ve birden fazla makine öğrenimi algoritması ile kıyaslama yapılarak en uygun yöntemin belirlenmesi bulunmaktadır.

Gelecek çalışmalarda, daha kapsamlı veri setleri, gerçek zamanlı tahmin sistemleri, açıklanabilir yapay zeka yaklaşımları ve insan-makine işbirliği sistemleri geliştirilerek, havacılık sektöründeki tahmin problemlerinin çözümünde daha da ileri adımlar atılabilir. Bu çalışmanın, havacılık sektöründeki operasyonel verimliliğin artırılmasına, yolcu deneyiminin iyileştirilmesine ve sektörün genel performansının yükseltilmesine katkı sağlaması beklenmektedir.



## KAYNAKLAR

- [1] B. Ghosh ve B. Tabrizi, "Machine learning approaches for flight delay prediction: A review," *International Journal of Aviation Studies*, 5(3), ss. 123-134, 2018.
- [2] D. Delahaye ve S. Puechmorel, "Weather impact on flight delay prediction: An AI-based approach," *Journal of Transportation Research*, 12(4), ss. 256-268, 2020.
- [3] N. V. Chawla, K. W. Bowyer, L. O. Hall, ve W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, 16, ss. 321-357, 2002.
- [4] H. Khaksar ve A. Sheikholeslami, "Airline delay prediction by machine learning algorithms," *Scientia Iranica*, 2017.
- [5] B. Ye, B. Liu, Y. Tian, ve L. Wan, "A methodology for predicting aggregate flight departure delays in airports based on supervised learning," *Sustainability*, 12(7), 2749, 2020.
- [6] M. C. Athioğlu, M. Bolat, M. Şahin, V. Tunalı, ve D. Kılınç, "Supervised learning approaches to flight delay prediction," *Sakarya University Journal of Science*, 2020.
- [7] M. S. Al-Tabbakh, H. M. Mohamed, ve Z. H. El, "Machine learning techniques for analysis of Egyptian flight delay," *International Journal of Data Mining & Knowledge Management Process*, 8(3), 01–14, 2018.
- [8] M. Kurt, "Flight Delay Prediction," Capstone Project, MEF University, İstanbul, 2019.
- [9] Y. Tang, "Airline Flight Delay Prediction Using Machine Learning Models," in *5th International Conference on E-Business and Internet*, Singapore, 2021, pp. 151-154.
- [10] O. T. Oladipupo, "Types of machine learning algorithms", *New Advances in Machine Learning*, 2010.

- [11] R. Kumar and N. Singh, “A Survey on Data Mining and Machine Learning Techniques for Flight Delay Prediction,” *International Journal of System Assurance Engineering and Management*, 2020.
- [12] M. Breiman, “Classification and Regression Trees”, Wadsworth and Brooks/Cole, 1984.
- [13] G. Ke, Q. Meng, T. Zhang, W. Chen, and T. Liu, “LightGBM: A Highly Efficient Gradient Boosting Decision Tree”, Microsoft Research, 2017.