



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



Doktora Tezi

**MAKİNE ÖĞRENMESİ TEKNİKLERİ İLE HAVACILIKTA YAKIT
TAŞIMACILIĞI KARARININ ALINMASI VE BİR UYGULAMA**

İlker Güven YILMAZ

Enformatik Anabilim Dalı

Enformatik Programı

**DANIŞMAN
Prof. Dr. Sevinç GÜLSEÇEN**

**II. DANIŞMAN
Dr. Elif KARTAL**

Ağustos, 2021

İSTANBUL

Bu çalışma, 12.08.2021 tarihinde ařağıdaki jüri tarafından Enformatik Anabilim Dalı, Enformatik Programı'nda Doktora tezi olarak kabul edilmiştir.

Tez Jürisi

Prof. Dr. Sevinç GÜLSEÇEN (Danışman)
İstanbul Üniversitesi
Enformatik Bölümü

Prof. Dr. Zuhâl TANRIKULU
Boğaziçi Üniversitesi
Uygulamalı Bilimler Yüksekokulu

Prof. Dr. Selim YAZICI
İstanbul Üniversitesi
Siyasal Bilgiler Fakültesi

Prof. Dr. Adem KARAHOCA
MEF Üniversitesi
Mühendislik Fakültesi

Doç. Dr. Aşkın DEMİR AĞ
Yeditepe Üniversitesi
Ticari Bilimler Fakültesi

İntihal Programı Beyanı

20.04.2016 tarihli Resmi Gazete’de yayımlanan Lisansüstü Eğitim ve Öğretim Yönetmeliğinin 9/2 ve 22/2 maddeleri gereğince; Bu Lisansüstü teze, İstanbul Üniversitesi’nin abonesi olduğu intihal yazılım programı kullanılarak Fen Bilimleri Enstitüsü’nün belirlemiş olduğu ölçütlere uygun rapor alınmıştır.

Tezden Üretilmiş Yayınların Künye Bilgileri

Yılmaz, İ.G., Kartal, E., Özen, Z., Gülseçen, S., 2021. Prediction of Fuel Tankering in Aviation Industry with Machine Learning Algorithms, *Journal of Aeronautics and Space Technologies*, 14(1), 19-34.



ÖNSÖZ

Tez çalışmamın her safhasında sunduğu çok değerli katkılar ile bana her zaman destek olan, bilgi ve tecrübelerinden yararlandığım çok kıymetli danışman hocam Prof. Dr. Sevinç GÜLSEÇEN'e ve tez çalışmamın her safhasında ikinci danışman hocam olarak büyük bir sabır ve özveri ile bana çok değerli katkılar sunan çok kıymetli hocam Dr. Elif KARTAL'a sonsuz teşekkür ederim.

Değerli katkıları ile bana her zaman destek olan çok kıymetli Dr. Zeki ÖZEN hocama teşekkürlerimi sunmayı bir borç bilirim. Ayrıca, İstanbul Üniversitesi Enformatik Bölümü'nün çok kıymetli hocalarına sonsuz teşekkür ederim.

Üzerimde büyük emekleri olan başta Sakarya Üniversitesi Fen-Edebiyat Fakültesi Matematik Bölümü Matematiğin Temelleri ve Matematik Lojik Anabilim Dalı Öğretim Üyesi Doç. Dr. Murat SARDUVAN hocam olmak üzere, Sakarya Üniversitesi Fen-Edebiyat Fakültesi Matematik Bölümü Uygulamalı Bilimler Anabilim Dalı Öğretim Üyesi Prof. Dr. Halim ÖZDEMİR hocama, Sakarya Üniversitesi Fen-Edebiyat Fakültesi İstatistik Bölümü İstatistik Teorisi Anabilim Dalı Öğretim Üyesi Doç. Dr. Nesrin GÜLER hocama ve Pamukkale Üniversitesi Fen-Edebiyat Fakültesi Matematik Bölümü Cebir ve Sayılar Teorisi Anabilim Dalı Öğretim Üyesi Prof. Dr. Serpil HALICI hocama sonsuz teşekkürlerimi sunarım.

Ayrıca, benden her zaman yardım ve desteklerini esirgemeyen eşime ve aileme sonsuz teşekkürlerimi sunarım.

Ağustos 2021

İlker Güven YILMAZ

İÇİNDEKİLER

Sayfa No

ÖNSÖZ	iv
İÇİNDEKİLER.....	v
ŞEKİL LİSTESİ	viii
TABLO LİSTESİ.....	x
SİMGE VE KISALTMA LİSTESİ	xiii
ÖZET	xv
SUMMARY	xvi
1. GİRİŞ	1
2. GENEL KISIMLAR.....	6
2.1.YAKIT TAŞIMACILIĞI	6
2.1.1.Yakıt Taşımacılığının Tarihçesi	12
2.1.2.Yakıt Taşımacılığı Uygulamaları	12
2.1.3.Yakıt Taşımacılığı ve Yönetmelikler	16
2.1.4.Yakıt Taşımacılığının Hava Yolu Maliyetleri Üzerindeki Etkisi.....	17
2.1.5.Yakıt Taşımacılığının Ekonomik Yararları ve Çevresel Etkileri	20
2.2.MAKİNE ÖĞRENMESİ	22
2.2.1.Öğrenme Stratejileri	29
2.2.1.1.Danışmanlı Öğrenme.....	29
2.2.1.2.Danışmansız Öğrenme.....	30
2.2.1.3.Pekiştirmeli Öğrenme	32
2.3.MAKİNE ÖĞRENMESİ SÜRECİ ADIMLARI	34
2.3.1. Problemin Tanımlanması (<i>Business Understanding</i>).....	37
2.3.2. Veriyi Anlama (<i>Data Understanding</i>).....	38
2.3.3. Veriyi Hazırlama (<i>Data Preparation</i>)	38
2.3.3.1. Aykırı Değer Tespiti.....	38
2.3.3.2. Tekrar Eden Gözlemlerin İncelenmesi	38
2.3.3.3. Kayıp Değerlerin Tamamlanması.....	39
2.3.3.4. Normalizasyon	39
2.3.4. Modelleme (<i>Modeling</i>).....	40
2.3.4.1. k-En Yakın Komşu Algoritması.....	41

2.3.4.2. Basit (Naive) Bayes Sınıflandırıcı	43
2.3.4.3. Sınıflandırma ve Regresyon Ağaçları	46
2.3.4.4. Rastgele Orman Algoritması	46
2.3.4.5. Yapay Sinir Ağları	47
2.3.4.6. Destek Vektör Makineleri	53
2.3.5. Model Değerlendirme ve Seçimi (Evaluation).....	55
2.3.5.1. Model Performans Değerlendirme Yöntemleri	56
2.3.5.2. Model Performans Değerlendirme Ölçütleri.....	57
2.3.6. Modelin Uygulamaya Geçirilmesi (Deployment)	61
3. MALZEME VE YÖNTEM.....	62
3.1.PROBLEMİN TANIMLANMASI	62
3.1.1. Sınırlılıklar.....	62
3.2. VERİYİ ANLAMA.....	62
3.3. VERİNİN HAZIRLANMASI.....	65
3.3.1. Aykırı Değerlerin İncelenmesi	66
3.3.2. Eksik Verinin Tamamlanması	68
3.3.3. Tekrar Eden Gözlemlerin Veri Setinden Çıkarılması	68
3.3.4. Nitelikler Arasındaki İlişkinin İncelenmesi.....	68
3.4. MODELLEME.....	71
4. BULGULAR.....	73
4.1. SINIFLANDIRMA AMACIYLA KULLANILAN ALGORİTMALARDAN ELDE EDİLEN BULGULAR.....	73
4.1.1. k-En Yakın Komşu Algoritmasından Elde Edilen Bulgular	73
4.1.2. Basit (Naive) Bayes Sınıflandırıcı Algoritmasından Elde Edilen Bulgular	76
4.1.3. Sınıflandırma ve Regresyon Ağaçları Algoritmasından Elde Edilen Bulgular.....	79
4.1.4. Rastgele Orman Algoritmasından Elde Edilen Bulgular.....	82
4.1.5. Yapay Sinir Ağları Algoritmasından Elde Edilen Bulgular	84
4.1.6. Destek Vektör Makineleri Algoritmasından Elde Edilen Bulgular.....	86
4.2. REGRESYON AMACIYLA KULLANILAN ALGORİTMALARDAN ELDE EDİLEN BULGULAR.....	92
4.2.1. k-En Yakın Komşu Algoritmasından Elde Edilen Bulgular	92
4.2.2. Sınıflandırma ve Regresyon Ağaçları Algoritmasından Elde Edilen Bulgular.....	94
4.2.3. Rastgele Orman Algoritmasından Elde Edilen Bulgular.....	95
4.2.4. Yapay Sinir Ağları Algoritmasından Elde Edilen Bulgular	96
4.2.5. Destek Vektör Makineleri Algoritmasından Elde Edilen Bulgular.....	97

4.3. SHINY UYGULAMASI.....	100
5. TARTIřMA VE SONUÇ	102
KAYNAKLAR.....	105
ÖZGEÇMİř	119



ŞEKİL LİSTESİ

Sayfa No

Şekil 2.1: Yakıt taşımacılığı uygulaması (Boeing, 2004).	7
Şekil 2.2: Hava yolu uçuş planlaması için temel yakıt yükleme kategorileri (Zhu ve Li, 2020).	9
Şekil 2.3: Yakıt taşımacılığı'nın yapılmadığı duruma ait taşınan ve yakılan yakıtlar (Eurocontrol, 2019).	11
Şekil 2.4: Yakıt taşımacılığı'nın yapıldığı duruma ait taşınan ve yakılan yakıtlar (Eurocontrol, 2019).	12
Şekil 2.5: U.S. körfez kıyısı gazyağı tipi jet yakıtı peşin fiyatı (<i>Galon başına dolar olarak</i>).	18
Şekil 2.6: Dünyada 1980 ve 2017 yılları arasındaki jet yakıtı tüketimi (U.S. Energy Information Administration, 2019b).	19
Şekil 2.7: Dünya'da 1980 ve 2017 yılları arasında yıllık petrol ve diğer likitler kaynaklı karbondioksit salınım miktarları (U.S. Energy Information Administration, 2020a).	21
Şekil 2.8: Türkiye'de 1980 ve 2017 yılları arasında yıllık petrol ve diğer likitler kaynaklı karbondioksit salınım miktarları (U.S. Energy Information Administration, 2020b).	22
Şekil 2.9: Danışmanlı öğrenme (Salian, 2018).	30
Şekil 2.10: Danışmansız öğrenme (IBM, 2021a).	31
Şekil 2.11: Bir ajanın eğitiminin genel şeması (Cheskidov vd., 2019).	33
Şekil 2.12: Standart pekiştirmeli öğrenme modeli (Kaelbling vd., 1996).	33
Şekil 2.13: Veri madenciliği yöntem ve teknikleri (Albayrak ve Koltan Yılmaz, 2009).	35
Şekil 2.14: Birçok disiplinin birleşimi olarak veri madenciliği (Han ve Kamber, 2006).	36
Şekil 2.15: Bilginin keşfi süreci (Han vd., 2012).	36
Şekil 2.16: Veri madenciliği için CRISP-DM işlem modeli (Wirth ve Hipp, 2000).	37
Şekil 2.17: k-En Yakın Komşu algoritması için akış diyagramı (Balaban ve Kartal, 2015).	43

Şekil 2.18: Rastgele Orman Algoritması (Javatpoint, 2020).	47
Şekil 2.19: Biyolojik sinir hücresi (Jarosz, 2009).	48
Şekil 2.20: Yapay sinir hücresi (Atalay ve Çelik, 2017).	49
Şekil 2.21: Çok katmanlı Yapay Sinir Ağı (Atalay ve Çelik, 2017).	51
Şekil 2.22: Tek ve çok katmanlı ileri beslemeli Yapay Sinir Ağları (Turkson vd., 2016).	51
Şekil 2.23: Geri yayımlı Yapay Sinir Ağı (Demirhan vd., 2010).	52
Şekil 2.24: Lineer olarak ayrılabilen durumlar için Destek Vektör Makineleri (Küçüksille ve Ateş, 2013)	54
Şekil 2.25: Lineer olarak ayrılamayan durumlar için Destek Vektör Makineleri (Küçüksille ve Ateş, 2013)	55
Şekil 2.26: 5-kat çapraz geçişleme (Koçoğlu vd., 2018).	57
Şekil 3.1: Veri setinde yer alan niteliklere ait özet bilgi.	65
Şekil 3.2: <i>yakitTasimaciligiDrm</i> hedef niteliğinin sınıf etiketi frekansları.	66
Şekil 3.3: <i>yakitTasimaciligiDrm</i> hedef niteliğinin mevsimlere göre dağılım grafiği.	66
Şekil 3.4: Veri setinde yer alan niteliklere ait kutu grafiği.	67
Şekil 3.5: Veri setinde yer alan niteliklere ait korelogram.	69
Şekil 3.6: <i>uzaklik</i> , <i>ucusSuresi</i> , <i>seyahatYakiti</i> ve <i>yakitTasimaciligiMktr</i> niteliklerine ait saçılım diyagramları.	69
Şekil 3.7: Yakıt taşımacılığı durumunu göz önünde bulundurarak <i>tplmTrafikYuku</i> , <i>sicaklik</i> , <i>uzaklik</i> , <i>ucusSuresi</i> , <i>seyahatYakiti</i> , <i>kalkisMydnYakitFyt</i> ve <i>varisMydnYakitFyt</i> niteliklerine ait kutu grafiği karşılaştırması.	70
Şekil 4.1: <i>k</i> komşu sayısına göre doğruluk değerleri.	75
Şekil 4.2: Sınıflandırma ve Regresyon Ağaçları Algoritması için sınıflandırma ağacı.	80
Şekil 4.3: Yapay Sinir Ağı modeli (<i>Sınıflandırma</i>).	84
Şekil 4.4: <i>k</i> komşu sayısına göre RMSE değerleri.	93
Şekil 4.5: Yapay Sinir Ağı modeli (<i>Regresyon</i>).	97
Şekil 4.6: Çevrimiçi Shiny uygulaması (Yılmaz, 2021).	101

TABLO LİSTESİ

Sayfa No

Tablo 2.1: Kullanılan algoritmaların kullanım amaçları.	41
Tablo 2.2: Yapay Sinir Ağlarının sınıflandırılması (Alp, 2019).	50
Tablo 2.3: Kontenjans tablosu (Costa vd., 2007; Akpınar, 2014).	58
Tablo 3.1: Boeing 737-800 ve Airbus A320 uçaklarının teknik özellikleri (Boeing, 2019a; Airbus, 2019a; Brady, 2020; Airbus, 2020a, 2020b).	63
Tablo 3.2: Veri setindeki niteliklere ilişkin özellikler.	64
Tablo 3.3: Shapiro-Wilk normallik testi sonuçları.	68
Tablo 4.1: k-En Yakın Komşu algoritması özeti (<i>Sınıflandırma</i>).	74
Tablo 4.2: k komşu sayısına göre elde edilen doğruluk değerleri.	74
Tablo 4.3: k-En Yakın Komşu modeli için kontenjans tablosu ($k=7$).	75
Tablo 4.4: k-En Yakın Komşu modeli için model performans değerlendirme ölçütleri ($k=7 - Sınıflandırma$).	75
Tablo 4.5: Basit (<i>Naive</i>) Bayes Sınıflandırıcı algoritması Özeti.	77
Tablo 4.6: Basit Bayes Sınıflandırıcı modeli için kontenjans tablosu.	77
Tablo 4.7: Bayes Sınıflandırıcı modeli için model performans değerlendirme ölçütleri.	78
Tablo 4.8: Sınıflandırma ve Regresyon Ağaçları algoritması özeti (<i>Sınıflandırma</i>).	79
Tablo 4.9: Sınıflandırma ve Regresyon Ağaçları algoritması modeli için kontenjans tablosu.	80
Tablo 4.10: Sınıflandırma ve Regresyon Ağaçları algoritması modeli için model performans değerlendirme ölçütleri (<i>Sınıflandırma</i>).	81
Tablo 4.11: Rastgele Orman algoritması özeti (<i>Sınıflandırma</i>).	82
Tablo 4.12: Rastgele Orman algoritması modeli için kontenjans tablosu.	82
Tablo 4.13: Rastgele Orman algoritması modeli için model performans değerlendirme ölçütleri (<i>Sınıflandırma</i>).	83
Tablo 4.14: Yapay Sinir Ağları algoritması özeti (<i>Sınıflandırma</i>).	84

Tablo 4.15: Yapay Sinir Ağları algoritması modeli için kontenjans tablosu.	85
Tablo 4.16: Yapay Sinir Ağları algoritması modeli için model performans değerlendirme ölçütleri (<i>Sınıflandırma</i>).	85
Tablo 4.17: Destek Vektör Makineleri algoritması özeti (<i>Sınıflandırma</i>).	86
Tablo 4.18: Destek Vektör Makinesi algoritması modeli için kontenjans tablosu (<i>Lineer</i>).	87
Tablo 4.19: Destek Vektör Makinesi algoritması modeli için model performans değerlendirme ölçütleri (<i>Lineer - Sınıflandırma</i>).	87
Tablo 4.20: Destek Vektör Makinesi algoritması modeli için kontenjans tablosu (<i>Polinomial</i>).	88
Tablo 4.21: Destek Vektör Makinesi algoritması modeli için model performans değerlendirme ölçütleri (<i>Polinomial - Sınıflandırma</i>).	89
Tablo 4.22: Destek Vektör Makinesi algoritması modeli için kontenjans tablosu (<i>Radyal</i>).	90
Tablo 4.23: Destek Vektör Makinesi algoritması modeli için model performans değerlendirme ölçütleri (<i>Radyal - Sınıflandırma</i>).	90
Tablo 4.24: Sınıflandırma amacıyla kullanılan algoritmalar için model performans karşılaştırması.	91
Tablo 4.25: k-En Yakın Komşu algoritması özeti (<i>Regresyon</i>).	92
Tablo 4.26: <i>k</i> komşu sayısına göre elde edilen RMSE ve <i>R</i> ² değerleri	93
Tablo 4.27: k-En Yakın Komşu modeli için model performans değerlendirme ölçütleri (<i>k=7 - Regresyon</i>).	94
Tablo 4.28: Sınıflandırma ve Regresyon Ağaçları algoritması özeti (<i>Regresyon</i>).	94
Tablo 4.29: Sınıflandırma ve Regresyon Ağaçları algoritması modeli için model performans değerlendirme ölçütleri (<i>Regresyon</i>).	95
Tablo 4.30: Rastgele Orman algoritması özeti (<i>Regresyon</i>).	95
Tablo 4.31: Rastgele Orman algoritması modeli için model performans değerlendirme ölçütleri (<i>Regresyon</i>).	96
Tablo 4.32: Yapay Sinir Ağları algoritması özeti (<i>Regresyon</i>).	96
Tablo 4.33: Yapay Sinir Ağları algoritması modeli için model performans değerlendirme ölçütleri (<i>Regresyon</i>).	97
Tablo 4.34: Destek Vektör Makineleri algoritması özeti (<i>Regresyon</i>).	98

Tablo 4.35: Destek Vektör Makineleri algoritması modeli için model performans değerlendirme ölçütleri (<i>Lineer - Regresyon</i>).	98
Tablo 4.36: Destek Vektör Makineleri algoritması modeli için model performans değerlendirme ölçütleri (<i>Polinomial - Regresyon</i>).	99
Tablo 4.37: Destek Vektör Makineleri algoritması modeli için model performans değerlendirme ölçütleri (<i>Radyal - Regresyon</i>).	99
Tablo 4.38: Regresyon amacıyla kullanılan algoritmalar için model performans karşılaştırması.	100



SİMGE VE KISALTMA LİSTESİ

Simgeler	Açıklama
\$	: Amerikan doları
°C	: Derece santigrat
bbl	: Varil
CO ₂	: Karbondioksit
kg	: Kilogram
lb	: Libre
lt.	: Litre
Mt.	: Milyon ton
Nm	: Nautical mile
R ²	: Belirleme katsayısı

Kısaltmalar	Açıklama
AID	: Otomatik Etkileşim Algılama
APU	: Yardımcı Güç Ünitesi
ARF	: Akut Romatizmal Ateş
ART	: Uyarlanabilir Rezonans Teorisi
ATAG	: Hava Taşımacılığı Faaliyet Grubu
ATM	: Hava Trafik Yönetimi
BDK	: İki-Yönlü Kohonen Ağı
CART	: Sınıflandırma ve Regresyon Ağaçları
CNS	: İletişim, Navigasyon ve Gözetim
DDoS	: Dağıtık Hizmet Engelleme
dn	: Doğru Negatif
dp	: Doğru Pozitif
DVM	: Destek Vektör Makineleri
ECAC	: Avrupa Sivil Havacılık Konferansı
EDA	: Keşfedici Veri Analizi
EIA	: Amerika Enerji Bilgi İdaresi
EU ETS	: Avrupa Birliği Emisyon Ticaret Sistemi

FAA	: Amerika Birleşik Devletleri Federal Havacılık Kurulu
FCOM	: Uçuş Ekibi Operasyonel Dokümanı
GPS	: Küresel Konumlama Sistemi
ICAO	: Uluslararası Sivil Havacılık Örgütü
ISA	: Uluslararası Standart Atmosfer
IATA	: Uluslararası Hava Taşımacılığı Birliği
KLM	: Hollanda Kraliyet Hava Yolları
kNN	: k-En Yakın Komşu
LVQ	: Öğrenmeli Vektör Nicelendirme
MIT	: Massachusetts Teknoloji Enstitüsü
ML	: Makine Öğrenmesi
MLW	: Minimum İniş Ağırlığı
MTOW	: Maksimum Kalkış Ağırlığı
NAA	: Ulusal Havacılık Otoritesi
NB	: Basit Bayes
OCC	: Operasyon Kontrol Merkezi
OFP	: Operasyonel Uçuş Planı
RAAF	: Avustralya Kraliyet Hava Kuvvetleri
RF	: Rastgele Orman
SOM	: Özdüzenleyici Haritalar
UPS	: United Parcel Service
U.S.	: Amerika Birleşik Devletleri
XYF	: XY Kaynaşık Kohonen Ağı
yn	: Yanlış Negatif
yp	: Yanlış Pozitif
YSA	: Yapay Sinir Ağları
YZ	: Yapay Zekâ

ÖZET

DOKTORA TEZİ

MAKİNE ÖĞRENMESİ TEKNİKLERİ İLE HAVACILIKTA YAKIT TAŞIMACILIĞI KARARININ ALINMASI VE BİR UYGULAMA

İlker Güven YILMAZ

İstanbul Üniversitesi

Fen Bilimleri Enstitüsü

Enformatik Anabilim Dalı

Danışman : Prof. Dr. Sevinç GÜLSEÇEN

II. Danışman : Dr. Elif KARTAL

Bu tez çalışmasının amacı havacılıkta yakıt taşımacılığı kararının makine öğrenmesi teknikleri ile alınmasıdır. Çalışmada kullanılan veri seti Türkiye’de hava taşımacılığı yapan ticari bir hava yolu firmasından temin edilmiştir. Literatürde yakıt taşımacılığı hesaplamasında kullanılan bazı temel özelleştirilebilir formüller/modeller mevcuttur. Bu çalışma ile bu formüllerden/modellerden bağımsız olarak ham veriden öğrenen makine öğrenmesi algoritmaları ile havacılık endüstrisinde yakıt taşımacılığı öngörüsünde bulunmak hedeflenmiştir. Çalışmada yakıt taşımacılığı durumunun “Var/Yok” şeklinde sınıflandırılması ve yakıt taşımacılığı miktarının tahmini için alternatif makine öğrenmesi modelleri oluşturulmuştur. k-En Yakın Komşu algoritması, Basit Bayes Sınıflandırıcı, Sınıflandırma ve Regresyon Ağaçları, Rastgele Orman algoritması, Yapay Sinir Ağları ve Destek Vektör Makinelerinden yararlanılmıştır. Sınıflandırma ve regresyon model performansları birbiriyle karşılaştırılmıştır. Analizler R programlama dili ile RStudio’da gerçekleştirilmiştir. En iyi performansı gösteren Rastgele Orman sınıflandırma (doğruluk=0,9646) ve regresyon ($R^2=0,8104$) modelleriyle çevrimiçi bir Shiny uygulaması geliştirilmiştir.

Ağustos 2021, 136 sayfa.

Anahtar kelimeler: Havacılık, makine öğrenmesi, maliyet azaltma, yakıt taşımacılığı

SUMMARY

Ph.D. THESIS

DECISION-MAKING IN FUEL TANKERING WITH MACHINE LEARNING TECHNIQUES IN AVIATION AND ITS APPLICATION

İlker Güven YILMAZ

İstanbul University

Institute of Graduate Studies in Sciences

Department of Informatics

Supervisor : Prof. Dr. Sevinç GÜLSEÇEN

Co-Supervisor : Dr. Elif KARTAL

In this thesis, it is aimed at decision-making in fuel tankering with machine learning techniques in aviation. The dataset used in this study is obtained from a commercial airline company based in Turkey. There are several basic customizable formulas/models used in the fuel tankering calculation referred in the literature. In this study, it is aimed to predict fuel tankering in the airline industry with machine learning algorithms that learn from raw data which is independent of these formulas/models. In this study, alternative machine learning models are created to classify the fuel tankering status as "Yes/No" and to predict the amount of fuel tankering. k-Nearest Neighbour algorithm, Naive Bayes Classifier, Classification and Regression Trees, Random Forest algorithm, Artificial Neural Networks, and Support Vector Machines algorithm are used. Performance of classification and regression models are compared with each other. Analyzes are carried out in RStudio with the R programming language. An online Shiny application is developed with the best performing Random Forest classification (Accuracy=0,9646) and regression ($R^2=0,8104$) models.

August 2021, 136 pages.

Keywords: Aviation, machine learning, cost reduction, fuel tankering

1. GİRİŞ

Havacılık sektörü oldukça rekabetçi bir pazardır ve bu nedenle hava yolu şirketlerinin mümkün olduğunca rekabetçi bilet fiyatlarını koruması için işletme maliyetlerini en aza indirmeye ihtiyacı vardır (Eurocontrol, 2019). Diğer sektörlerle benzer şekilde havacılık sektöründeki şirketlerin de maliyeti azaltmak için dikkat ettiği bazı parametreler vardır. Bunlar; bakım masrafları, yakıt giderleri, parça ve işçilik giderleri, havaalanı ücretleri, vergiler, pazarlama ve promosyon giderleri, seyahat acentesi komisyonları ve yolcu giderleridir (FAA, 2017). Her uçuş bir görev olarak düşünülürse bu görevin tamamlanması için - yani bir uçağın A noktasından B noktasına gitmesi için - gerekli olan yakıtı uçağa alması şarttır. Alınan yakıt miktarı o uçuşun uzaklığı, uçaktaki toplam trafik yükü, uçağın performansı ve kısıtlamaları, varış meydanı hava durumu, varış meydanı kısıtlamaları, üst geçişlerdeki kısıtlamalar vb. durumlara bağlı olarak uçuş harekât uzmanları tarafından hesaplanmaktadır. Uluslararası Sivil Havacılık Örgütü (*International Civil Aviation Organization - ICAO*) tarafından uluslararası standartları belirlemek amacıyla uçuş operasyonu esnasında bir uçağın taşınması gereken yakıtlara ait tavsiyeler ICAO EK 6 (*ICAO Annex 6*) Bölüm 4'te 4.3.6 Yakıt Gereklilikleri (*Fuel Requirements*) başlığı altında Madde 4.3.6.1 ve Madde 4.3.6.2 ile sunulmuştur (ICAO, 2010). Öte yandan yakıt fiyatları günlük ve hatta anlık olarak değişebilmektedir. Bu nedenle uçağa alınan yakıtın miktarı oldukça önemlidir. Alınan her bir kilogram ekstra yakıt ekstra gider demektir ve o uçuş için maliyeti artırmaktadır. Havacılık sektöründe yakıt alımında maliyet kaleminin göz önünde bulundurulduğu ve şirketlerin sıklıkla kullandığı uygulamalardan biri yakıt taşımacılığı (*fuel tankering*)'dır.

Örneğin; A ve B noktaları arasında gidiş-dönüş bir uçuş görevini ele alalım. A noktasındaki yakıt fiyatı B noktasına göre daha ucuz olsun. Bu durumda, uygun şartlar altında B noktasından A noktasına dönüş için gerekli olan yakıtın belirli bir miktarının veya tamamının A noktasından alınması mümkündür. Böylelikle, toplam yakıt maliyeti azaltılmaktadır. Burada; A ve B noktalarındaki yakıt fiyatı A noktasından B noktasına olan uçuş öncesi uçuş planlanma safhasında (*pre-flight*) bilinmektedir ve A noktasından alınarak B noktasına taşınacak yakıtın kâr sağlayıp/sağlamama durumuna karar verilmektedir.

Yakıt taşımacılığı durumuna (“*Evet*” *yakıt taşımacılığı yapılsın* veya “*Hayır*” *yakıt taşımacılığı yapılmassın*) karar vermek için çeşitli yöntemlerden faydalanılmaktadır. Bunlardan ilki, detayları bu tez çalışmasının GENEL KISIMLAR bölümünde verilmiş olan çeşitli matematiksel eşitsizlikleri temel almaktadır. Havacılıkla ilgili yeterli yetkinlik ve teknik bilgiye sahip alanında uzman kişilerce bu eşitsizlikler yorumlanabilir. Uçuş sayısının az olduğu durumda kolay kullanılabilir olması ve maliyetinin olmaması bu yöntemin avantajlarından. Öte yandan bu yöntemde insan faktörü devreye girdiği için hesaplamalarda hatalı sonuçlarla karşılaşılabilme riski bu yöntemin bir dezavantajı olarak görülebilir. Ek olarak, uçuş sayısının fazla olduğu durumlarda hesaplamaların da uzun sürebilmesi bu yöntemin bir diğer dezavantajı olarak görülebilir. Ayrıca, bu eşitsizliklerden bazılarında parametrelerin özelleştirilebilir olması, hesaplamayı yapan tarafından farklı parametre tercihlerini ortaya çıkardığından elde edilecek hesaplama sonuçları farklılık gösterebilmektedir.

İkinci alternatif yöntem ise şirketlerin her uçuş öncesi ilgili uçuşta yakıt taşımacılığı durumuna karar vermek ve yakıt taşımacılığı miktarını hesaplamak için çeşitli ticari yazılımlardan faydalanmasıdır. Bu yazılımlar, insan faktörü yerine sahip oldukları algoritmalar aracılığıyla hesaplamaları gerçekleştirmektedir. Uçuş sayısının fazla olduğu durumlarda hesaplama sürelerini kısaltması ve insan tarafından temel formüller kullanılarak elle yapılacak hesaplama hatalarını minimuma indirmesi bu yöntemin avantajlarındadır. Öte yandan bu ticari yazılımların belirli süreye sahip lisanslama giderlerinin uçuş firmalarına bir maliyet kalemi oluşturması bu yöntemin dezavantajıdır. Örneğin; Etiyopya Hava Yolları, Lufthansa tarafından geliştirilen Lido/Flight 4D uçuş planlama yazılımı ile Lido Uçuş Planlama Servislerini (*Lido Flight Planning Services / (Lido/FPLS)*) kullanması karşılığında Lufthansa ile beş yıl süreli olmak üzere 3.380,000 Euro bedelli bir anlaşma imzalamıştır (TAL Aviation Group, 2021).

Değişken piyasa koşullarının işletmelere getirdiği rekabet koşulu havacılık sektöründeki işletmeleri de ekonomik açıdan zor durumda bırakmaktadır. Bilindiği gibi günümüzde bilgi bir işletmenin elindeki en önemli sermaye olarak görülmektedir. İşletmeler bilgiyi kendilerine yarar sağlayacak stratejiler belirlemek, katma değer yaratmak için kullanmaya çalışmaktadır. Bilginin ham maddesi olan verinin gelişen teknoloji sayesinde hızlı biçimde üretimi ve depolanması büyük veri kavramını da beraberinde getirmiştir. IBM (2021a) büyük veriyi (*big data*), klasik-ilişkisel veritabanlarının düşük sürede gerçekleştirdikleri veriyi yakalama, yönetme ve işleme yeteneğinin ötesinde bir büyüklük ve türe sahip olan veri kümeleri olarak

tanımlamıştır. Büyük veri hem veri hacmini hem veri oluşum hızını hem de verideki çeşitliliği ifade etmektedir (Atalay ve Çelik, 2017). IBM (2021a) büyük verinin; yüksek hacim (*high volume*), yüksek hız (*high velocity*) ve yüksek çeşitlilik (*high variety*) özelliklerine sahip olduğunu belirtmiştir. Jin vd. (2015) ise veri ile karşılaştırıldığında, büyük verinin 5V olarak adlandırılan; yüksek hacim (*huge volume*), yüksek hız (*high velocity*), yüksek çeşitlilik (*high variety*), düşük doğruluk (*low veracity*) ve yüksek değer (*high value*) özellikleri ile karakterize edilebileceğini ifade etmiştir. Ayrıca; büyük verinin karakteristik özelliği bakımından yapılandırılmamış veri (*unstructured data*) içerdiğinden belirsizliklere sahip olduğunu, öte yandan verinin yapılandırılması ile ortaya değerli veri çıkabileceğinden yüksek değere sahip olduğunu ifade etmiştir. Wu vd. (2014) büyük verinin sahip olduğu özelliklerin büyük veriden yararlı bilgiler keşfedilmesini son derece zorlaştırdığını ifade etmiştir. Casolla vd. (2020) büyük veriyi, verinin bulut ortamında sıklıkla kümelenmiş ve durgun bir yığını olarak tanımlamışlardır. Ayrıca bu tür veriyi kullanmak için verinin yeniden yapılandırılmasının ve sorgulanmasının gerekli olduğunu belirtmişlerdir. Bu veriyi sorgulamak ve faydalı bilgiler üretmek için tahmin modelleri oluşturmak amacıyla veriyi kategorize edebilen ve sınıflandırabilen algoritmalara da ihtiyaç olduğunu belirtmişlerdir.

Verinin işe yarar anlamlı bilgiye dönüştürülmesi ve karar süreçlerinde kullanılması oldukça önemlidir. Mair vd. (2000) her gün tahminlere dayalı değerli kaynakların nasıl ayrılacağına karar vermenin işletmeler için bir gereklilik olduğunu belirtmiştir. Bu aşamada, makine öğrenmesi (*machine learning*) devreye girmektedir. Flach (2012) makine öğrenmesini, bilgilerini veya performansını deneyim ile geliştiren algoritma ve sistemlerin sistematik bir biçimde çalışması olarak tanımlamıştır. Bu tanımdaki “deneyim” makineye ilgili görev için sunulan veri olarak düşünülebilir.

Yapay zekâ ve onun alt çalışma alanlarından biri olan makine öğrenmesinin hemen hemen her alanda uygulamasını bulmak mümkündür. Örneğin; Arslan (2020)’ın eğitimde yapay zekâ ve uygulamaları çalışması, Akalın ve Veranyurt (2020)’un sağlıkta dijitalleşme ve yapay zekâ çalışması vb. Havacılık da bu alanlardan biridir. Korvesis (2017) havacılıkta öngörüye dayalı bakım için makine öğrenmesinin kullanıldığı bir çalışma ortaya koymuştur. Rahim Taleqani vd. (2018) havacılıkta siber güvenlik için makine öğrenmesi yaklaşımını ortaya koymuşlardır. Bununla birlikte uçuş planlama (Kurklu vd., 2008), havacılık türbülansını teşhis etme (Williams, 2014) gibi makine öğrenmesini doğrudan havacılığa entegre eden çalışmalar da

mevcuttur. Collins ve Thomas (2012) basit bir dinamik hava yolu fiyatlandırma oyununu çözenin yolu olarak makine öğrenmesi stratejilerinden biri olan pekiştirmeli öğrenme (*reinforcement learning*) kullanımını araştırmıştır. Ali (2010) KLM Cityhoppers (KLC) Fokker 70/100 filosu için mümkün olduğunca yakıt taşımacılığını kullanarak (*yakıt taşımacılığı ile alınabilir olan yakıt miktarı*) ve yüksek buzlanma (*de-icing*) riskinden kaçınarak kârı optimize eden bir yaklaşım sunmuştur. Ek olarak, yüksek buzlanmayı tahmin etmede önemli faktörleri tanımlamak için keşfedici veri analizini (*Exploratory Data Analysis - EDA*) kullanmıştır. Son olarak, en iyi yakıt taşımacılığı miktarını tahmin etmeyi hesaplayan bir Excel sayfası oluşturmuştur.

Literatür incelendiğinde; büyük veri ve makine öğrenmesini içeren havacılık sektörü için geliştirilmiş bazı yazılımların da mevcut olduğu görülmektedir. Örneğin; OpenAirlines (2020a, 2020b) şirketinin geliştirdiği SkyBreathe isimli yazılım büyük veri algoritmaları, yapay zekâ ve makine öğrenmesi gibi son dijital teknolojileri kullanarak hava yollarına yakıt tüketimini azaltma noktasında çözümler sunmaktadır. Bu şirketin geliştirdiği bir diğer yazılım olan MyFuelCoach isimli mobil yazılım; pilotlara yaklaşan uçuşlarına hazırlanma, potansiyel yakıt tasarruflarını belirleme, kendi ilerlemelerini takip etme, operasyonel uçuş planı ile karşılaştırma, uçuşunu 4D ile tekrar izleme vb. gibi imkânlar sunmaktadır. Honeywell şirketi hava yolu kârlılığını arttırmak için büyük veri ve makine öğrenmesini kullanan Honeywell Forge for Airlines isimli yazılım platformunu geliştirmiştir (Honeywell, 2020a). Bu platformun uçuş verimliliği modülü (*Honeywell Forge Flight Efficiency*) uçuş verimliliğini artırma noktasında hava yollarına ve pilotlara çözümler sunmaktadır (Honeywell, 2020b). Öyle ki bu modül; verimliliği arttırmak ve maliyetleri azaltmak amacıyla uçuş planları, hava durumu, seyrüsefer haritaları, uçak performansı ve daha fazlasını içeren tüm uçuş değişkenlerini tek bir noktada bir araya getirerek müşterilerine tüm filo verisine kapsamlı bir bakış atma olanağı sunmaktadır.

Görülmektedir ki; günümüzde yapay zekâ pek çok alana girdiği gibi havacılıkta da kullanılmakta olup yapay zekânın bir alt çalışma alanı olarak makine öğrenmesi teknikleriyle geliştirilecek modellerin karar vericilere fayda sağlaması hedeflenmektedir. Yakıt taşımacılığı özelinde son nihaî karar verici uçuş planlamasını gerçekleştiren alanında uzman kişilerdir (*uçuş harekât uzmanları*); ancak insanın doğası gereği çevresiyle olan etkileşimi karar verme süreçlerini de etkilemektedir. Nabiye (2005) yapay zekâ ve doğal zekâ arasındaki birtakım

farkları ortaya koyarken, yapay zekânın; daha kalıcı, kolay kopyalanabilir ve geniş kitlelere aktarılabilir, maliyeti düşük, tutarlı ve belgelenebilir olduğundan bahsetmiştir. Yakıt taşımacılığı durumu ya da miktarını tahmin edecek şekilde geliştirilen makine öğrenmesi modellerinin ya da bu modellerin entegre edildiği sistemlerin, Nabiyev (2005)'in de belirttiği yapay zekânın getirdiği avantajlarla birlikte alınacak kararları daha rasyonel hale getireceğine ve ortaya çıkabilecek hatayı en aza indirgeyerek karar vericilerin daha doğru kararlar alabilmesini sağlayabileceğine inanılmaktadır.

Verilen tüm bu bilgiler ışığında bu tez çalışmasında; makine öğrenmesi teknikleriyle havacılıkta kâr sağlayan yakıt taşımacılığı durumunun ve miktarının tahmin edilmesi amaçlanmaktadır. Bu kapsamda geliştirilen makine öğrenmesi tahmin modellerinin havacılık alanında faaliyet gösteren şirketlere maliyetleri azaltma yönünde destek olunacağı düşünülmektedir. Çalışmada kullanılan veri seti Türkiye'de hava taşımacılığı yapan ticari bir hava yolu firmasından temin edilmiştir. Yakıt taşımacılığı durumu ve miktarının tahmininde sırasıyla sınıflandırma ve regresyon tabanlı farklı makine öğrenmesi algoritmalarıyla alternatif modeller oluşturulmuştur. Yakıt taşımacılığı durumu ve miktarının tahmininde en iyi performansı sergileyen makine öğrenmesi modelleri seçilmiş, bu modellerle çevrimiçi bir Shiny uygulaması geliştirilmiştir.

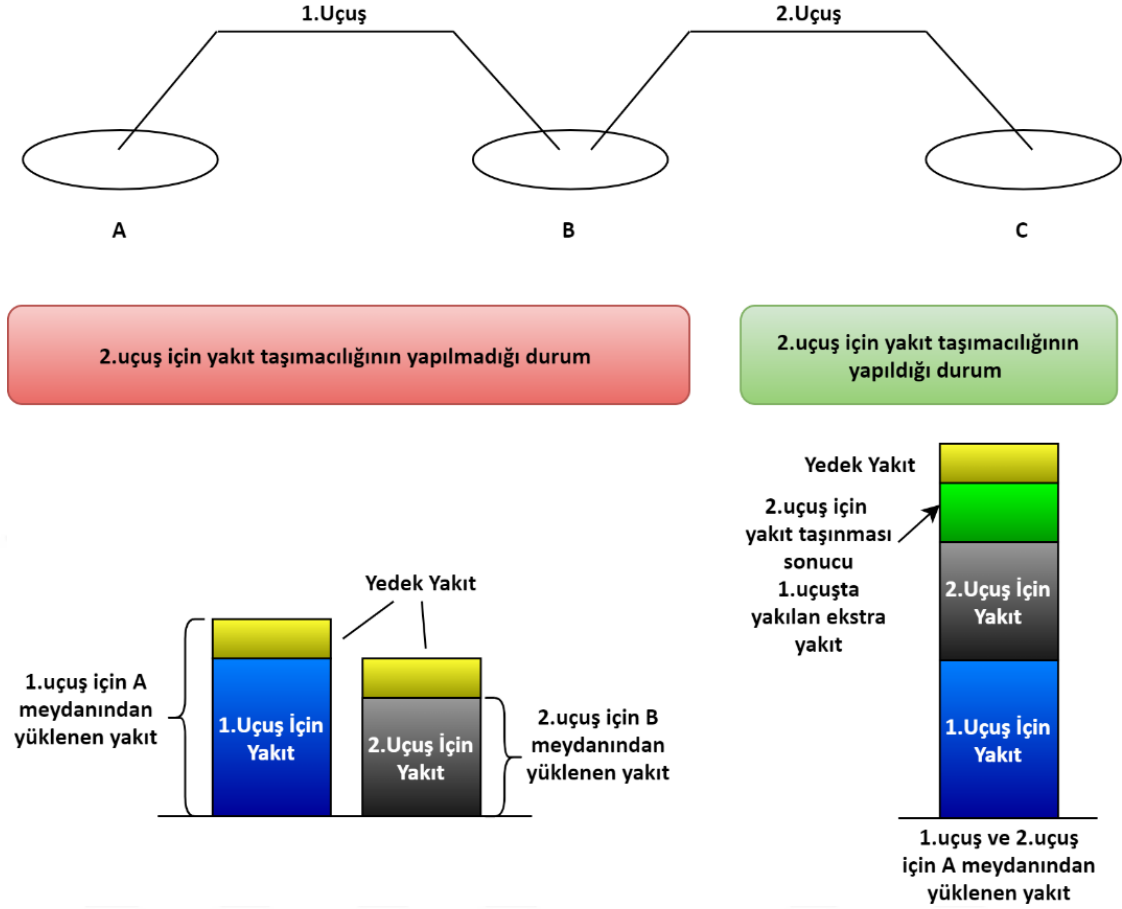
2. GENEL KISIMLAR

Bu bölümde, çalışmanın genel kısımlarını oluşturacak yakıt taşımacılığı, makine öğrenmesi ve makine öğrenmesi süreci adımları konuları sırasıyla ele alınmıştır. Literatürde yakıt taşımacılığı ile ilgili yer alan bilgiler sunulurken, yakıt taşımacılığının tarihçesi, yakıt taşımacılığı uygulamaları, yakıt taşımacılığı ve yönetmelikler, yakıt taşımacılığının hava yolu maliyetleri üzerindeki etkisi ve yakıt taşımacılığının ekonomik yararları ve çevresel etkileri konuları ayrıntılı bir şekilde incelenmiştir. Ardından makine öğrenmesi konusu ayrıntılı olarak incelenerek, tez çalışmasının MALZEME VE YÖNTEM kısmını oluşturacak veri analizleri için makine öğrenmesi süreci adımları ele alınmıştır.

2.1.YAKIT TAŞIMACILIĞI

“Yakıt taşımacılığı” veya İngilizce adıyla “*fuel tankering*” bir sonraki uçuş (veya uçuşlar) için varış meydanından uçağa yüklenecek yakıt miktarını azaltmak adına, belirli bir uçuş için gerekli olan yakıttan daha fazlasının kalkış meydanından uçağa yüklenerek taşınması uygulamasıdır (Boeing, 2004; Eurocontrol, 2019).

Şekil 2.1 ile toplam iki uçuşa sahip bir görev için yakıt taşımacılığı uygulaması gösterilmiştir. Örneğin, bir uçağın A noktasından B noktasına gittiği ve sonrasında B noktasından C noktasına devam ettiği durum ele alınsın. Şekil 2.1’den de görüldüğü üzere, A noktasından B noktasına giderken kalkış noktasından (A), A-B uçuşu (1.uçuş) için gerekli yakıt alınır. Çeşitli nedenlerden ötürü B noktasından C noktasına gitmek (2.uçuş) için gerekli olan yakıtın tamamının veya uygun miktarının A noktasından alınması ve B noktasına taşınması mümkündür. Bu uygulama yakıt taşımacılığı olarak ifade edilmektedir. Burada amaç, bir sonraki uçuş (B’den C’ye) için gerekli olan ve B noktasından alınacak olan yakıt miktarını azaltmak veya sıfırlamaktır.



Şekil 2.1: Yakıt taşımacılığı uygulaması (Boeing, 2004).

Yakıt taşımacılığında taşınacak yakıtın miktarını sınırlayan durumlar şunlardır (Boeing, 2004):

- Onaylı Maksimum Kalkış Ağırlığı (*Certified Maximum Take-off Weight – MTOW*),
- Performans Limitli Maksimum Kalkış Ağırlığı (*Performance-limited MTOW*),
- Onaylı Maksimum İniş Ağırlığı (*Certified Maximum Landing Weight – MLW*),
- Performans Limitli Maksimum İniş Ağırlığı (*Performance-limited MLW*),
- Yakıt Tankı Kapasitesi (*Fuel Capacity*).

Boeing (2004)'e göre yakıt alımının nedenleri;

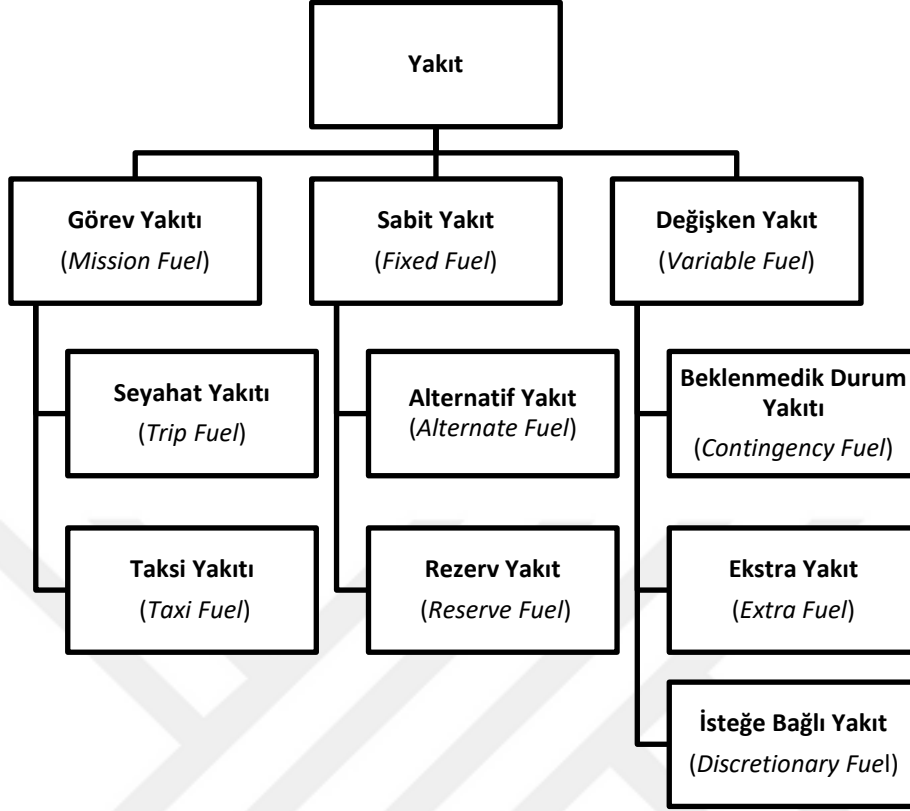
- Varış meydanındaki kısa yükleme-boşaltma süresi,
- Varış meydanında sınırlı miktarda yakıtın olması,
- Varış meydanındaki güvenilirmez havalimanı servisleri,
- Varış meydanındaki yakıtın kalitesi,
- Yakıt fiyatı farklılıkları,

olmasına rağmen, yakıt taşımacılığı için genellikle asıl neden birden çok uçuş için toplam yakıt maliyetini azaltmaktır. Bu aynı zamanda, yakıt taşımacılığının hava yolları için önemli olmasının nedenidir. Filippone (2012) bazı ticari işleticilerin varış meydanında yüksek fiyata yakıt satın alma maliyetini dengelemek için görev profiline göre gerekli olandan fazla yakıt yüklemesi yaparak yakıt taşımacılığı uygulamasını kullandığını belirtmiştir.

Bu çalışmada, “kalkış ve varış meydanları arasındaki yakıt fiyatı farklılıkları”ndan ötürü yapılan yakıt taşımacılığı uygulaması ele alınmıştır. Buna göre; bir uçağa bir sonraki uçuş için kalkış meydanından alınan fazla; ancak varış meydanına göre kârlı yakıtın varış meydanına taşınmasıdır. Taşınan bu yakıt, toplam yakıt maliyetini azalttığı (*cost reduction*) için kârlı bir yakıttır ve yapılan taşımacılık işlemi “kâr sağlayan yakıt taşımacılığı” olarak da ifade edilebilir.

Kheng (2021) uçuş planlama sürecinin hava yolu uçuş operasyonun en temel parçası olduğunu ve uçuş harekât uzmanlarının her bir uçuşun güvenliğini sağlamak için birçok parametreyi; örneğin hava trafiği, hava sahaları, havalimanları ve hava durumu ile uçak performans verisini analiz etmek zorunda olduğunu belirtmiştir. Ayrıca her bir uçuş için en uygun uçuş rotasının hesaplanmasının gerektiğini ifade etmiştir. Uçağa yüklenecek yakıtın hesaplanması da tüm bu hesaplama sürecinin bir parçasıdır.

Uçuş planlama aşamasında bir uçuş için uçağa yüklenen toplam yakıt; görev yakıtı (*mission fuel*), alternatif yakıt (*alternate fuel*), yedek yakıt (*reserve fuel*), teknik arıza vb. acil durumlar için beklenmedik durum yakıtı (*contingency fuel*) vb. gibi yakıtları içermektedir (Zhu ve Li, 2020). Şekil 2.2’de hava yolu uçuş planlama aşamasında uçağa yüklenen yakıtlar kategorilere ayrılarak verilmiştir.



Şekil 2.2: Hava yolu uçuş planlaması için temel yakıt yükleme kategorileri (Zhu ve Li, 2020).

Uçağa yüklenen yakıtlar için detaylı açıklamalar aşağıda sunulmuştur (SKYbrary, 2020; ICAO, 2010):

Seyahat yakıtı (*trip fuel*); seyahat yakıtı kalkış meydanında uçağın ilk fren bıraktığı andan varış meydanına teker koyduğu ana kadar gerekli yakıt miktarıdır. Bu yakıt; kalkış, uçuş seviyesine tırmanma, uçuş seviyelerine göre planlanan tırmanma ve alçalmanın olduğu uçuş, inişin başlangıcından yaklaşmanın başlangıcına kadar olan uçuş, yaklaşma ve varış meydanına iniş için gerekli olan yakıtları içermektedir.

Taksi yakıtı (*taxi fuel*); kalkıştan önce kullanılan yakıttır ve genellikle çalıştırma öncesi APU (*Auxiliary Power Unit*) tüketimi, motor çalıştırma ve taksi yakıtını içerir. Taksi yakıtı genellikle ortalama bir taksi süresi için sabit bir miktardır; ancak kalkış meydanındaki ortalama taksi süresi, normal yer gecikmeleri ve beklenen buz çözme (*de-icing*) gecikmeleri gibi yerel koşullar dikkate alınmalı ve taksi yakıtı buna göre ayarlanmalıdır.

Alternatif yakıt (*alternate fuel*) veya yedek meydan yakıtı; varış meydanı için kaçırılan yaklaşma noktasından alternatif meydana inişe kadar gereken yakıt miktarıdır. Alternatif yakıt; varış meydanı için yaklaşmanın kaçırılması (*missed approach*), alternatif meydana (*enroute alternate*) gitmek için gereken yüksekliğe tırmanma, seyir ve alçalma, alternatif meydana yaklaşma ve alternatif meydana iniş için gerekli olan yakıtı içermektedir.

Son rezerv yakıt (*final reserve fuel*), **sabit rezerv yakıtı** (*fixed reserve fuel*) veya **bekleme yakıtı** (*holding fuel*) alternatif meydanın veya alternatif meydanın gerekli olmadığı durumda varış meydanının 1500 feet üzerinde ISA (*International Standard Atmosphere*) koşullarında 30 dakika boyunca uçmak için gereken minimum yakıttır. **Rezerv yakıt** (*reserve fuel*); alternatif yakıt ve son rezerv yakıtın toplamıdır.

Beklenmedik durum yakıtı (*contingency fuel*); rüzgâr, yönlendirme değişiklikleri veya ATM: ATM/CNS (*Communication, Navigation and Surveillance/Air Traffic Management*) kısıtlamalarından kaynaklanan ek yol boyu yakıt tüketimini hesaba katmak için taşınır. ICAO (2010) tarafından yayınlanan Uluslararası Sivil Havacılık Sözleşmesi Ek 6'ya göre, önerilen minimum beklenmedik yakıt, seyahat yakıtının %5'i veya hesaplanan varış ağırlığına göre hesaplanan varış meydanı pistinin 1500 feet yüksekliğinde 5 dakikalık bekleme yakıtıdır. Bununla birlikte, bazı düzenleyiciler minimum zaman gereksinimini ortadan kaldırmış, bazıları ise ulusal düzenlemelerinde önerilen zaman aralığını artırmıştır. Ayrıca bazı düzenleyiciler, yol boyu yedek meydanların kullanılması ve operatör tarafından gösterilen performans kriterlerine bağlı olarak beklenmedik durum yakıtının seyahat yakıtının %3'üne kadar azaltılmasına veya belirli zaman artışlarına izin vermektedir. En az bir otorite, çok özel şartlar altında, beklenmedik yakıtın %0'a düşürülmesine izin verir. Tüm durumlarda, bir operatör mürettebatını Ulusal Havacılık Otoritesi (NAA – National Aviation Authority) tarafından talep edilenden fazla beklenmedik yakıt taşıması konusunda yönlendirebilir.

Ekstra yakıt (*extra fuel*), kaptan pilotun veya uçuş harekât uzmanının takdir yetkisine bağlı olarak eklenen yakıttır.

Yakıt taşımacılığı uygulamasında uçağa fazladan alınan yakıt da ekstra yakıttır. Operasyonel uçuş planında (*Operational Flight Plan - OFP*) ekstra yakıt (*tankering*) şeklinde gösterilmektedir.

Yakıt taşımacılığının yapıldığı/yapılmadığı durumlarda taşınan ve harcanan yakıtlar Şekil 2.3 ve Şekil 2.4'te gösterilmiştir. Örneğin, bir uçağın A noktasından B noktasına gittiği ve B noktasından A noktasına tekrar geri döndüğü durum ele alınsın. Şekil 2.3'ten de görüldüğü üzere A noktasından B noktasına giderken kalkış noktasından (A), o uçuşta gerekli yakıt (A-B *seyahat yakıtı*) alınır. A'dan B'ye giderken A-B seyahat yakıtı yakılır ve B noktasına varıldığında geriye yedek yakıt (*reserve fuel*) kalır. B noktasından tekrar A'ya dönmek için gerekli olan B-A seyahat yakıtı B noktasından alınır. B'den A'ya giderken B-A seyahat yakıtı yakılır ve A noktasında geriye yedek yakıt kalır.



Şekil 2.3: Yakıt taşımacılığı'nın yapılmadığı duruma ait taşınan ve yakılan yakıtlar (Eurocontrol, 2019).

Öte yandan B noktasından A noktasına dönmek için gerekli olan yakıtın tamamının veya belirli bir miktarının yukarıda da ele alınan nedenlerden ötürü A noktasından alınması mümkündür. Şekil 2.4'ten de görüldüğü üzere kalkış noktasından (A) alınan ve varış noktasına (B) giderken taşınan bu yakıt (B-A *seyahat için taşınan yakıt*) ile “yakıt taşımacılığı” yapılmış olunur.



Şekil 2.4: Yakıt taşımacılığı'nın yapıldığı duruma ait taşınan ve yakılan yakıtlar (Eurocontrol, 2019).

Her ne kadar bu tez çalışması kapsamında ele alınmasa da yakıt taşımacılığının çoklu versiyonları da düşünülebilir. Örneğin; bir uçağın A noktasından B noktasına, B noktasından da C noktasına gittiği durum ele alınsın. A'dan B'ye gitmek için gerekli olan seyahat yakıtına ek olarak B'den C'ye gitmek için gerekli olan seyahat yakıtının tamamı veya uygun bir miktarı, kârlılık verdiği durumda A noktasından alınabilir. Bu durum çoklu yakıt taşımacılığı (*multiple fuel tankering*) olarak ifade edilmektedir.

2.1.1. Yakıt Taşımacılığının Tarihçesi

Havacılıkta yakıt taşımacılığı uygulamasının ilk olarak ne zaman kullanılmaya başlandığına dair literatürde kesin bir bilgi elde edilememiştir. Havacılık sektörünün gelişmesine paralel olarak uçakların gelişmesi, yakıt kapasitelerinin artması, menzillerinin artması gibi şirketleri olumlu etkileyen gelişmelerin yanında yakıt fiyatlarındaki farklılıkların getirdiği maliyetler ve şirketlerin maliyetlerini azaltma politikaları sonucu ortaya çıktığı düşünülmektedir.

2.1.2. Yakıt Taşımacılığı Uygulamaları

Gürsoy (2020) hava yollarının fazladan yakıt taşıma stratejisini modelleyen bir doğrusal olmayan programlama modeli geliştirmiş ve oluşturduğu doğrusal olmayan programlama problemini çözmek için genel indirgenmiş gradyan (*generalized reduced gradient*) metodundan yararlanmış ve Microsoft-Excel (*MS-EXCEL*) çözücü eklentisini kullanmıştır. Tasarladığı senaryolar ile yakıt taşımacılığı yapılmayan seferlere kıyasla yakıt taşımacılığı yapılan

seferlerde yakıt sarfiyatının %2,5-3 arttığını fakat gidiş-dönüş seferleri için toplam yakıt maliyetinden %1-%47 arasında tasarruf sağlanabileceğini tespit etmiştir.

Hubert vd. (2015) verilen bir görevde maliyetten kaçınmak için yakıt alınıp alınmayacağına karar vermede; fazla yakıt taşımak için uygun kapasiteyi, yakıt yanma penaltısını, gelecek görev bacağı için gerekli yakıt miktarını ve kalkış ve varış yerlerindeki yakıt fiyatı farkını dört temel faktör olarak tanımlamıştır. Ek olarak, yakıt alınıp alınmayacağı ve ne kadar alınacağına karar vermek için karar verme sürecini gösteren bir algoritma vermiştir.

Yakıt taşımacılığına karar vermek için havacılık endüstrisinde kullanılmakta olan birçok model vardır. Lesinski (2011) yakıt taşımacılığı seçeneğini Hava Harekât Komutanlığı'nda (*Air Mobility Command - AMC*), Amerika Birleşik Devletleri Hava Kuvvetleri'nde ve Savunma Departmanı'nda uygun bir maliyet tasarrufu girişimi olarak incelemiştir ve bir yakıt taşımacılığı modeli geliştirmiştir.

Lindgren ve Brynhagen (2012) yakıt taşımacılığından kâr elde edilip edilmediğini incelemek için aşağıdaki eşitsizliği kullanmışlardır:

$$P_{dep} \times (F + F_B (F)) < P_{des} \times F \quad (2.1)$$

Eşitsizlik 2.1 göz önünde bulundurulursa, eğer sağ taraf sol taraftan büyük ise, para kazanılır. P_{dep} ve P_{des} sırasıyla kalkış ve iniş meydanı yakıt fiyatları olsun. F alınan ekstra yakıtı (*yakıt taşımacılığı kaynaklı*) göstermektedir. F_B ekstra yakılan yakıtı hesaplayan bir fonksiyondur ve bu sayı uçağın F ile birlikteki ağırlığı, loaded onboard, rota, hava durumu vb. gibi birçok değişken kullanılarak hesaplanmaktadır. Bu formülün yakıt taşımacılığı için temel formül olduğu söylenebilir ve bu formülün literatürde başka gösterimlerini bulmak da mümkündür.

Airbus (2004)'a göre başa baş noktası (*breakeven point*) titizlikle belirlenmelidir. Bunun altında yatan temel mantık ve sebep uçağa fazladan yakıt yüklendiğinde yakıt tüketiminin artacak olması iddiasıdır.

K taşımacılık katsayısını belirtmek üzere;

- ΔTOW kalkış ağırlığındaki değişim,
- ΔLW iniş ağırlığındaki değişim,

- ΔT uçuş süresindeki değişim,
- P_d kalkış meydanı yakıt fiyatı,
- P_a varış meydanı yakıt fiyatı,
- C_h saatlik maliyet,

olsun. Buna göre; K eşitlik 2.2’de gösterildiği üzere hesaplanmaktadır:

$$K = \frac{\Delta TOW}{\Delta LW} \quad (2.2)$$

Örneğin; $K = 1.3$ ve kalkış meydanından alınan yakıt 1300 kilogram (kg) olsun. Varış meydanında geriye alınan bu yakıtın 1000 kilogramı kalacaktır. Yani, bir ton yakıtın taşınması 300 kg. daha fazla yakıt yakılması maliyeti oluşturacaktır.

Eşitlik 2.3, 2.4 ve 2.5 ile verilen formüller sırasıyla, kalkış meydanından alınan ekstra yakıtın maliyetini, taşınan yakıtın maliyet tasarrufunu ve uçuş süresindeki olası bir artıştan kaynaklanan maliyeti göstermektedir.

$$\Delta TOW \times P_d = \Delta LW \times K \times P_d \quad (2.3)$$

$$\Delta LW \times P_a \quad (2.4)$$

$$\Delta T \times C_h \quad (2.5)$$

Airbus (2004); maliyet tasarrufu miktarının, yüklenen ekstra yakıt maliyeti ve ekstra süre maliyeti toplamını aştığı durumda ekstra yakıt taşımanın kârlı olduğunu belirtmiş ve bunu eşitsizlik 2.6 ile göstermiştir.

$$(\Delta LW \times P_a) > (\Delta LW \times K \times P_d) + (\Delta T \times C_h) \quad (2.6)$$

Eşitsizlik 2.6 aşağıda verilen Eşitsizlik 2.7 şeklinde de ifade edilebilir:

$$\Delta LW(P_a - K \times P_d) - (\Delta T \times C_h) > 0 \quad (2.7)$$

Böylece, $\Delta T = 0$ olduğunda, kalkış meydanı yakıt fiyatının varış meydanı yakıt fiyatına oranı taşımacılık katsayısı K ’dan büyük ise ekstra yakıt taşımak kârlıdır.

$$\frac{P_d}{P_a} > K \quad (2.8)$$

Airbus (2004), taşınacak optimum yakıt miktarını ilk kalkış ağırlığı (*initial take-off weight/ek yakıt (additional fuel) hariç*), menzil uzunluğu (*stage length*), seyir seviyesi (*cruise flight level*) ve yakıt fiyatı oranının (*fuel price ratio*) bir fonksiyonu olarak belirlemek için uçuş ekibi operasyonel dokümanında (*Flight Crew Operating Manuel - FCOM*) bu grafikleri ifade etmiştir.

Geliştirilen bu formüllerin yanı sıra, yakıttan elde edilecek kazancın hesaplanması için çevrimiçi uygulamalar da mevcuttur. Örneğin; FuelerLinx (FuelerLinx, 2019) uçuş planlanması ve çoklu (*multi-leg*) yakıt taşımacılığı sağlamaktadır. Bu uygulama yakıt taşımacılığının var olup olmadığı durumlardaki tahmini harcanan ve kazanılanı göz önüne alarak yakıt taşımacılığının verimli olup olmayacağını belirlemek için yardımcı olmaktadır. iFuel (Cockpit Apps LLC., 2019) yakıt taşımacılığı uygulaması 3 meydana kadar yakıt fiyatlarını, taksi yakıtını (*taxi fuel*), yedek yakıtı (*reserve fuel*), alınacak yakıt miktarını (*yakıt taşımacılığı durumunda*), yanma oranını (*burn rate*), yer giderlerini (*ramp fee (opsiyonel)*), yer giderinden feragat etmek için alınacak minimum yakıtı (*opsiyonel*) hesaplama faktörleri olarak kullanan ve seyahat yakıtı maliyetini, iniş yakıtını (*yakıt taşımacılığı durumunda*), yakıt taşımacılığı maliyetini veya yakıt taşımacılığı tasarrufunu, net yakıt taşımacılığı tasarrufunu (*yer giderleri sonrası*) çıktı olarak sunan bir hesaplama uygulamasıdır.

Lesinski (2011) işletmenin adını, işletme tarafından uçuş planlaması ve yakıt taşımacılığı için kullanılan yazılımı, yazılımın sahibini, uçuş sayısını, yakıt taşımacılığı sayısını ve uçuş yüzdelerini içeren gerçek hayattan örnekleri vermiştir. Sunduğu örneklerden biri olan Fedex, Jeppesen Uçuş Planlama yazılımını kullanmaktadır ve tahmini olarak yakıt taşımacılığı ile 7,5 milyon dolar tasarruf elde etmiştir. Sunduğu diğer bir örnekte, United Parcel Service (*UPS*), Lufthansa Systems tarafından geliştirilmiş olan Lido/Flight yazılımını kullanmaktadır ve yaklaşık olarak 7,8 milyon dolar tasarruf elde etmiştir. Sunduğu diğer bir örnekte, Continental Hava Yolları uçuş planlama ve yakıt taşımacılığı yazılımı kullanmaktadır ve 334.000 uçuştan 5,5 milyon dolar tasarruf elde etmiştir.

Yılmaz vd. (2021) havacılık endüstrisinde yakıt taşımacılığı durumunun makine öğrenmesi teknikleri ile tahmini için ham veriden öğrenerek yakıt taşımacılığı öngörüsünde bulunan bir çalışma gerçekleştirmiştir. Yakıt taşımacılığı durumunu sınıflandırmak için k-En Yakın Komşu algoritması, C4.5 Karar Ağacı algoritması, Basit Bayes Sınıflandırıcı ve Yapay Sinir Ağları (YSA) algoritmalarından yararlanmışlardır. Algoritmaların model performans değerlerini karşılaştırarak Yapay Sinir Ağları algoritmasını %83,8 doğrulukla en iyi performans değerine sahip algoritma olarak bulmuşlardır.

2.1.3.Yakıt Taşımacılığı ve Yönetmelikler

Yakıt taşımacılığı metodu için bir uçağın mevcut yakıt tankı kapasitesi bir sınır olmakla birlikte bunun yanında ülkelerin kanunlar ile belirledikleri yasal sınırlamalar da mevcuttur. Örneğin; T.C. Gümrük ve Ticaret Bakanlığı (2009) tarafından belirlenen ve 07.10.2009 tarihinde resmî gazetede (*Resmî Gazete Sayısı: 27369 (Mükerrer)*) yayınlanan Gümrük Yönetmeliği Birinci Kitap İkinci Kısım (*Değişik kısım başlığı: RG-12/6/2012-28321*) Taşıtların Kontrolü, Gümrüklü Sahaların Gözetimi ve İşletmelerin Yükümlülükleri, (*Ek bölüm: RG-12/6/2012-28321*) Birim Bölümü Taşıtların Kontrolü - Madde 72/Ö (*Gelen hava taşıtlarının kontrolü*) ve Madde 72/P (*Giden hava taşıtlarının kontrolü*) uygulanmaktadır. Bu yönetmeliğe göre; “Yurtdışından gelen hava yolu taşıtları için işleticisi veya temsilcisi tarafından EK-10/E’de yer alan Hava yolu Beyan Formu (*Ek ibare: RG-22/3/2018-30368*) veri işleme tekniği ile doldurularak diğer belgelerle birlikte (*Değişik ibare: RG-22/3/2018-30368*) gümrük idaresine elektronik ortamda beyan edilir. Beyan Formu ile ekindeki yükleme veya taşıma belgeleri, mürettebat eşya listesi ile kumanya ve yakıt listeleri gibi belgeler arasında çelişki olması veya ihbar ya da şüphe halinde hava taşıtı ayrıca kontrol edilebilir.”. Yine aynı yönetmeliğin birinci maddesinde yönetmeliğin amaç ve kapsamı belirtilmiştir. Buna göre; “(1) Bu Yönetmeliğin amacı; 4458 sayılı Gümrük Kanunu’nun yönetmelikle düzenlenmesini öngördüğü konuları düzenlemek ve uygulamaya yön verecek hususlara açıklık getirmektir. (2) Bu Yönetmelik, gümrük işlemlerine ilişkin usul ve esasları kapsar.”

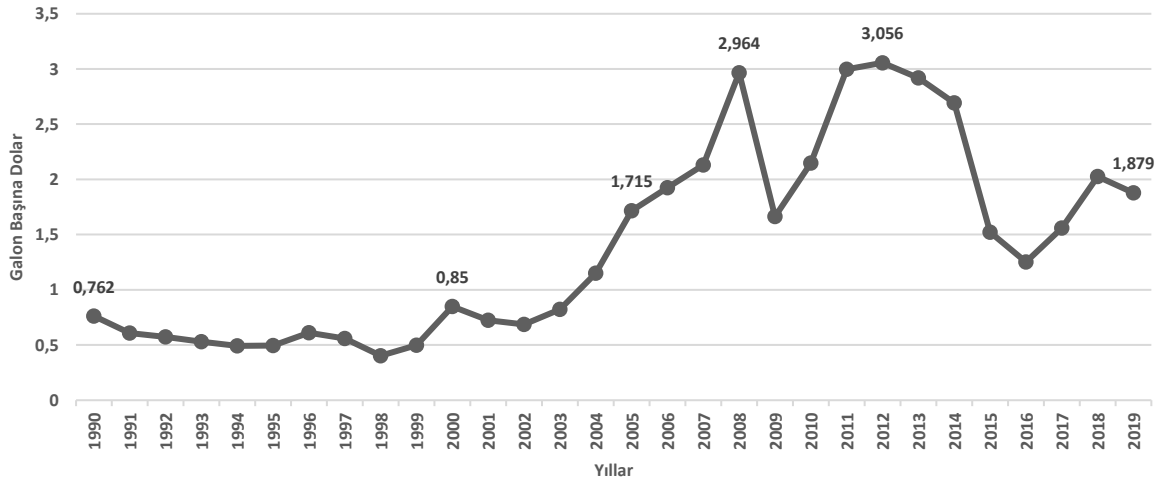
Türkiye’de hava taşımacılığı yapan milli bir şirket, yurt dışından Türkiye’ye gerçekleştireceği herhangi bir uçuşta ineceği meydana belli miktarda yakıt ile inebilir (yaklaşık 4500 kg.). Eğer bu miktarın üzerindeki yakıtı yurt dışından alarak Türkiye’ye iniş gerçekleşirse, bu yakıtın millileştirilmesi gerekmektedir. Aksi takdirde bu yakıtın iç hatta kullanılması yasaktır ve T.C. Gümrük ve Ticaret bakanlığı tarafından cezaî yaptırım uygulanır. Buradaki amaç Türkiye’ye

göre yakıtın daha ucuz olduğu herhangi bir ülkeden yakıtı alarak vergi ödemeksizin bu yakıtın iç hatta kullanılmasını önlemektir. Bu durumun ihlali yakıt kaçakçılığı olarak ifade edilebilir.

2.1.4.Yakıt Taşımacılığının Hava Yolu Maliyetleri Üzerindeki Etkisi

Orville ve Wilbur Wright ilk uçuşunu gerçekleştirdiğinden beri; hava taşımacılığı endüstrisi küresel ekonominin önemli bir sektörü haline gelmiştir (Barnhart vd., 2003). Buna bağlı olarak uçak sayısı her geçen gün artmaya devam etmektedir. Airbus, 2018 yılında dünya çapında 93 müşteriye toplam 800 ticari uçak teslim ettiğini belirtmiştir (Airbus, 2019b). Boeing ise endüstrinin 2037 yılına kadar 42.000'den fazla uçağa ihtiyaç duyacağını belirtmiştir (Boeing, 2019b). Havacılık endüstrisi, global ekonominin bir parçası olarak personel, sigorta, yiyecek, havalimanı iniş harçları gibi birçok işletme giderine sahiptir. Ancak yakıt giderleri bu gider listesinin en başında yer almaktadır. Jet yakıtı, bir hava yolunun giderlerinin üçte birini oluşturmaktadır (Chandran ve Anandarajan, 2020).

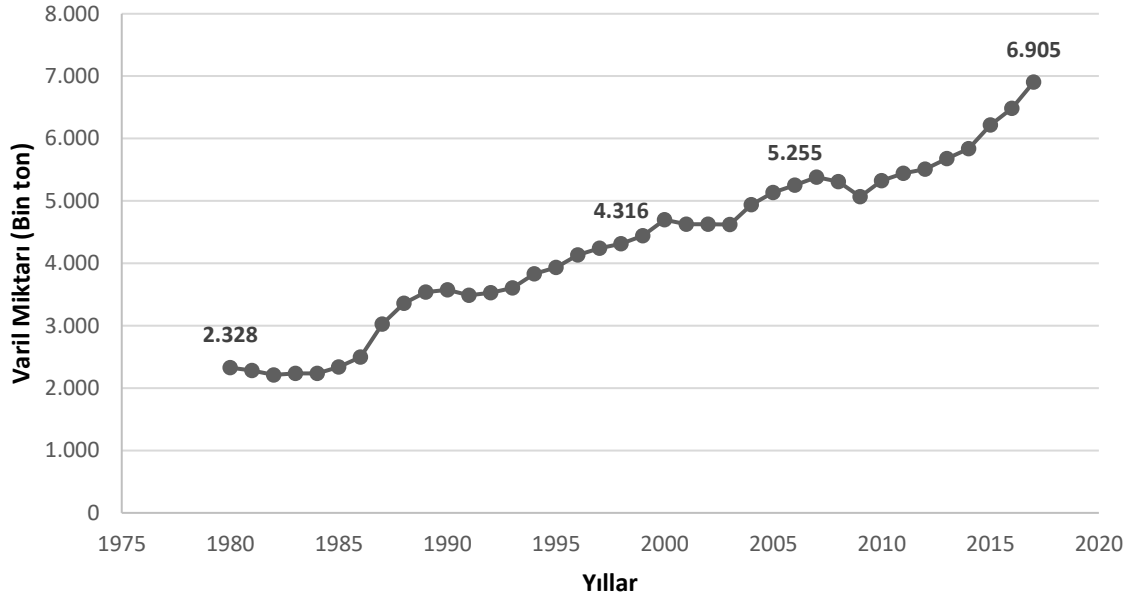
Hava yolu endüstrisinin büyümesiyle yakıt tüketimi daha önemli bir hale gelmektedir ve hava taşımacılığı endüstrisindeki şirketleri direkt olarak etkilemektedir (Singh ve Sharma, 2015). Naumann ve Suhl (2013) jet yakıtı maliyetinin, hava yolu şirketlerinin günden güne artan maliyetlerinin bir parçası olduğunu açıkladı. Amerika Enerji Bilgi İdaresi (EIA) (U.S. Energy Information Administration, 2019a)'nın açıkladığı rakamlara göre Şekil 2.5'ten de görüldüğü üzere U.S.'nin 1990 ve 2019 yılları arasındaki peşin jet yakıtı fiyatları bunu desteklemektedir.



Şekil 2.5: U.S. körfez kıyısı gazyağı tipi jet yakıtı peşin fiyatı (*Galon başına dolar olarak*).

Öte yandan, IATA (2019) jet yakıt fiyatı gelişmelerinin, 2019 endüstrisinin toplam yakıt faturası üzerindeki etkisi ile ilgili olarak (*yer hizmetleri giderleri hariç*) 2019 yılı için yakıt fiyatı ücret averajını 80,4\$/bbl ve 2019 yakıt faturası üzerindeki etkisini -8,1 milyar dolar olarak açıklamıştır.

Günümüzde uçak sayısı ile doğru orantılı olarak uçuş sayıları da giderek artmaktadır. FAA (2019) U.S.'de bir günde 44.000'den fazla ve bir yılda 16.100,000'den fazla uçuşun gerçekleştiğini belirtmiştir. Modern uçaklar her 100 yolcu-km için 3.5 lt. veya her galon için 67 yolcu-mil yakıt harcamaktadır (Air Transport Action Group, 2011). Rao (1999) önemli hava yolları tarafından 2.5 milyar galon yakıtın bir yıl içerisinde tüketildiğini iddia etmiştir. Öte yandan yakıt fiyatları sürekli olarak değişkenlik göstermektedir. Amerika Enerji Bilgi İdaresi'ne göre dünyada 2017 yılında 6905 bin ton galon jet yakıtı tüketilmiştir. Şekil 2.6 dünyada 1980-2017 yılları arasındaki jet yakıtı tüketimini göstermektedir (U.S. Energy Information Administration, 2019b).



Şekil 2.6: Dünyada 1980 ve 2017 yılları arasındaki jet yakıtı tüketimi (U.S. Energy Information Administration, 2019b).

IATA (2018) 2019 yılında hava yollarının yakıt harcaması tutarının 200 milyar \$(*işletme giderlerinin ortalama %24,2'si*)'a ulaşacağını öngörmektedir. Ryerson vd. (2011), çevresel kaygılar ve yakıt fiyatlarındaki değişikliklerin yanı sıra, havacılık topluluğu için yakıt tüketimini azaltmanın önemli bir hedef olduğunu vurgulamıştır. Yakıt tüketimini azaltmak için uçakların yakıt verimliliği, toplam trafik yükünün (*payload*) azaltılması, uçuş rotalarının optimizasyonu, düzenli uçak bakımı ve yakıt taşımacılığı gibi değişik çözümler bulunmaktadır (Lindgren ve Brynhagen, 2012). Singh vd. (2012) uçuş operasyonlarında iyi planlama, doğru uçak yükleme, uygun bakım, uçuş prosedürleri ve yakıt taşımacılığı vb. uçağın yakıt tüketimini etkilediğini vurgulamaktadır.

Yakıt taşımacılığı yakıt kaynaklı işletme giderlerini azaltan ve sonraki uçuş sektörü için ekstra yakıt taşınması ile kâr getiren bir uygulamadır. Yakıt taşımacılığı bir görevi gerçekleştirmek için gerekli yakıttan fazlasını almayı tanımlayan bir terimdir (Filippone, 2015). Bu maliyet tasarrufu tekniğinin arkasındaki asıl düşünce kalkış ve varış meydanları arasındaki yakıt fiyatları farklılığıdır.

2.1.5.Yakıt Taşımacılığının Ekonomik Yararları ve Çevresel Etkileri

Guerreiro Fregnani vd. (2013) fazla yakıt taşımanın uçağı ağırlaştırdığını ve bu nedenle yakıt alınan sektördeki yakıt tüketimi arttırdığını ifade etmişlerdir. Ayrıca ekstra yakıtın neden olduğu yakıt maliyetini hesaplamanın önemli olduğunu ve genellikle hava yollarının her bir uçak için yakıt taşımanın ekonomik uygulugunu sektör-sektör analiz ettiğini belirtmişlerdir.

Uçağın taşıyacağı her bir ekstra ağırlık ekstra yakıt tüketimine yol açmaktadır. Bu durum atmosfere salınan (CO_2) miktarını da artırmaktadır. Yakıt taşımacılığı kullanımı da uçağı ağırlaştırdığından fazla yakıt tüketilmesine ve dolayısıyla atmosfere daha fazla (CO_2) salınmasına yol açmaktadır. Örneğin; Frost (2019) British Airways'in İtalya'ya olan en az bir uçuşunda 40 sterlin tasarruf etmek için uçağın yaklaşık 3,3 ton ek yakıt taşıdığını ancak bunun 600kg. (1320lb) (CO_2) yanmasına neden olduğunu belirtmiştir.

ICAO (2020), 2015 yılında uluslararası havacılığın yaklaşık olarak 160 megaton (Mt) yakıt tükettiğini belirtmiştir. Öte yandan teknoloji ve ATM (*Air Traffic Management*) senaryolarına bağlı olarak 2045'e kadar uluslararası hava trafiğinde beklenen 3,3 kez büyümenin (*ton-kilometre geliri/Revenue tonne-kilometre*), yakıt tüketiminde 2,2 ile 3,3 arasında artışa yol açacağını tahmin edildiğini ifade etmiştir.

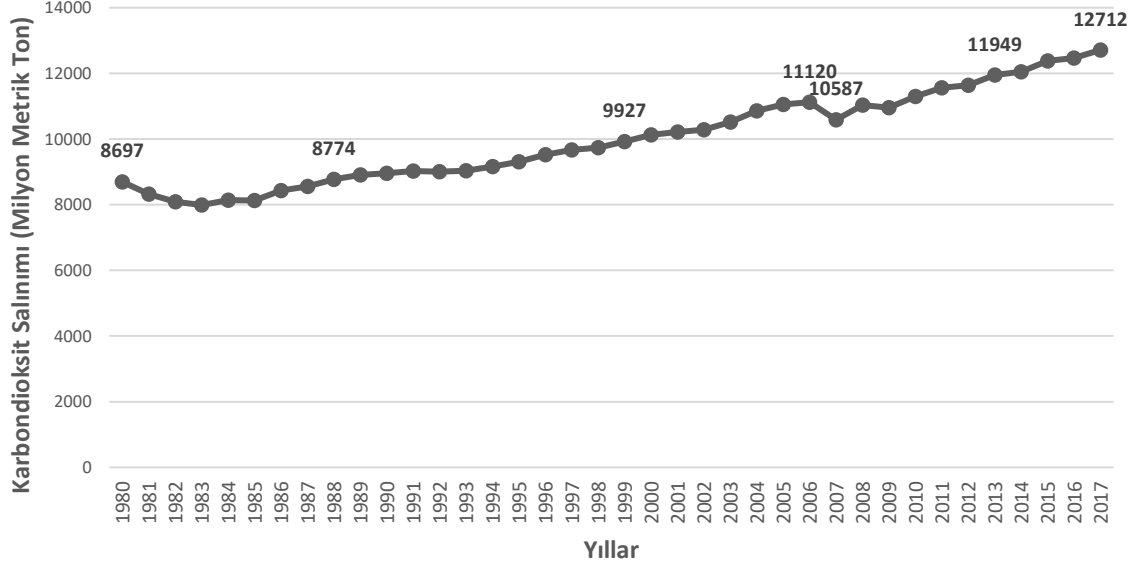
Eurocontrol (2019) gerçekleştirmiş olduğu bir simülasyon ile yakıt taşımacılığının ECAC (*European Civil Aviation Conference*) bölgesinde 286.000 ton ekstra yakıt yanmasına ve bunun ek 157 milyon Euro maliyet ve 901.000 ton ekstra karbondioksit (CO_2) salınımına (*18 milyon Euro ticari maliyet*) yol açacağını, fakat yılda net 265 milyon Euro tasarruf sağlayabileceğini göstermiştir.

Avrupa içerisinde yer alan havalimanlarına ve bu meydanlardan diğer yerlere yapılan tüm uçuşlar için 2012 yılından itibaren Avrupa Birliği Emisyon Ticaret Sistemi (EU ETS) yürürlüğe girmiştir (European Commission, 2016a). Bu sistem AB'nin iklim değişikliği ile mücadele politikasının temel taşıdır ve sera gazı emisyonları maliyetini etkin bir şekilde azaltmaya yönelik temel bir araçtır (European Commission, 2016b).

Kaya (2007) küresel ısınmayı atmosfere verilen gazların sera etkisi yaratması sonucunda, dünya yüzeyinde sıcaklığın artması olarak tanımlayarak sera gazlarının karbon dioksit (CO_2) gazı,

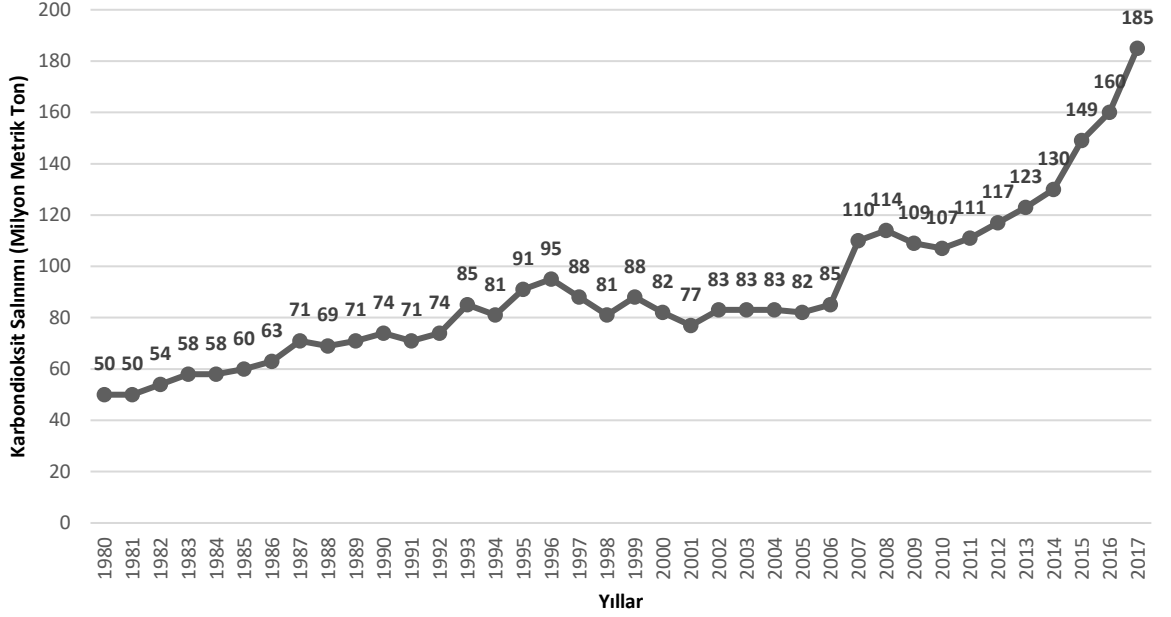
metan (CH_4) ve karbon monoksit (CO) gazları, azot oksit (N_2O) gazı, halokarbon gazları olduğunu belirtmiştir.

Dünya’da ve Türkiye’de petrol ve diğer likitler kaynaklı karbon dioksit salınım miktarları Şekil 2.7 ve Şekil 2.8 ile gösterilmiştir.



Şekil 2.7: Dünya’da 1980 ve 2017 yılları arasında yıllık petrol ve diğer likitler kaynaklı karbondiyoksit salınım miktarları (U.S. Energy Information Administration, 2020a).

Şekil 2.7’den de görüldüğü üzere Dünya’da 1980 yılında petrol ve diğer likitlerden kaynaklı yıllık karbondiyoksit salınımı toplamı 8697 milyon metrik ton iken, 2017 yılında bu rakam 12.712 milyon metrik tona ulaşmıştır.



Şekil 2.8: Türkiye’de 1980 ve 2017 yılları arasında yıllık petrol ve diğer likitler kaynaklı karbondioksit salınım miktarları (U.S. Energy Information Administration, 2020b).

Öte yandan bu durum Türkiye ölçeğinde incelendiğinde ise Şekil 2.8’den de görüldüğü üzere 1980 yılında petrol ve diğer likitlerden kaynaklı yıllık karbondioksit salınımı toplamı 50 milyon metrik ton iken, 2017 yılında bu rakam 185 milyon metrik tona ulaşmıştır.

2.2.MAKİNE ÖĞRENMESİ

Sürücüsüz araçlarla, fabrikalarda insansız üretim yapan makinelerle, hastanede ameliyat yapabilen robotlarla ve daha birçok alt çalışma alanı ve uygulamalarıyla Yapay Zekâ - kısaca YZ - (*Artificial Intelligence - AI*) her geçen gün biraz daha hayatımızın vazgeçilmez bir parçası haline gelmektedir. Yapay zekâ, bilim ve mühendislikte yeni alanlardan biridir (Russell ve Norvig, 2010); ancak tarihi çok eskiye dayanmaktadır. Eski Yunan mitolojisinde rüzgâr tanrısı zannedilen Daedalus’un “yapay-insan” teşebbüsü buna bir örnek olarak gösterilebilir (Pirim, 2006). Yapay zekânın kısa bir kronolojik tarihçesi aşağıda sunulmuştur:

1789 yılında; Joseph Friedrich Freiherr zu Racknitz tarafından bir insan ile satranç oynayan bir robotun mücadelesi, gravür sanatı ile oyma resim olarak resmedilmiştir. Esere “*The Turk*” adı verilmiştir (Buchanan, 2005). 1942 yılında; Amerikan bilim kurgu yazarı Isaac Asimov “*Runaround*” isimli kısa hikayeyi kaleme almıştır. Hikayenin konusu, Gregory Powell ve Mike Donovan adlı iki mühendis tarafından geliştirilmiş olan robottur. Asimov’un çalışması robotik,

yapay zekâ ve bilgisayar bilimi alanındaki bilim insanlarına nesiller boyu ilham vermiştir (Haenlein ve Kaplan, 2019). 1950 yılında; Alan Turing, akıllı makinelerin nasıl oluşturulacağını ve özellikle de zekâlarının nasıl test edileceğini anlattığı “Computing Machinery and Intelligence” adlı makale çalışmasını yayınlamıştır (Haenlein ve Kaplan, 2019). Alan Turing’in bir bilgisayar programının zeki olup/olmadığına karar vermek için önerdiği bu tanımlama “Turing Testi” olarak da ifade edilmektedir (Benko ve Lányi, 2009). 1956 yılında; “Yapay Zekâ” kelimesi resmi olarak, Marvin Minsky ve John McCarthy (*Standford Üniversitesi’nde bilgisayar bilimci*) tarafından yönetilen ve Amerika Birleşik Devletleri’nin New Hampshire eyaletindeki Dartmouth kolejinde düzenlenen, yaklaşık sekiz hafta süren Dartmouth Yapay Zekâ Üzerine Yaz Araştırma Projesi (*Dartmouth Summer Research Project on Artificial Intelligence - DSRPAI*) atölye çalışmasında icat edilmiştir. DSRPAI, insan zekâsını taklit edebilen makineler inşa etmeyi amaçlayan yeni bir araştırma alanı oluşturmak için çeşitli alanlardan araştırmacıları yeniden bir araya getirmeyi amaçlamıştır (Haenlein ve Kaplan, 2019). 1964-1966 yılları arasında; Massachusetts Teknoloji Enstitüsü (*MIT - Massachusetts Institute of Technology*)’nden Joseph Weizenbaum, ELIZA adı verilen bir doğal dil işleme aracı olan bilgisayar programı geliştirmiştir. ELIZA insan ile olan bir konuşmayı simüle etme yeteneğine sahiptir. Turing Testi’ni geçmeye çalışan ilk bilgisayar programlarından biridir ve bir uzman sistem (*expert system*)’dir (Haenlein ve Kaplan, 2019). 1997 yılında; IBM tarafından geliştirilen ve bir uzman sistem olan Deep Blue satranç oynama programı dünya şampiyonu Gary Kasparov yenmeyi başarmıştır (Haenlein ve Kaplan, 2019). 2005 yılında; Stanford yarış takımının otonom robotik arabası “Stanley” Savunma İleri Araştırma Projeleri Ajansı (*Defense Advanced Research Projects Agency - DARPA*)’nın büyük mücadelesini kazandı. Stanley, Amerika Birleşik Devletleri’nin Las Vegas eyaletinin güneybatısındaki off-road çöl parkurunu yedi saatten biraz daha kısa bir sürede geçmiştir (Buchanan, 2005). 2015 yılında; Google tarafından geliştirilen AlphaGo, satrançtan daha karmaşık olan Go oyununda dünya şampiyonunu yenmeyi başarmıştır. AlphaGo, en yüksek performansına yapay sinir ağlarının özel bir türü olan derin öğrenme ile ulaşmıştır (Haenlein ve Kaplan, 2019). 2021 yılında; Amerika Birleşik Devletleri’nin uzay programından sorumlu olan Ulusal Havacılık ve Uzay Dairesi (*National Aeronautics and Space Administration – NASA*)’nin “Perseverance” adını verdiği yapay zekâ sistemlerinin kullanıldığı robot uzayda toplamda 3,9 milyon kilometre (*km.*) yol kat ederek 18 Şubat 2021 tarihinde kızıl gezegen Mars’a başarılı bir şekilde iniş gerçekleştirmiştir. Perseverance, geçmişte Mars’ta hiç yaşam olup olmadığı sorusuna cevap

aramak için Mars'ın Jezero kraterinde araştırmalarını yürütecektir (NASA, 2021). Benzer şekilde bir çok alanda yapay zekâ çalışmaları yürütülmektedir. ProMare isimli kâr amacı gütmeyen bir kurum tarafından yapay zekâyâ sahip “Mayflower 400” adını verdikleri mürettebatsız bir otonom gemiyi (*The Mayflower Autonomous Ship - MAS*) geliştirilmiştir. 1620 yılında Atlantik geçişini tamamlayan Mayflower isimli geminin gerçekleştirdiği tarihi yolculuktan esinlenerek üretilen bu gemi ile bu yolculuğun bu kez mürettebatsız olarak yeniden canlandırılması amaçlamaktadır (ProMare, 2021).

Li ve Du (2017) yapay zekâyı hafıza, duygu, yargılama, muhakeme, ispat, tanıma, anlama, iletişim, tasarım, düşünme, öğrenme, unutma, yaratma vb. yapay olarak makine, sistem veya ağ tarafından gerçekleştirilebilir insan davranışlarının bir çeşidi olarak tanımlamıştır. Huysal ve Erkoç (2013) yapay zekâyı insan zekâsına benzeyen ve benzer işlemleri gerçekleştirebilen bilgisayar sistemlerinin geliştirilebilmesine çalışan çok yönlü bir disiplin olarak tanımlamıştır. Yani bir başka ifadeyle yapay zekâ bilgisayar veya bilgisayar kontrolündeki makinelerin çeşitli faaliyetleri “insanlarla benzer şekilde” yerine getirme yeteneğidir. Nilsson (1980) bilgisayar programları yazmanın, matematik yapmanın, ortak akıl yürütmeye katılmanın, dili anlamının ve hatta bir otomobil kullanmanın vb. gibi birçok insani zihinsel aktivitenin “zekâ” talep ettiğini söylemiştir. “Zekâ” kelime olarak Arapça kökenli olup; insanın düşünme, akıl yürütme, objektif gerçekleri algılama, yargılama ve sonuç çıkarma yeteneklerinin tamamı olarak ifade edilmiştir (Türk Dil Kurumu, 2019).

Makine öğrenmesi üzerine belki de her şey ünlü bir matematikçi ve şifre bilimci olan Alan Turing'in “Makineler düşünebilir mi?” sorusuyla başlamıştır (Avaner, 2018). 1950 yılında Mind dergisinde yayınlanan “Computer Machinery and Intelligence” adlı makalesinde Alan Turing, bugün Turing Testi olarak bilinen ve bir zeki sistemin başarıya ulaşp ulaşmadığını değerlendiren testi açıklamıştır (Huysal ve Erkoç, 2013). Buna göre bu test aşağıdaki düzeneğe sahiptir:

- Sorgulayıcının (*bir insan*) görmeyeceği bir şekilde bilgisayar ve insan aynı ortama saklanır,
- Sorgulayıcı, yalnızca soru sorarak hangisinin bilgisayar olduğunu saptamaya çalışır,
- Sorgulayıcının soracağı sorular ve aldığı yanıtlar, bir klavye sisteminde yazılarak veya ekran gösterilerek yani sessiz bir şekilde verilir,

- Sorgulayıcıya her iki taraf hakkında soru/cevap oturumundan elde ettiği bilgiler dışında başka hiçbir bilgi verilmez,
- İnsan denek kendisinin insan diğer deneğin bilgisayar olduğuna sorgulayıcıyı ikna etmek için soruları içtenlikle yanıtlar verir. Öte yandan bilgisayar ise “yalan” söylemeye programlanmış olduğu için kendisinin insan olduğuna sorgulayıcıyı inandırmaya çalışır.

John Searle, makinelerin Turing testini geçse dahi düşünüp anlama yeteneğine sahip olamayacağını savunmuş ve Turing testine alternatif Çin Odası Testini önermiştir (Huysal ve Erkoç, 2013). Turing testini geçen bir yapay zekânın varlığı günümüzde halen bir tartışma konusudur. Günümüzde yapay zekânın ulaştığı boyut ise bu soruyu cevaplamayı her geçen gün daha da zorlaştırmaktadır. Çünkü yeni teknolojiler, artık insanla aynı dilde konuşabilmekte, sorunlarına çözüm üretebilmekte, günlük rutin işlerde insanın yerini alabilmektedir. Bu yönden bakıldığında yapay zekâların geleceği insanlardan ayırt edilemeyecekleri bir dönemi işaret etmektedir.

Dünyada ve hayatımızda ses getiren yapay zekâ örneklerini incelediğimizde tarihte ilk defa vatandaşlık alan ve nüfus cüzdanına sahip ilk robot olan dünyanın en insansı robotu Sophia ile karşılaşmaktayız (BBC, 2017). Yine bir başka etkileyici yapay zekâ örneği ise IBM Watson’dır (IBM, 2019). Watson doktorların bile öngöremeyeceği kadar veriyi işleyerek teşhis ve tanıda tavsiyelerde bulunabilmektedir. Hatta daha önce gördüğü hastalardan yola çıkarak tedavinin de en uygun nasıl yapılabileceği konusunda önerilerde bulunmaktadır. Kısa bir süre sonra hastanelerde hastalara bakacak, tanıyı koyacak ve tedavi edecek doktorlardan biri de Watson olabilir.

Yapay zekâ hâlihazırda yaşamın her alanında kullanılmaktadır. Günlük hayatta en çok kullandığımız yapay zekâ örnekleri arasında ise Siri, Google Asistan, Google Translate, Microsoft Cortana, akıllı buzdolapları vb. gibi birçok uygulama bulunmaktadır. Tüm bu yapay zekâ uygulamaları bazen basit gibi gözükse de hayatımıza birçok fayda sunmaktadır. Örneğin; modern küresel konumlama sistemleri (*Global Positioning Systems – GPS*) sayesinde gidilecek bir yere en kısa zamanda ulaşılabilir. Firmaların yapay zekâ tabanlı müşteri hizmetleri sayesinde birçok soruna pratik bir şekilde çözüm getirilebilmektedir. Üretimde yapay zekâyâ sahip makineler kullanılarak zaman ve güç gerektiren işlerde yüksek verimlilik sağlanabilmektedir. Tüm bunların ötesinde uzaya yapay zekâyâ sahip robotlar

gönderilmektedir. Bu açıdan bakıldığında; gelecek, insanlığın hala cevabını bilmediği soruların yanıtını, geliştirilmeye devam eden yapay zekâlardan alacağı bir döneme işaret etmektedir.

Yapay zekâ ve yapay zekânın alt çalışma alanları bankacılık, sağlık, güvenlik, spordan müşteri ilişkileri yönetimine kadar birçok sektörde kullanılmaktadır. Emre vd. (2019) çocukluk döneminde geçirilmiş olan akut romatizmal ateş (*Acute Rheumatic Fever - ARF*)’in kalp hastalıkları temelinde etkilerini Naive Bayes Sınıflandırıcı (*Naive Bayes Classifier*), Sınıflandırma ve Regresyon Ağaçları (*Classification and Regression Trees – CART*), C4.5, C5.0 ve C5.0 Yükseltilmiş Karar Ağacı Algoritmaları (*Boosted Decision Tree Algorithms*) ve Rastgele Orman Algoritması (*Random Forest Algorithm*) gibi veri madenciliği yöntemlerini kullanarak analiz etmiştir. Kartal ve Balaban (2018) kalp ameliyatı sırasında veya sonrasında hastaların ölüm riskini makine öğrenmesi tekniklerini (*k-En Yakın Komşu algoritması (k-Nearest Neighbor algorithm)*, *Naive Bayes Sınıflandırıcı (Naive Bayes Classifier)*, *ID3 ve C4.5 Karar Ağacı algoritmaları (ID3 and C4.5 Decision Tree Algorithms)* ve *Logistik Regresyon (Logistic Regression)*) kullanarak tahmin etmiştir. Smeureanu vd. (2013) özel bankacılık sektöründe müşteri segmentasyonu için Yapay Sinir Ağları (*Artificial Neural Networks*) ve Destek Vektör Makinelerini (*Support Vector Machines*) kullanmıştır. Özen (2016) davranışsal biyometrik kimlik doğrulama yöntemlerinden tuş vuruş dinamiklerini Yapay Sinir Ağı algoritmaları (*Çok katmanlı ileri beslemeli Yapay Sinir Ağı Geri Yayılım algoritması, Radyal Tabanlı Fonksiyon Yapay Sinir Ağı, Öğrenmeli Vektör Nicemleme Yapay Sinir Ağı (LVQ1, LVQ2, LVQ3 ve Optimize LVQ1 algoritmaları)*, *Kendini Örgütleyen Haritalar Yapay Sinir Ağı (SOM (Kendini Örgütleyen Haritalar - (Self-Organizing Maps))*, *BDK (İki-Yönlü Kohonen Ağı - (Bi-directional Kohonen Network))* ve *XYF (XY Kaynaşık Kohonen Ağı - (XY Fused Network)) algoritmaları*), *Olasılıksal Yapay Sinir Ağı, Uyarlanabilir Rezonans Teorisi Yapay Sinir Ağı ailesinden ART2 (Adaptive Resonance Theory) algoritması, Öz-İlişkili Yapay Sinir Ağı algoritmaları*) ile kullanılarak kimlik doğrulamada güvenliği artırmaya yönelik bir çalışma gerçekleştirmiştir. Koçoğlu (2017) “Aşırı Öğrenme Makineleri (*Extreme Learning Machines*)”nin bir sınıflandırma problemi olarak müşteri kayıp analizi probleminin çözümündeki etkinliğinin ve belirli koşullar altında en iyi çözümü üreten parametrelerini araştırarak Aşırı Öğrenme Makineleri ile elde edilen performans değerlendirme ölçü değerlerini, k-En Yakın Komşu algoritması, Naive Bayes Algoritması ve Destek Vektör Makinelerinin kullanımı ile elde edilen performans değerlendirme ölçü değerleri ile

karşılaştırılmıştır. Benzer uygulamaların gelecek yıllarda daha da yaygınlaşacağı öngörülmektedir.

Havacılık sektöründe de gelecekte otonom uçakların kullanılması öngörülmektedir. Her yıl 3,8 milyar insanın uçakla seyahat ettiği düşünülürse bu gelişme kayda değer değişimlere yol açacaktır. Bilgisayar sistemleri ve yapay zekâdaki ilerlemeler bu otonom uçakların kalkış ve inişine destek vermekte kullanılacaktır. Bu konu ile ilgili olarak, Boeing; Avustralya Kraliyet Hava Kuvvetleri (*Royal Australian Air Force - RAAF*) filosunda kullanılacak ve mürettebatlı uçaklarla birlikte çalışmak üzere tasarlanmış pilotsuz savaş uçağı benzeri bir jet geliştirmiş ve bu jet ilk test uçuşunu başarıyla tamamlanmıştır (Wigman, 2021).

Bir araba modeli hakkında araştırma yaparken girilen web sitesinde beğenebilecek diğer araba modellerinin önerilmesi, şarkılardan filmlere, kitaplardan haberlere kadar birçok konuda uygulamalar insanlara çeşitli öneriler sunmaktadır. Akıllı telefonlar; bir insanın günde kaç saat uyuduğundan, kaç adım attığına kadar çeşitli sağlık verilerini tutarak insanları çeşitli konularda bilgilendirmektedir. Bahsedilen bu gelişmeler Makine Öğrenmesi (*Machine Learning*) sayesinde. Makine öğrenmesi yapay zekânın bir alt dalıdır ve hesaplama sistemlerine yapay zekâ konseptini uygulamak için basit ve karmaşık birbirinden farklı pek çok algoritma kullanır. Yani makine öğrenmesi tıpkı insanlar gibi deneyimlerinden öğrenerek tahminlerde bulunur; ancak makine öğrenmesinin deneyimlerini veri oluşturmaktadır. İşte bu veriden çıkarımlar yapan ve tahminlerde bulunan yöntemler bütününe makine öğrenmesi denir. Makine öğrenmesini kullanan sistemler verilen görevleri veriden otomatik olarak öğrenebilir (Domingos, 2012).

Çoğu zaman farkına varılmasa da günlük hayatta makine öğrenmesi örnekleri sıkça kullanılmaktadır. Örneğin, bir e-posta geldiğinde ilgili e-postanın içeriği taranarak zararlı olup olmadığını anlayan makine öğrenmesi uygulamaları sayesinde istenmeyen (*spam*) e-postalar gelen kutunuza düşmemektedir. İstenmeyen e-postalar dışında sosyal medya uygulamalarından biri olan Facebook'un tanıyan olabileceğiniz kişileri, Youtube'un beğenebileceğiniz videoları, Amazon'un öne çıkan öneriler bölümleri de günlük yaşantımızda sürekli karşılaştığımız birer makine öğrenmesi örneğidir. Örnekler öneri sunan birçok sistemle çoğaltılabilir. Hareketli nesneleri tanıma, parmak izi, yüz tanıma gibi görüntü işleme uygulamaları da makine öğrenmesinin alanıdır. Alpaydın (2020) makine öğrenmesi uygulamalarının örneklerini; birliktelik kuralları (*association rules*), sınıflandırma (*classification*), regresyon (*regression*),

danışmansız öğrenme (*unsupervised learning*), pekiştirmeli öğrenme (*reinforcement learning*) olarak ifade etmiştir.

“Eğer bir bilgisayar programının; G görevlerinde P ile ölçülen performansı, deneyim D ile artıyorsa, o bilgisayar programının bazı G görevlerinin sınıflarına ve performans ölçüsü P 'ye göre deneyim D 'den öğrendiği söylenmektedir” (Balaban ve Kartal, 2015; Mitchell, 1997). Bu tezde; yakıt taşımacılığı durumunun tahmin edildiği durumda,

- G : Yakıt taşımacılığı durumunun tahmin edilmesi,
- D : Yakıt taşımacılığı durumunun tahmin edilmesi için oluşturulan veri seti,
- P : Doğruluk, hata oranı, duyarlılık, belirleyicilik, kesinlik (*pozitif öngörü değeri*), negatif öngörü değeri, F -ölçüsü model performans değerlendirme ölçütleri,

yakıt taşımacılığı miktarının tahmin edildiği durumda,

- G : Yakıt miktarının tahmin edilmesi,
- D : Yakıt taşımacılığı miktarının tahmin edilmesi için oluşturulan veri seti,
- P : Ortalama mutlak hata, ortalama kare hata ve kök ortalama kare hata model performans değerlendirme ölçütleridir.

Mitchell (1997) ayrıca makine öğrenmesinin; yapay zekâ, olasılık ve istatistik, hesaplamalı karmaşıklık, bilgi teorisi, psikoloji ve nörobiyoloji, kontrol teorisi ve felsefeyi içeren çeşitli disiplinlerden fikirler aldığını belirtmiştir. Makine öğrenmesi; üretim ve imalat, perakende, seyahat, finans ve mali hizmetler, enerji, sağlık gibi birçok sektörde kullanılmaktadır. Buna göre;

- Finasta; bankalar kredi hesaplamalarında, sahtekârlık tespitinde ve borsada kullanmak için modeller inşa etmede kendi geçmiş verilerini analiz etmektedirler,
- İmalatta; öğrenme modelleri optimizasyon, kontrol ve sorun gidermek için kullanılmaktadır,
- Sağlıkta; tıbbi teşhis için öğrenen programlar kullanılmaktadır,
- Telekomünikasyonda; arama örüntüleri şebeke optimizasyonu ve maksimum servis kalitesi için analiz edilmektedir,
- Bilimde; fizik, astronomi ve biyoloji deki fazla sayıdaki veri bilgisayarlar ile yeterince hızlı analiz edilmektedir (Alpaydın, 2010).

2.2.1.Öğrenme Stratejileri

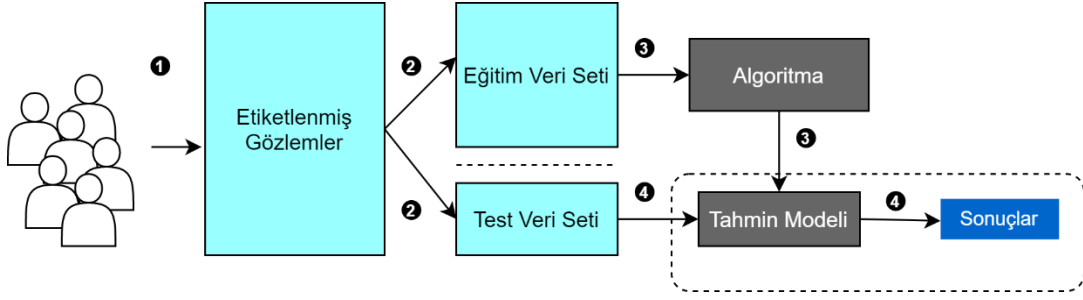
Lison (2012), makine öğrenmesinde kullanılan öğrenme yönteminin eldeki verinin türüne bağlı olduğunu ve temelde üç çeşit öğrenme türünün olduğunu belirtmiştir. Bunlar aşağıdaki gibi sıralanabilir:

- Danışmanlı öğrenme (*Supervised learning*),
- Danışmansız öğrenme (*Unsupervised learning*),
- Pekiştirmeli öğrenme (*Reinforcement learning*).

2.2.1.1.Danışmanlı Öğrenme

Danışmanlı öğrenme; gözetimli, gözetmenli ya da denetimli öğrenme olarak da bilinmektedir. Danışmanlı öğrenmede, görülmemiş durumlardaki hedef nitelik (*target feature*) değerinin tahmin edilmesi (*prediction*) hedeflenmektedir ve öğrenmiş model “tahmin edici (*predictor*)” olarak da adlandırılır. Tahmin edici, etiket (*label / tahmin niteliklerinin değerlerine göre tahmin edilen hedef nitelik değerleri*) bilgisi bilinmeyen bir örneğin (*instance*) etiketini tahmin etmektedir. Eğer etiket kategorik ise görev “sınıflandırma (*classification*)” olarak ve öğrenen “sınıflandırıcı (*classifier*)” olarak adlandırılır. Eğer etiket nümerik ise görev “regresyon (*regression*)” ve öğrenen “uydurulmuş regresyon modeli (*fitted regression model*)” olarak adlandırılmaktadır. Her iki durumda da öğrenme işlemi etiket bilgilerini içeren veri setleri üzerinden yürütülmektedir. Burada, etiketi bilinen durum “örnek (*example*)” olarak da adlandırılmaktadır (Zhou, 2012).

Danışmanlı öğrenme, tahmin modelleri üretmek için sınıflandırma algoritmaları ve regresyon tekniklerini kullanmaktadır (Shobha vd., 2018). Danışmanlı makine öğrenmesi teknikleri, geçmiş ve mevcut veriyi kullanarak etiketler yardımıyla öngöründe bulunmaktadır. Buna göre, girdi değerlerine karşılık gelen çıktı değerlerini öngören bir fonksiyon geliştirilerek, yeterli eğitimin ardından girdiye ait sonuçlar ortaya konmaktadır. Algoritma sonuçlara bağlı olarak modeli değiştirmek için elde edilen sonuçlar ile gerçek ve beklenen sonuçları karşılaştırarak hatayı belirlemektedir (Saravanan ve Sujatha, 2018). Şekil 2.9 ile standart bir danışmanlı öğrenme modeli sunulmuştur.



Şekil 2.9: Danışmanlı öğrenme (Salian, 2018).

Talabis vd. (2015) birçok farklı sınıflandırma tekniğinin ve sınıflandırıcının var olduğunu fakat bunlardan yaygın olarak kullanılanlarının; Logistik Regresyon (*Logistic Regression*), Lineer Diskriminat Analizi (*Linear Discriminant Analysis*), k-En Yakın Komşu (*k-Nearest Neighbors*), Karar ağaçları (*Decision Trees*), Yapay Sinir Ağları ve Destek Vektör Makineleri olduğunu belirtmiştir.

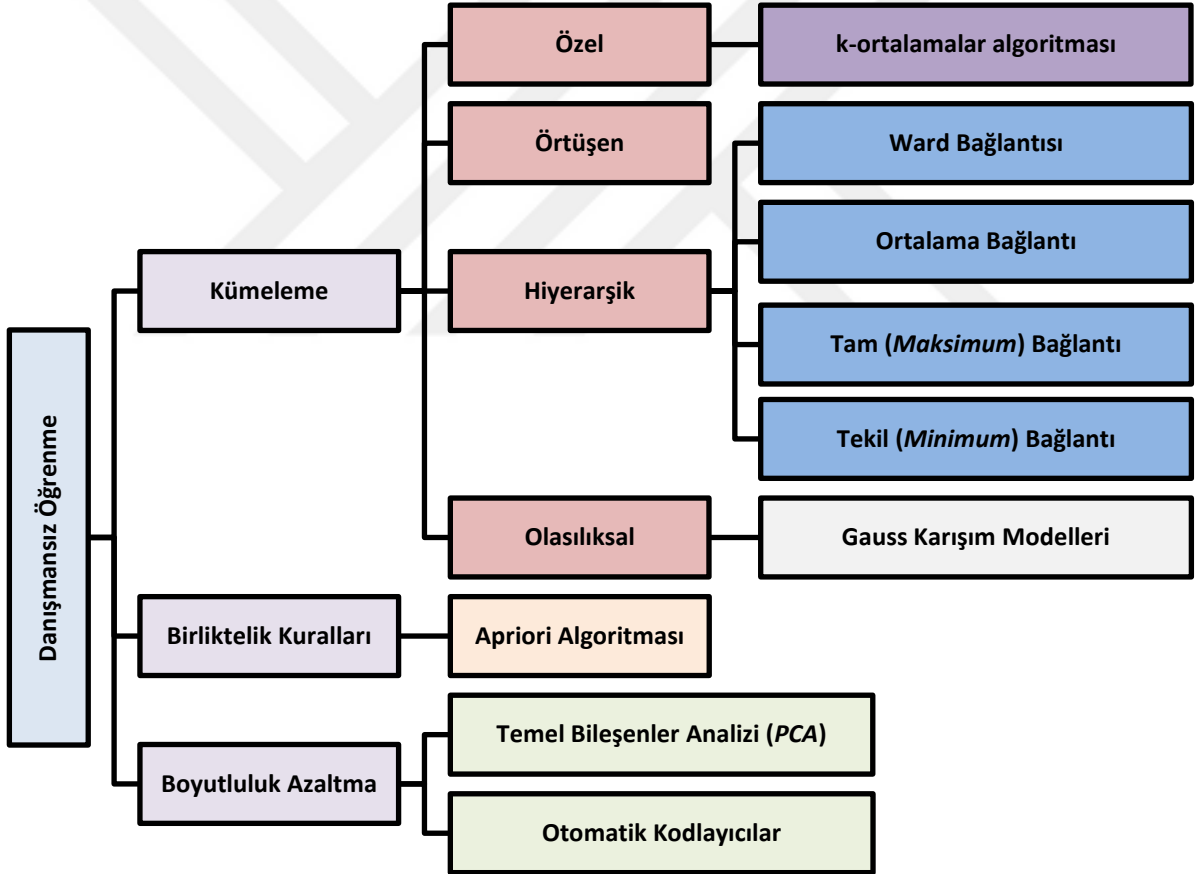
Sınıflandırma problemlerinde hedef niteliğin iki sınıf etiketine sahip olması halinde, problem “ikili sınıflandırma problemi” olarak ifade edilmektedir. Zhou (2012) ikili sınıflandırmada, iki sınıfın etiketini göstermek için “pozitif” ve “negatif” terimlerinin kullanıldığını belirtmiştir. Sınıflandırma problemlerinin çözümünde makine öğrenmesi tekniklerinden yararlanılmaktadır. Örneğin; Alpaydın (2020) bankalar için bir krediyle ilişkili riskin önceden tahmin edilebilmesinin önemli olduğunu ve makine öğrenmesi sistemi ile geçmiş veriye dayalı bir modele uyularak yeni kredi başvurularının düşük veya yüksek riskli şeklinde sınıflandırılabilirdiğini ve bu durumun bankaya başvuruyu reddetme veya kabul etme olanağı sağladığını belirtmiştir.

2.2.1.2. Danışmansız Öğrenme

Danışmansız öğrenme literatürde gözetimsiz veya denetimsiz öğrenme olarak da ifade edilmektedir. Bu öğrenme yönteminde, danışmanlı öğrenmeden farklı olarak tahmin etme işlemi için hedef veya sonuç değişkeni veri setinde bulunmamaktadır (Abdi, 2016). Yani veri seti etiketlenmemiş veriden oluşmaktadır. Bu nedenle danışmansız öğrenmede, veri setindeki ilgili örnekler gözlemlenerek ve bu örneklerin özellikleri arasındaki benzerliklerden yola çıkarak sınıfların tanımlanması amaçlanmaktadır (Terzi vd., 2011). Danışmansız öğrenme etiket bilgilerine dayanmamaktadır. Bunun amacı veri setinde yer alan bazı doğal dağılım

bilgilerini keşfetmektir. Genel görev, veri noktalarının küme yapısını keşfetmeyi amaçlayan kümeleme (*clustering*)'dir (Zhou, 2012).

IBM (2021), danışmansız öğrenmenin danışmansız makine öğrenmesi olarak da bilindiğini ve etiketlenmemiş veri kümelerini analiz etmek ve kümelemek için makine öğrenmesi algoritmalarını kullandığını belirtmiştir. Bu algoritmaların insan müdahalesine gerek kalmaksızın gizli örüntüleri keşfettiğini veya veriyi gruplandırıldığını ifade etmiştir. Şekil 2.10'dan da görüleceği üzere danışmansız öğrenme modellerinin kümeleme (*clustering*), birliktelik (*association*) ve boyutluluk azaltma (*dimensionality reduction*) olmak üzere üç temel görev için kullanıldığını belirtmiştir.



Şekil 2.10: Danışmansız öğrenme (IBM, 2021a).

Danışmansız öğrenme, makine öğrenmesi alanında; kümeleme (*clustering*), yoğunluk tahmini (*density estimation*) ve örüntü tanıma (*pattern recognition*) dahil olmak üzere yaygın olarak kullanılmaktadır (Lee vd., 2020). Örneğin; Shriram ve Sivasankar (2019) uzay mekiği verisindeki anormallikleri tespit etmek için danışmansız öğrenme tekniklerini kullanmıştır.

Silahtaroglu ve Canbolat (2020) sepsis enfeksiyonunun gelişimini etkileyen önemli parametrelerden olan laktat ve ph değerlerini kullanarak yeni bir danışmansız makine öğrenmesi modeli gerçekleştirmeyi amaçlamışlardır. Çalışma kapsamında, sepsis tanısı konmuş ve sepsis tanısı konmamış hastalar üzerinde Bulanık C-Ortalamlar algoritması ile çalışmalarını gerçekleştirmişlerdir. Bektaş (2020) kalkış, iniş, tırmanış, düz uçuş, yatay düzlemde dönüş vb. uçuş aşamalarında hava aracından elde edilen verilerin farklılık gösterdiğini ve sinyal değerlerinin içinde bulundukları çevre ve uygulamalara göre değiştiğini belirterek, aynı grupta bulunan uçuş veri kümelerinin diğer veri kümelerine göre daha ayırt edici olan karakteristik özelliklerine göre sınıflandırılması ve uçuş rejimlerinin bu kümeler amacıyla saptanması amaçlayan bir çalışma gerçekleştirmiştir. Buna göre; bölümlendirme işlemini yapay sinir ağlarının danışmansız bir eğitim biçimini kullanan "Özdüzenleyici Haritalar (*Self-Organising Maps – SOM*)" yöntemini kullanarak gerçekleştirmiştir.

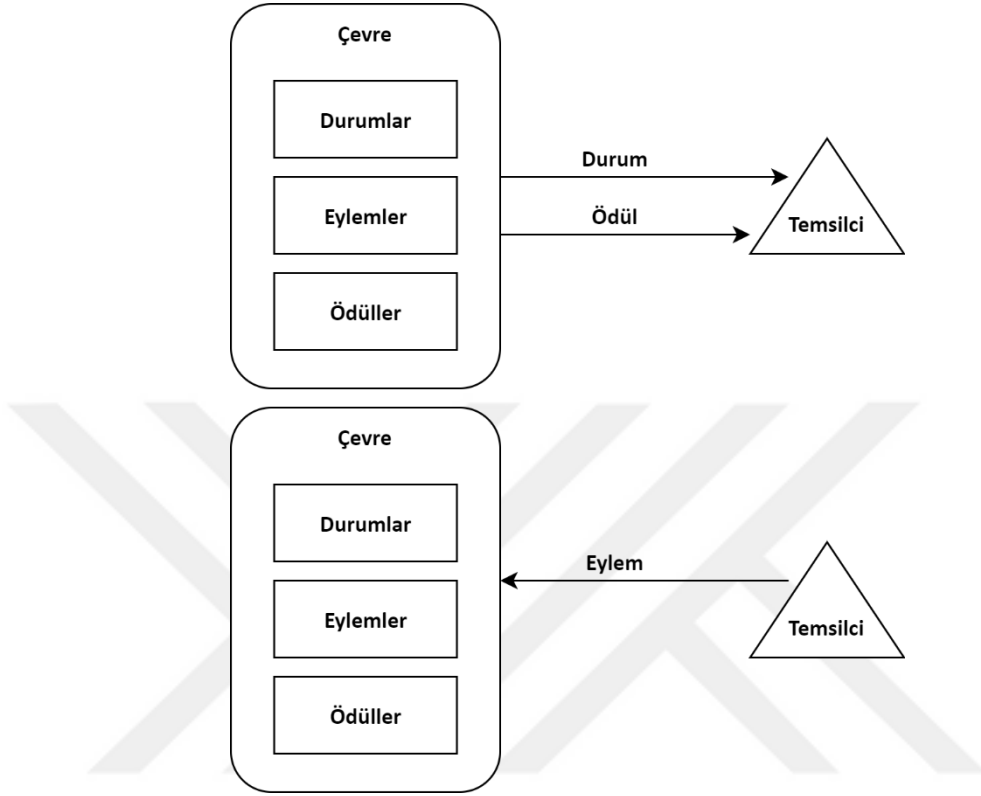
2.2.1.3. Pekiştirmeli Öğrenme

Pekiştirmeli öğrenme (*Reinforcement Learning*) çevresini algılayarak hareket eden otonom bir ajanın (autonomous agent) hedeflerine ulaşması için en uygun eylemleri nasıl öğrenebileceği sorusunu ele alır (Mitchell, 1997). Ajana öğretilmek istenen davranış dinamik bir ortamda deneme-yanılma etkileşimleriyle gerçekleşmektedir (Kaelbling vd., 1996).

Pekiştirmeli öğrenmeye ait terminoloji Cheskidov vd. (2019) tarafından aşağıdaki gibi açıklanmıştır:

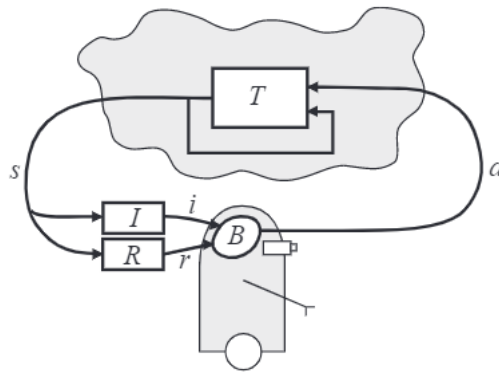
- **Çevre** (*Environment*): Ajanın var olduğu ve etkileşime girdiği yerdir.
- **Durumlar** (*States*): Durum, bu dünyayı karakterize eden bilginin tek bir parçasının bile eksik olmadığı bir dünyanın tam bir tanımıdır. Sabit veya dinamik bir durum olabilir.
- **Eylem** (*Action*): Eylem genellikle çevresel koşullara bağlıdır ve farklı ortamlarda ajan farklı eylemler gerçekleştirecektir. Birçok geçerli ajan eylemi “aksiyon uzayı” adı verilen uzaya kaydedilir. Tipik olarak, uzaydaki eylem sayısı sonludur.
- **Ödüller ve Kazanımlar** (*Rewards and Winnings*): Pekiştirmelerle öğrenme yaparken ödül fonksiyonu R’yi sürekli olarak izlemeniz gerekir. Bir algoritma kurarken, optimize ederken ve ayrıca öğrenmeyi bıraktığınızda çok önemlidir. Bu, dünyanın mevcut durumuna, yapılan eyleme ve dünyanın bir sonraki durumuna bağlıdır.

- **Stratejiler (Strategies):** Strateji, ajanın bir sonraki eylemi seçmesine göre bir kuraldır. Stratejilerin kümesi, ajanın "beyni" olarak da adlandırılır.



Şekil 2.11: Bir ajanın eğitiminin genel şeması (Cheskidov vd., 2019).

Bir standart pekiştirmeli öğrenme modeli Şekil 2.12’de sunulmuştur.



Şekil 2.12: Standart pekiştirmeli öğrenme modeli (Kaelbling vd., 1996).

Şekil 2.12’den de görüldüğü üzere bir standart pekiştirmeli öğrenme modeli şunları içermektedir:

- Çevresel s durumlarının ayrık kümesi (S),
- Ajanın a eylemlerinin ayrık kümesi (A),
- Ajanın davranışı B ,
- Sayısal pekiştirme sinyali r ,
- Girdi i ,
- Sayısal pekiştirme sinyallerinin kümesi, genellikle $\{0,1\}$ veya gerçekte sayılar,
- Ajanın çevresel durumu nasıl gördüğünü belirleyen girdi fonksiyonu (I),
- Ödül fonksiyonu R ,
- Durum dönüşüm fonksiyonu T (Kaelbling vd., 1996).

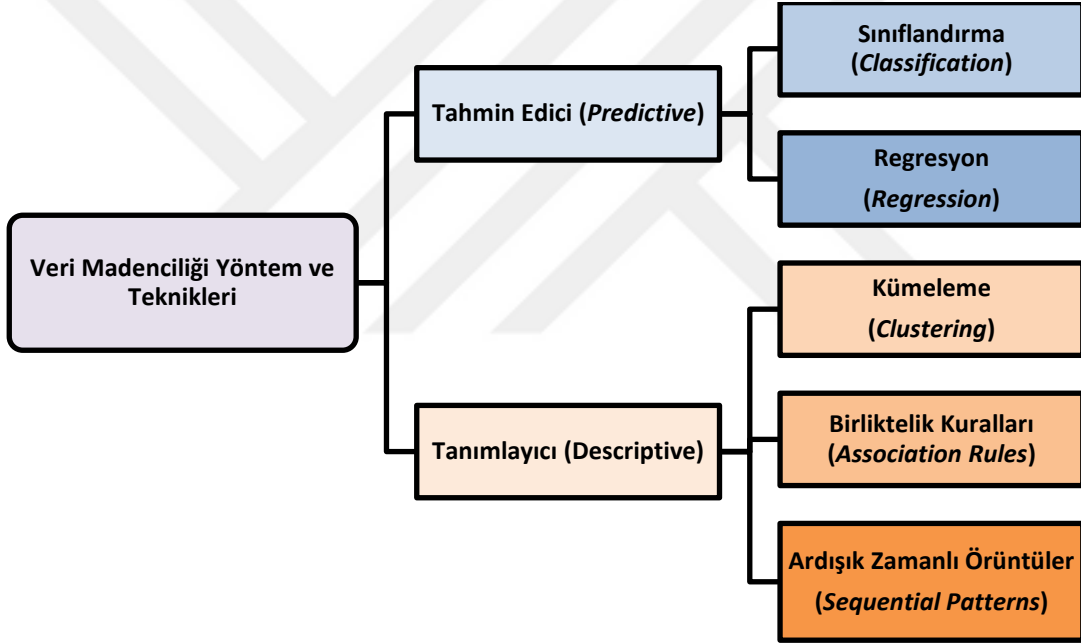
Bir standart pekiştirmeli öğrenme modelinde etkileşimin her bir adımında ajan çevrenin mevcut durumuna ait bazı son göstergelerini (s), girdi (i) olarak almaktadır. Sonrasında ajan bir eylemi (a) çıktı olarak üretmek için seçmektedir. Eylem çevrenin durumuna göre değişmektedir ve bu eylemin değişimi bir sayısal pekiştirme sinyali (r) ile ajana iletilmektedir. Ajanın davranışı (B), pekiştirme sinyali değerlerinin uzun vadedeki toplamını artırma eğiliminde olan eylemleri seçmelidir. Bunu yapmayı, çok çeşitli algoritmalar tarafından yönlendirilerek sistematik deneme yanılma yoluyla zaman içinde öğrenebilir (Kaelbling vd., 1996).

Bartın (2017) pekiştirmeli öğrenme yöntemini kullanarak araçların kritik aralık kabul kararlarının basit bir T-kavşakta modellenmesini sunmuştur. Hachaichi vd. (2020) tedarik zinciri yöntemi alanında yapmış oldukları çalışma ile doğru miktarlarda ve doğru zamanda en uygun siparişleri verebilen pekiştirmeli öğrenmenin son başarılarına dayalı akıllı bir temsilci geliştirmeyi amaçlamışlardır. Cheskidov vd. (2019) siber güvenlik alanında yapmış oldukları çalışma ile pekiştirmeli öğrenme metotlarından biri kullanılarak eğitilen üretici-rekabetçi ağları kullanarak dağıtık hizmet engelleme (*Distributed Denial of Service - DDoS*) saldırılarının modellenmesi ihtimalini değerlendirmeyi önermiştir.

2.3.MAKİNE ÖĞRENMESİ SÜRECİ ADIMLARI

Değersiz taşlar ve değerli altın parçacıkları birleşerek altın madenlerini oluşturmaktadır; madencilikteki esas amaç ise altını bulabilmektir. Bunun için belirli teknikler ile değersiz taşlar değerli altın parçacıklarından ayrıştırılarak altın bulma işlemi gerçekleştirilir. Veri için de durum benzerdir. Günümüzde teknolojinin hızla gelişmesine paralel olarak bilgisayarların veri depolama ve saklama kapasiteleri her geçen gün artmaktadır. Büyüyen bu kapasiteler daha fazla

verinin depolanmasına imkân tanımaktadır. Öte yandan depolanan bu veri, kendi başına bir anlam ifade etmemektedir. Bu noktada veri analizi kavramı devreye girmektedir. Veri analizi ile büyük miktardaki veri belli bir amaç doğrultusunda belirli teknikler aracılığıyla işlendiğinde anlamlı hale dönüşmektedir. İşte bu aşamada ham verinin anlamlı bilgiye dönüştürülmesi veri madenciliği (*data mining*) kavramını ortaya çıkarmıştır. Savaş vd. (2012) veri madenciliğini, geleceğe yönelik tahminlerde bulunmak için bilgisayar programları aracılığıyla büyük miktarlardaki verinin içinden anlamlı ve yararlı bağlantı ve kuralların aranması ve analizi olarak tanımlamışlardır. Witten vd. (2017) ise veri madenciliğini verideki örüntüleri keşfetme süreci olarak tanımlamış ve yararlı örüntülerin yeni veriden aşikâr olmayan tahminler yapmamıza imkân vereceğini belirtmişlerdir.

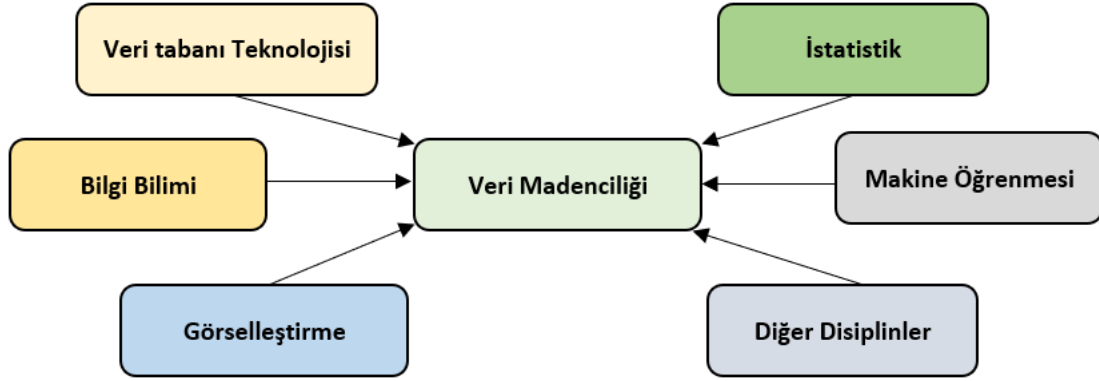


Şekil 2.13: Veri madenciliği yöntem ve teknikleri (Albayrak ve Koltan Yılmaz, 2009).

Albayrak ve Koltan Yılmaz (2009) veri madenciliğinde kullanılan modellerin, Tahmin Edici (*Predictive*) ve Tanımlayıcı (*Descriptive*) olmak üzere iki ana başlık altında incelendiğini ve bu modelleri gördükleri işlevlere göre Sınıflandırma (*Classification*) ve Regresyon (*Regression*) modelleri, Kümeleme (*Clustering*), Birliktelik Kuralları (*Association Rules*) ve Ardışık Zamanlı Örüntüler (*Sequential Patterns*) olmak üzere üç ana başlık altında incelemenin mümkün olduğunu belirtmiştir (Şekil 2.13).

Han ve Kamber (2006) veri madenciliğini büyük miktarda veriden bilginin çıkarılması veya madenciliği (*mining*) olarak tanımlamış ve Şekil 2.14'ten de görüldüğü üzere veri

madenciliğinin veri tabanı sistemleri, istatistik, makine öğrenmesi, görselleştirme, bilgi bilimi ve diğer disiplinlerin birleşiminden oluşan disiplinler arası bir alan olduğunu belirtmiştir.



Şekil 2.14: Birçok disiplinin birleşimi olarak veri madenciliği (Han ve Kamber, 2006).

Öte yandan Han vd. (2012) bilginin keşfi sürecini Şekil 2.15’te de belirtilen yedi adım şeklinde tanımlamış ve veri madenciliğinin de bu adımların bir parçası olduğunu ifade etmiştir.

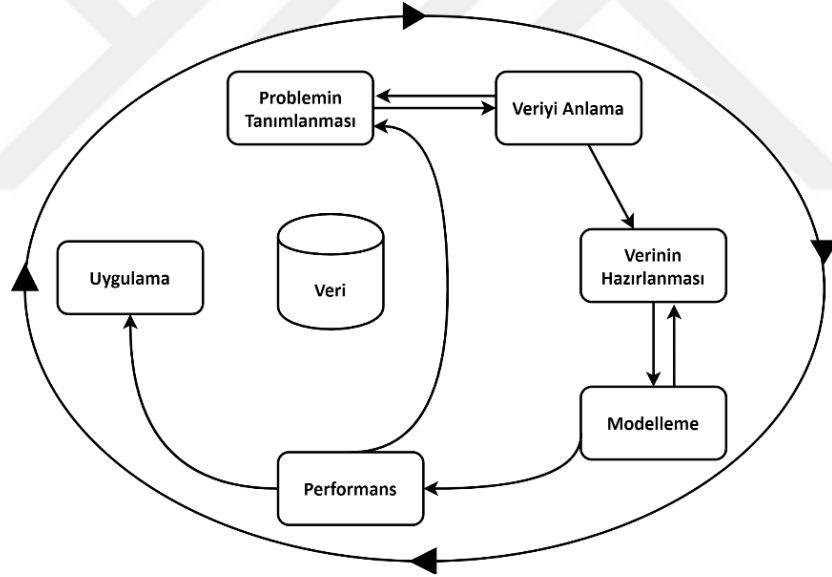


Şekil 2.15: Bilginin keşfi süreci (Han vd., 2012).

Ticari faaliyetler için veri madenciliği uygulamalarını ilk benimseyen şirketler ve tedarikçilerden oluşan konsorsiyum (*Daimler–Chrysler (sonradan Daimler–Benz AG,*

Almanya), NCR Systems Engineering Copenhagen (Danimarka), SPSS/Integral Solutions Ltd (İngiltere) ve OHRA Verzegeingen en Bank Groep B.V.I. (Hollanda)) Veri Madenciliği için Çapraz Endüstri Standart Süreç Modeli (CRoss Industry Standart Process for Data Mining – CRISP-DM) olarak adlandırılan bir talimatlar kümesi geliştirmişlerdir (Chapman vd., 2000; Edelstein, 1999; Becerra-Fernandez vd., 2002). Bu model ile endüstride veri madenciliği süreçlerinin standartlaştırılması amaçlanmıştır. Veri madenciliği süreçlerinden en yaygın olarak kullanılanı CRISP-DM'dir (Küçüksille, 2009).

Balaban ve Kartal (2015) makine öğrenmesi ve veri madenciliği arasında yakın bir ilişkinin olduğunu belirtmiştir. Bu yönüyle düşünüldüğünde veri madenciliği sürecinde izlenen yolun benzeri makine öğrenmesinde de izlenebilir kanaatine varılmıştır. Bu çalışmada CRISP-DM (Shearer, 2000) örnek model olarak seçilmiş ve bu sürecin adımları Şekil 2.16'da verilerek sırasıyla açıklanmıştır.



Şekil 2.16: Veri madenciliği için CRISP-DM işlem modeli (Wirth ve Hipp, 2000).

2.3.1. Problemin Tanımlanması (*Business Understanding*)

CRISP-DM modelinin ilk aşaması problemin amaçlarını ve gereksinimlerini iş perspektifinden anlamaya ve bu bilgiyi veri madenciliği problemi tanımına ve hedeflere ulaşmak için tasarlanmış bir ön proje planına dönüştürmeye odaklanır (Wirth ve Hipp, 2000).

2.3.2. Veriyi Anlama (*Data Understanding*)

Problemin tanımlanması sonrası problemin çözümünde kullanılacak olan veri setinin ele alındığı aşamadır. Veri seti içerisinde yer alan tüm değişkenler incelenir. Problemin çözümünde hangi nitelik/niteliklerden yararlanılacağı belirlenir.

2.3.3. Veriyi Hazırlama (*Data Preparation*)

Han vd. (2012) gerçek dünya verilerinin eksik (*missing*), gürültülü (*noisy*) ve tutarsız (*inconsistent*) olma eğiliminde olduğunu belirterek, veri temizleme (*data cleaning*) işlemlerinin, veride yer alan eksik değerleri tamamlamayı, aykırı değerleri belirleyerek gürültüyü gidermeyi ve verideki tutarsızlıkları düzeltmeyi gerçekleştirdiğini ifade etmiştir. Veriye çeşitli işlemler uygulanarak analizlere uygun hale getirilir. Bu işlemler sırasıyla şu şekilde sıralanmaktadır:

- Analizlerde kullanılmayacak olan niteliklerin veri setinden çıkarılması,
- Aykırı değerlerin (*uç noktaların*) veri setinden çıkartılması,
- Tekrar eden gözlemlerin veri setinden çıkartılması,
- Veri setinde yer alan kayıp değerlerin tamamlanması,
- Nitelikler arasındaki ilişkinin incelenmesi.

2.3.3.1. Aykırı Değer Tespiti

Veri seti içerisinde aynı kategorideki nitelikler içinde çok büyük ve çok küçük bir değere sahip olan nitelikler analiz sonuçlarının sağlıklı olmasını etkileyecektir. Literatürde bu özellikteki nitelikler aykırı değer/uç nokta (*outlier*) ile ifade edilmektedir. Veri setinde bulunan nümerik niteliklere ait kutu grafiği çizimleri ile uç noktalar tespit edilebilir (Balaban ve Kartal, 2015).

2.3.3.2. Tekrar Eden Gözlemlerin İncelenmesi

Balaban ve Kartal (2015) veri setinde tekrar eden gözlemlerin analizler sırasında vakit kaybı yaşattığını ve veri setinin gereksiz şişmesine neden olacağını ifade etmiştir. Öte yandan, tekrar eden gözlemlerin var olması veri setinde benzer vak'aların olabileceği anlamına geldiğini belirtmiştir.

2.3.3.3. Kayıp Değerlerin Tamamlanması

Veri setinde yer alan niteliklerden herhangi birine ait bir değer ölçülmemiş olabilir ve bu durum literatürde kayıp değer (*missing value*) olarak ifade edilmekle beraber analizlerde kayıp değerlerden kaynaklı bazı problemleri beraberinde getirmiştir (Balaban ve Kartal, 2015). Bu durum ile ilgili olarak, Edelstein (1999) bir alanı eksik olan her kaydın atılmasının; verinin küçülmesine veya tüm verinin yanlış bir resmi ile karşılaşılmamasına yol açabileceğini belirtmiştir. Bu nedenle bir değer eksik olması durumunun kendi içerisinde önemli olduğunu ifade etmiştir. Han vd. (2012) bazı durumlarda, veride kayıp değerlerin yer almasının verinin hatalı olduğu anlamına gelmediğini belirtmiştir. Örneğin, bir kredi kartı başvurusu için adaylardan sürücü belgesi numaraları istenebilir. Adaylardan sürücü belgesine sahip olmayanlar bu alanı boş geçecektir. Han vd. (2012) kayıp değerlerin tamamlanması ile ilgili olarak aşağıdaki metotların uygulanabileceğini ifade etmiştir:

- Kayıp ögeyi önemsememek,
- Kayıp değeri elle tamamlamak,
- Kayıp değer için genel bir sabit kullanmak,
- Kayıp değeri tamamlamak için niteliğin ortalama değerini kullanmak,
- Kayıp değer ait olduğu nitelikteki tüm örneklerle niteliğin ortalama değerini vermek,
- Kayıp değeri tamamlamak için en olası değeri kullanmak.

2.3.3.4. Normalizasyon

Balaban ve Kartal (2015) veri setinde yer alan nümerik değerlerin tamamının aynı aralıkta bulunamayabileceğini belirtmiş ve bu durumun, analizler sırasında özellikle uzaklık gibi hesaplamalarda bazı niteliklerin analizi domine edebileceğini belirtmiştir. Bu nedenle, eğitim giriş setinde yer alan her bir parametrenin modelin tahmin işlemine eşit ölçüde katkı sağlaması adına verinin normalizasyonu işlemi gerçekleştirilir (Yüksek vd., 2007). Literatürde verinin normalizasyonu için birçok teknik bulunmaktadır (Jayalakshmi ve Santhakumaran, 2011). Han vd. (2012) bu tekniklerden bazılarını aşağıdaki şekilde ifade etmiştir:

- **Min-maks. Normalizasyon** (*Min-max Normalization*):

$$v' = \frac{v - \min_A}{\max_A - \min_A} (\text{yeni_maks}_A - \text{yeni_min}_A) + \text{yeni_min}_A \quad (2.9)$$

min_A : A niteliğinin minimum değeri

$maks_A$: A niteliğinin maksimum değeri

v : Normalize edilecek değer

v' : Normalize edilen değer

- **Z-skor Normalizasyon** (*Z-score Normalization*):

$$v' = \frac{v - \bar{A}}{\sigma_A} \quad (2.10)$$

$\bar{A}$: A niteliğinin ortalama değeri

σ_A : A niteliğinin standart sapması

- **Ondalık Ölçekleme** (*Decimal Scaling*):

$$v' = \frac{v}{10^j} \quad (2.11)$$

$j, Maks(|v'|) < 1$ Olacak şekilde en küçük tam sayı

2.3.4. Modelleme (*Modeling*)

Blaha (2010) modeli, bir problemin bazı yönlerinin bir soyutlaması olarak tanımlamış ve modellerin geliştiricilere problemi nasıl anladıklarını göstermesi ve bunu başkalarına (*teknoloji uzmanları, iş uzmanları, kullanıcılar, yöneticiler ve diğer proje paydaşları*) aktarmasına izin verdiğini ifade etmiştir. Ek olarak modellerin veri modelleri, durum-geçiş modelleri ve veri-akış modelleri gibi birçok çeşidinin bulunduğunu ve bunların veri tabanları, programlama ve diğer amaçlar için kullanıldığını belirtmiştir.

Bir model genellikle bir tahmin modeli veya veri setini kullanarak oluşturulmak ya da keşfedilmek istenen verinin yapısının modelidir, örneğin; karar ağacı, sinir ağı, destek vektör makinesi vb. Bir öğrenme algoritması tarafından veriden modellerin üretilmesi işlemi “öğrenme” (*learning*) ya da “eğitim” (*training*) olarak adlandırılmaktadır. Öğrenmiş model (*learner*), hipotez olarak da adlandırılabilir. Birçok öğrenme yöntemi olmakla beraber literatürde en sık karşılaşılanları tezin GENEL KISIMLAR bölümünde de bahsedildiği gibi danışmanlı ve danışmansız öğrenmedir.

Modelleme safhasında, ele alınan probleme uygun öğrenme yöntemi ve algoritmalar belirlendikten sonra modeller kurulur. Problemin çözümü için en uygun (*en iyi performansı sergileyen*) model elde edilmeye çalışılır.

Bu tez çalışmasında, sınıflandırma işlemini gerçekleştirmede yakıtTasimaciligi hedef niteliği ikili kategorik (*binary*) (0: *Yok*, 1: *Var*) olarak (*yakitTasimaciligiDrm*), regresyon işlemini gerçekleştirmede ise yakıtTasimaciligi hedef niteliği nümerik olarak (*yakitTasimaciligiMktr*) kullanılmıştır. Sınıflandırma amacıyla kullanıldığında; yakıt taşımacılığı durumunun ‘**Var**’ veya ‘**Yok**’ şeklinde sınıflandırılması, regresyon amacıyla kullanıldığında ise yakıt taşımacılığı durumunun ‘**Var**’ olduğu durumlarda taşınacak yakıt miktarının tahmin edilmesi sağlanmıştır. Tablo 2.1’de bu tez çalışmasında kullanılan algoritmaların kullanım amaçları belirtilmiştir.

Tablo 2.1: Kullanılan algoritmaların kullanım amaçları.

Sınıflandırma	Regresyon
k-En Yakın Komşu Algoritması	k-En Yakın Komşu Algoritması
Naive (<i>Basit</i>) Bayes Sınıflandırıcı	-
Sınıflandırma ve Regresyon Ağaçları	Sınıflandırma ve Regresyon Ağaçları
Yapay Sinir Ağları	Yapay Sinir Ağları
Destek Vektör Makineleri	Destek Vektör Makineleri
Rastgele Orman Algoritması	Rastgele Orman Algoritması

2.3.4.1. k-En Yakın Komşu Algoritması

k-En Yakın Komşu (*k-Nearest Neighbor Algorithm - kNN*) algoritmasının temelleri ilk olarak Fix ve Hodges (1951)’un ortaya koyduğu k-En Yakın Komşu kuralı ile atılmıştır. Basu vd. (2012) k-En Yakın Komşu kuralının çok popüler ve basit bir sınıflandırma tekniği olduğunu belirtmiştir.

k-En Yakın Komşu algoritması statik bir model vermeyip, analiz sırasında her denemede uzaklık hesaplamasına bağlıdır. Kolay uygulanması, veri seti küçükken hızlı çalışması ve eğitim veri setinin yapısı hakkında ön bilgiye gerek duymaması bu algoritmanın bazı avantajlarıdır (Sarkar ve Leong, 2000). Harrington (2012) k-En Yakın Komşu algoritmasını şu şekilde ifade etmiştir:

1. Eğitim örnekleri ve verilen etiketsiz yeni örnek arasındaki (veya hedef niteliği bilinmeyen bir sınıf etiketine sahip olan) uzaklıklar hesaplanır,
2. Yeni örneğin en yüksek frekansa (çoğunluk sınıfı) sahip k en yakın komşusunun sınıf etiketine göre yeni örneğin sınıf etiketi belirlenir.

Balaban ve Kartal (2015) k-En Yakın Komşu algoritmasında yeni bir örneğin sınıfının, örneğin belirlenen bir k değerine göre, mevcut örneklem içindeki örneklerle olan uzaklığı hesaplanarak (k tane en yakın komşusu bulunarak) tespit edildiğini belirtmiştir. Shen (2010) algoritmalarında kullanılan bazı uzaklık fonksiyonlarını;

- Öklid Uzaklığı (*Euclidean Distance*),
- Manhattan Uzaklığı (*Manhattan Distance*),
- Hamming Uzaklığı (*Hamming Distance*),
- Kosinüs Benzerliği (*Cosine Similarity*),

olarak ifade etmiş ve bu uzaklık fonksiyonları sırasıyla denklem 2.12, 2.13, 2.14 ve 2.15'te verilmiştir:

$d(x_i, x_j)$, n adet nitelik ve x_i ve x_j gibi iki örnek arasındaki uzaklık olmak üzere;

$$d_{\text{öklid}}(x_i, x_j) = \sqrt{\sum_{s=1}^n (a_s(x_i) - a_s(x_j))^2} \quad (2.12)$$

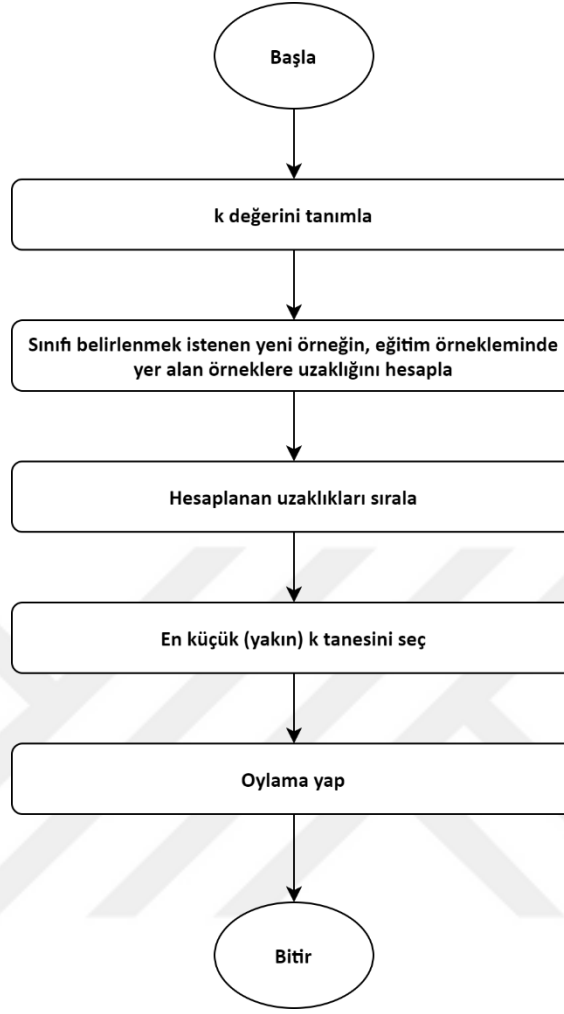
$$d_{\text{manhattan}}(x_i, x_j) = \sum_{s=1}^n |a_s(x_i) - a_s(x_j)| \quad (2.13)$$

$$= \begin{cases} 0, & \text{eğer } x = y \\ 1, & \text{eğer } x \neq y \end{cases}$$

$$d_{\text{hamming}}(x_i, x_j) = \sum_{s=1}^n I(a_s(x_i), a_s(x_j)); I(x, y) \quad (2.14)$$

$$d_{\text{kosinüs}}(x_i, x_j) = \frac{x_i \cdot x_j}{\|x_i\| \|x_j\|} = \cos \theta \quad (2.15)$$

k-En Yakın Komşu algoritması için Şekil 2.17'de bir akış diyagramı sunulmuştur.



Şekil 2.17: k-En Yakın Komşu algoritması için akış diyagramı (Balaban ve Kartal, 2015).

2.3.4.2. Basit (Naive) Bayes Sınıflandırıcı

Basit (Naive) Bayes Sınıflandırıcı danışmanlı sınıflandırma metodudur (Nahayo ve Arı, 2016). Temeli, istatistikteki Bayes teoremine dayanan ve sahip olunan verinin belirlenmiş olan sınıflara ait olma olasılıklarını öngören bir algoritmadır (Kökver vd., 2014). Bayes Teoremi, olasılık ve karar teorisinde ilk çalışmaları yapan Thomas Bayes'ten adını almıştır (Han vd., 2012). Bayes Teoremini aşağıdaki şekilde açıklamak mümkündür (Akdeniz, 2010):

A ve B , bir S örnek uzayında iki olay olsun. $P(A|B)$ 'ye B verilmişken, A olayının koşullu olasılığı (veya B verilmişken A nın gerçekleşme olasılığı) denir ve

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, (P(B) \neq 0 \text{ ise}) \quad (2.16)$$

şeklinde tanımlanır. Benzer şekilde A verilmişken, B nin gerçekleşme olasılığı,

$$P(B|A) = \frac{P(A \cap B)}{P(A)}, (P(A) \neq 0 \text{ ise}) \quad (2.17)$$

dir. Denklem 2.16'nın her iki yanını $P(B)$ ile çarparsak,

$$P(A \cap B) = P(B).P(A|B) \quad (2.18)$$

buluruz. $A \cap B = B \cap A$ olduğundan, ayrıca,

$$P(A \cap B) = P(A).P(B|A) \quad (2.19)$$

yazılabilir.

$B_1, B_2, \dots, B_k$ lar bir S örnek uzayının bir parçalanışı olsun. $B_1, B_2, \dots, B_k$ olayları ayrık ve birleşimleri örnek uzay olduğundan,

$$S = B_1 \cup B_2 \cup \dots \cup B_k \quad (2.20)$$

ve $B_i \cap B_j = \emptyset, i = 1, 2, \dots, k; j = 1, 2, \dots, k; i \neq j$ yazılır.

A, S içinde $P(A) \neq 0$ olan bir olay ise, A olayı, $B_1 \cap A, B_2 \cap A, \dots, B_k \cap A$ ayrık olaylarının birleşimidir. Öyle ki,

$$A = (B_1 \cap A) \cup (B_2 \cap A) \cup \dots \cup (B_k \cap A) \quad (2.21)$$

dir. $B_1 \cap A, B_2 \cap A, \dots, B_k \cap A$ lar ayrık olduklarından,

$$P(A) = P(B_1 \cap A) + \dots + P(B_k \cap A) \quad (2.22)$$

dır. Koşullu olasılık tanımını uygulayarak,

$$\begin{aligned} P(B_r|A) &= \frac{P(B_r \cap A)}{P(A)} = \frac{P(B_r \cap A)}{P(B_1 \cap A) + \dots + P(B_k \cap A)} \\ &= \frac{P(B_r \cap A)}{\sum_{i=1}^k P(B_i \cap A)}, r = 1, 2, \dots, k \end{aligned} \quad (2.23)$$

buluruz. Öte yandan denklem 2.23,

$$P(B_r|A) = \frac{P(B_r)P(A|B_r)}{\sum_{i=1}^k P(B_i)P(A|B_i)}, r = 1, 2, \dots, k \quad (2.24)$$

şeklinde de yazılır. Bu denklem Bayes Teorimi olarak ifade edilir.

Basit Bayes Sınıflandırıcı aşağıdaki biçimde özetlenebilir (Han vd., 2012):

$X = \{x_1, x_2, \dots, x_n\}$ örnekler kümesi, $A_k = \{A_1, A_2, \dots, A_n\}$ nitelikler kümesi ve $C_1, C_2, \dots, C_m$ m adet sınıf olsun. Amaç, $P(C_i|X)$ 'i maksimize etmektir. C_i sınıfı için $P(C_i|X)$ 'in maksimize edilmesi maksimum sonsal hipotez (*maximum posteriori hypothesis*) olarak adlandırılmaktadır. Bayes teoreminden,

$$P(C_i|X) = \frac{P(X|C_i) \cdot P(C_i)}{P(X)} \quad (2.25)$$

olarak yazılır. $P(X)$ tüm sınıflar için sabit olduğundan, yalnızca $P(X|C_i) \cdot P(C_i)$ maksimize edilmelidir. Birçok niteliğe sahip veri kümeleri verildiğinde, $P(X|C_i)$ 'yi hesaplamak oldukça zor olacaktır. Bu nedenle $P(X|C_i)$ 'nin hesaplanmasını kolaylaştırmak için sınıf koşullu bağımsızlık basit varsayımı (*naive assumption of class conditional independence*) yapılmıştır. Bu varsayıma göre nitelikler arasında hiçbir bağımlı ilişki bulunmamaktadır. Böylece;

$$P(X|C_i) = \prod_{k=1}^n P(x_k|C_i) = P(x_1|C_i) \times P(x_2|C_i) \times \dots \times P(x_n|C_i) \quad (2.26)$$

olarak yazılabilir. x_k, A_k niteliği için nitelik değeri olduğundan, A_k 'nin kategorik veya nümerik olmasına göre $P(X|C_i)$ aşağıdaki biçimde hesaplanmaktadır:

- A_k **kategorik** ise, $P(x_k|C_i)$, x_k nitelik değerine sahip C_i sınıfına ait veri sayısının, C_i sınıfına ait tüm veri sayısına oranıdır.
- A_k **nümerik** ise, nümerik değere sahip bir niteliğin genel olarak μ ortalamalı ve σ standart sapmalı Gauss dağılımına sahip olduğu varsayılmaktadır ve

$$g(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (2.27)$$

olarak tanımlanmaktadır. Böylece,

$$P(x_k|C_i) = g(x_k, \mu_{C_i}, \sigma_{C_i}) \quad (2.28)$$

olarak bulunur.

2.3.4.3. Sınıflandırma ve Regresyon Ağaçları

Sınıflandırma ve Regresyon Ağaçları veya (*Classification and Regression Trees - CART*) algoritması 1984 yılında Breiman vd. tarafından önerilmiştir (Breiman vd., 1984). Morgan ve Sonquist (1963)'in AID (*Automatic Interaction Detection*) adlı karar ağacı algoritmasının devamı niteliğindedir. Sınıflandırma ve Regresyon Ağaçları algoritması, sınıflandırma ve regresyon problemlerinin çözümünde kullanılabilir (Akçapınar Sezer vd., 2010).

Sınıflandırma ve Regresyon Ağaçları yöntemi, en uygun bölünme noktasını Gini indeksini kullanarak bulmaktadır (Yerlikaya-Özkurt ve Askan, 2020). D örneklemini k tane sınıfa ait n adet veri içersin. p_i , i sınıfında yer alan örnek sayısının D örneklemindeki örnek sayısına oranı (D veri setindeki i sınıfının olasılığı) olmak üzere,

$$Gini(D) = 1 - \sum_{i=1}^k p_i^2 \quad (2.29)$$

olarak hesaplanır (Aiping Wang vd., 2009). Uzbaş (2017) Gini indeksinin her özellik için ikili bölünmeyi değerlendirdiğini belirterek, bir A niteliği için en iyi ikili bölünmeyi belirlemek amacıyla olası tüm alt grupların incelendiğini ifade etmiştir. A niteliği için D veri seti D_1 ve D_2 şeklinde iki alt veri setine bölündüğünde, ayrım yapıldıktan sonraki gini indeksi denkleminin;

$$Gini_A(D) = \frac{|D_1|}{|D|} Gini(D_1) + \frac{|D_2|}{|D|} Gini(D_2) \quad (2.30)$$

olduğunu belirtmiştir. A niteliğinin ikili bölünmesiyle elde edilen safsızlıktaki azalmayı,

$$\Delta Gini(A) = Gini(D) - Gini_A(D) \quad (2.31)$$

denklemini ile ifade ederek en düşük $Gini_A(D)$ değerine sahip alt kümenin bölme alt kümesi olarak seçileceğini ifade etmiştir.

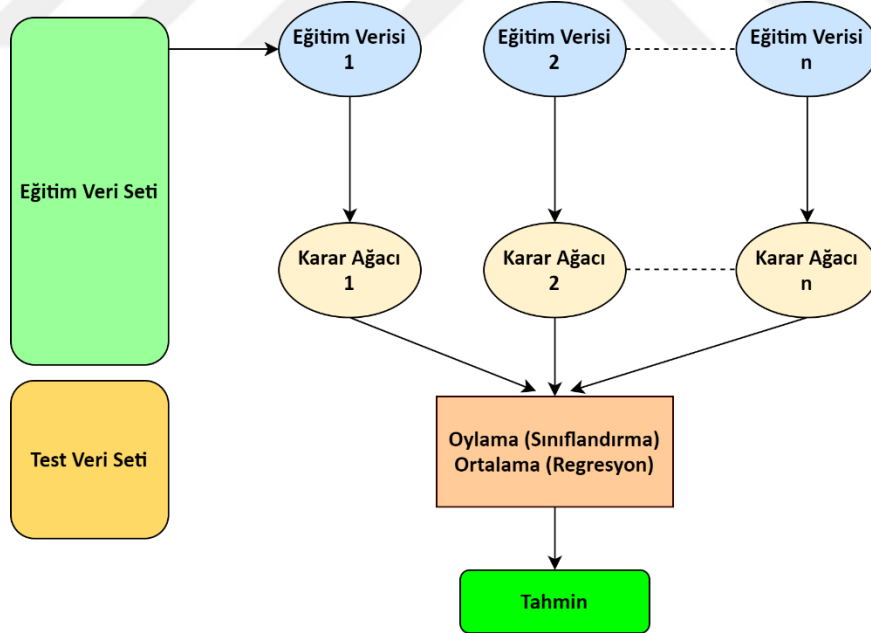
2.3.4.4. Rastgele Orman Algoritması

Rastgele Orman (*RF*) 2001 yılında Breiman tarafından önerilen ve yüksek sınıflandırma performansına sahip yeni bir sınıflandırıcı kombinasyon modelidir (Liu vd., 2019). Rastgele

Orman, bir topluluk öğrenmesi (*ensemble learning*) yöntemidir. Toplu öğrenme veya topluluk öğrenmesi fikri bir topluluğun genelleme yeteneğinin genellikle tek bir öğreneninkinden çok daha güçlü olduğu teorisi temeline dayanmaktadır (Huang vd., 2009). Gürsakal (2018) topluluk öğrenmesinin, iki veya daha fazla benzer veya benzer olmayan türden algoritmayı birleştirme tekniği olduğunu ve makine öğrenmesi sonuçlarını iyileştirdiğini belirtmiştir.

Topluluk yöntemleri hem regresyon hem de sınıflandırma problemlerine uygulanmaktadır ve bu yöntemler birçok durumda diğer tekli model yöntemlerinden daha iyi performans göstermesi nedeniyle popülerlik kazanmıştır (Goudos ve Athanasiadou, 2019). Bu tez çalışmasında, Rastgele Orman algoritması hem sınıflandırma hem de regresyon amacıyla kullanılmıştır.

Şekil 2.18’de Rastgele Orman algoritmasının çalışma şekli gösterilmiştir. Buna göre, Rastgele Orman algoritması eğitim veri setindeki eğitim verisini kullanarak birden fazla karar ağacı üretmektedir. Üretilen bu karar ağaçlarının her birinden elde edilen tahminler, algoritma sınıflandırma amacıyla kullanılıyorsa oylama (*majority voting*) yapılarak, regresyon amacıyla kullanılıyorsa ortalama (*averaging*) alınarak nihai tahmin elde edilmektedir.



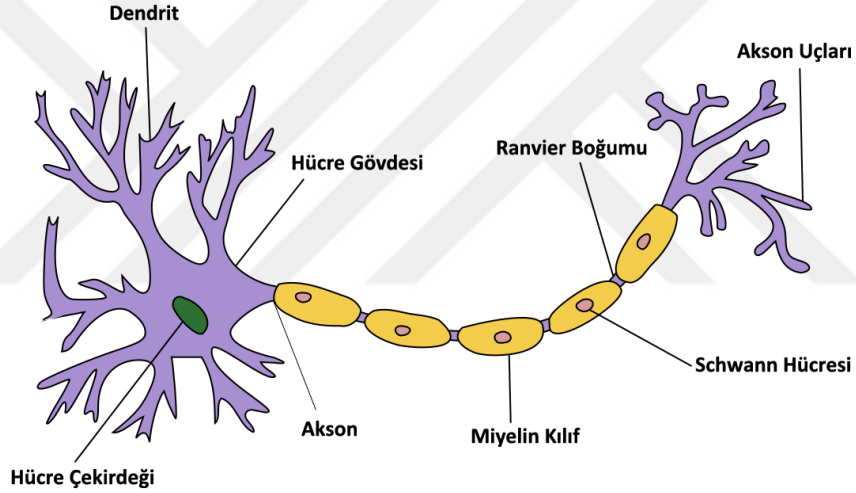
Şekil 2.18: Rastgele Orman Algoritması (Javatpoint, 2020).

2.3.4.5. Yapay Sinir Ağları

Yapay Sinir Ağları, insan beyninin en temel özelliği olan öğrenme fonksiyonunu örnekler yardımı ile gerçekleştiren, bulundukları ortama uyum sağlayan, adaptif, eksik bilgi ile

çalışabilen, belirsizlikler altında karar verebilen ve hatalara karşı toleranslı hesaplamalar gerçekleştirebilen bilgisayar sistemleridir (Öztemel, 2012). Yapay Sinir Ağları, beynin herhangi bir görevi yerine getirme yönteminin modellenmesi için tasarlanmış bir sistemdir (Ataseven, 2013).

Biyolojik sinir hücrelerinin diğer bir adıyla nöronların (*neuron*) birbirleriyle bağlantılar kurması ve etkileşime geçmesi ile biyolojik sinir sistemi oluşmaktadır (Alp, 2019). Şekil 2.19’da bir biyolojik sinir hücresinin yapısı sunulmuştur. Bir insanda birbiriyle bağlantılı yaklaşık 100 milyar nöron vardır ve görevleri sinirsel uyarıları elektriksel sinyaller biçiminde iletterek vücuttan beyne, beyinden vücuda ve beynin kendi içinde bilgi taşımaktır (TÜBİTAK, 2014).



Şekil 2.19: Biyolojik sinir hücresi (Jarosz, 2009).

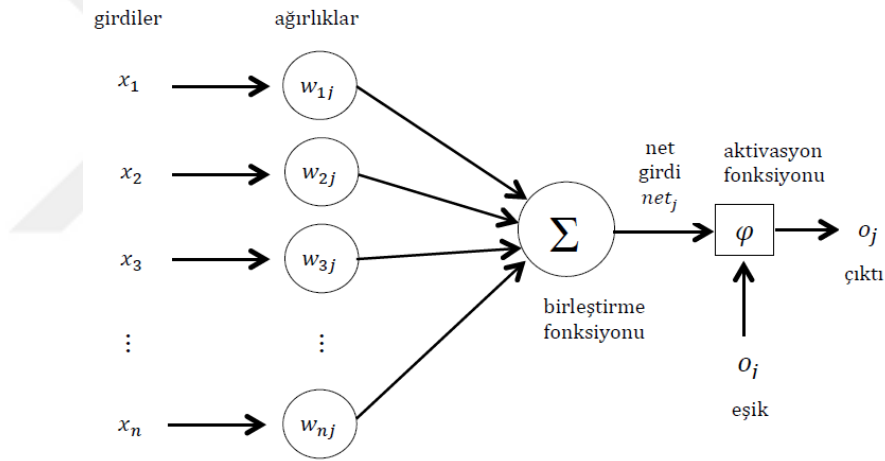
Nöronlar işlevleri bakımından duyuşal nöron, motor nöron ve aracı nöron olmak üzere üç sınıfa ayrılmaktadır:

- **Duyuşal Nöron:** Duyulardan merkezi sinir sistemine (*beyne*) bilgi taşıyan nöronlardır,
- **Motor Nöron:** Beyinden genel bilgileri kaslara ve dokulara ileten nöronlardır,
- **Aracı Nöron:** Duyuşal nöronlardan motor nöronlara ya da etrafındaki diğer aracı nöronlara mesaj ileten beyin hücreleridir (Ankara Üniversitesi, 2019).

Ersoy ve Karal (2012), insanda öğrenmenin nöronlar arasındaki sinaptik (*synaptic*) bağlantıların ayarlanması ile gerçekleştiğini belirtmiş ve aşağıdaki basit anlatımın nöronlar için model oluşturduğunu ifade etmişlerdir. Buna göre;

- Öncelikle nöronlar çok sayıda dendritleri aracılığıyla harekete geçirici (*tetikleyici*) veya yasaklayıcı giriş uyarısını alır.
- Alınan bu girişler toplanır ve nöron gövdesine yerleştirilir.
- Bu giriş, belirli bir eşik değerini aştığı zaman, hücreden diğer hücrelere aksonu aracılığıyla bir etki iletilir.

Biyolojik sinir hücrelerine benzer şekilde, yapay sinir hücreleri dışarıdan gelen bilgileri bir toplama fonksiyonu ile toplar ve aktivasyon fonksiyonundan geçirerek çıktıyı üretip ağırlık bağlantılarının üzerinden diğer hücrelere (*proses elemanlarına*) göndermektedir (Öztemel, 2012). Yapay sinir hücresinin girdiler, ağırlıklar, toplama işlevi, aktivasyon işlevi ve çıktılar olmak üzere 5 temel ögesi bulunmaktadır (Krenker vd., 2011). Şekil 2.20 ile bir yapay sinir hücresine ait mimari gösterilmiştir.



Şekil 2.20: Yapay sinir hücresi (Atalay ve Çelik, 2017).

Nöron, dış dünyadan ya da bir başka nöronun gelen girdileri alır. Her girdiye ait bir ağırlık değeri bulunmaktadır. Nöronun girdileri ve bu girdilere ait ağırlıklar çarpım, maksimum, minimum, çoğunluk, kümülatif toplam gibi toplama fonksiyonu yardımı ile nöronun net girdisi bulunmuş olunur. Hücrenin net girdisi kullanılacak bir aktivasyon fonksiyonu ile çıktıya dönüştürülür. Üretilen çıktı hücre girdisinde olduğu gibi ya dış dünyaya ya da bir başka nörona iletilir (Özen, 2016; Öztemel, 2012).

Yapay sinir hücreleri üç katman halinde (*girdi katmanı (input layer)*, *ara (gizli) katmanlar (hidden layer)*, *çıkıtı katmanı (output layer)*) ve her katman içinde paralel olarak bir araya gelerek yapay sinir ağını oluştururlar (Öztemel, 2012). Girdi katmanı ağına gelen girdileri ara

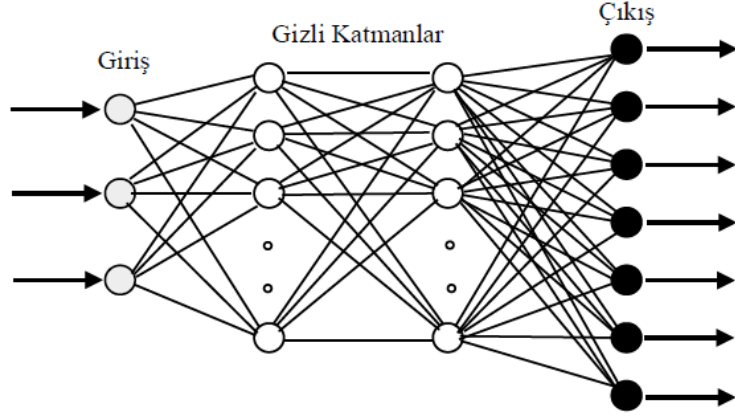
katman(lar)a iletir. Ađın yapısına bađlı olarak bir ađda birden ok ara katman bulunabilir. Ara katmana gelen girdiler iřlenerek ıktı katmanına iletilir. ıktı katmanını ara katman(lar)dan gelen bilgiyi iřleyerek ađın ıktısını dıř dnyaya iletir (zen, 2016). te yandan, bir ađın eđitimine bařlamadan nce, girdi katmanında ka adet nron bulunacađı, ara katmanda ka adet nron bulunacađı (*eđer ara katman sayısı birden fazla ise her bir ara katmanda ka adet nron bulunacađı*) ve ıktı katmanında ka adet nron bulunacađı belirlenerek, ađın topolojisine karar verilmelidir. Ek olarak, gizli katmanın en iyi olacak řekilde ka adet nrondan oluřacađına dair kesin bir kural bulunmamaktadır (Han vd., 2012).

Yapay Sinir Ađları, katman sayısına ve ađın alıřma yapısına gre Tablo 2.2’de belirtildiđi řekilde sınıflandırılabilir.

Tablo 2.2: Yapay Sinir Ađlarının sınıflandırılması (Alp, 2019).

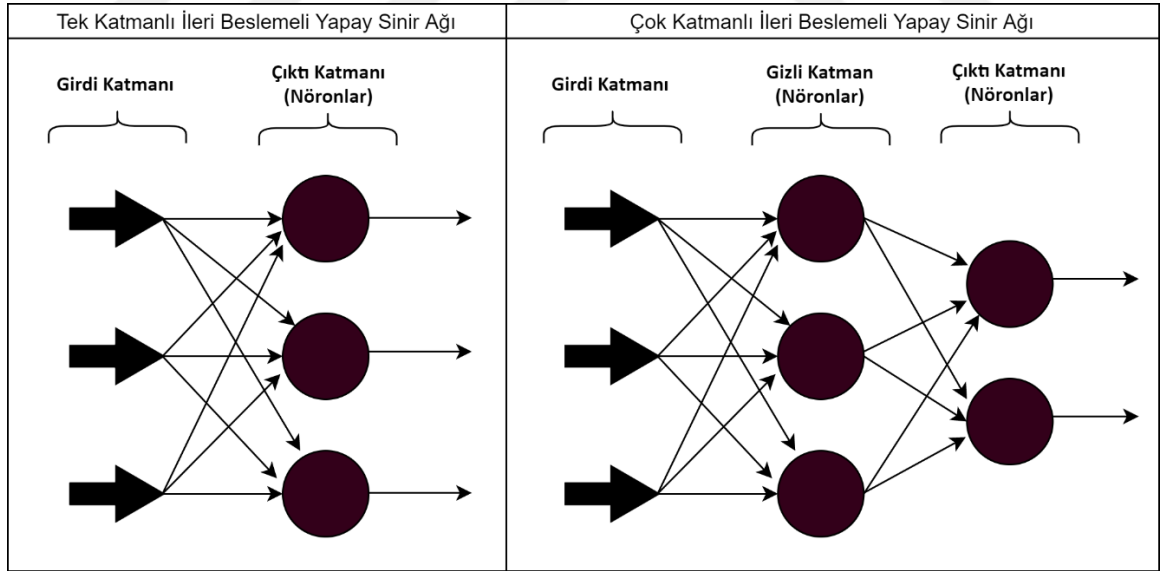
Sınıflandırma Tr	Sınıf
Katman Sayısına Gre	Tek Katmanlı Yapay Sinir Ađları • Perceptron • ADALINE/MADALINE
	ok Katmanlı Yapay Sinir Ađları
Ađın alıřma Yapısına Gre	İleri Beslemeli Yapay Sinir Ađları
	Geri Beslemeli/Yayılmalı Yapay Sinir Ađları

Alp (2019) Tek Katmanlı Yapay Sinir Ađlarının yalnızca girdi ve ıktı katmanından oluřtuđunu ve gizli katman(lar)ın bulunmadıđını, ok Katmanlı Yapay Sinir Ađlarının ise girdi, ara (*gizli*) ve ıktı katmanlarından oluřtuđunu belirtmiřtir. řekil 2.21 ile ok Katmanlı Bir Yapay Sinir Ađına ait mimari sunulmuřtur.



Şekil 2.21: Çok katmanlı Yapay Sinir Ağı (Atalay ve Çelik, 2017).

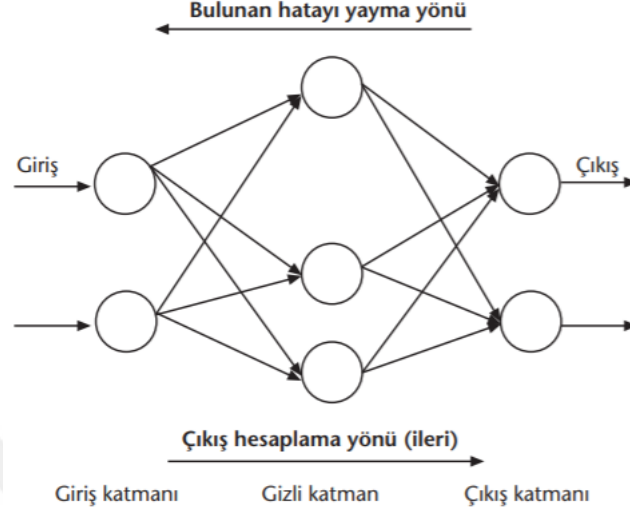
Sinyallerin iletim yönüne bağlı olarak yapay sinir ağları, ileri beslemeli (*feed forward*) ve geri yayımlı olmak üzere ikiye ayrılır. Sinyal iletimi; ileri beslemeli ağlarda bir katmandaki nöronlardan, bir sonraki katmandaki nöronlara doğru yapılır ve aynı katmandaki nöronlar arası bağlantı bulunmaz. Şekil 2.22 ile Tek Katmanlı İleri Beslemeli Yapay Sinir Ağı (*Single-Layer Feedforward Neural Network*) ve Çok Katmanlı İleri Beslemeli Yapay Sinir Ağı (*Multi-Layer Feedforward Neural Network*)’na ait mimari sunulmuştur.



Şekil 2.22: Tek ve çok katmanlı ileri beslemeli Yapay Sinir Ağları (Turkson vd., 2016).

Geri yayımlı (*backpropagation*) ağlarda ise, iki yönlü (*ileri ve geri*) olarak da mümkündür ve bir çıkış sinyali, giriş ve çıkış değerlerini dinamik bir şekilde kontrol edebilmektedir (Demirhan vd., 2010). Geri yayılma algoritması; girdi sinyallerinin ileri yayılması, hata sinyalinin (*ağın tahmini ve hedef arasındaki fark*) hesaplanması ve birkaç eğitim iterasyonu (*epochs - devir*)

için hata sinyali minimize edilene kadar hatanın geri yayılmasından oluşmaktadır (Turkson vd., 2016). Şekil 2.23 ile geri yayılımlı bir yapay sinir ağına ait mimari sunulmuştur.



Şekil 2.23: Geri yayılımlı Yapay Sinir Ağı (Demirhan vd., 2010).

Çok sayıda Yapay Sinir Ağı çeşidi ve Yapay Sinir Ağı Algoritması olmakla beraber, bunlar içerisinde en popüler olan algoritma 1980’li yıllarda itibar kazanan Geri Yayılım Algoritması (*Backpropagation Algorithm*)’dır. Geri yayılım algoritması, bir Yapay Sinir Ağı öğrenme algoritmasıdır. Bu algorithmada öğrenme, veri setindeki eğitim verisinin yinelemeli olarak işlenmesi ve her bir örnek için ağı tahminini bilinen hedef değer ile karşılaştırması ile gerçekleşmektedir (Han vd., 2012). Ağı çıktığı değerleri (y_l) ile hedef değerleri (t_l) arasındaki karesel farkların toplamı hata fonksiyonunu vermektedir ve kare hata formülü ile hesaplanmaktadır (Lillicrap vd., 2020):

$$E = \frac{1}{2} \sum_l (y_l - t_l)^2 \quad (2.32)$$

Ağı tahmini ve hedef değer arasındaki ortalama hatalar karesini minimize etmek adına her bir örnek için ağırlıklar (*nörondan nörona olan bağlantıların gücünü belirleyen ağ parametreleri*) yenilenir. Burada yenilemenin yönü, çıktı katmanından ilk gizli katmana olacak şekilde geriye (*backward*) doğrudur (Han vd., 2012).

η öğrenme katsayısı (*learning rate*) ve w_{ij} i nöronundan j nöronuna olan ağırlık değeri olmak üzere; denklem 2.18’de verildiği şekilde her bir w_{ij} ağırlığı eğim azaltma (*gradient decent*) metodu kullanılarak güncellenir (Lillicrap vd., 2020):

$$\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}} \quad (2.33)$$

n .inci öğrenme adımında w_{ij} ağırlığı $w_{ij}(n)$ ve ağırlığındaki değişim miktarı $\Delta w_{ij}(n)$ olmak üzere; w_{ij} ağırlığının güncellenmiş hali denklem 2.19'daki gibi hesaplanmaktadır (Schiffmann vd., 1994):

$$w_{ij}(n + 1) = \Delta w_{ij}(n) + w_{ij}(n) \quad (2.34)$$

2.3.4.6. Destek Vektör Makineleri

Koçoğlu vd. (2018) Destek Vektör Makineleri (DVM)'nin temellerinin Vladimir Vapnik ve Alexey Chervonenkis tarafından Vapnik-Chervonenkis teorisi kapsamında ortaya atıldığını ve 1992 yılında Bernhard Boser, Isabelle Guyon ve Vladimir Vapnik tarafından geliştirildiğini belirterek birçok alanda başarıyla uygulanabilen bir yöntem olduğunu ifade etmiştir. Gürsakal (2018) Destek Vektör Makineleri'nin sınıflandırma ve regresyon için uygulanabilecek bir danışmanlı öğrenme algoritması olduğunu ve sıklıkla sınıflandırma amacıyla kullanıldığını belirtmiştir.

Destek Vektör Makineleri ile sınıflandırma, verinin lineer olarak ayrılabilme (*linearly separable*) ve lineer olarak ayrılamama (*linearly inseparable*) durumuna göre ikiye ayrılır (Han vd., 2012).

D adet niteliğe sahip (*yani D-boyutlu*) her bir x_i girdisi, $y_i = -1$ ya da $y_i = +1$ sınıflarından birine ait olmak üzere, L adet eğitim noktası olsun. Veri setindeki örneklerin lineer olarak birbirinden ayrıldığı kabulü altında eğitim verisi;

$$\{x_i, y_i\} \text{ öyle ki } i = 1, \dots, L, y_i \in \{-1, 1\}, x \in \mathbb{R}^D \quad (2.35)$$

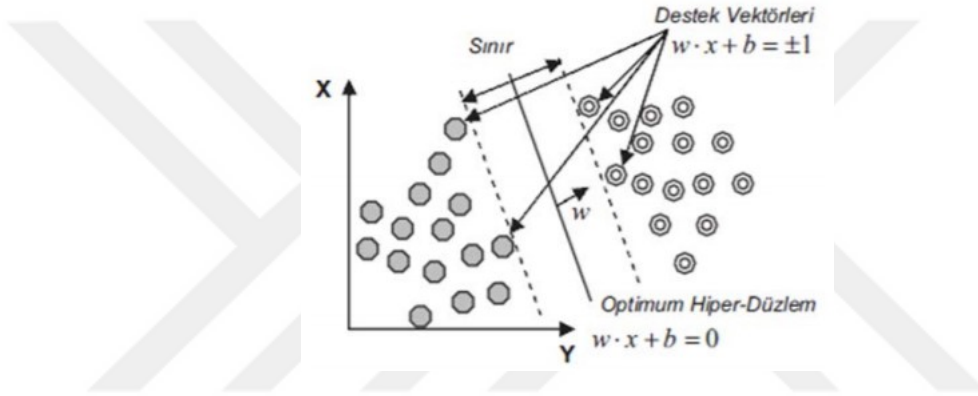
olarak ifade edilebilir (Fletcher, 2008).

İki sınıf arasındaki karar sınırı (*decision boundary*) iki boyutlu uzayda doğru (*separating line*), üç boyutlu uzayda düzlem (*separating plane*), üçten daha fazla boyutlu uzaylarda ise hiperdüzlem (*separating hyperplane*)'dir (Han vd., 2012).

w hiperdüzlemin normali ve $\frac{-b}{\|w\|}$ hiperdüzlemden orijine olan dik uzaklık olmak üzere bu hiperdüzlem denklem 2.21'deki gibi ifade edilebilir (Fletcher, 2008):

$$w \cdot x + b = 0 \quad (2.36)$$

Destek Vektör Makineleri, sınıflar arasındaki en iyi karar sınırını (*optimum hiperdüzlemi*) belirlemek için en büyük marjine (*margin*) sahip hiperdüzlemi (*maximum marginal hyperplane* - *MMH*) aramaktadır. Marjin, hiperdüzlemden her iki sınıfa olan en kısa uzaktır (Han vd., 2012). Basit vektör geometrisine göre marjin, $\frac{1}{\|w\|}$ 'ya eşittir (Fletcher, 2008).



Şekil 2.24: Lineer olarak ayrılabilen durumlar için Destek Vektör Makineleri (Küçüksille ve Ateş, 2013)

Destek vektör makinelerinin uygulanması, w ve b 'nin seçimidir, buna göre eğitim veri seti denklem 2.37 ve 2.38'deki şekilde yazılabilir (Fletcher, 2008):

$$y_i = +1 \text{ için } x_i \cdot w + b \geq +1 \quad (2.37)$$

$$y_i = -1 \text{ için } x_i \cdot w + b \leq -1 \quad (2.38)$$

Denklem 2.37 ve 2.38 birleştirilerek denklem 2.39 yazılabilir (Fletcher, 2008):

$$y_i(x_i \cdot w + b) - 1 \geq 0 \quad (2.39)$$

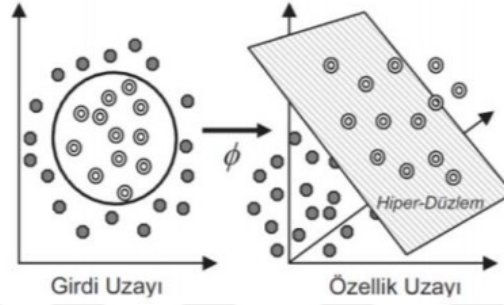
Ayrımcı hiperdüzleme her iki sınıftan en yakın olan nokta(lar)ı içeren düzlemler denklem 2.40 ve denklem 2.41'deki şekilde tanımlanabilir (Fletcher, 2008):

$$H_1: x_i \cdot w + b = +1 \quad (2.40)$$

$$H_2: x_i \cdot w + b = -1 \quad (2.41)$$

H_1 ve H_2 hiperdüzlemlerinin üzerinde kalan eğitim örnekleri destek vektörleri (*support vectors*) olarak adlandırılır (Han vd., 2012).

Verinin doğrusal ayrılamadığı durumda, Şekil 2.25'deki gibi girdi uzayı kernel (çekirdek) fonksiyonları (ϕ) aracılığıyla daha yüksek boyutlu bir uzaya taşınır ve burada lineer ayırma işlemleri uygulanır (Küçüksille ve Ateş, 2013).



Şekil 2.25: Lineer olarak ayrılamayan durumlar için Destek Vektör Makineleri (Küçüksille ve Ateş, 2013)

$k(x_i, x_j) = x_i^T x_j$ lineer (*linear*) kernel olarak bilinmektedir ve bunun dışında a ve b kernel davranışı olmak üzere diğer kernel fonksiyonları şunlardır (Fletcher, 2008):

Radyal tabanlı (*Radial Basis*) kernel:

$$k(x_i, x_j) = e^{-\left(\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)} \quad (2.42)$$

Polinomial (*Polynomial*) kernel:

$$k(x_i, x_j) = (x_i \cdot x_j + a)^b \quad (2.43)$$

Sigmoid (*Sigmoidal*) kernel:

$$k(x_i, x_j) = \tan(ax_i \cdot x_j - b) \quad (2.44)$$

2.3.5. Model Değerlendirme ve Seçimi (*Evaluation*)

Bu adımda, daha önce modelleme adımında kurulan modelin/modellerin performansı ortaya konmaktadır. Model performans değerlendirmesi için literatürde geliştirilen çok sayıda farklı

yöntem ve ölçüt bulunmaktadır. Bu nedenle bu başlık aşağıdaki şekilde iki alt başlıkta incelenecektir.

2.3.5.1. Model Performans Değerlendirme Yöntemleri

Bir modelin performansını belirlemek ve diğer modeller ile karşılaştırmak için model performansı değerlendirilir. Literatürde birçok performans değerlendirme yöntemi mevcuttur. Bu yöntemlerden bazıları şöyle sıralanabilir (Balaban ve Kartal, 2015):

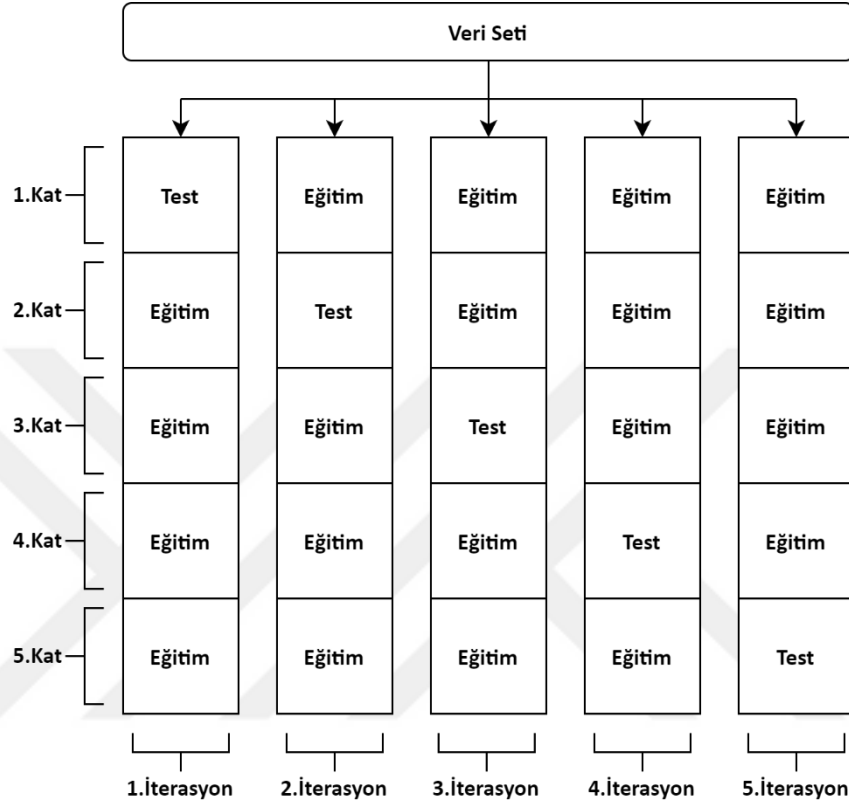
- Hold-out (*Dışarıda tutma*),
- Tekrarlı Holdout (*Repeated Holdout*),
- Tabakalı Örneklem (*Stratified Sampling*),
- Üçlü Ayırma (*Three-way Split*),
- Çapraz Geçerleme (*Cross Validation*),
- Bootstrap Örneklem (*Bootstrap Sampling*).

Hold-out yöntemi, verinin rastgele olarak eğitim ve test şeklinde iki parçaya ayrıldığı yöntemdir. Genellikle verinin üçte ikisi eğitim, üçte biri test olarak ayrılır ve eğitim verisi kullanılarak üretilen modelin doğruluğu test verisi ile değerlendirilir. Hold-out yönteminin k kez tekrarlanması **tekrarlı hold-out** (*repeated hold-out veya random subsampling*) olarak adlandırılmaktadır. **Tabakalı örneklem** ise, verinin aynı sınıfa ait veriler bir arada olacak şekilde “tabaka” adı verilen alt gruplarda toplandığı ve bu alt gruplardan rastgele seçilen örneklemelerin birleşmesiyle oluşan örneklem yöntemi. **Bootstrap** yöntemi ise, verilen eğitim demetlerini homojen bir biçimde değiştirerek örneklem yapmaktadır. Yani, seçilen bir demetin tekrar seçilmesi ve eğitim veri setine eklenmesi olasıdır (Han vd., 2012).

Üçlü ayırma yöntemi ile veri seti eğitim, test ve doğrulama olmak üzere üçe ayrılmaktadır. Eğitim-doğrulama çifti model parametrelerinin ayarlanması ve model seçimi için kullanılırken, bağımsız olarak bir kenarda tutulan test veri seti model değerlendirmesi için kullanılmaktadır (Raschka, 2018).

Çapraz geçerleme yönteminde veri seti rastgele olarak k tane yaklaşık olarak eşit büyüklükte olan ve “kat - fold” adı verilen ayrık alt kümeler ayrılmaktadır (Han vd., 2012). Bir kat modelin testi, kalan katlar modelin eğitimi için kullanılmaktadır ve tüm katlar mutlaka bir kez test veri seti olacak şekilde k defa yani kat sayısı kadar analizler tekrarlanmaktadır. Modelin

performansı, her bir kat için elde edilen performans ölçülerinin ortalaması alınarak bulunmaktadır (Koçoğlu vd., 2018). Bu çalışmada, 5-kat çapraz geçerleme yöntemi kullanılmıştır. Şekil 2.26 ile 5-kat çapraz geçerlemenin yapısı sunulmuştur.



Şekil 2.26: 5-kat çapraz geçerleme (Koçoğlu vd., 2018).

2.3.5.2. Model Performans Değerlendirme Ölçütleri

Genel olarak, sınıflandırma problemlerindeki değerlendirme ölçütleri, her sınıf için doğru ve yanlış sınıflandırılan örnek sayılarının bulunduğu karışıklık matrisi (*confusion matrix*) adlı bir matristen belirlenir (Costa vd., 2007). Bu matrise kontenjans tablosu (*contingency table*) da denmektedir (Balaban ve Kartal, 2015).

Kontenjans tablosu, model performansını değerlendirmek amacıyla kullanılmaktadır ve bir modelin performansının tam bir görüntüsünü sunmaktadır (Sun vd., 2020). Kontenjans tablosu dört bileşenden oluşmaktadır (Sun vd., 2020; Sokolova ve Lapalme, 2009):

- **Doğru Pozitif (TP - True Positives):** Pozitif olarak doğru tahmin edilmiş pozitifler,
- **Doğru Negatif (TN – True Negatives):** Negatif olarak doğru tahmin edilmiş negatifler,

- **Yanlış Pozitif** ($FP - False Positives$): Pozitif olarak yanlış tahmin edilmiş negatifler,
- **Yanlış Negatif** ($FN - False Negatives$): Negatif olarak yanlış tahmin edilmiş pozitifler.

Tablo 2.3 ile bir ikili sınıflandırma problemi (*pozitif ve negatif olmak üzere iki sınıfa sahip*) için kontenjans tablosu verilmiştir.

Tablo 2.3: Kontenjans tablosu (Costa vd., 2007; Akpınar, 2014).

Tahmini Sınıf	Gerçek Sınıf		Değerlendirme Ölçüleri
	Pozitif	Negatif	
Pozitif	Doğru Pozitif (dp)	Yanlış Pozitif (yp)	<i>Kesinlik (Precision) (ppv)</i> $\frac{dp}{dp + yp}$
Negatif	Yanlış Negatif (yn)	Doğru Negatif (dn)	<i>Negatif Öngörü Değeri (npv)</i> $\frac{dn}{dn + yn}$
Değerlendirme Ölçüleri	<i>Duyarlılık (Sensitivity) (sen)</i> $\frac{dp}{dp + yn}$	<i>Belirleyicilik (Specificity) (spe)</i> $\frac{dn}{dn + yp}$	<i>Doğruluk (Accuracy) (acc)</i> $\frac{dp + dn}{dp + yp + yn + dn}$

Tablo 2.3'te yer alan dp , yp , yn ve dn kavramları şu şekilde açıklanmaktadır:

- **dp** : Gerçekte pozitif, algoritma/model tarafından pozitif tahmin edilmiş,
- **yn** : Gerçekte pozitif, ancak algoritma/model tarafından negatif şeklinde tahmin edilmiş,
- **yp** : Gerçekte negatif, ancak algoritma/model tarafından pozitif tahmin edilmiş,
- **dn** : Gerçekte negatif, algoritma/model tarafından da negatif olarak tahmin edilmiş (Costa vd., 2007).

Kontenjans tablosuna göre; doğruluk, hata oranı, duyarlılık, belirleyicilik, kesinlik, negatif öngörü değeri ve f -ölçüsü model performans değerlendirme ölçütleri olarak hesaplanmaktadır ve aşağıdaki şekilde ifade edilmektedir (Balaban ve Kartal, 2015):

- **Doğruluk (Accuracy - ACC)**: Bir modelin doğruluğu; $m = dp + yp + yn + dn$ olmak üzere;

$$ACC = \frac{dp + dn}{m} \quad (2.45)$$

olarak hesaplanmaktadır.

- **Hata Oranı** (*Error rate – ERR*): Bir modelin hata oranı,

$$ERR = 1 - ACC \quad (2.46)$$

olarak hesaplanmaktadır.

- **Duyarlılık** (*Sensitivity, True Positive Rate veya Recall - TPR*): Doğru sınıflandırılan pozitif örneklerin toplam pozitif örnek sayısına oranıdır.

$$TPR = \frac{dp}{dp + yn} \quad (2.47)$$

- **Belirleyicilik** (*Specificity, True Negative Rate – SPC veya TNR*): Doğru sınıflandırılan negatif örneklerin toplam negatif örnek sayısına oranıdır.

$$TNR = \frac{dn}{dn + yp} \quad (2.48)$$

- **Kesinlik** (*Precision/Positive Predictive Value - PPV*): Doğru sınıflandırılan pozitif örneklerin toplam pozitif tahmin edilen örneklere oranıdır.

$$PPV = \frac{dp}{dp + yp} \quad (2.49)$$

- **Negatif Öngörü Değeri** (*Negative Predictive Value - NPV*): Doğru sınıflandırılan negatif sınıf etiketine sahip örneklerin toplam negatif tahmin edilen örneklere oranıdır.

$$NPV = \frac{dn}{dn + yn} \quad (2.50)$$

- **F-ölçüsü** (*F-measure*): Kesinlik ve duyarlılık performans ölçütlerinin harmonik ortalamasıdır.

$$F = \frac{2 * PPV * TPR}{PPV + TPR} \quad (2.51)$$

Regresyon amacıyla kullanılan algoritmaların model performansının değerlendirilmesi için hesaplanan bazı performans ölçütleri aşağıda sunulmuştur (Hyndman ve Koehler, 2006; Botchkarev, 2018; Fernandes, 2020; Touzani vd., 2018):

A_j – Gerçek değerler,

P_j – Tahmin değerleri,

$e_j = A_j - P_j$ (Hata),

n – veri setinin büyüklüğü,

olmak üzere;

- **Ortalama Hata** (*Mean Error - ME*):

$$ME = \frac{1}{n} \sum_{j=1}^n e_j \quad (2.52)$$

- **Ortalama Mutlak Hata** (*Mean Absolute Error – MAE*):

$$MAE = \frac{1}{n} \sum_{j=1}^n |e_j| \quad (2.53)$$

- **Ortalama Kare Hata** (*Mean Square Error – MSE*):

$$MSE = \frac{1}{n} \sum_{j=1}^n e_j^2 \quad (2.54)$$

- **Kök Ortalama Kare Hata** (*Root Mean Square Error – RMSE*):

$$RMSE = \sqrt{\frac{\sum_{j=1}^n e_j^2}{n}} \quad (2.55)$$

veya

$$RMSE = \sqrt{MSE} \quad (2.56)$$

- **Ortalama Yüzde Hata** (*Mean Percentage Error – MPE*):

$$MPE = \frac{100}{n} \sum_{j=1}^n \frac{e_j}{A_j} \quad (2.57)$$

- **Ortalama Mutlak Yüzde Hata** (*Mean Absolute Percentage Error – MAPE*):

$$MAPE = \frac{100}{n} \sum_{j=1}^n \frac{|e_j|}{|A_j|} \quad (2.58)$$

- **R-Kare** (*R-Squared - R²*)

R^2 , regresyonun öngörü gücünü ifade etmektedir. Belirleme (*determinasyon*) katsayısı olarak da bilinmektedir ve denklem 2.59’da verildiği şekilde hesaplanmaktadır:

$$R^2 = \left[1 - \frac{\frac{1}{n} \sum_{j=1}^n (A_j - P_j)^2}{var(A)} \right] \times 100 \quad (2.59)$$

R^2 değeri 0 ile 1 arasında değişmektedir ve R^2 değerinin yüksek olması model uyumunun iyi olduğunu göstermektedir.

2.3.6. Modelin Uygulamaya Geçirilmesi (*Deployment*)

Bir modelin sonuçları doğrudan bir veritabanına yazılabilir veya veritabanından kayıtlar çıkarılarak uygulanabilir ve model yeni durumlara uygulanmak istendiğinde, modeli bir programa dâhil etmek için API (*Application Programming Interface - Uygulama Programlama Arayüzü*) veya veri madenciliği programı ile üretilmiş kod kullanılması gereklidir (Edelstein, 1999). Örneğin; RStudio (2021) ortamında üretilen veri madenciliği model(*ler*)i için Chang vd. (2021) tarafından geliştirilen Shiny R paketi ile etkileşimli web uygulamaları oluşturulabilmektedir. Shiny, doğrudan RStudio programı kullanılarak etkileşimli web uygulamaları oluşturmayı kolaylaştıran bir R paketidir (RStudio, 2021b). Oluşturulan bir Shiny uygulamasının çevrimiçi olarak paylaşılabilmesi için ücretsiz/ücretli barındırma hizmeti sunan shinyapps.io kullanılabilir (RStudio, 2021c).

3. MALZEME VE YÖNTEM

Tezin malzeme ve yöntem bölümünde; çalışmada kullanılan veri setinin toplanma sürecinden, analizlerde yer alan niteliklerden, veri ön işleme adımlarından bahsedilmiştir. Tezde izlenen süreç Bölüm 2.3'te de belirtildiği üzere Veri Madenciliği için Çapraz Endüstri Standart Süreç Modeli (*CRISP-DM*) adımları çerçevesinde açıklanmıştır.

3.1.PROBLEMİN TANIMLANMASI

Bu tez çalışmasında esas olarak “makine öğrenmesi teknikleri ile havacılıkta yakıt taşımacılığı kararının alınması” probleminin ele alınmasına karar verilmiştir. Probleme ilişkin tüm detaylar tezin giriş ve literatür taraması bölümlerinde verilmiştir. Bir sonraki bölümde analizlerde kullanılan veri seti hakkında bilgiler yer almaktadır.

3.1.1. Sınırlılıklar

Bu çalışmada, çoklu yakıt taşımacılığı (*birden fazla uçuş için alınan ekstra yakıt neticesinde yapılan yakıt taşımacılığı*) durumu ele alınmamıştır. Çalışmanın MALZEME VE YÖNTEM kısmındaki analizlerde kullanılan veri seti, Bölüm 2.3.2. Veriyi Anlama (*Data Understanding*) açıklanan şekliyle sınırlandırılmıştır. Analizlerde kullanılan algoritmalar; k-En Yakın Komşu algoritması, Basit Bayes Sınıflandırıcı, Sınıflandırma ve Regresyon Ağaçları, Rastgele Orman algoritması, Yapay Sinir Ağları ve Destek Vektör Makineleri ile sınırlandırılmıştır.

3.2. VERİYİ ANLAMA

Bu tezde analizlerde kullanılan veri seti, Türkiye’de hava taşımacılığı yapan ticari bir hava yolu firmasından temin edilmiştir. Veri seti, ilgili şirketin uçuş operasyonunun yönetiminin gerçekleştirildiği Operasyon Kontrol Merkezi (*Operations Control Center-OCC*) departmanında lisanslı havacılık personeli olan uçuş harekât uzmanları (*flight dispatcher*) tarafından kullanılan ve bir uçuşa ait tüm planlamanın yapıldığı uçuş planlama sisteminden toplanan uçuşlar ile oluşturulmuştur.

Veri seti 2017 ve 2018 yılları içerisinde yer alan toplam 7 farklı aya ve 4 farklı mevsime ait gerçekleşmiş 1130 uçuştan oluşmaktadır. Veri setinde yer alan her bir uçuşa ait veri, ilgili uçuş öncesi (*pre-flight*) planlama aşamasında alınmıştır. Geliştirilecek makine öğrenmesi

modellerinin verdiği kararın belirli mevsimlerle sınırlı kalmasını engellemek için uçuşların her mevsimden de alınmasına dikkat edilmiştir. Veri setinde yer alan uçuşlar Boeing 737-800 ve Airbus A320 (*A320* ve *A320neo*) tipli uçaklarla gerçekleştirilmiştir. İlgili uçakların teknik özellikleri Tablo 3.1’de sunulmuştur. Boeing 737-800 tipi ve Airbus A320 tipi uçaklar dünya genelinde en fazla kullanılmakta olan uçak tiplerindendir.

Tablo 3.1: Boeing 737-800 ve Airbus A320 uçaklarının teknik özellikleri (Boeing, 2019a; Airbus, 2019a; Brady, 2020; Airbus, 2020a, 2020b).

Özellik	Uçak Tipi		
	Boeing 737-800	Airbus A320	Airbus A320neo
Maksimum koltuk kapasitesi	189	180	180
Maksimum kalkış ağırlığı	79002 kg.	78000 kg.	79000 kg.
Maksimum iniş ağırlığı	66349 kg.	66000 kg.	67400 kg.
Yakıt tankı kapasitesi	21000 kg.	27200 lt.	26730 lt.
Maksimum uçuş menzili	4000 nm.	3350 nm.	3400 nm
Uzunluk	39.5 m (129 ft 6 in)	37.57 m (123.27 ft)	37.57 m (123.27 ft)
Kanat genişliği	35.8 m (117 ft 5 in)	35.80 m (117.45 ft)	35.80 m (117.45 ft)
Yükseklik	12.5 m (41 ft 3 in)	11.76 m (38.58 ft)	11.76 m (38.58 ft)
Motor	CFM-56	CFM-56	CFM-56

Veri setinde kullanılan niteliklere ait özellikler Tablo 3.2’de sunulmuştur.

Veri setinde tahmin için kullanılan nitelikler 7 adettir. Tahmin için kullanılan niteliklerden *kalkisMydnYakitFyt* ve *varisMydnYakitFyt* Lindgren ve Brynhagen (2012) tarafından verilen formüle dayandırılarak seçilmiştir. *tplmTrafikYuku*, *sicaklik*, *uzaklik*, *ucusSuresi* ve *seyahatYakiti* veri setinin oluşturulması aşamasında uçuş harekât uzmanlardan uzman görüşü olarak karşılıklı yapılan görüşmeler sonucu seçilmiştir.

- *tplmTrafikYuku*, uçağa yüklenen yüklerinin toplam ağırlığını ifade etmektedir. Bu ağırlık; yolcu ağırlığı, bagaj ağırlığı ve kargo ağırlıklarının toplanması ile bulunmaktadır.
- *sicaklik*, ile ifade edilen tahmin niteliği, varış meydanı hava sıcaklığıdır.
- *uzaklik*, ile ifade edilen tahmin niteliği, bir noktadan diğer noktaya hava koşullarının da dahil olduğu kuş uçuşu en kısa yolun uzunluğudur.

- **ucusSuresi**, bir uçuşta kalkıştan itibaren inişe kadar geçen süreyi ifade etmektedir.
- **seyahatYakiti**, her bir uçuşta kalkış ve iniş arasında harcanan yakıtı ifade etmektedir.
- **kalkisMydnYakitFyt**, uçağın kalkış meydanındaki yakıtın fiyatını ifade etmektedir.
- **varisMydnYakitFyt**, uçağın varış meydanındaki yakıtın fiyatını ifade etmektedir.

Tablo 3.2: Veri setindeki niteliklere ilişkin özellikler.

TAHMİN İÇİN KULLANILAN NİTELİKLER			
Nitelik Adı	Kısaltma	Veri Tipi	Birim
Toplam Trafik Yüğü (Payload)	<i>tplmTrafikYuku</i>	Nümerik	Kilogram (kg)
Sıcaklık (Temperature)	<i>sicaklik</i>	Nümerik	Celcius (°C)
Uzaklık (Air Distance)	<i>uzaklik</i>	Nümerik	Deniz Mili (Nautical Mile - Nm)
Uçuş Süresi (Flight Time)	<i>ucusSuresi</i>	Nümerik	Dakika
Seyahat Yakıtı (Trip Fuel)	<i>seyahatYakiti</i>	Nümerik	Kilogram (kg)
Kalkış Meydanı Yakıt Fiyatı (Departure Fuel Price)	<i>kalkisMydnYakitFyt</i>	Nümerik	Dolar (\$)
Varış Meydanı Yakıt Fiyatı (Destination Fuel Price)	<i>varisMydnYakitFyt</i>	Nümerik	Dolar (\$)
HEDEF NİTELİK			
Nitelik Adı	Kısaltma	Veri Tipi	Kodlama Biçimi
Yakıt Taşımacılığı	<i>yakitTasimaciligiDrm</i>	İkili Kategorik (binary)	0: Yok, 1: Var
	<i>yakitTasimaciligiMktr</i>	Nümerik	Kilogram (kg)

Hedef nitelik olarak alınan **yakitTasimaciligiDrm** niteliği, herhangi bir uçuş için o uçuşta, gelecek uçuş veya uçuşlar için taşınacak ekstra yakıtın uçağa alınıp/alınmayacağı durumunu ifade etmektedir. Uçuşlarda yakıt taşımacılığı durumunun olup/olmadığının genel öngörüsü için (sınıflandırma analizlerinde) ikili kategorik (binary) (0: Yok, 1: Var) biçiminde kodlanmıştır. Burada “0” ilgili uçuşta yakıt taşımacılığının yapılmayacağı durumu, “1” ise yapılacağı durumu ifade etmektedir.

Hedef nitelik olarak alınan **yakitTasimaciligiMktr** niteliği, herhangi bir uçuş için o uçuşta taşınacak kârlılık veren yakıtın miktarıdır. Uçuşlarda yakıt taşımacılığı miktarının doğrudan tahmini için (regresyon analizlerinde) nümerik (kilogram) biçimde kullanılmıştır. Niteliğin 0

değeri alması bir öncekine benzer şekilde uçuşta yakıt taşımacılığının yapılmayacağı durumu göstermektedir. Şekil 3.1 ile veri setinde yer alan niteliklere ait özet bilgi sunulmuştur.

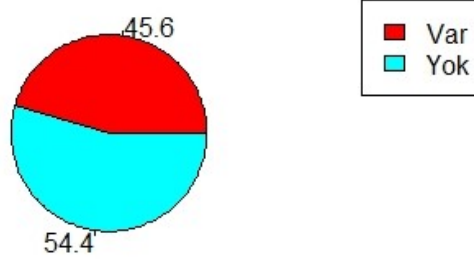
tpImTrafikYuku	sicaklik	uzaklik	ucusSuresi	seyahatYakiti	kalkisMydnYakitFyt	varisMydnYakitFyt
Min. : 6033	Min. : -10.00	Min. : 184.0	Min. : 34.00	Min. : 1280	Min. : 479.7	Min. : 482.8
1st Qu.:14336	1st Qu.: 14.00	1st Qu.: 349.0	1st Qu.: 57.00	1st Qu.: 2486	1st Qu.:516.1	1st Qu.: 542.5
Median :15756	Median : 19.00	Median : 469.0	Median : 73.00	Median : 3097	Median :535.2	Median : 561.5
Mean :15529	Mean : 19.51	Mean : 619.1	Mean : 92.92	Mean : 3963	Mean :535.9	Mean : 590.4
3rd Qu.:16759	3rd Qu.: 26.00	3rd Qu.: 922.8	3rd Qu.:133.00	3rd Qu.: 4984	3rd Qu.:540.5	3rd Qu.: 622.0
Max. :20470	Max. : 43.00	Max. :1828.0	Max. :252.00	Max. :11189	Max. :678.2	Max. :1092.3
yakitTasimaciligiDrm	yakitTasimaciligiMktr					
Var:515	Min. : 0					
Yok:615	1st Qu.: 0					
	Median : 0					
	Mean :1516					
	3rd Qu.:3248					
	Max. :7471					

Şekil 3.1: Veri setinde yer alan niteliklere ait özet bilgi.

3.3. VERİNİN HAZIRLANMASI

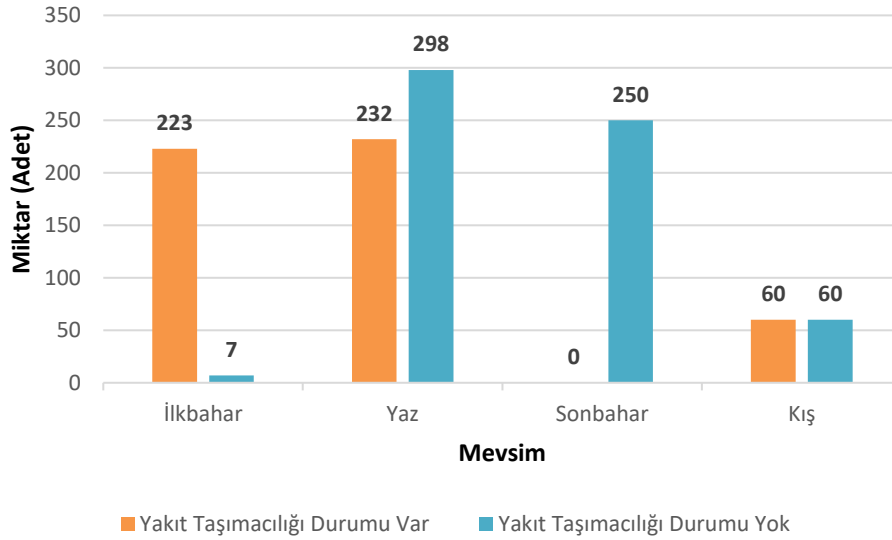
Eurocontrol (2019) uçak performansı, Maksimum Kalkış Ağırlığı (*Maximum Take-Off Weight – MTOW*), Maksimum İniş Ağırlığı (*Maximum Landing Weight – MLW*), yakıt tankı kapasitesi, en az yasal yakıt ikmali miktarı ve bazı hava yollarının anlaşmalı yakıt fiyatlarını hesaba katarak yakıt taşımacılığının ne sıklıkta kullanıldığını tahmin etmek için simülasyonlar gerçekleştirmiştir. Gerçekleştirilen bu simülasyonlar sonucuna göre, her yıl Avrupa Sivil Havacılık Konferansı (*European Civil Aviation Conference - ECAC*) bölgesinde uçuşların %16,5’inde yakıt taşımacılığı, buna karşılık %4,5’inde kısmi yakıt taşımacılığı yapılabilir. Simülasyonlara ek olarak, Eurocontrol tarafından hava yollarından birçok pilot, genel havacılık uçuş harekât uzmanları ve yer hizmetleri temsilcileri ile yürütülen görüşmeler sonucunda uçuşların %15’inde yakıt taşımacılığı, %15’inde kısmi yakıt taşımacılığı gerçekleştirildiği belirtilmiştir.

Eurocontrol (2019) tarafından belirtilen bu duruma paralel olarak veri seti için örneklem seçimi yaparken yakıt taşımacılığının olduğu durumlar diğer uçuşlara oranlara çok daha az olduğu görülmüştür. Bu tez çalışmasında veri toplama aşamasında bu durum göz önünde bulundurulmuş, kullanılan makine öğrenmesi algoritmalarından elde edilecek modellerin performansının olumsuz etkilenmemesi için hedef nitelik olan yakıt taşımacılığı niteliğinin sınıflarının dengeli dağılmasına özen gösterilmiştir. Bu nedenle, kullanılan veri setinde dengesiz veri (*imbalanced data*) durumu söz konusu değildir. Buna göre, Şekil 3.2’den de görüleceği üzere 1130 uçuştan 615 tanesinde (%54,4) *yakitTasimaciligiDrm* hedef niteliği “0” yani yakıt taşımacılığı “Yok”, 515 tanesinde (%45,6) “1” yani “Var” sınıf etiketine sahiptir.



Şekil 3.2: *yakitTasimaciligiDrm* hedef niteliğinin sınıf etiketi frekansları.

yakitTasimaciligiDrm hedef niteliğinin mevsimlere göre dağılımını içeren grafik Şekil 3.3'te sunulmuştur. Buna göre İlkbahar (*Var*: 223, *Yok*:7), Yaz (*Var*:232, *Yok*:298), Sonbahar (*Var*:0, *Yok*:250), Kış (*Var*:60, *Yok*:60) şeklinde dağılım vardır.

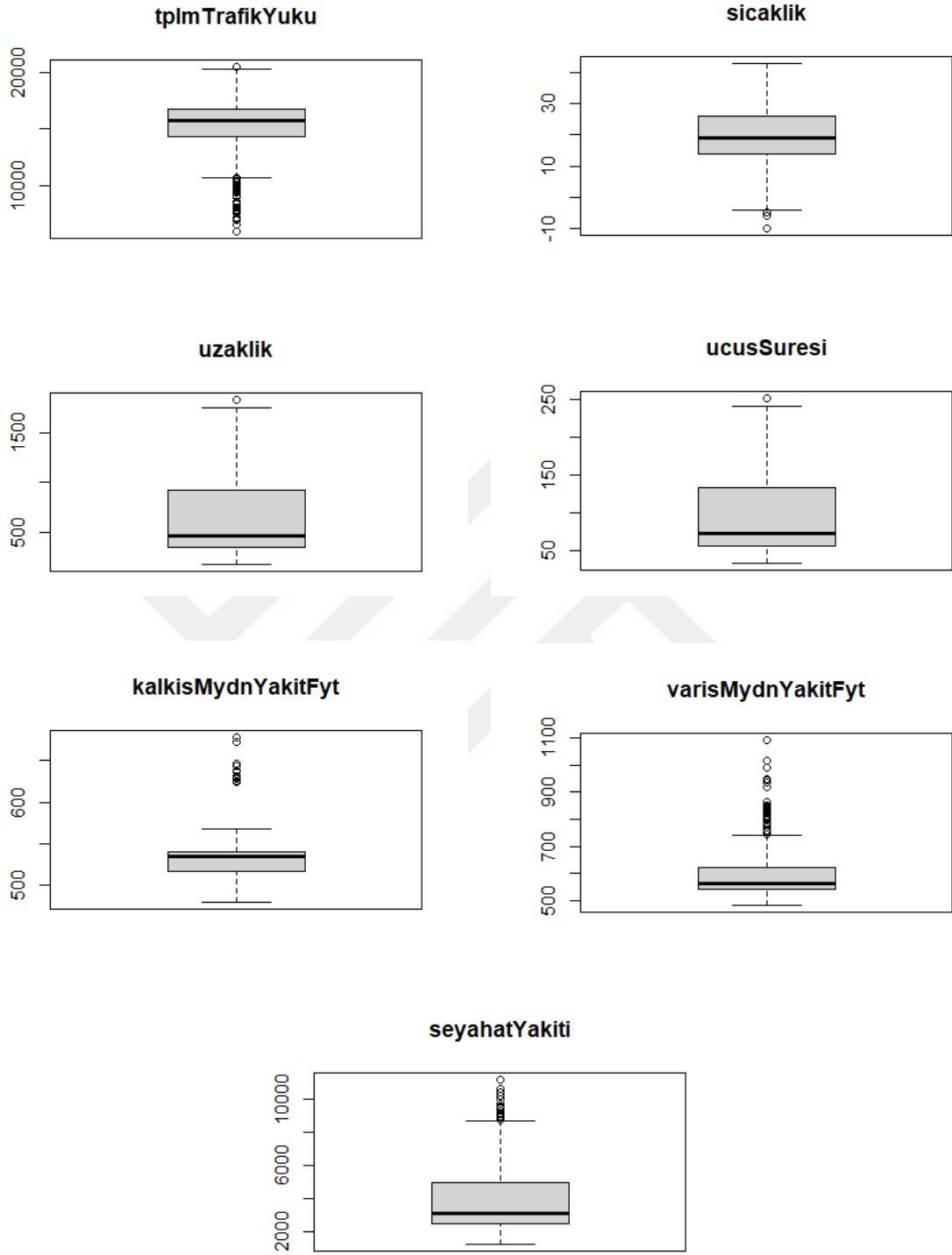


Şekil 3.3: *yakitTasimaciligiDrm* hedef niteliğinin mevsimlere göre dağılım grafiği.

Bölüm 2.3.3'te verinin hazırlanması ile ilgili verilen bilgilerden faydalanılarak veri analizine başlamadan önce analizlerin sonuçlarına etki edebilecek durumları ortadan kaldırarak veri setinin analizlere hazır hale getirilmesi için sırasıyla aşağıdaki işlemler uygulanmıştır.

3.3.1. Aykırı Değerlerin İncelenmesi

Veri setinde yer alan niteliklerdeki olası aykırı değerler kutu grafiği (*whisker-box plot*) çizimleri Şekil 3.4 yardımıyla incelenmiştir.



Şekil 3.4: Veri setinde yer alan niteliklere ait kutu grafiği.

Şekil 3.4 incelendiğinde tüm nitelikler için aykırı değerlerin varlığı tespit edilmiştir. İlgili grafiklerde gözlemlenen aykırı değerler yakıt taşımacılığı durumuna ait analizler için farklı ve önemli durumları temsil ediyor olabileceği düşüncesiyle veri setinde bırakılmıştır.

3.3.2. Eksik Verinin Tamamlanması

Analizler için toplanan uçuş verisi ilgili sistemlerden manuel biçimde tek tek aktarıldığından veri setinde herhangi bir eksik değer mevcut değildir. Tüm veri tamdır.

3.3.3. Tekrar Eden Gözlemlerin Veri Setinden Çıkarılması

Veri setinde tekrar eden gözlemler incelenmiş; ancak tekrar eden herhangi bir uçuş verisine rastlanmamıştır.

3.3.4. Nitelikler Arasındaki İlişkinin İncelenmesi

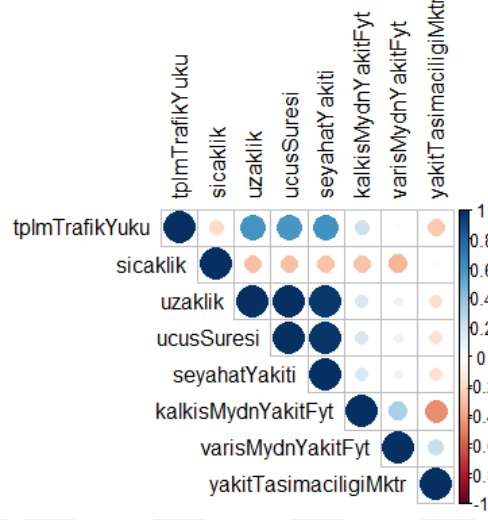
Veri setindeki tahmin nitelikleri arasındaki ilişkiyi incelemek için ilk olarak her bir niteliğin normal dağılım gösterip/göstermediği tespit edilmiştir. Bunun için Shapiro-Wilk normallik testi uygulanmıştır. Tablo 3.3'te verilen sonuçlara göre tüm p değerleri 0,05'ten küçüktür. Bu sonuçlar veri setindeki tahmin niteliklerinin normal dağılıma sahip olmadığını belirtmektedir (SAS Institute Inc., 2017). Bu nedenle, bu nitelikler arasındaki ilişkiyi incelemek için Spearman korelasyon katsayısı hesaplanmıştır (Laerd Statistics, 2013).

Tablo 3.3: Shapiro-Wilk normallik testi sonuçları.

Nitelik	W	p değeri
<i>tplmTrafikYuku</i>	0,96462	5.98e-16
<i>sicaklik</i>	0,98983	4.711e-07
<i>uzaklik</i>	0,83093	< 2.2e-16
<i>ucusSuresi</i>	0,83095	< 2.2e-16
<i>seyahatYakiti</i>	0,84245	< 2.2e-16
<i>kalkisMydnYakitFyt</i>	0,8196	< 2.2e-16
<i>varisMydnYakitFyt</i>	0,79957	< 2.2e-16

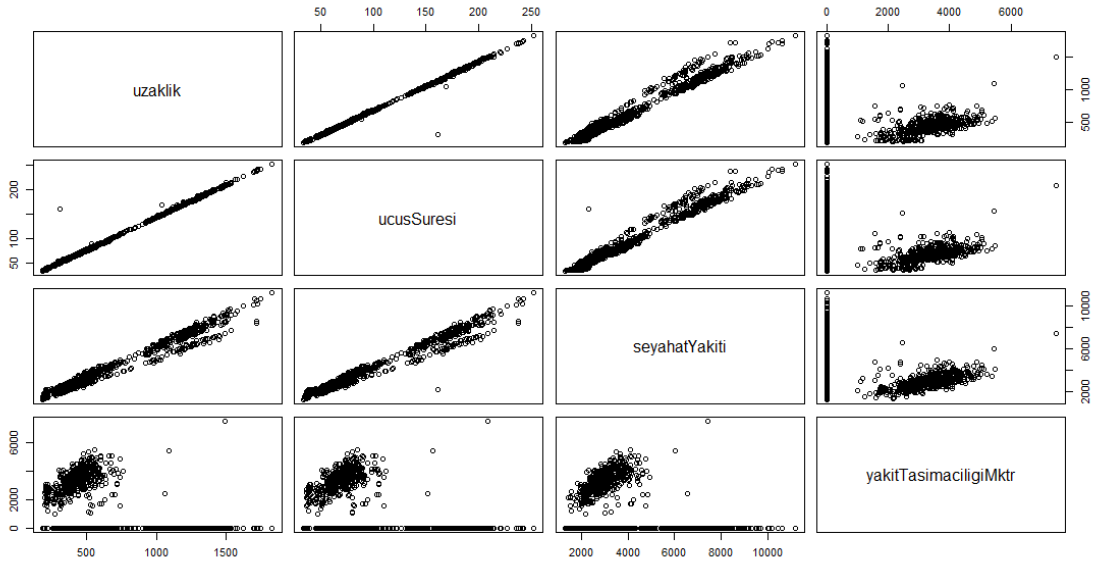
Veri setindeki niteliklerin birbiriyle arasındaki ilişkiyi ve bu ilişkinin derecesini göstermek için Şekil 3.5'teki korelogram verilmiştir (Wei ve Simko, 2016). Mavi renk iki nitelik arasında pozitif, kırmızı ise negatif ilişkinin varlığını temsil etmektedir. Ek olarak, dairelerin boyutu ise ilişkinin kuvvetini göstermektedir. Daha büyük daireler iki değişken arasındaki ilişkinin daha güçlü, daha küçük daireler ise iki değişken arasındaki ilişkinin daha zayıf olduğu anlamına gelir. Pratikte, toplam trafik yükü arttığında/azaldığında seyahat yakıtı artar/azalır. Benzer

şekilde uzaklık arttığında/azaldığında seyahat yakıtı ve uçuş süresi artar/azalır. Bu fikirler Şekil 3.5 tarafından desteklenmektedir.



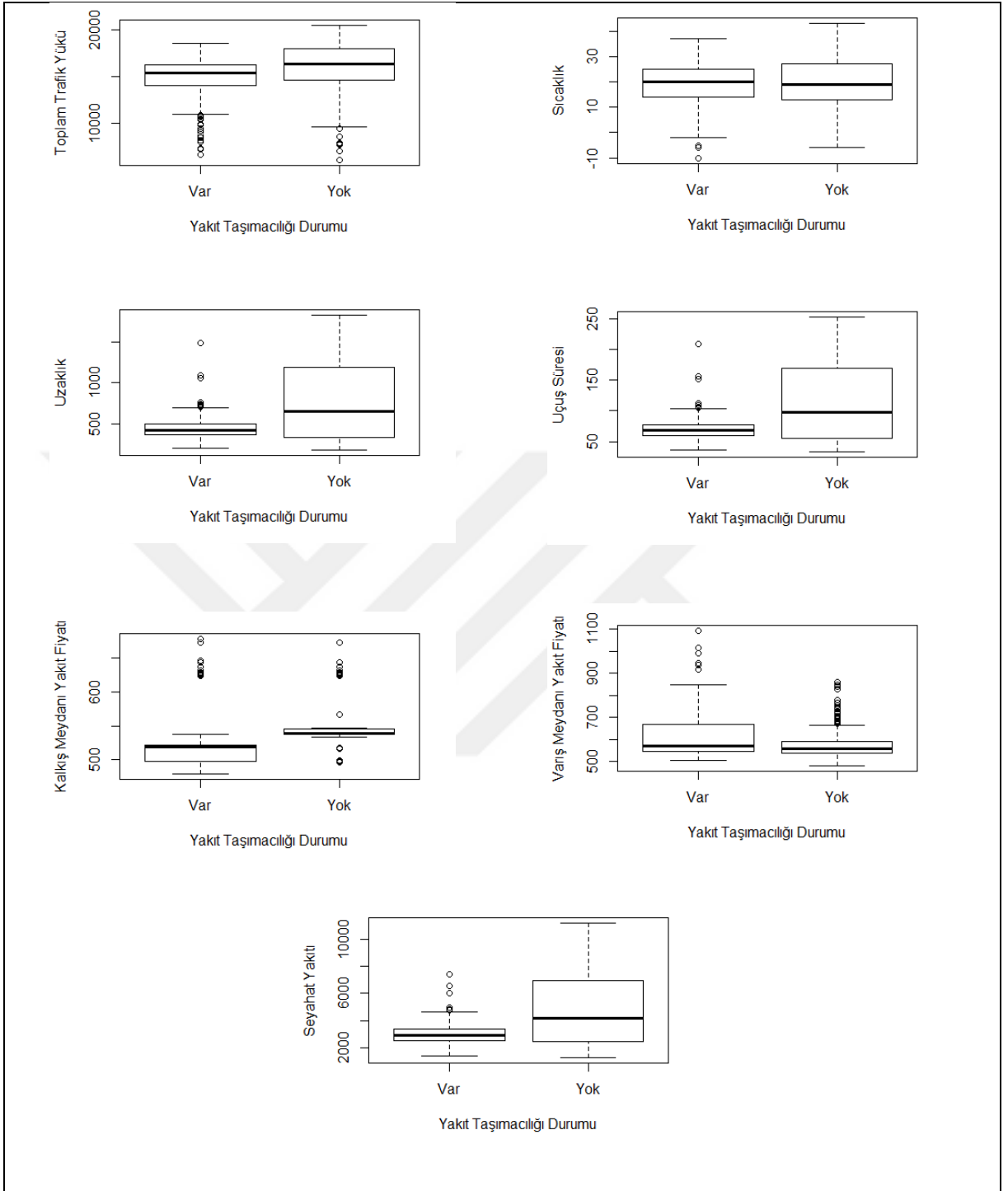
Şekil 3.5: Veri setinde yer alan niteliklere ait korelogram.

Şekil 3.5'e göre veri setindeki uzaklık, uçuş süresi, seyahat yakıtı ve yakıt taşımacılığı miktarı Şekil 3.6'daki serpilme diyagramları ile değerlendirilmiştir.



Şekil 3.6: uzaklik, ucuSuresi, seyahatYakiti ve yakitTasimaciligiMktr niteliklerine ait saçılım diyagramları.

Şekil 3.6'ya göre; uzaklik ve ucuSuresi, uzaklik ve seyahatYakiti, ucuSuresi ve seyahatYakiti arasında pozitif yönde doğrusal bir ilişkinin varlığı tespit edilmiş, diğer nitelikler arasında böyle bir ilişki gözlemlenmemiştir.



Şekil 3.7: Yakıt taşımacılığı durumunu göz önünde bulundurarak *toplTrafikYuku*, *sicaklik*, *uzaklik*, *ucusSuresi*, *seyahatYakiti*, *kalkisMydnYakitFyt* ve *varisMydnYakitFyt* niteliklerine ait kutu grafiği karşılaştırması.

Şekil 3.7’ye göre *yakitTasimaciligiDrm* hedef niteliğinin “0” yani “Yok” olduğu durumdaki *toplTrafikYuku*, *uzaklik*, *ucusSuresi* ve *seyahatYakiti* tahmin niteliği değerleri,

yakitTasimaciligi hedef niteliğinin “1” yani “Var” olduğu durumdaki *tplmTrafikYuku*, *uzaklik*, *ucusSuresi* ve *seyahatYakiti* tahmin niteliği değerlerinden daha fazladır.

3.4. MODELLEME

Bu tez çalışmasında havacılıkta kâr sağlayan yakıt taşımacılığı durumunun ve miktarının tahmini için alternatif modeller oluşturulmuştur. Bu modellerin kurulmasında, Bölüm 2.3.4’te detayları verilen k-En Yakın Komşu algoritması, Basit (*Naive*) Bayes Sınıflandırıcı, Sınıflandırma ve Regresyon Ağaçları algoritması, Rastgele Orman algoritması, Yapay Sinir Ağları ve Destek Vektör Makineleri algoritmalarından yararlanılmıştır.

Algoritmaların karşılaştırılması için Bölüm 2.3.5. Model Değerlendirme ve Seçimi (*Evaluation*) verilen performans değerlendirme yöntemleri ve ölçütleri kullanılmıştır. Bu tez çalışmasında sınıflandırma yaparken model performans değerlendirme ölçütleri hesaplanırken;

- **Pozitif sınıf:** Yakıt taşımacılığı yapılma durumu,
- **Negatif sınıf:** Yakıt taşımacılığı yapılmama durumu,

biçiminde ele alındığından kontenjans tablosu oluşturulurken faydalanan **dp**, **yn**, **yp** ve **dn** aşağıdaki gibidir:

- **dp:** Gerçekte yakıt taşımacılığı yapılan uçuşların algoritma/model tarafından da yakıt taşımacılığı yapılsın şeklinde doğru tahmin edilmesi,
- **yn:** Gerçekte yakıt taşımacılığı yapılan uçuşların algoritma/model tarafından yakıt taşımacılığı yapılmaması şeklinde yanlış tahmin edilmesi,
- **yp:** Gerçekte yakıt taşımacılığı yapılmayan uçuşların algoritma/model tarafından yakıt taşımacılığı yapılsın şeklinde yanlış tahmin edilmesi,
- **dn:** Gerçekte yakıt taşımacılığı yapılmayan uçuşların algoritma/model tarafından da yakıt taşımacılığı yapılmaması şeklinde doğru tahmin edilmesi,

Model performans değerlendirmesi için üçlü ayırma ve tabakalı 5-kat çapraz geçerleme yöntemlerinden faydalananılmıştır. Veri setinin %80’lik kısmı üzerinde tabakalı 5-kat çapraz geçerleme yöntemi ile modelin eğitim performansı incelenmiştir. Parametrelili modellerin parametre ayarlaması yapılmıştır. Kalan %20’lik kısım test veri seti olarak ayrılmış, eğitilen modellerin performansı modellerin daha önce hiç görmediği bu veri seti üzerinde incelenmiştir.

Çalışma kapsamında tüm analizler, R programlama dili ile RStudio ortamında gerçekleştirilmiştir (R Development Core Team, 2021; RStudio, 2021a). Analizlerde aşağıdaki R paketlerinden yararlanılmıştır:

- xlsx (Dragulescu ve Arendt, 2018),
- plyr (Wickham, 2019),
- corrplot (Wei ve Simko, 2016),
- pacman (Rinker ve Kurkiewicz, 2017),
- caret (Kuhn, 2020),
- rpart (Therneau ve Atkinson, 2019),
- rpart.plot (Milborrow, 2020),
- e1071 (Meyer vd., 2020),
- kernlab (Karatzoglou vd., 2004),
- randomForest (Liaw ve Wiener, 2002),
- shiny (Chang vd., 2021),
- shinythemes (Chang, 2021),
- rsconnect (Allaire, 2021),
- shinyjs (Attali, 2021),
- NeuralNetTools (Beck, 2018),
- klaR (Weihs vd., 2005).

Analizlerde elde edilen tüm bulgular Bölüm 4'te sunulmuştur.

4. BULGULAR

Bu bölümde, sırasıyla sınıflandırma ve regresyon amacıyla kullanılan algoritmalarından elde edilen bulgular verilmiştir.

4.1. SINIFLANDIRMA AMACIYLA KULLANILAN ALGORİTMALARDAN ELDE EDİLEN BULGULAR

Bu bölümde sınıflandırma amacıyla kullanılan k-En Yakın Komşu algoritması, Basit Bayes Sınıflandırıcı, Sınıflandırma ve Regresyon Ağaçları algoritması, Rastgele Orman algoritması, Yapay Sinir Ağları ve Destek Vektör Makineleri algoritmalarından elde edilen bulgular verilmiştir.

Model eğitiminde en iyi algoritma parametre değerini otomatik olarak ayarlayan *tuneLength* fonksiyonu kullanılmıştır. *tuneLength* ile her ayar parametresi için denenecek farklı değerlerin sayısı gösterilmektedir. Örneğin; Rastgele Orman Algoritması için *mtry* ayar parametresidir. *tuneLength=5* olsun. Bu “5 farklı *mtry* değerini dene ve bu 5 farklı değeri temel alarak en uygun modeli bul” demektir.

Buna göre; k-En Yakın Komşu algoritması, Basit Bayes Sınıflandırıcı, Sınıflandırma ve Regresyon Ağaçları Algoritması, Rastgele Orman Algoritması ve Yapay Sinir Ağları için model eğitiminde *tuneLength=10* olarak seçilmiştir. Hesaplama süresi uzunluğundan ötürü Destek Vektör Makineleri algoritması için *tuneLength=3* olarak seçilmiştir.

4.1.1. k-En Yakın Komşu Algoritmasından Elde Edilen Bulgular

k-En Yakın Komşu algoritmasının özeti Tablo 4.1 ile sunulmuştur.

Tablo 4.1: k-En Yakın Komşu algoritması özeti (*Sınıflandırma*).

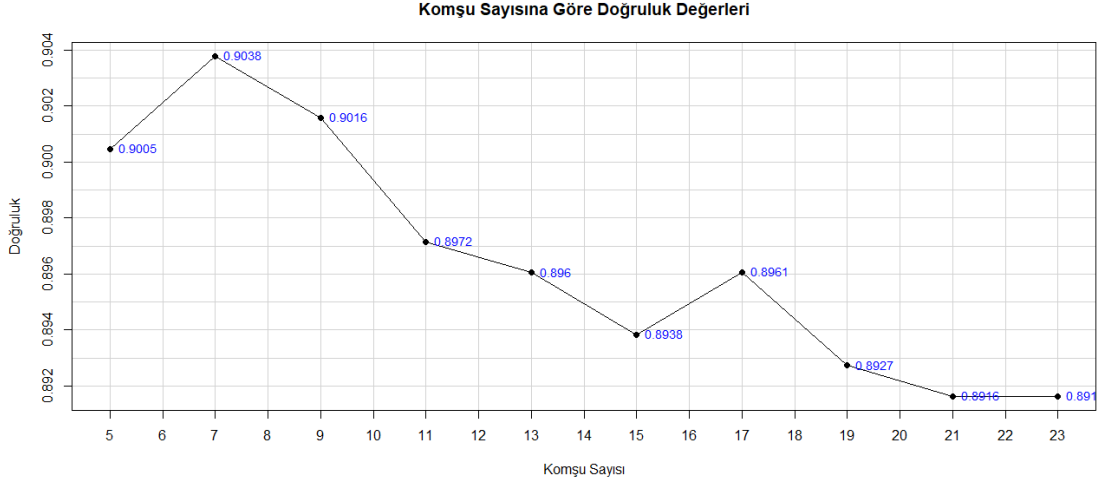
Veri Seti	fueltankering.RData
Hedef Nitelik	<i>yakitTasimaciligiDrm</i> (0: Yok, 1: Var)
Performans Değerlendirme Yöntemi	5-kat çapraz geçerleme
Kullanılan R Kütüphaneleri	caret (Kuhn, 2020)

k-En Yakın Komşu algoritmasının performansını diğer algoritmalarla karşılaştırmak için en iyi model performansına sahip k-En Yakın Komşu modelini veren k (*komşu sayısı*) parametre değeri belirlenmiştir. Buna göre, en iyi k parametresini elde edebilmek için eğitim veri seti kullanılarak tabakalı 5-kat çapraz geçerleme ile algoritma $k=5, 7, \dots, 23$ için ayrı ayrı uygulanmıştır. Burada algoritmanın hızını ve doğru sınıflandırma performansını arttırmak için k değeri tek sayılar olarak belirlenmiştir (Hassanat vd., 2014). Buna göre; k komşu sayısına göre elde edilen *doğruluk* değerleri Tablo 4.2’de sunulmuştur.

Tablo 4.2: k komşu sayısına göre elde edilen doğruluk değerleri.

k Komşu Sayısı	Doğruluk
5	0,9004
7	0,9037
9	0,9015
11	0,8971
13	0,8960
15	0,8938
17	0,8960
19	0,8927
21	0,8916
23	0,8916

Ayrıca, $k=5, 7, \dots, 23$ için elde edilen doğruluk değerleri Şekil 4.1’de verilmiştir.



Şekil 4.1: k komşu sayısına göre doğruluk değerleri.

Elde edilen k -En Yakın Komşu modellerinin eğitim performansları, doğruluk temel alınarak karşılaştırılmıştır. En yüksek doğruluk değeri k komşu sayısı 7 ile 0,9037 olarak elde edilmiştir. Bu nedenle k -En Yakın Komşu algoritmasının parametre ayarlamasında modelin k parametresi 7 olarak belirlenmiştir. Ardından $k=7$ ile oluşturulan model, test performansının bulunması için kullanılmıştır. Modelin tahmin performansı için Tablo 4.3'teki kontenjans tablosu oluşturularak Tablo 4.4'te verilen performans değerlendirme ölçütleri elde edilmiştir.

Tablo 4.3: k -En Yakın Komşu modeli için kontenjans tablosu ($k=7$).

		Gerçek		
		Yakıt Taşımacılığı Durumu = Var	Yakıt Taşımacılığı Durumu = Yok	Toplam
Tahmin	Yakıt Taşımacılığı Durumu = Var	100	17	117
	Yakıt Taşımacılığı Durumu = Yok	3	106	109
	Toplam	103	123	226

Tablo 4.4: k -En Yakın Komşu modeli için model performans değerlendirme ölçütleri ($k=7$ – Sınıflandırma).

k	Ölçüt						
	Doğruluk	Hata	Duyarlılık	Belirleyicilik	Pozitif Öngörü Değeri	Negatif Öngörü Değeri	F -Ölçüsü
7	0,9115	0,0885	0,9709	0,8618	0,8547	0,9725	0,9091

Tablo 4.4 ile verilen ölçütlere göre k-En Yakın Komşu modeli performansı aşağıdaki biçimde yorumlanmıştır:

k-En Yakın Komşu modeli yakıt taşımacılığı durumunu %91,15 doğru tahmin etmektedir ($Hata=0,0885$).

Modelin yakıt taşımacılığı yapılsın ($yakitTasimaciligiDrm=Var$) biçiminde doğru tahmin ettiği uçuşların, gerçekte yakıt taşımacılığı yapılan tüm uçuşlara oranı 0,9709'dur. k-En Yakın Komşu modeli yakıt taşımacılığı yapılan uçuşları %97,09 doğru tahmin etmektedir.

Modelin yakıt taşımacılığı yapılmazın ($yakitTasimaciligiDrm=Yok$) biçiminde doğru tahmin ettiği uçuşların gerçekte yakıt taşımacılığı yapılmayan tüm uçuşlara oranı 0,8618'dir. k-En Yakın Komşu modeli yakıt taşımacılığı yapılmayan uçuşları %86,18 doğru tahmin etmektedir.

Modelin yakıt taşımacılığı yapılsın ($yakitTasimaciligiDrm=Var$) biçiminde doğru tahmin ettiği uçuşların, yakıt taşımacılığı yapılsın şeklindeki tüm tahminlerine oranı 0,8547'dir. k-En Yakın Komşu modeli yakıt taşımacılığı yapılsın biçimindeki tüm tahminleri arasından, gerçekte yakıt taşımacılığı yapılan uçuşları %85,47 doğru tahmin etmektedir.

Modelin yakıt taşımacılığı yapılmazın ($yakitTasimaciligiDrm=Yok$) biçiminde doğru tahmin ettiği uçuşların, yakıt taşımacılığı yapılmazın şeklindeki tüm tahminlerine oranı 0,9725'tir. k-En Yakın Komşu modeli yakıt taşımacılığı yapılmazın biçimindeki tüm tahminleri arasından, gerçekte yakıt taşımacılığı yapılmayan uçuşları %97,25 doğru tahmin etmektedir.

Modelin F -ölçüsü %90,91 olarak ölçülmüştür. Bu değer k-En Yakın Komşu modelinin doğruluğunu desteklemektedir. Yani, modelin yakıt taşımacılığı yapılmayan uçuşları yakıt taşımacılığı yapılan uçuşlar ve yakıt taşımacılığı yapılan uçuşları yakıt taşımacılığı yapılmadı şeklinde tahmin etme durumu çok azdır.

4.1.2. Basit (Naive) Bayes Sınıflandırıcı Algoritmasından Elde Edilen Bulgular

Basit (*Naive*) Bayes Sınıflandırıcı algoritmasının özeti Tablo 4.5 ile sunulmuştur.

Tablo 4.5: Basit (*Naive*) Bayes Sınıflandırıcı algoritması Özeti.

Veri Seti	fueltankering.RData
Hedef Nitelik	<i>yakitTasimaciligiDrm</i> (0: Yok, 1: Var)
Performans Değerlendirme Yöntemi	5-kat çapraz geçerleme
Kullanılan R Kütüphaneleri	caret (Kuhn, 2020) klaR (Weihs vd., 2005)

Basit Bayes Sınıflandırma algoritmasının performansını diğer algoritmalarla karşılaştırabilmek için en iyi performansı veren *fL* (*Laplace düzeltmesi – Naive Bayes ’teki sıfır olasılık sorununu çözmeye yardımcı olan bir düzeltme tekniği*), *usekernel* (*Kernel yoğunluk fonksiyonu – kernel density estimation*) ve *adjust* (*usekernel=True olduğu durumlarda kullanılır ve kernel yoğunluk fonksiyonundaki dağılım miktarını temsil eder*) parametre değerleri belirlenmiştir. Buna göre, en iyi Bayes modeli *fL=0*, *usekernel=True* ve *adjust=1* parametre değerleri ile elde edilmiştir. Oluşturulan modelin tahmin performansı önceden ayrılan test veri seti kullanılarak test edilmiştir. Buna göre; model performans değerlendirmesi için Tablo 4.6’da ki kontenjans tablosu oluşturulmuş ve Tablo 4.7’de verilen performans değerlendirme ölçütleri elde edilmiştir.

Tablo 4.6: Basit Bayes Sınıflandırıcı modeli için kontenjans tablosu.

		Gerçek		
		Yakıt Taşımacılığı Durumu = Var	Yakıt Taşımacılığı Durumu = Yok	Toplam
Tahmin	Yakıt Taşımacılığı Durumu = Var	100	26	126
	Yakıt Taşımacılığı Durumu = Yok	3	97	100
	Toplam	103	123	226

Tablo 4.7: Bayes Sınıflandırıcı modeli için model performans değerlendirme ölçütleri.

Ölçüt						
Doğruluk	Hata	Duyarlılık	Belirleyicilik	Pozitif Öngörü Değeri	Negatif Öngörü Değeri	F-Ölçüsü
0,8717	0,1283	0,9709	0,7886	0,7937	0,9700	0,8734

Tablo 4.7 ile verilen ölçütlere göre Basit Bayes modeli performansı aşağıdaki biçimde yorumlanmıştır:

Basit Bayes Sınıflandırıcı modeli yakıt taşımacılığı durumunu %87,17 doğru tahmin etmektedir ($Hata = \%12,83$).

Modelin yakıt taşımacılığı yapılsın ($yakitTasimaciligiDrm = Var$) biçiminde doğru tahmin ettiği uçuşların, gerçekte yakıt taşımacılığı yapılan tüm uçuşlara oranı 0,9709'dur. Basit Bayes modeli yakıt taşımacılığı yapılan uçuşları %97,09 doğru tahmin etmektedir.

Modelin yakıt taşımacılığı yapılmazın ($yakitTasimaciligiDrm = Yok$) biçiminde doğru tahmin ettiği uçuşların gerçekte yakıt taşımacılığı yapılmayan tüm uçuşlara oranı 0,7886'dır. Basit Bayes modeli yakıt taşımacılığı yapılmayan uçuşları %78,86 doğru tahmin etmektedir.

Modelin yakıt taşımacılığı yapılsın ($yakitTasimaciligiDrm = Var$) biçiminde doğru tahmin ettiği uçuşların, yakıt taşımacılığı yapılsın şeklindeki tüm tahminlerine oranı 0,7937'dir. Basit Bayes modeli yakıt taşımacılığı yapılsın biçimindeki tüm tahminleri arasından, gerçekte yakıt taşımacılığı yapılan uçuşları %79,37 doğru tahmin etmektedir.

Modelin yakıt taşımacılığı yapılmazın ($yakitTasimaciligiDrm = Yok$) biçiminde doğru tahmin ettiği uçuşların, yakıt taşımacılığı yapılmazın şeklindeki tüm tahminlerine oranı 0,9700'tir. Basit Bayes modeli yakıt taşımacılığı yapılmazın biçimindeki tüm tahminleri arasından, gerçekte yakıt taşımacılığı yapılmayan uçuşları %97,00 doğru tahmin etmektedir.

Modelin F-ölçüsü %87,34 olarak ölçülmüştür. Bu değer basit bayes modelinin doğruluğunu desteklemektedir. Yani, modelin yakıt taşımacılığı yapılmayan uçuşları yakıt taşımacılığı yapılan uçuşlar ve yakıt taşımacılığı yapılan uçuşları yakıt taşımacılığı yapılmadı şeklinde tahmin etme durumu çok azdır.

4.1.3. Sınıflandırma ve Regresyon Ağaçları Algoritmasından Elde Edilen Bulgular

Sınıflandırma ve Regresyon Ağaçları algoritmasının özeti Tablo 4.8 ile sunulmuştur.

Tablo 4.8: Sınıflandırma ve Regresyon Ağaçları algoritması özeti (*Sınıflandırma*).

Veri Seti	fueltankering.RData
Hedef Nitelik	<i>yakitTasimaciligi</i> (0: Yok, 1: Var)
Performans Değerlendirme Yöntemi	5-kat çapraz geçerleme
Kullanılan R Kütüphaneleri	caret (Kuhn, 2020) rpart (Therneau ve Atkinson, 2019) rpart.plot (Milborrow, 2020)

Sınıflandırma ve Regresyon Ağaçları algoritmasının performansını diğer algoritmalarla karşılaştırmak için en iyi performansı veren *cp* (*Karmaşıklık parametresi (cp)*, *karar ağacının boyutunu kontrol etmek ve optimum ağaç boyutunu seçmek için kullanılır. Varsayılan değer 0,01'dir. cp ne kadar yüksekse ağaç o kadar küçüktür.*) parametre değeri belirlenmiştir. Buna göre, en iyi Sınıflandırma ve Regresyon Ağacı modeli $cp=0$ parametre değeri ile elde edilmiştir. En iyi model kullanılarak, Şekil 4.2'de Sınıflandırma ve Regresyon Ağaçları algoritması için sınıflandırma ağacı sunulmuştur. Bu ağaç yapısından yararlanarak kuralların türetilmesi mümkündür. Örneğin; kalkış meydanı yakıt fiyatı (*kalkisMydnYakitFyt*) 535,3918 dolardan küçük VE varış meydanı yakıt fiyatı (*varisMydnYakitFyt*) 536,8482 dolardan büyük eşit VE uzaklık (*uzaklik*) 622,5 deniz milinden küçük VE varış meydanı yakıt fiyatı (*varisMydnYakitFyt*) 538,7954 dolar VE sıcaklık (*sicaklik*) 28,5 °C'den küçük İSE yakıt taşımacılığı yapılmaktadır (*yakitTasimaciligiDrm=Var*).

n= 904

node), split, n, loss, yval, (yprob)
* denotes terminal node

```

1) root 904 412 Yok (0.455752212 0.544247788)
  2) kalkisMydnYakitFyt< 535.3918 452 95 Var (0.789823009 0.210176991)
    4) varisMydnYakitFyt>=536.8482 319 22 Var (0.931034483 0.068965517)
      8) uzaklik< 622.5 294 12 Var (0.959183673 0.040816327)
        16) varisMydnYakitFyt>=538.7954 284 9 Var (0.968309859 0.031690141)
          32) sicaklik< 28.5 237 3 Var (0.987341772 0.012658228) *
            33) sicaklik>=28.5 47 6 Var (0.872340426 0.127659574)
              66) tplmTrafikYuku>=13854.5 40 2 Var (0.950000000 0.050000000) *
                67) tplmTrafikYuku< 13854.5 7 3 Yok (0.428571429 0.571428571) *
              17) varisMydnYakitFyt< 538.7954 10 3 Var (0.700000000 0.300000000) *
            9) uzaklik>=622.5 25 10 Var (0.600000000 0.400000000)
              18) kalkisMydnYakitFyt>=518.6254 7 0 Var (1.000000000 0.000000000) *
                19) kalkisMydnYakitFyt< 518.6254 18 8 Yok (0.444444444 0.555555556) *
          5) varisMydnYakitFyt< 536.8482 133 60 Yok (0.451127820 0.548872180)
            10) kalkisMydnYakitFyt< 496.9733 47 0 Var (1.000000000 0.000000000) *
              11) kalkisMydnYakitFyt>=496.9733 86 13 Yok (0.151162791 0.848837209)
                22) kalkisMydnYakitFyt< 510.7347 37 13 Yok (0.351351351 0.648648649)
                  44) varisMydnYakitFyt>=523.8726 16 3 Var (0.812500000 0.187500000) *
                    45) varisMydnYakitFyt< 523.8726 21 0 Yok (0.000000000 1.000000000) *
                  23) kalkisMydnYakitFyt>=510.7347 49 0 Yok (0.000000000 1.000000000) *
              3) kalkisMydnYakitFyt>=535.3918 452 55 Yok (0.121681416 0.878318584)
                6) varisMydnYakitFyt>=651.5726 100 46 Var (0.540000000 0.460000000)
                  12) uzaklik< 452 46 9 Var (0.804347826 0.195652174)
                    24) ucusSuresi>=51.5 38 3 Var (0.921052632 0.078947368) *
                      25) ucusSuresi< 51.5 8 2 Yok (0.250000000 0.750000000) *
                  13) uzaklik>=452 54 17 Yok (0.314814815 0.685185185)
                    26) tplmTrafikYuku< 15438 16 6 Var (0.625000000 0.375000000) *
                      27) tplmTrafikYuku>=15438 38 7 Yok (0.184210526 0.815789474)
                        54) varisMydnYakitFyt>=743.4735 7 3 Var (0.571428571 0.428571429) *
                          55) varisMydnYakitFyt< 743.4735 31 3 Yok (0.096774194 0.903225806) *
                        7) varisMydnYakitFyt< 651.5726 352 1 Yok (0.002840909 0.997159091) *

```

Şekil 4.2: Sınıflandırma ve Regresyon Ağaçları Algoritması için sınıflandırma ağacı.

Oluşturulan modelinin tahmin performansı önceden ayrılan test veri seti kullanılarak test edilmiştir. Buna göre; model performans değerlendirmesi için Tablo 4.9’da ki kontenjans tablosu oluşturulmuş ve Tablo 4.10’da verilen performans değerlendirme ölçütleri elde edilmiştir.

Tablo 4.9: Sınıflandırma ve Regresyon Ağaçları algoritması modeli için kontenjans tablosu.

		Gerçek		
		Yakıt Taşımacılığı Durumu = Var	Yakıt Taşımacılığı Durumu = Yok	Toplam
Tahmin	Yakıt Taşımacılığı Durumu = Var	99	10	109
	Yakıt Taşımacılığı Durumu = Yok	4	113	107
	Toplam	103	123	226

Tablo 4.10: Sınıflandırma ve Regresyon Ağaçları algoritması modeli için model performans değerlendirme ölçütleri (*Sınıflandırma*).

Ölçüt						
Doğruluk	Hata	Duyarlılık	Belirleyicilik	Pozitif Öngörü Değeri	Negatif Öngörü Değeri	F-Ölçüsü
0,9381	0,0619	0,9612	0,9187	0,9083	0,9658	0,9340

Tablo 4.10 ile verilen ölçütlere göre Sınıflandırma ve Regresyon Ağacı modeli performansı aşağıdaki biçimde yorumlanmıştır:

Sınıflandırma ve Regresyon Ağacı modeli yakıt taşımacılığı durumunu %93,81 doğru tahmin etmektedir (*Hata*=%0,0619).

Modelin yakıt taşımacılığı yapılsın (*yakitTasimaciligiDrm=Var*) biçiminde doğru tahmin ettiği uçuşların, gerçekte yakıt taşımacılığı yapılan tüm uçuşlara oranı 0,9612’dir. Sınıflandırma ve Regresyon Ağacı modeli yakıt taşımacılığı yapılan uçuşları %96,12 doğru tahmin etmektedir.

Modelin yakıt taşımacılığı yapılmazın (*yakitTasimaciligiDrm=Yok*) biçiminde doğru tahmin ettiği uçuşların gerçekte yakıt taşımacılığı yapılmayan tüm uçuşlara oranı 0,9187’dir. Sınıflandırma ve Regresyon Ağacı modeli yakıt taşımacılığı yapılmayan uçuşları %91,87 doğru tahmin etmektedir.

Modelin yakıt taşımacılığı yapılsın (*yakitTasimaciligiDrm=Var*) biçiminde doğru tahmin ettiği uçuşların, yakıt taşımacılığı yapılsın şeklindeki tüm tahminlerine oranı 0,9083’tür. Sınıflandırma ve Regresyon Ağacı modeli yakıt taşımacılığı yapılsın biçimindeki tüm tahminleri arasından, gerçekte yakıt taşımacılığı yapılan uçuşları %90,83 doğru tahmin etmektedir.

Modelin yakıt taşımacılığı yapılmazın (*yakitTasimaciligiDrm= Yok*) biçiminde doğru tahmin ettiği uçuşların, yakıt taşımacılığı yapılmazın şeklindeki tüm tahminlerine oranı 0,9658’tir. Sınıflandırma ve Regresyon Ağacı modeli yakıt taşımacılığı yapılmazın biçimindeki tüm tahminleri arasından, gerçekte yakıt taşımacılığı yapılmayan uçuşları %96,58 doğru tahmin etmektedir.

Modelin *F*-ölçüsü %93,40 olarak ölçülmüştür. Bu değer Sınıflandırma ve Regresyon Ağacı modelinin doğruluğunu desteklemektedir. Yani, modelin yakıt taşımacılığı yapılmayan uçuşları

yakıt taşımacılığı yapılan uçuşlar ve yakıt taşımacılığı yapılan uçuşları yakıt taşımacılığı yapılmadı şeklinde tahmin etme durumu çok azdır.

4.1.4. Rastgele Orman Algoritmasından Elde Edilen Bulgular

Rastgele Orman algoritmasının özeti Tablo 4.11 ile sunulmuştur.

Tablo 4.11: Rastgele Orman algoritması özeti (*Sınıflandırma*).

Veri Seti	fueltankering.RData
Hedef Nitelik	<i>yakitTasimaciligiDrm</i> (0: Yok, 1: Var)
Performans Değerlendirme Yöntemi	5-kat çapraz geçerleme
Kullanılan R Kütüphaneleri	caret (Kuhn, 2020) randomForest (Liaw ve Wiener, 2002)

Rastgele Orman Algoritmasının performansını diğer algoritmalarla karşılaştırabilmek için en iyi performansı veren *mtry* (*her bölünmede aday olarak rastgele örneklenen değişkenlerin sayısı*) parametre değeri belirlenmiştir. Buna göre, en iyi Rastgele Orman modeli *mtry*=6 parametre değeri ile elde edilmiştir. Oluşturulan modelin tahmin performansı önceden ayrılan test veri seti kullanılarak test edilmiştir. Buna göre; model performans değerlendirmesi için Tablo 4.12’de ki kontenjans tablosu oluşturulmuş ve Tablo 4.13’te verilen performans değerlendirme ölçütleri elde edilmiştir.

Tablo 4.12: Rastgele Orman algoritması modeli için kontenjans tablosu.

		Gerçek		
		Yakıt Taşımacılığı Durumu = Var	Yakıt Taşımacılığı Durumu = Yok	Toplam
Tahmin	Yakıt Taşımacılığı Durumu = Var	102	7	109
	Yakıt Taşımacılığı Durumu = Yok	1	116	117
	Toplam	103	123	226

Tablo 4.13: Rastgele Orman algoritması modeli için model performans değerlendirme ölçütleri (Sınıflandırma).

Ölçüt						
Doğruluk	Hata	Duyarlılık	Belirleyicilik	Pozitif Öngörü Değeri	Negatif Öngörü Değeri	F-Ölçüsü
0,9646	0,0354	0,9903	0,9431	0,9358	0,9915	0,9623

Tablo 4.13 ile verilen ölçütlere göre Rastgele Orman modeli performansı aşağıdaki biçimde yorumlanmıştır:

Rastgele Orman modeli yakıt taşımacılığı durumunu %96,46 doğru tahmin etmektedir ($Hata=0,0354$).

Modelin yakıt taşımacılığı yapılsın ($yakitTasimaciligiDrm=Var$) biçiminde doğru tahmin ettiği uçuşların, gerçekte yakıt taşımacılığı yapılan tüm uçuşlara oranı 0,9903'tür. Rastgele Orman modeli yakıt taşımacılığı yapılan uçuşları %99,03 doğru tahmin etmektedir.

Modelin yakıt taşımacılığı yapılmasın ($yakitTasimaciligiDrm=Yok$) biçiminde doğru tahmin ettiği uçuşların gerçekte yakıt taşımacılığı yapılmayan tüm uçuşlara oranı 0,9431'dir. Rastgele Orman modeli yakıt taşımacılığı yapılmayan uçuşları %94,31 doğru tahmin etmektedir.

Modelin yakıt taşımacılığı yapılsın ($yakitTasimaciligiDrm=Var$) biçiminde doğru tahmin ettiği uçuşların, yakıt taşımacılığı yapılsın şeklindeki tüm tahminlerine oranı 0,9358'dir. Rastgele Orman modeli yakıt taşımacılığı yapılsın biçimindeki tüm tahminleri arasından, gerçekte yakıt taşımacılığı yapılan uçuşları %93,58 doğru tahmin etmektedir.

Modelin yakıt taşımacılığı yapılmasın ($yakitTasimaciligiDrm=Yok$) biçiminde doğru tahmin ettiği uçuşların, yakıt taşımacılığı yapılmasın şeklindeki tüm tahminlerine oranı 0,9915'tir. Rastgele Orman modeli yakıt taşımacılığı yapılmasın biçimindeki tüm tahminleri arasından, gerçekte yakıt taşımacılığı yapılmayan uçuşları %99,15 doğru tahmin etmektedir.

Modelin F-ölçüsü %96,23 olarak ölçülmüştür. Bu değer rastgele orman modelinin doğruluğunu desteklemektedir. Yani, modelin yakıt taşımacılığı yapılmayan uçuşları yakıt taşımacılığı yapılan uçuşlar ve yakıt taşımacılığı yapılan uçuşları yakıt taşımacılığı yapılmadı şeklinde tahmin etme durumu çok azdır.

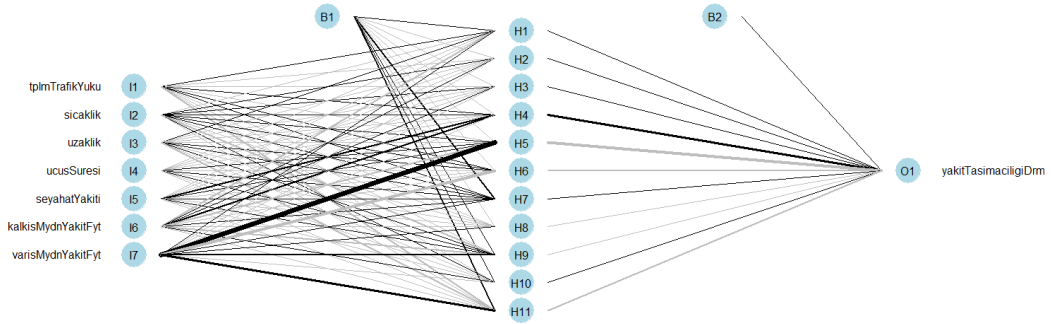
4.1.5. Yapay Sinir Ağları Algoritmasından Elde Edilen Bulgular

Yapay Sinir Ağları algoritmasının özeti Tablo 4.14 ile sunulmuştur.

Tablo 4.14: Yapay Sinir Ağları algoritması özeti (*Sınıflandırma*).

Veri Seti	fueltankering.RData
Hedef Nitelik	<i>yakitTasimaciligiDrm</i> (0: Yok, 1: Var)
Performans Değerlendirme Yöntemi	5-kat çapraz geçerleme
Kullanılan R Kütüphaneleri	caret (Kuhn, 2020) NeuralNetTools (Beck, 2018)

Yapay Sinir Ağları algoritmasının performansını diğer algoritmalarla karşılaştırabilmek için en iyi model performansına sahip Yapay Sinir Ağı modelini veren *size* (gizli katmandaki nöron sayısı) ve *decay* (aşırı uyumu engellemek adına kullanılan düzenleme parametresi) parametre değerleri belirlenmiştir. Buna göre, en iyi Yapay Sinir Ağı modeli *size*=11 ve *decay*=0,001333521 parametre değerleri ile elde edilmiştir. Şekil 4.3 ile ele edilen Yapay Sinir Ağı modeline ait mimari sunulmuştur. Yapay Sinir Ağı modeli tek gizli katmandan oluşmaktadır. Gizli katmandaki nöron sayısı 11'dir.



Şekil 4.3: Yapay Sinir Ağı modeli (*Sınıflandırma*).

Oluşturulan modelin tahmin performansı önceden ayrılan test veri seti kullanılarak test edilmiştir. Buna göre; model performans değerlendirmesi için Tablo 4.15'te ki kontenjans

tablosu oluşturulmuş ve Tablo 4.16’da verilen performans değerlendirme ölçütleri elde edilmiştir.

Tablo 4.15: Yapay Sinir Ağları algoritması modeli için kontenjans tablosu.

		Gerçek		
		Yakıt Taşımacılığı Durumu = Var	Yakıt Taşımacılığı Durumu = Yok	Toplam
Tahmin	Yakıt Taşımacılığı Durumu = Var	98	10	108
	Yakıt Taşımacılığı Durumu = Yok	5	113	118
	Toplam	103	123	226

Tablo 4.16: Yapay Sinir Ağları algoritması modeli için model performans değerlendirme ölçütleri (Sınıflandırma).

Ölçüt						
Doğruluk	Hata	Duyarlılık	Belirleyicilik	Pozitif Öngörü Değeri	Negatif Öngörü Değeri	F-Ölçüsü
0,9336	0,0664	0,9515	0,9187	0,9074	0,9576	0,9289

Tablo 4.16 ile verilen ölçütlere göre Yapay Sinir Ağı modeli performansı aşağıdaki biçimde yorumlanmıştır:

Yapay Sinir Ağı modeli yakıt taşımacılığı durumunu %93,36 doğru tahmin etmektedir ($Hata=0,0664$).

Modelin yakıt taşımacılığı yapılsın ($yakitTasimaciligiDrm=Var$) biçiminde doğru tahmin ettiği uçuşların, gerçekte yakıt taşımacılığı yapılan tüm uçuşlara oranı 0,9515’tir. Yapay Sinir Ağı modeli yakıt taşımacılığı yapılan uçuşları %95,15 doğru tahmin etmektedir.

Modelin yakıt taşımacılığı yapılmayın ($yakitTasimaciligiDrm=Yok$) biçiminde doğru tahmin ettiği uçuşların gerçekte yakıt taşımacılığı yapılmayan tüm uçuşlara oranı 0,9187’dir. Yapay Sinir Ağı modeli yakıt taşımacılığı yapılmayan uçuşları %91,87 doğru tahmin etmektedir.

Modelin yakıt taşımacılığı yapılsın ($yakitTasimaciligiDrm=Var$) biçiminde doğru tahmin ettiği uçuşların, yakıt taşımacılığı yapılsın şeklindeki tüm tahminlerine oranı 0,9074’tür. Yapay Sinir

Ağı modeli yakıt taşımacılığı yapılsın biçimindeki tüm tahminleri arasından, gerçekte yakıt taşımacılığı yapılan uçuşları %90,74 doğru tahmin etmektedir.

Modelin yakıt taşımacılığı yapılmasın (*yakitTasimaciligiDrm= Yok*) biçiminde doğru tahmin ettiği uçuşların, yakıt taşımacılığı yapılmasın şeklindeki tüm tahminlerine oranı 0,9576'dır. Yapay Sinir Ağı modeli yakıt taşımacılığı yapılmasın biçimindeki tüm tahminleri arasından, gerçekte yakıt taşımacılığı yapılmayan uçuşları %95,76 doğru tahmin etmektedir.

Modelin *F*-ölçüsü %92,89 olarak ölçülmüştür. Bu değer yapay sinir ağı modelinin doğruluğunu desteklemektedir. Yani, modelin yakıt taşımacılığı yapılmayan uçuşları yakıt taşımacılığı yapılan uçuşlar ve yakıt taşımacılığı yapılan uçuşları yakıt taşımacılığı yapılmadı şeklinde tahmin etme durumu çok azdır.

4.1.6. Destek Vektör Makineleri Algoritmasından Elde Edilen Bulgular

Destek Vektör Makineleri algoritmasının özeti Tablo 4.17 ile sunulmuştur.

Tablo 4.17: Destek Vektör Makineleri algoritması özeti (*Sınıflandırma*).

Veri Seti	fueltankering.RData
Hedef Nitelik	<i>yakitTasimaciligiDrm</i> (0: Yok, 1: Var)
Performans Değerlendirme Yöntemi	5-kat çapraz geçерleme
Kullanılan R Kütüphaneleri	caret (Kuhn, 2020) e1071 (Meyer vd., 2020) kernlab (Karatzoglou vd., 2004)

Eğitim veri seti ve Destek Vektör Makinesi algoritması kullanılarak tabakalı 5-kat çapraz geçерleme yöntemi ile lineer, polinomial ve radyal çekirdek fonksiyonlarına göre Destek Vektör Makinesi sınıflandırma modelleri oluşturulmuştur. Destek Vektör Makinesi algoritmasının performansını diğer algoritmalarla karşılaştırabilmek için en iyi performansı veren parametre değerleri belirlenmiştir. Bu parametreler; çekirdek fonksiyonu lineer

olduğunda c (*cost - olası yanlış sınıflandırmaları belirleyen parametredir. Varsayılan değeri 1'dir.*), çekirdek fonksiyonu polinomial olduğunda *degree* (*polinom derecesi – polynomial degree*), *scale* (*ölçek*) ve c (*cost*), çekirdek fonksiyonu radyal olduğunda *sigma* (*sigma*), c (*cost*)'dir. Buna göre, en iyi Destek Vektör Makinesi modeli çekirdek fonksiyonu lineer olduğunda $c=1$ ile, çekirdek fonksiyonu polinomial olduğunda *degree*=3, *scale*=0,1 ve $c=1$ ile, çekirdek fonksiyonu radyal olduğunda *sigma*= 0,2645369 ve $c=1$ parametre değerleri ile elde edilmiştir. Oluşturulan modellerin tahmin performansı önceden ayrılan test veri seti kullanılarak test edilmiştir. Buna göre; model performans değerlendirmesi için Tablo 4.18, Tablo 4.20 ve Tablo 4.22'de ki kontenjans tabloları oluşturulmuş ve Tablo 4.19,

Tablo 4.21 ve Tablo 4.23'te verilen performans değerlendirme ölçütleri elde edilmiştir.

Tablo 4.18: Destek Vektör Makinesi algoritması modeli için kontenjans tablosu (*Lineer*).

		Gerçek		
		Yakıt Taşımacılığı Durumu = Var	Yakıt Taşımacılığı Durumu = Yok	Toplam
Tahmin	Yakıt Taşımacılığı Durumu = Var	81	20	101
	Yakıt Taşımacılığı Durumu = Yok	22	103	125
	Toplam	103	123	226

Tablo 4.19: Destek Vektör Makinesi algoritması modeli için model performans değerlendirme ölçütleri (*Lineer - Sınıflandırma*).

Ölçüt						
Doğruluk	Hata	Duyarlılık	Belirleyicilik	Pozitif Öngörü Değeri	Negatif Öngörü Değeri	F-Ölçüsü
0,8142	0,1858	0,7864	0,8374	0,8020	0,8240	0,7941

Tablo 4.19 ile verilen ölçütlere göre Destek Vektör Makinesi (*linear*) modeli performansı aşağıdaki biçimde yorumlanmıştır:

Destek Vektör Makinesi (*linear*) modeli yakıt taşımacılığı durumunu %81,42 doğru tahmin etmektedir (*Hata*=%0,1858).

Modelin yakıt taşımacılığı yapılsın (*yakitTasimaciligiDrm=Var*) biçiminde doğru tahmin ettiği uçuşların, gerçekte yakıt taşımacılığı yapılan tüm uçuşlara oranı 0,7864'tür. Destek Vektör Makinesi (*lineer*) modeli yakıt taşımacılığı yapılan uçuşları %78,64 doğru tahmin etmektedir.

Modelin yakıt taşımacılığı yapılmazın (*yakitTasimaciligiDrm=Yok*) biçiminde doğru tahmin ettiği uçuşların gerçekte yakıt taşımacılığı yapılmayan tüm uçuşlara oranı 0,8374'tür. Destek Vektör Makinesi (*lineer*) modeli yakıt taşımacılığı yapılmayan uçuşları %83,74 doğru tahmin etmektedir.

Modelin yakıt taşımacılığı yapılsın (*yakitTasimaciligiDrm=Var*) biçiminde doğru tahmin ettiği uçuşların, yakıt taşımacılığı yapılsın şeklindeki tüm tahminlerine oranı 0,8020'dir. Destek Vektör Makinesi (*lineer*) modeli yakıt taşımacılığı yapılsın biçimindeki tüm tahminleri arasından, gerçekte yakıt taşımacılığı yapılan uçuşları %80,20 doğru tahmin etmektedir.

Modelin yakıt taşımacılığı yapılmazın (*yakitTasimaciligiDrm= Yok*) biçiminde doğru tahmin ettiği uçuşların, yakıt taşımacılığı yapılmazın şeklindeki tüm tahminlerine oranı 0,8240'tır. Destek Vektör Makinesi (*lineer*) modeli yakıt taşımacılığı yapılmazın biçimindeki tüm tahminleri arasından, gerçekte yakıt taşımacılığı yapılmayan uçuşları %82,40 doğru tahmin etmektedir.

Modelin *F*-ölçüsü %79,41 olarak ölçülmüştür. Bu değer Destek Vektör Makinesi (*lineer*) modelinin doğruluğunu desteklemektedir. Yani, modelin yakıt taşımacılığı yapılmayan uçuşları yakıt taşımacılığı yapılan uçuşlar ve yakıt taşımacılığı yapılan uçuşları yakıt taşımacılığı yapılmadı şeklinde tahmin etme durumu çok azdır.

Tablo 4.20: Destek Vektör Makinesi algoritması modeli için kontenjans tablosu (*Polynomial*).

		Gerçek		
		Yakıt Taşımacılığı Durumu = Var	Yakıt Taşımacılığı Durumu = Yok	Toplam
Tahmin	Yakıt Taşımacılığı Durumu = Var	99	17	116
	Yakıt Taşımacılığı Durumu = Yok	4	106	110
	Toplam	103	123	226

Tablo 4.21: Destek Vektör Makinesi algoritması modeli için model performans değerlendirme ölçütleri (*Polinomial - Sınıflandırma*).

Ölçüt						
Doğruluk	Hata	Duyarlılık	Belirleyicilik	Pozitif Öngörü Değeri	Negatif Öngörü Değeri	F-Ölçüsü
0,9071	0,0929	0,9612	0,8618	0,8534	0,9636	0,9041

Tablo 4.21 ile verilen ölçütlere göre Destek Vektör Makinesi (*Polinomial*) modeli performansı aşağıdaki biçimde yorumlanmıştır:

Destek Vektör Makinesi (*Polinomial*) modeli yakıt taşımacılığı durumunu %90,71 doğru tahmin etmektedir ($Hata = \%0,0929$).

Modelin yakıt taşımacılığı yapılsın (*yakitTasimaciligiDrm=Var*) biçiminde doğru tahmin ettiği uçuşların, gerçekte yakıt taşımacılığı yapılan tüm uçuşlara oranı 0,96,12'dir. Destek Vektör Makinesi (*Polinomial*) modeli yakıt taşımacılığı yapılan uçuşları %96,12 doğru tahmin etmektedir.

Modelin yakıt taşımacılığı yapılmazın (*yakitTasimaciligiDrm=Yok*) biçiminde doğru tahmin ettiği uçuşların gerçekte yakıt taşımacılığı yapılmayan tüm uçuşlara oranı 0,8618'dir. Destek Vektör Makinesi (*Polinomial*) modeli yakıt taşımacılığı yapılmayan uçuşları %86,18 doğru tahmin etmektedir.

Modelin yakıt taşımacılığı yapılsın (*yakitTasimaciligiDrm=Var*) biçiminde doğru tahmin ettiği uçuşların, yakıt taşımacılığı yapılsın şeklindeki tüm tahminlerine oranı 0,8534'tür. Destek Vektör Makinesi (*Polinomial*) modeli yakıt taşımacılığı yapılsın biçimindeki tüm tahminleri arasından, gerçekte yakıt taşımacılığı yapılan uçuşları %85,34 doğru tahmin etmektedir.

Modelin yakıt taşımacılığı yapılmazın (*yakitTasimaciligiDrm= Yok*) biçiminde doğru tahmin ettiği uçuşların, yakıt taşımacılığı yapılmazın şeklindeki tüm tahminlerine oranı 0,9636'dır. Destek Vektör Makinesi (*Polinomial*) modeli yakıt taşımacılığı yapılmazın biçimindeki tüm tahminleri arasından, gerçekte yakıt taşımacılığı yapılmayan uçuşları %96,36 doğru tahmin etmektedir.

Modelin F-ölçüsü %90,41 olarak ölçülmüştür. Bu değer Destek Vektör Makinesi (*Polinomial*) modelinin doğruluğunu desteklemektedir. Yani, modelin yakıt taşımacılığı yapılmayan uçuşları

yakıt taşımacılığı yapılan uçuşlar ve yakıt taşımacılığı yapılan uçuşları yakıt taşımacılığı yapılmadı şeklinde tahmin etme durumu çok azdır.

Tablo 4.22: Destek Vektör Makinesi algoritması modeli için kontenjans tablosu (*Radyal*).

		Gerçek		
		Yakıt Taşımacılığı Durumu = Var	Yakıt Taşımacılığı Durumu = Yok	Toplam
Tahmin	Yakıt Taşımacılığı Durumu = Var	99	12	111
	Yakıt Taşımacılığı Durumu = Yok	4	111	115
	Toplam	103	123	226

Tablo 4.23: Destek Vektör Makinesi algoritması modeli için model performans değerlendirme ölçütleri (*Radyal - Sınıflandırma*).

Ölçüt						
Doğruluk	Hata	Duyarlılık	Belirleyicilik	Pozitif Öngörü Değeri	Negatif Öngörü Değeri	F-Ölçüsü
0,9292	0,0708	0,9612	0,9024	0,8919	0,9652	0,9252

Tablo 4.23 ile verilen ölçütlere göre Destek Vektör Makinesi (*Radyal*) modeli performansı aşağıdaki biçimde yorumlanmıştır:

Destek Vektör Makinesi (*Radyal*) modeli yakıt taşımacılığı durumunu %92,92 doğru tahmin etmektedir (*Hata*=%0,0708).

Modelin yakıt taşımacılığı yapılsın (*yakitTasimaciligiDrm=Var*) biçiminde doğru tahmin ettiği uçuşların, gerçekte yakıt taşımacılığı yapılan tüm uçuşlara oranı 0,96,12'dir. Destek Vektör Makinesi (*Radyal*) modeli yakıt taşımacılığı yapılan uçuşları %96,12 doğru tahmin etmektedir.

Modelin yakıt taşımacılığı yapılmazın (*yakitTasimaciligiDrm=Yok*) biçiminde doğru tahmin ettiği uçuşların gerçekte yakıt taşımacılığı yapılmayan tüm uçuşlara oranı 0,9024'tür. Destek Vektör Makinesi (*Radyal*) modeli yakıt taşımacılığı yapılmayan uçuşları %90,24 doğru tahmin etmektedir.

Modelin yakıt taşımacılığı yapılsın (*yakitTasimaciligiDrm=Var*) biçiminde doğru tahmin ettiği uçuşların, yakıt taşımacılığı yapılsın şeklindeki tüm tahminlerine oranı 0,8919'dur. Destek Vektör Makinesi (*Radyal*) modeli yakıt taşımacılığı yapılsın biçimindeki tüm tahminleri arasından, gerçekte yakıt taşımacılığı yapılan uçuşları %89,19 doğru tahmin etmektedir.

Modelin yakıt taşımacılığı yapılmazın (*yakitTasimaciligiDrm= Yok*) biçiminde doğru tahmin ettiği uçuşların, yakıt taşımacılığı yapılmazın şeklindeki tüm tahminlerine oranı 0,9652'dir. Destek Vektör Makinesi (*Radyal*) modeli yakıt taşımacılığı yapılmazın biçimindeki tüm tahminleri arasından, gerçekte yakıt taşımacılığı yapılmayan uçuşları %96,52 doğru tahmin etmektedir.

Modelin *F*-ölçüsü %92,52 olarak ölçülmüştür. Bu değer Destek Vektör Makinesi (*Radyal*) modelinin doğruluğunu desteklemektedir. Yani, modelin yakıt taşımacılığı yapılmayan uçuşları yakıt taşımacılığı yapılan uçuşlar ve yakıt taşımacılığı yapılan uçuşları yakıt taşımacılığı yapılmadı şeklinde tahmin etme durumu çok azdır.

Sınıflandırma amacıyla kullanılan algoritmalar için model performans karşılaştırma tablosu Tablo 4.24 ile sunulmuştur.

Tablo 4.24: Sınıflandırma amacıyla kullanılan algoritmalar için model performans karşılaştırması.

Algoritma	Ölçüt						
	Doğruluk	Hata	Duyarlılık	Belirleyicilik	Pozitif Öngörü Değeri	Negatif Öngörü Değeri	<i>F</i> -Ölçüsü
kNN	0,9115	0,0885	0,9709	0,8618	0,8547	0,9725	0,9091
NB	0,8717	0,1283	0,9709	0,7886	0,7937	0,9700	0,8734
CART	0,9381	0,0619	0,9612	0,9187	0,9083	0,9658	0,9340
RF	0,9646	0,0354	0,9903	0,9431	0,9358	0,9915	0,9623
YSA	0,9336	0,0664	0,9515	0,9187	0,9074	0,9576	0,9289
DVM (Lineer)	0,8142	0,1858	0,7864	0,8374	0,8020	0,8240	0,7941
DVM (Polinomial)	0,9071	0,0929	0,9612	0,8618	0,8534	0,9636	0,9041
DVM (Radyal)	0,9292	0,0708	0,9612	0,9024	0,8919	0,9652	0,9252

Tablo 4.24'e göre sınıflandırma algoritmaları içinde en yüksek doğruluk değerine (%96,46) sahip algoritma Rastgele Orman algoritması olarak bulunmuştur. Sonrasında en yüksek doğruluğa sahip algoritmalar sırasıyla; Sınıflandırma ve Regresyon Ağaçları (%93,81), Yapay Sinir Ağları (%93,36), Destek Vektör Makineleri (%92,92), k-En Yakın Komşu algoritması (%91,15) ve Basit Bayes Sınıflandırıcı (%87,17)'dir.

4.2. REGRESYON AMACIYLA KULLANILAN ALGORİTMALARDAN ELDE EDİLEN BULGULAR

Bu bölümde regresyon amacıyla kullanılan k-En Yakın Komşu algoritması, Sınıflandırma ve Regresyon Ağaçları algoritması, Rastgele Orman algoritması, Yapay Sinir Ağları ve Destek Vektör Makineleri algoritmalarından elde edilen bulgular verilmiştir.

4.2.1. k-En Yakın Komşu Algoritmasından Elde Edilen Bulgular

k-En Yakın Komşu algoritmasının özeti Tablo 4.25 ile sunulmuştur.

Tablo 4.25: k-En Yakın Komşu algoritması özeti (*Regresyon*).

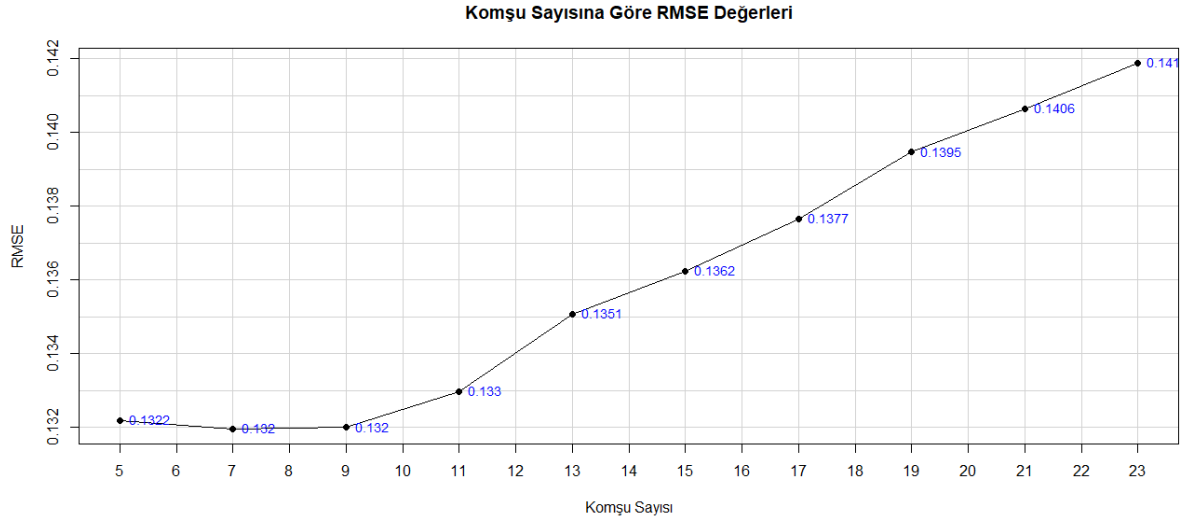
Veri Seti	fueltankering.RData
Hedef Nitelik	<i>yakitTasimaciligiMktr</i> (Miktar / kg.)
Performans Değerlendirme Yöntemi	5-kat çapraz geçerleme
Kullanılan R Kütüphaneleri	caret (Kuhn, 2020)

k-En Yakın Komşu algoritmasının performansını diğer algoritmalarla karşılaştırabilmek için öncelikle en iyi performansı veren k değeri seçilmiştir. Buna göre, en iyi k parametresini elde edebilmek için eğitim veri seti kullanılarak tabakalı 5-kat çapraz geçerleme ile algoritma $k=5, 7, \dots, 23$ için ayrı ayrı uygulanmıştır. Buna göre; k komşu sayısına göre elde edilen RMSE ve R^2 değerleri Tablo 4.26'da sunulmuştur.

Tablo 4.26: k komşu sayısına göre elde edilen RMSE ve R^2 değerleri

k Komşu Sayısı	RMSE	R^2
5	0,1321	0,6823
7	0,1319	0,6816
9	0,1319	0,6814
11	0,1329	0,6776
13	0,1350	0,6674
15	0,1362	0,6619
17	0,1376	0,6551
19	0,1394	0,6465
21	0,1406	0,6407
23	0,1418	0,6346

Ayrıca, $k=5, 7, \dots, 23$ için elde edilen RMSE değerleri Şekil 4.4'te verilmiştir.

**Şekil 4.4:** k komşu sayısına göre RMSE değerleri.

Elde edilen modellerin eğitim performansları, RMSE ölçüsü temel alınarak karşılaştırılmıştır. En düşük RMSE değeri 0,1319 olarak bulunmuştur. Burada; $k=7$ ve $k=9$ iken RMSE değerleri aynıdır. Ancak, en yüksek R^2 değeri $k=7$ ile elde edilmiştir. Bu nedenle k en yakın komşu algoritmasının parametre ayarlamasında modelin k parametresi 7 olarak belirlenmiştir. Ardından $k=7$ ile oluşturulan model, test veri setindeki örneklerin tahmininde kullanılmıştır. Modelin tahmin performansı için Tablo 4.27'de verilen performans değerlendirme ölçütleri elde edilmiştir.

Tablo 4.27: k-En Yakın Komşu modeli için model performans değerlendirme ölçütleri ($k=7$ - Regresyon).

Performans Değerlendirme Ölçütü			
R^2	MSE	RMSE	MAE
0,6610	0,0195	0,1396	0,0787

Tablo 4.27 ile verilen model performans ölçütlerine göre; k-En Yakın Komşu modeli yakıt taşımacılığı miktarını %66,10 ile doğru tahmin etmektedir. Modelin yakıt taşımacılığı miktarını hatalı tahmin ettiği durumlar için hesaplanan hatalar; Ortalama Kare Hata (0,0195), Kök Ortalama Kare Hata (0,1396) ve Ortalama Mutlak Hata (0,0787) olarak bulunmuştur.

4.2.2. Sınıflandırma ve Regresyon Ağaçları Algoritmasından Elde Edilen Bulgular

Sınıflandırma ve Regresyon Ağaçları algoritmasının özeti Tablo 4.28 ile sunulmuştur.

Tablo 4.28: Sınıflandırma ve Regresyon Ağaçları algoritması özeti (Regresyon).

Veri Seti	fueltankering.RData
Hedef Nitelik	yakitTasimaciligiMktr (Miktar / kg.)
Performans Değerlendirme Yöntemi	5-kat çapraz geçерleme
Kullanılan R Kütüphaneleri	caret (Kuhn, 2020) rpart (Therneau ve Atkinson, 2019) rpart.plot (Milborrow, 2020)

Regresyon amacıyla kullanılan Sınıflandırma ve Regresyon Ağaçları algoritmasının performansını diğer algoritmalarla karşılaştırmak için öncelikle en iyi performansı veren parametre değerleri seçilmiştir. Buna göre, en iyi Sınıflandırma ve Regresyon Ağacı modeli $cp=0,008378939$ ile elde edilmiştir. Oluşturulan eğitim modelinin tahmin performansı önceden ayrılan test veri seti kullanılarak test edilmiştir. Buna göre; model performans değerlendirmesi için Tablo 4.29’da verilen performans değerlendirme ölçütleri elde edilmiştir.

Tablo 4.29: Sınıflandırma ve Regresyon Ağaçları algoritması modeli için model performans değerlendirme ölçütleri (*Regresyon*).

Performans Değerlendirme Ölçütü			
R^2	MSE	RMSE	MAE
0,7407	0,0146	0,1211	0,0608

Tablo 4.29 ile verilen model performans ölçütlerine göre; Sınıflandırma ve Regresyon Ağacı modeli yakıt taşımacılığı miktarını %74,07 ile doğru tahmin etmektedir. Modelin yakıt taşımacılığı miktarını hatalı tahmin ettiği durumlar için hesaplanan hatalar; Ortalama Kare Hata (0,0146), Kök Ortalama Kare Hata (0,1211) ve Ortalama Mutlak Hata (0,0608) olarak bulunmuştur.

4.2.3. Rastgele Orman Algoritmasından Elde Edilen Bulgular

Rastgele Orman algoritmasının özeti Tablo 4.30 ile sunulmuştur.

Tablo 4.30: Rastgele Orman algoritması özeti (*Regresyon*).

Veri Seti	fueltankering.RData
Hedef Nitelik	<i>yakitTasimaciligiMktr</i> (Miktar / kg.)
Performans Değerlendirme Yöntemi	5-kat çapraz geçерleme
Kullanılan R Kütüphaneleri	caret (Kuhn, 2020) randomForest (Liaw ve Wiener, 2002)

Rastgele Orman algoritmasının performansını diğer algoritmalarla karşılaştırabilmek için öncelikle en iyi performansı veren parametre değerleri seçilmiştir. Buna göre, en iyi Rastgele Orman modeli $mtry=5$ ile elde edilmiştir. Oluşturulan eğitim modelinin tahmin performansı önceden ayrılan test veri seti kullanılarak test edilmiştir. Buna göre; model performans değerlendirmesi için Tablo 4.31’de verilen performans değerlendirme ölçütleri elde edilmiştir.

Tablo 4.31: Rastgele Orman algoritması modeli için model performans değerlendirme ölçütleri (*Regresyon*).

Performans Değerlendirme Ölçütü			
R^2	MSE	RMSE	MAE
0,8104	0,0107	0,1036	0,0530

Tablo 4.31 ile verilen model performans ölçütlerine göre; Rastgele Orman algoritması modeli yakıt taşımacılığı miktarını %81,04 ile doğru tahmin etmektedir. Modelin yakıt taşımacılığı miktarını hatalı tahmin ettiği durumlar için hesaplanan hatalar; Ortalama Kare Hata (0,0107), Kök Ortalama Kare Hata (0,1036) ve Ortalama Mutlak Hata (0,0530) olarak bulunmuştur.

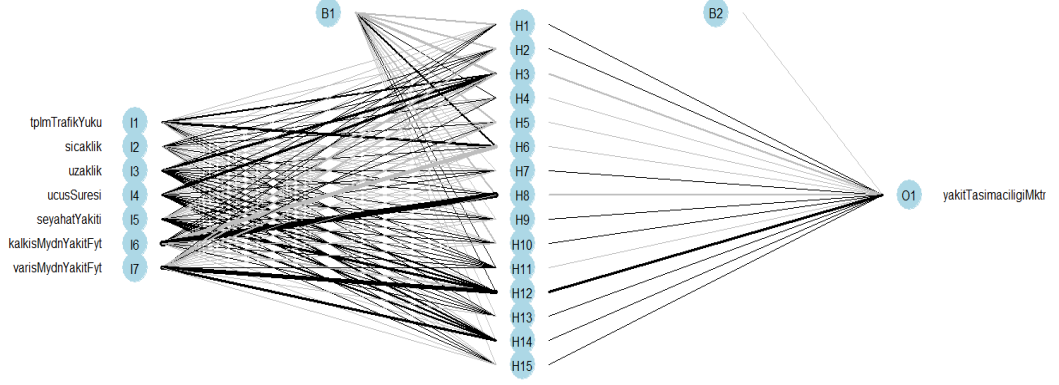
4.2.4. Yapay Sinir Ağları Algoritmasından Elde Edilen Bulgular

Yapay Sinir Ağları algoritmasının özeti Tablo 4.32 ile sunulmuştur.

Tablo 4.32: Yapay Sinir Ağları algoritması özeti (*Regresyon*).

Veri Seti	fueltankering.RData
Hedef Nitelik	yakitTasimaciligiMktr (Miktar / kg.)
Performans Değerlendirme Yöntemi	5-kat çapraz geçерleme
Kullanılan R Kütüphaneleri	caret (Kuhn, 2020) NeuralNetTools (Beck, 2018)

Yapay Sinir Ağları algoritmasının performansını diğer algoritmalarla karşılaştırabilmek için öncelikle en iyi performansı veren parametre değerleri seçilmiştir. Buna göre, Yapay Sinir Ağı modeli $size=15$ (gizli katmandaki nöron sayısı) ve $decay= 0,0005623413$ ile elde edilmiştir. Şekil 4.5 ile Yapay Sinir Ağı modeline ait mimari sunulmuştur.



Şekil 4.5: Yapay Sinir Ağı modeli (*Regresyon*).

Oluşturulan eğitim modelinin tahmin performansı önceden ayrılan test veri seti kullanılarak test edilmiştir. Buna göre; model performans değerlendirmesi için Tablo 4.33'te verilen performans değerlendirme ölçütleri elde edilmiştir.

Tablo 4.33: Yapay Sinir Ağları algoritması modeli için model performans değerlendirme ölçütleri (*Regresyon*).

Performans Değerlendirme Ölçütü			
R^2	MSE	RMSE	MAE
0,6978	0,0171	0,1307	0,0744

Tablo 4.33 ile verilen model performans ölçütlerine göre; Yapay Sinir Ağı modeli yakıt taşımacılığı miktarını %69,78 ile doğru tahmin etmektedir. Modelin yakıt taşımacılığı miktarını hatalı tahmin ettiği durumlar için hesaplanan hatalar; Ortalama Kare Hata (0,0171), Kök Ortalama Kare Hata (0,1307) ve Ortalama Mutlak Hata (0,0744) olarak bulunmuştur.

4.2.5. Destek Vektör Makineleri Algoritmasından Elde Edilen Bulgular

Destek Vektör Makineleri algoritmasının özeti Tablo 4.34 ile sunulmuştur.

Tablo 4.34: Destek Vektör Makineleri algoritması özeti (*Regresyon*).

Veri Seti	fueltankering.RData
Hedef Nitelik	<i>yakitTasimaciligiMktr</i> (Miktar / kg.)
Performans Değerlendirme Yöntemi	5-kat çapraz geçerleme
Kullanılan R Kütüphaneleri	caret (Kuhn, 2020) e1071 (Meyer vd., 2020) kernlab (Karatzoglou vd., 2004)

Eğitim veri seti ve Destek Vektör Makineleri algoritması kullanılarak tabakalı 5-kat çapraz geçerleme yöntemi ile lineer, polinomial ve radyal çekirdek fonksiyonlarına göre Destek Vektör Makinesi regresyon modelleri oluşturulmuştur. Destek Vektör Makineleri algoritmasının performansını diğer algoritmalarla karşılaştırabilmek için öncelikle en iyi performansı veren parametre değerleri seçilmiştir. Buna göre, en iyi Destek Vektör Makinesi modeli çekirdek fonksiyonu lineer olduğunda $c=1$ ile, çekirdek fonksiyonu polinomial olduğunda $degree=3$, $scale=0,1$, $c=0,25$ ile ve çekirdek fonksiyonu radyal olduğunda $sigma=0,2702702$, $c=2$ ile elde edilmiştir. Oluşturulan eğitim modellerinin tahmin performansı önceden ayrılan test veri seti kullanılarak test edilmiştir. Buna göre; model performans değerlendirmesi için Tablo 4.35, Tablo 4.36 ve Tablo 4.37’de verilen performans değerlendirme ölçütleri elde edilmiştir.

Tablo 4.35: Destek Vektör Makineleri algoritması modeli için model performans değerlendirme ölçütleri (*Lineer - Regresyon*).

Performans Değerlendirme Ölçütü			
R^2	MSE	RMSE	MAE
0,3164	0,0413	0,2033	0,1513

Tablo 4.35 ile verilen model performans ölçütlerine göre; Destek vektör makinesi (*lineer*) modeli yakıt taşımacılığı miktarını %31,64 ile doğru tahmin etmektedir. Modelin yakıt taşımacılığı miktarını hatalı tahmin ettiği durumlar için hesaplanan hatalar; Ortalama Kare Hata

(0,0413), Kök Ortalama Kare Hata (0,2033) ve Ortalama Mutlak Hata (0,1513) olarak bulunmuştur.

Tablo 4.36: Destek Vektör Makineleri algoritması modeli için model performans değerlendirme ölçütleri (*Polynomial - Regresyon*).

Performans Değerlendirme Ölçütü			
R^2	MSE	RMSE	MAE
0,5204	0,0271	0,1647	0,1203

Tablo 4.36 ile verilen model performans ölçütlerine göre; Destek Vektör Makinesi (*polynomial*) modeli yakıt taşımacılığı miktarını %52,04 ile doğru tahmin etmektedir. Modelin yakıt taşımacılığı miktarını hatalı tahmin ettiği durumlar için hesaplanan hatalar; Ortalama Kare Hata (0,0271), Kök Ortalama Kare Hata (0,1647) ve Ortalama Mutlak Hata (0,1203) olarak bulunmuştur.

Tablo 4.37: Destek Vektör Makineleri algoritması modeli için model performans değerlendirme ölçütleri (*Radial - Regresyon*).

Performans Değerlendirme Ölçütü			
R^2	MSE	RMSE	MAE
0,6949	0,0175	0,1323	0,0829

Tablo 4.37 ile verilen model performans ölçütlerine göre; Destek Vektör Makinesi modeli (*radial*) modeli yakıt taşımacılığı miktarını %69,49 ile doğru tahmin etmektedir. Modelin yakıt taşımacılığı miktarını hatalı tahmin ettiği durumlar için hesaplanan hatalar; Ortalama Kare Hata (0,0175), Kök Ortalama Kare Hata (0,1323) ve Ortalama Mutlak Hata (0,0829) olarak bulunmuştur.

Regresyon amacıyla kullanılan algoritmalar için model performans karşılaştırma tablosu Tablo 4.38 ile sunulmuştur.

Tablo 4.38: Regresyon amacıyla kullanılan algoritmalar için model performans karşılaştırması.

Algoritma		Ölçüt			
		R^2	MSE	RMSE	MAE
k-En Yakın Komşu Algoritması		0,6610	0,0195	0,1396	0,0787
Sınıflandırma ve Regresyon Ağaçları		0,7407	0,0146	0,1211	0,0608
Rastgele Orman Algoritması		0,8104	0,0107	0,1036	0,0530
Yapay Sinir Ağları		0,6978	0,0171	0,1307	0,0744
Destek Vektör Makineleri	Lineer	0,3164	0,0413	0,2033	0,1513
	Polinomial	0,5204	0,0271	0,1647	0,1203
	Radyal	0,6949	0,0175	0,1323	0,0829

Tablo 4.38'e göre en düşük RMSE değeri (0,1036) Rastgele Orman algoritması ile elde edilmiştir. Öte yandan; en yüksek R^2 değeri 0,8104 ile Rastgele Orman algoritması ile elde edilmiştir. Bu nedenle; Rastgele Orman algoritması yakıt taşımacılığının yapıldığı durumlarda miktar tahmininde en başarılı model performansı gösteren algoritma olarak bulunmuştur.

4.3. SHINY UYGULAMASI

Bu tez çalışmasının BULGULAR kısmında yakıt taşımacılığı durumunun sınıflandırılmasında en iyi model performansına sahip sınıflandırma modeli Rastgele Orman algoritması ile elde edilmiştir. Öte yandan, yakıt taşımacılığı miktarını öngören en iyi model performansına sahip regresyon modeli Rastgele Orman algoritması ile elde edilmiştir. Buna göre; en iyi performansa sahip bu sınıflandırma ve regresyon modeli kullanılarak çevrimiçi bir Shiny uygulaması geliştirilmiştir. Uygulamanın geliştirilmesinde, RStudio programı kullanılmış ve Shiny R paketinden yararlanılmıştır (RStudio, 2021a, 2021b). Geliştirilen Shiny uygulaması, shinyapps.io'da ücretsiz bir hesap açılarak web ortamına aktarılmıştır (RStudio, 2021c). Uygulamanın kullanıcı arayüzü Şekil 4.6'da verilmiştir. Uygulamaya, <https://ilkerguvenyilmaz.shinyapps.io/fueltankering/> web adresinden erişilebilmektedir.

Prediction Of Profitable Fuel Tankering Status With Machine Learning Techniques In Aviation

Data Summary:

Total Traffic Load: 6033
 Temperature at arrival airport: -10
 Air distance: 184
 The time between take-off and landing: 34
 Trip fuel: 1280
 The fuel price at departure airport: 479.66020996
 The fuel price at destination airport: 482.75668191

Prediction Result: Tankering 1693.284 kgs.

How to Cite:
 Yılmaz, I.G., Kartal, E., Ozen, Z., & Gulsecen, S. (2021). Prediction of Fuel Tankering in Aviation Industry with Machine Learning Algorithms. Journal of Aeronautics and Space Technologies, 14(1), 19-34.

Şekil 4.6: Çevrimiçi Shiny uygulaması (Yılmaz, 2021).

Şekil 4.6’den de görüldüğü üzere kullanıcı arayüzünün sol kısmında yer alan giriş alanları ile kullanıcıdan toplam trafik yükü, varış meydanı hava sıcaklığı, uzaklık, uçuş süresi, seyahat yakıtı miktarı, kalkış meydanı yakıt fiyatı ve varış meydanı yakıt fiyatı değerleri alınmaktadır. Kullanıcı arayüzünün sağ kısmında “Data Summary” bölümü ile kullanıcıdan alınan parametre değerleri sunulmaktadır. Yine aynı alanda “Prediction Result” bölümü ile kullanıcıdan alınan parametre değerlerine göre modeller aracılığıyla elde edilen sonuçlar sunulmaktadır. Burada ilk sonuç yakıt taşımacılığı durumunu “Var/Yok”, ikinci sonuç eğer yakıt taşımacılığı yapılıyorsa taşınabilecek yakıt miktarını ifade etmektedir.

5. TARTIŞMA VE SONUÇ

Bu tez çalışmasında; makine öğrenmesi teknikleriyle havacılıkta yakıt taşımacılığı kararının alınması amaçlanmıştır. Bu kapsamda, problem iki farklı açıdan incelenmiştir. Öncelikle yakıt taşımacılığı durumunun belirlenmesi için sınıflandırma modelleri geliştirilmiştir ve “Var/Yok” şeklinde uçuşlarda yakıt taşımacılığı öngörüsünde bulunulmuştur. Sonrasında tahminlerin daha hassas hale getirilmesi için sadece “Var/Yok” şeklinde durum tahmini değil, regresyon modelleri kullanılarak yakıt taşımacılığı miktarı tahmin edilmeye çalışılmıştır. Analizler sonucunda en iyi performans gösteren sınıflandırma ve regresyon modellerinin kullanılabilirliğini sağlamak amacıyla çevrimiçi bir Shiny uygulaması geliştirilmiştir.

Bu tez çalışmasının BULGULAR kısmından elde edilen sonuçlara göre, çalışmada tercih edilen makine öğrenmesi algoritmalarından sınıflandırma amacıyla kullanılanlar içerisinde en yüksek performans sahip modeli üreten algoritma (*doğruluk oranı %96,46*) Rastgele Orman Algoritması olarak bulunmuştur. Sonrasında sırasıyla; Sınıflandırma ve Regresyon Ağaçları (*doğruluk oranı %93,81*), Yapay Sinir Ağları (*doğruluk oranı %93,36*), Destek Vektör Makineleri (*Radyal*) (*doğruluk oranı %92,92*), k-En Yakın Komşu algoritması (*doğruluk oranı %91,15*) ve Basit Bayes Sınıflandırıcı (*doğruluk oranı %87,17*)’dır. Regresyon amacıyla kullanılan algoritmalar içerisinde en yüksek performansa sahip modeli üreten algoritma ise Rastgele Orman algoritması ($R^2=0,8104$, $MSE=0,0107$, $RMSE=0,1036$, $MAE=0,0530$) olarak bulunmuştur. Buna göre; Rastgele Orman algoritmasının yakıt taşımacılığı durumunun sınıflandırılmasında ve yakıt taşımacılığı miktar tahminin gerçekleştirilmesi amacıyla kullanılmasında (*regresyon*) en iyi makine öğrenmesi algoritması olduğu söylenebilir.

Bu tezde önerilen makine öğrenmesi modellerinin kârlı ulaşım açısından yapay zekânın havacılık sektörüne entegrasyonunun en önemli örneklerinden biri olacağı düşünülmektedir. Ayrıca bu tez çalışmasında önerilen makine öğrenmesi modelleri, şirketlere sahip oldukları ham verinin katma değere dönüştürülmesi sürecinde destek ve fikirler sunacaktır. Şirketlere ürettikleri ve depoladıkları yakıt taşımacılığı verisini, kendi yakıt taşımacılığı politikalarına uygun bir biçimde makine öğrenmesi teknikleriyle modelleyebilmesi noktasında ilham verecektir. Çalışma kapsamında geliştirilen çevrimiçi Shiny uygulaması, şirket bünyesinde geliştirilecek uygulamalara makine öğrenmesi tahmin modellerinin entegrasyonunu sağlayarak, üçüncü parti uygulamalara olan bağımlılıklarının azaltılabileceğine örnek teşkil etmektedir.

Bu çalışmada hem sınıflandırma ve hem regresyon amacıyla kullanılan makine öğrenmesi algoritmaları içerisinde Rastgele Orman algoritmasının en iyi performans gösteren algoritma olarak bulunması sonucu; literatürde topluluk öğrenmesi yöntemine dayalı algoritmaların diğer algoritmalara göre daha iyi performans gösterdiği sonucunu desteklemektedir (Gürsaka, 2018; Goudos ve Athanasiadou, 2019). Dolayısıyla gelecek çalışmalarda Adaboost, Stochastic Gradient Boosting vb. diğer topluluk öğrenmesi yöntemleri kullanılarak çalışmanın sonuçları genişletilebilir.

Bu tez çalışmasında yakıt taşımacılığı problemi yakıt taşımacılığı durumunun tahmini (*sınıflandırma*) ve yakıt taşımacılığı yapıldığı durumlarda miktarın tahmini (*regresyon*) olacak şekilde iki açıdan ele alınmıştır. Yakıt taşımacılığı durumunun tahminin modellenmesi için sınıflandırma algoritmaları (*k-En Yakın Komşu algoritması, Basit (Naive) Bayes Sınıflandırıcı, Sınıflandırma ve Regresyon Ağaçları, Yapay Sinir Ağları, Destek Vektör Makineleri ve Rastgele Orman algoritması*), yakıt taşımacılığının yapıldığı durumlarda miktar tahminin modellenmesi için regresyon algoritmaları (*k-En Yakın Komşu algoritması, Sınıflandırma ve Regresyon Ağaçları, Yapay Sinir Ağları, Destek Vektör Makineleri ve Rastgele Orman algoritması*) kullanılmıştır. Gelecekte yapılacak çalışmalarda, yakıt taşımacılığı durumun tahmini için k-Ortalamlar algoritması, Bulanık C-Ortalama Kümeleme algoritması vb. gibi danışmansız öğrenme algoritmaları kullanılarak uçuşlar kümelenebilir. Böylelikle veri setinde herhangi bir hedef niteliğe bağlı kalınsızın uçuşlar otomatik biçimde gruplandırılabilir.

Makine öğrenmesi çalışmalarında veri seti büyüklüğü oldukça önemlidir. En iyi performansa sahip modelin belirlenmesi kullanılan veri setine göre değişmektedir. Bu çalışmada, veri seti 1130 uçuş içermektedir. Çalışmanın sonuçlarını genelleştirmek ve model performanslarını iyileştirmek için bu uçuş adedi daha da arttırılabilir. Öte yandan, bu çalışmanın veri seti iki uçak tipine ait uçaklardan elde edilen uçuş verisini içermektedir. Bu nedenle, uçak tipi ve veri setindeki parametreler değiştiğinde en iyi performansa sahip modelde değişecektir. Ayrıca veri setinde kullanılan tahmini sağlayan nitelikler yeniden yapılandırılabilir (*ekleme/çıkarma*) ve analizler yeni veri setiyle tekrarlanabilir.

Bu çalışma da karşılaşılan zorluklardan bir tanesi de makine öğrenmesi kullanılarak yakıt taşımacılığı probleminin çözümü için açık kaynak kodlu bir uygulama/yazılımın bulunmamasıdır. Ayrıca, yakıt taşımacılığı probleminin çözümünde kullanılabilecek herkese açık bir veri seti bulunamamıştır. Bu durumun, şirketlerin finansal politikaları (*borsaya açık*

olma durumu vb.) ve rekabet politikaları (*mali verilerin gizliliği vb.)* ile ilgili olduđuna inanılmaktadır. Tm bu nedenlerden tr bu alıřmanın sonularını karřılařtırmak iin benzer alıřmalar bulunamamıřtır. Bu durumun stesinden gelebilmek iin hava yolu řirketleri ile niversiteler arasında ok disiplinli bir yapıya sahip iřbirliđi modellerinin geliřtirilmesi teřvik edilmelidir.



KAYNAKLAR

- Abdi, A., 2016, Three Types of Machine Learning Algorithms, <http://rgdoi.net/10.13140/RG.2.2.26209.10088>, [Ziyaret Tarihi: 20.04.2020].
- Aiping Wang, Guowei Wan, Zhiqian Cheng, Sikun Li, 2009, An incremental extremely random forest classifier for online learning and tracking, *2009 16th IEEE International Conference on Image Processing (ICIP)*, 7-10 November 2009 Egypt, IEEE, 1449-1452.
- Air Transport Action Group (ATAG), 2011, *The Economic & Social Benefits of Air Transport*, http://www.icao.int/Meetings/wrdss2011/Documents/JointWorkshop2005/ATAG_SocialBenefitsAirTransport.pdf, [Ziyaret Tarihi: 10.02.2017].
- Airbus, 2020a, *A320ceo Technical Details*, <https://www.airbus.com/aircraft/passenger-aircraft/a320-family/a320ceo.html>, [Ziyaret Tarihi: 31.05.2021].
- Airbus, 2020b, *A320neo Technical Details*, <https://www.airbus.com/aircraft/passenger-aircraft/a320-family/a320neo.html>, [Ziyaret Tarihi: 31.05.2021].
- Airbus, 2019a, *Aircraft Characteristics Airport And Maintenance Planning*, https://www.airbus.com/content/dam/corporate-topics/publications/backgrounders/techdata/aircraft_characteristics/Airbus-Commercial-Aircraft-AC-A320.pdf, [Ziyaret Tarihi: 05.09.2019].
- Airbus, 2019b, *Building on success (2019)*, <https://www.airbus.com/company/history/aircraft-history/building-on-success--2019-.html>, [Ziyaret Tarihi: 07.08.2019].
- Airbus, 2004, Fuel For Transportation, *Getting to grips with Fuel Economy*, 4, 19-20.
- Akalın, B., Veranyurt, Ü., 2020, Sağlıkta Dijitalleşme ve Yapay Zekâ, *Süleyman Demirel Üniversitesi Sağlık Yönetimi Dergisi*, 2 (2), 128–137.
- Akçapınar Sezer, E., Bozkır, A.S., Yağız, S., Gökçeoğlu, C., 2010, Karar Ağacı Derinliğinin CART Algoritmasında Kestirim Kapasitesine Etkisi: Bir Tünel Açma Makinesinin İlerleme Hızı Üzerinde Uygulama, *Akıllı Sistemlerde Yenilikler ve Uygulamaları Sempozyumu*, 2010 Kayseri.
- Akdeniz, F., 2010, *Olasılık ve İstatistik*, Nobel Kitabevi, Adana, ISBN: 978-605-397-066-8.
- Akpınar, H., 2014, *Data: Veri Madenciliği - Veri Analizi*, Papatya Bilim Yayınevi, İstanbul, ISBN: 978-605-4220-81-6.
- Albayrak, A.S., Koltan Yılmaz, Ş., 2009, Veri Madenciliği: Karar Ağacı Algoritmaları ve İMKB Verileri Üzerine Bir Uygulama, *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 14 (1), 31–52.
- Ali, S.A.H.O., 2010, *Optimization of Economic Fuelling for The Fokker 70/100*, https://beta.vu.nl/nl/Images/stageverslag-aithaddououali_tcm235-159729.doc, [Ziyaret Tarihi: 28.08.2019].

- Allaire, J. J., 2021, *rsconnect: Deployment Interface for R Markdown Documents and Shiny Applications*, <https://cran.r-project.org/web/packages/rsconnect/index.html>, [Ziyaret Tarihi: 07.07.2021].
- Alp, S., 2019, *Yapay Sinir Ağları*, Makine Öğrenmesinde Sınıflandırma Yöntemleri ve R Uygulamaları, In: Alp, S., Öz, E. (ed.), Bölüm 5, Nobel, Ankara, ISBN: 978-605-033-129-5, 113-144.
- Alpaydın, E., 2020, *Introduction to Machine Learning*, 4th ed, MIT Press, London, ISBN: 978-0-262-04379-3.
- Alpaydın, E., 2010, *Introduction to Machine Learning*, 2nd ed. MIT Press, London, ISBN: 978-0-262-01243-0.
- Ankara Üniversitesi, 2019, *Sinir Sistemi*, https://acikders.ankara.edu.tr/pluginfile.php/19311/mod_resource/content/1/4.%20S%C4%B0N%C4%B0R%20S%C4%B0STEM%C4%B0.pdf, [Ziyaret Tarihi: 05.11.2019].
- Arslan, K., 2020. Eğitimde Yapay Zeka ve Uygulamaları, *Batı Anadolu Eğitim Bilimleri Dergisi*, 11 (1), 71–88.
- Atalay, M., Çelik, E., 2017, Büyük Veri Analizinde Yapay Zekâ Ve Makine Öğrenmesi Uygulamaları - Artificial Intelligence And Machine Learning Applications In Big Data Analysis, *Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 9 (22), 155–172.
- Ataseven, B., 2013, Yapay Sinir Ağları ile Öngörü Modellemesi, *Öneri Dergisi*, 10 (39), 101–115.
- Attali, D., 2021, *shinyjs: Easily Improve the User Experience of Your Shiny Apps in Seconds*, <https://cran.r-project.org/web/packages/shinyjs/index.html>, [Ziyaret Tarihi: 07.07.2021].
- Avaner, E.B., 2018, Multidisciplinary Approach to Thought Through the Turing Test III, *Turkish Journal of Bioethics*, 5 (4), 183–192.
- Balaban, M.E., Kartal, E., 2015, *Veri Madenciliği ve Makine Öğrenmesi Temel Algoritmaları ve R Dili ile Uygulamaları*, Çağlayan Kitabevi, İstanbul, ISBN: 9789754360899.
- Barnhart, C., Belobaba, P., Odoni, A.R., 2003, Applications of Operations Research in the Air Transport Industry, *Transportation Science*, 37 (4), 368–391.
- Bartın, B.O., 2017, Araçların Kritik Aralık Kabul Kararlarının Pekiştirmeli Öğrenmeyle Simülasyonu, *Uludağ University Journal of The Faculty of Engineering*, 22 (2), 161–178.
- Basu, T., Murthy, C.A., Chakrabarty, H., 2012, A Tweak on K-Nearest Neighbor Decision Rule, *Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition*, 16-19 July 2012 Las Vegas – USA.
- BBC, 2017, *Dünya'nın ilk robot vatandaşı Suudi Arabistanlı - BBC News Türkçe*, <https://www.bbc.com/turkce/haberler-dunya-41780346>, [Ziyaret Tarihi: 04.11.2019].

- Becerra-Fernandez, I., Zanakis, S.H., Walczak, S., 2002, Knowledge discovery techniques for predicting country investment risk, *Computers & Industrial Engineering*, 43 (4), 787–800.
- Beck, M.W., 2018, NeuralNetTools: Visualization and Analysis Tools for Neural Networks, *Journal of Statistical Software*, 85 (11), 1–20.
- Bektaş, O., 2020, Uçuş Aşamalarının Bölümlendirilmesi: Havacılık Verilerinde Gözetimsiz Öğrenme Uygulaması, *Afyon Kocatepe Üniversitesi Fen Ve Mühendislik Bilimleri Dergisi*, 20 (6), 1178–1186.
- Benko, A., Lányi, C.S., 2009, *History of artificial intelligence*, Encyclopedia of Information Science and Technology, Second Edition, IGI Global, 1759–1762.
- Blaha, M., 2010, *Patterns of Data Modeling*, Taylor & Francis Group, Baton Rouge, United States, ISBN: 978-1-4398-1990-6.
- Boeing, 2019a, *Technical Specs*, <https://www.boeing.com/commercial/737ng/>, [Ziyaret Tarihi: 05.09.2019].
- Boeing, 2019b, *Boeing: Commercial Market Outlook*, <http://www.boeing.com/commercial/market/commercial-market-outlook/>, [Ziyaret Tarihi: 03.03.2019].
- Boeing, 2004, *Fuel Conservation*, http://www.phenom.aero/resources/library/fuel_conservation.pdf, [Ziyaret Tarihi: 31.01.2017].
- Botchkarev, A., 2018, Evaluating Performance of Regression Machine Learning Models Using Multiple Error Metrics in Azure Machine Learning Studio, *Social Science Research Network* (SSRN Scholarly Paper No. ID 3177507), 1-16.
- Brady, C., 2020, *Boeing 737 Detailed Technical Data*, <http://www.b737.org.uk/techspecs/techspecs.htm>, [Ziyaret Tarihi: 31.05.2020].
- Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A., 1984, *Classification and Regression Trees*, Taylor & Francis, USA, ISBN: 978-0-412-04841-8.
- Buchanan, B.G., 2005, A (Very) Brief History of Artificial Intelligence, *AI Magazine*, 26 (4), 53–60.
- Casolla, G., Cuomo, S., Cola, V.S. di, Piccialli, F., 2020, Exploring Unsupervised Learning Techniques for the Internet of Things, *IEEE Transactions on Industrial Informatics*, 16 (4), 2621–2628.
- Chandran, S., Anandarajan, M., 2020, Decision Support System for Selecting Sustainable Alternatives to Conventional Jet Fuel: Impact of Emissions, Production Costs and Carbon Pricing, *Journal of Management and Sustainability*, 10 (1), 83–95.
- Chang, W., 2021, *shinythemes: Themes for Shiny*, <https://cran.r-project.org/web/packages/shinythemes/index.html>, [Ziyaret Tarihi: 07.07.2021].

- Chang, W., Cheng, J., Allaire, J.J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., Borges, B., 2021, *shiny: Web Application Framework for R*, <https://cran.r-project.org/web/packages/shiny/index.html>, [Ziyaret Tarihi: 28.02.2021].
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R., 2000, *CRISP-DM 1.0 Step-by-step data mining guide*, <https://www.kde.cs.uni-kassel.de/wp-content/uploads/lehre/ws2012-13/kdd/files/CRISPWP-0800.pdf>, [Ziyaret Tarihi: 16.11.2019].
- Cheskidov, P., Nikolskaia, K., Minbaleev, A., 2019, Choosing the Reinforcement Learning Method for Modeling DDos Attacks, *2019 International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon)*, 01-04 October 2019 Vladivostok – Russia, IEEE, ISBN: 978-1-7281-0062-3, 1-4.
- Cockpit Apps LLC., 2019, *iFuel – Fuel Tankering App*, <http://www.cockpitapps.com/aviation-apps/ifuel-fuel-tankering-app>, [Ziyaret Tarihi: 28.08.2019].
- Collins, A., Thomas, L., 2012, Comparing reinforcement learning approaches for solving game theoretic models: a dynamic airline pricing game example, *Journal of the Operational Research Society*, 63 (8), 1165–1173.
- Costa, E.P., Lorena, A.C., Carvalho, A.C.P.L.F., Freitas, A.A., 2007, A Review of Performance Evaluation Measures for Hierarchical Classifiers, *Evaluation Methods for Machine Learning II: Papers from the AAAI-2007 Workshop*, USA, AAAI Press, 182–196.
- Demirhan, A., Kiliç, Y.A., Güler, İ., 2010, Tıpta Yapay Zeka Uygulamaları, *Yoğun Bakım Dergisi*, 9 (1), 31–41.
- Domingos, P., 2012, A few useful things to know about machine learning, *Communications of the ACM*, 55 (10), 78–87.
- Dragulescu, A.A., Arendt, C., 2018, *xlsx: Read, Write, Format Excel 2007 and Excel 97/2000/XP/2003 Files*, <https://CRAN.R-project.org/package=xlsx>, [Ziyaret Tarihi: 19.01.2020].
- Edelstein, H.A., 1999, *Introduction to Data Mining and Knowledge Discovery*, 3rd ed, Two Crows Corporation, Potomac, MD, USA, ISBN: 1-892095-02-5.
- Emre, İ.E., Erol, N., Ayhan, Y.İ., Özkan, Y., Erol, Ç., 2019, The analysis of the effects of acute rheumatic fever in childhood on cardiac disease with data mining, *International Journal of Medical Informatics*, 123, 68–75.
- Ersoy, E., Karal, Ö., 2012, Yapay Sinir Ağları ve İnsan Beyni, *İnsan ve Toplum Bilimleri Araştırmaları Dergisi*, 1 (2), 188–205.
- Eurocontrol, 2019, *Fuel Tankering: economic benefits and environmental impact*, https://www.eurocontrol.int/archive_download/all/node/11476, [Ziyaret Tarihi: 17.11.2019].

- European Commission, 2016a, *Reducing emissions from aviation*, https://ec.europa.eu/clima/policies/transport/aviation_en, [Ziyaret Tarihi: 31.05.2020].
- European Commission, 2016b, *EU Emissions Trading System (EU ETS)*, https://ec.europa.eu/clima/policies/ets_en, [Ziyaret Tarihi: 31.05.2020].
- FAA, 2019., *Air Traffic By The Numbers*, https://www.faa.gov/air_traffic/by_the_numbers/, [Ziyaret Tarihi: 08.08.2019].
- FAA, 2017, Aircraft Operating Costs, https://www.faa.gov/regulations_policies/policy_guidance/benefit_cost/media/ECONOMICVALUESFORFAAINVESTMENTANDREGULATORYDECISIONS10032007.pdf, [Ziyaret Tarihi: 03.10.2020].
- Fernandes, L., 2020, *Evaluate the Performance of a Regression Model*, <https://openclassrooms.com/en/courses/6401081-improve-the-performance-of-a-machine-learning-model/6519016-evaluate-the-performance-of-a-regression-model>, [Ziyaret Tarihi: 25.03.2021].
- Filippone, A., 2015, *Case Study Fuel Tankering: Is it Worth it?*, http://www.flight.mace.manchester.ac.uk/aircraft-performance-software/case-studies/by-solution/11_CS_Fuel_Tankering.pdf, [Ziyaret Tarihi: 13.02.2017].
- Filippone, A., 2012, *Advanced Aircraft Flight Performance*, Cambridge University Press, USA, ISBN: 978-1-139-78966-0.
- Fix, E., Hodges, J.L., 1951, *Discriminatory analysis-nonparametric discrimination: consistency properties*, USAF School of Aviation Medicine, Texas.
- Flach, P., 2012, *Machine Learning: The Art and Science of Algorithms that Make Sense of Data*, 1st ed, Cambridge University Press, Glasgow, UK, ISBN: 978-1-107-09639-4.
- Fletcher, T., 2008, Support Vector Machines Explained, https://cling.csd.uwo.ca/cs860/papers/SVM_Explained.pdf, [Ziyaret Tarihi: 27.02.2021].
- Frost, N., 2019, *A common airline cost-saving practice trashes the planet in the process*, <https://qz.com/1746196/fuel-tankering-saves-airlines-money-but-costs-the-planet/>, [Ziyaret Tarihi: 30.05.2020].
- FuelerLinx, 2019, *FuelerLinx - Aviation Jet Fuel Planning & Tankering Software for Enhanced Flight Planning*, <https://fuelerlinx.com/Landing2017/index.html>, [Ziyaret Tarihi: 28.08.2019].
- Goudos, S.K., Athanasiadou, G., 2019, Application of an Ensemble Method to UAV Power Modeling for Cellular Communications, *IEEE Antennas and Wireless Propagation Letters*, 18 (11), 2340–2344.
- Guerreiro Fregnani, J.A.T., Müller, C., Correia, A.R., 2013, A fuel tankering model applied to a domestic airline network, *Journal of Advanced Transportation*, 47 (4), 386–398.

- Gürsakar, N., 2018, *Makine Öğrenmesi*, 1st ed. Dora Yayınevi, Bursa, ISBN: 978-605-247-063-3.
- Gürsoy, N.C., 2020, *Havayolu Fazladan Yakıt Taşıma (Tankering) Stratejisinin Doğrusal Olmayan Programlama İle Modellenmesi ve Bir Uygulama*, (Yüksek Lisans Tezi). Anadolu Üniversitesi, Eskişehir.
- Hachaïchi, Y., Chemingui, Y., Affes, M., 2020. A Policy Gradient Based Reinforcement Learning Method for Supply Chain Management, *4th International Conference on Advanced Systems and Emergent Technologies*, 15-18 December 2020 Tunisia, IEEE, ISBN: 978-1-7281-6357-4, 135-140.
- Haenlein, M., Kaplan, A., 2019, A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence, *California Management Review*, 61 (4), 5–14.
- Han, J., Kamber, M., Pei, J., 2012, *Data Mining: Concepts and Techniques*, 3rd ed., Morgan Kaufmann Publishers, Waltham, MA, USA, ISBN: 978-0-12-381479-1.
- Harrington, P., 2012, *Machine Learning in Action*, Manning Publications, USA, ISBN: 978-1-61729-018-3.
- Hassanat, A.B., Abbadi, M.A., Hassanat, A.A., 2014, Solving the Problem of the K Parameter in the KNN Classifier Using an Ensemble Learning Approach, *International Journal of Computer Science and Information Security*, 12 (8), 33-39.
- Honeywell, 2020a, *Introducing Honeywell Forge*, <https://www.honeywell.com/content/honeywell/us/en/honeywell-forge/aircraft.html>, [Ziyaret Tarihi: 29.05.2020].
- Honeywell, 2020b, *Honeywell Forge Flight Efficiency*, <https://aerospace.honeywell.com/content/aero/en/us/home/learn/services/flight-services/flight-efficiency.html>, [Ziyaret Tarihi: 29.05.2020].
- Huang, F., Xie, G., Xiao, R., 2009, Research on Ensemble Learning, *2009 International Conference on Artificial Intelligence and Computational Intelligence*, 07-08 November 2009 China, IEEE, 249-252.
- Hubert, T., Guo, C., Mouton, C.A., Powers, J.D., 2015, *Tankering fuel on U.S. Air Force transport aircraft: an assessment of cost savings*, http://www.rand.org/content/dam/rand/pubs/research_reports/RR800/RR837/RAND_RR837.pdf, [Ziyaret Tarihi: 13.02.2017].
- Huysal, K., Erkoç, M.F., 2013, *Bilgi Yönetimi İçin Teknolojiler: Yapay Zeka ve Bilgi Tabanlı Sistemler*, Bilgi ve Bilginin Yönetimi, In: Gülseçen, A. (ed.), Bölüm 6, Papatya, İstanbul, ISBN: 978-605-4220-38-0, 145-180.
- Hyndman, R.J., Koehler, A.B., 2006, Another look at measures of forecast accuracy, *International Journal of Forecasting*, 22 (4), 679–688.
- IATA, 2019, *Jet Fuel Price Monitor*, <https://www.iata.org/publications/economics/fuel-monitor/pages/index.aspx>, [Ziyaret Tarihi: 07.08.2019].

- IATA, 2018, *Economic performance of the airline industry*, <http://www.iata.org/publications/economics/Reports/Industry-Econ-Performance/IATA-Economic-Performance-of-the-Industry-end-year-2018-report.pdf>, [Ziyaret Tarihi: 22.05.2021].
- IBM, 2021a, *What is Unsupervised Learning?*, <https://www.ibm.com/cloud/learn/unsupervised-learning>, [Ziyaret Tarihi: 21.02.2021].
- IBM, 2021b, *Big Data Analytics*, <https://www.ibm.com/analytics/hadoop/big-data-analytics>, [Ziyaret Tarihi: 13.03.2021].
- IBM, 2019, *About | IBM*, <https://www.ibm.com/watson/about>, [Ziyaret Tarihi: 04.11.2019].
- ICAO, 2020, *Trends in Emissions that affect Climate Change*, https://www.icao.int/environmental-protection/Pages/ClimateChange_Trends.aspx, [Ziyaret Tarihi: 31.05.2020].
- ICAO, 2010, *Annex 6 - Operation Of Aircraft Part I, International Commercial Air Transport - Aeroplanes*, 9th ed., International Civil Aviation Organization, Canada.
- Jarosz, Q., 2009, *Structure of A Neuron*, https://commons.wikimedia.org/wiki/File:Neuron_Hand-tuned.svg, [Ziyaret Tarihi: 01.03.2020].
- Javatpoint, 2020, *Machine Learning Random Forest Algorithm – Javatpoint*, <https://www.javatpoint.com/machine-learning-random-forest-algorithm>, [Ziyaret Tarihi: 07.09.2020].
- Jayalakshmi, T., Santhakumaran, A., 2011, Statistical Normalization and Back Propagation for Classification, *International Journal of Computer Theory and Engineering*, 3 (1), 89–93.
- Jin, X., Wah, B.W., Cheng, X., Wang, Y., 2015, Significance and Challenges of Big Data Research. Big Data Research, *Visions on Big Data*, 2 (2), 59–64.
- Kaelbling, L.P., Littman, M.L., Moore, A.W., 1996, Reinforcement Learning: A Survey, *Journal of Artificial Intelligence Research*, 4, 237–285.
- Karatzoglou, A., Smola, A., Hornik, K., Zeileis, A., 2004, *kernlab - An S4 Package for Kernel Methods in R*. *Journal of Statistical Software*, 11 (9), 1–20.
- Kartal, E., Balaban, M.E., 2018, Machine learning techniques in cardiac risk assessment, *Turkish Journal of Thoracic and Cardiovascular Surgery*, 26 (3), 394–401.
- Kaya, T., 2007, *Küresel ısınma : etkileri ve önlemleri*, Ferman Yayınları, İstanbul, ISBN: 978-975-01287-0-7.
- Kheng, Capt.G.K., 2021, *The optimum route to significant savings*, <https://www.lhsystems.com/static/fa7779d81c7c1a345956aa374edda6d3/case-study-singapore-en.pdf>, [Ziyaret Tarihi: 04.05.2021].
- Koçoğlu, F.Ö., 2017, *Müşteri Kayıp Analizi Probleminin Çözümünde Analitik Yaklaşımlar*, Doktora Tezi, İstanbul Üniversitesi.

- Koçoğlu, F.Ö., Özcan, T., Baray, Ş.A., 2018, *Ayrılma Eğilimli Müşterilerin Belirlenmesi Probleminde Veri Madenciliği Algoritmalarının Performans Analizi*, Yönetim Bilişim Sistemlerinde Gündem, In: Gülseçen, S., Selçukhan Erol, Ç., Ayvaz Reis, Z., Gezer M. (ed.), Çağlayan Kitabevi, İstanbul, ISBN: 978-975-436-152-0, 203-221.
- Kökver, Y., Barışçi, N., Çiftçi, A., Ekmekçi, Y., 2014, Hipertansiyona Etki Eden Faktörlerin Veri Madenciliği Yöntemleriyle İncelenmesi, *Engineering Sciences*, 9 (2), 15–25.
- Korvesis, P., 2017, *Machine Learning for Predictive Maintenance in Aviation*, Doktora Tezi, Université Paris-Saclay.
- Krenker, A., Bester, J., Kos, A., 2011, *Introduction to the Artificial Neural Networks*, Artificial Neural Networks - Methodological Advances and Biomedical Applications, In: Suzuki, K. (ed.), InTech, ISBN: 978-953-307-243-2.
- Küçüksille, E., 2009, *Veri Madenciliği Süreci Kullanılarak Portföy Performansının Değerlendirilmesi ve İMKB Hisse Senetleri Piyasasında Bir Uygulama*, Doktora Tezi, Süleyman Demirel Üniversitesi Sosyal Bilimler Enstitüsü.
- Küçüksille, E.U., Ateş, N., 2013, Destek Vektör Makineleri ile Yaramaz Elektronik Postaların Filtrelenmesi Spam e-mail Filtering Using Support Vector Machine, *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*, 6 (1), 81–87.
- Kuhn, M., 2020, *caret: Classification and Regression Training*, <https://cran.r-project.org/web/packages/caret/caret.pdf>, [Ziyaret Tarihi: 10.05.2020].
- Kurklu, E., Morris, R.A., Oza, N., 2008, Machine Learning for Earth Observation Flight Planning Optimization, *NASA Earth Science Technology Conference*.
- Laerd Statistics, 2013, *Spearman's Correlation in Stata*, <https://statistics.laerd.com/stata-tutorials/spearmans-correlation-using-stata.php>, [Ziyaret Tarihi: 28.06.2017].
- Lee, S., Farley, C., Shim, S., Yoo, W.-S., Zhao, Y., Choi, W., 2020, Unsupervised Learning of Deep-Learned Features from Breast Cancer Images, *2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE)*, 26-28 October 2020 USA, IEEE, ISBN: 978-1-7281-9575-9, 740-745.
- Lesinski, W.J., 2011, *Tankering Fuel: A Cost Saving Initiative*, Graduate Research Paper, Graduate School of Engineering and Management, Air Force Institute of Technology, Air University, Air Education and Training Command.
- Li, D., Du, Y., 2017, *Artificial intelligence with uncertainty*, 2nd ed., CRC Press, Taylor & Francis Group, Boca Raton.
- Liaw, A., Wiener, M., 2002, Classification and Regression by randomForest, *R News*, 2 (3), 18–22.
- Lillicrap, T.P., Santoro, A., Marris, L., Akerman, C.J., Hinton, G., 2020, Backpropagation and the brain, *Nature Reviews Neuroscience*, 21 (6), 335–346.

- Lindgren, M., Brynhagen, J., 2012, *Fuel Tankering - for Flygprestanda AB's flight planning system*, Bachelor Dissertation, Department of Computer Engineering, Lund University, Helsingborg.
- Lison, P., 2012, *An introduction to machine learning*, <http://folk.uio.no/plison/pdfs/talks/machinelearning.pdf>, [Ziyaret Tarihi: 30.04.2019].
- Liu, J., Lv, F., Di, P., 2019, Identification of Sunflower Leaf Diseases Based on Random Forest Algorithm, *2019 International Conference on Intelligent Computing, Automation and Systems (ICICAS)*, 06-08 December 2019 China, IEEE, ISBN: 978-1-7281-6107-5, 459-463.
- Mair, C., Kadoda, G., Lefley, M., Phalp, K., Schofield, C., Shepperd, M., Webster, S., 2000, An investigation of machine learning based prediction systems, *Journal of Systems and Software*, 53 (1), 23–29.
- Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., 2020, *e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071)*, TU Wien, <https://cran.r-project.org/web/packages/e1071/index.html>, [Ziyaret Tarihi: 17.05.2020].
- Milborrow, S., 2020, *rpart.plot: Plot "rpart" Models: An Enhanced Version of "plot.rpart."*, <https://CRAN.R-project.org/package=rpart.plot>, [Ziyaret Tarihi: 26.03.2021].
- Mitchell, T.M., 1997, *Machine Learning*, 1st ed., McGraw-Hill Education, USA, ISBN: 0-07-042807-7.
- Morgan, J.N., Sonquist, J.A., 1963, Problems in the Analysis of Survey Data, and a Proposal, *Journal of the American Statistical Association*, 58 (302), 415–434.
- Nabiyev, V.V., 2005, *Yapay Zekâ*, Seçkin Yayıncılık, Ankara.
- Nahayo, Y., Arı, S., 2016, Performance of svm, k-nn and nbc classifiers for text-independent speaker identification with and without modelling through merging models, *Sakarya Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 20 (1), 1–6.
- NASA, 2021, *NASA's Next Mars Rover Is Ready for the Most Precise Landing Yet*, <https://mars.nasa.gov/news/8860/nasas-next-mars-rover-is-ready-for-the-most-precise-landing-yet>, [Ziyaret Tarihi: 15.03.2021].
- Naumann, M., Suhl, L., 2013, How does fuel price uncertainty affect strategic airline planning?, *Operational Research*, 13 (3), 343–362.
- Nilsson, N.J., 1980, *Principles of Artificial Intelligence*, Morgan Kaufmann Publishers, USA, ISBN: 0-934613-10-9.
- OpenAirlines, 2020a, *SkyBreathe Fuel Management Software*, <https://openairlines.com/fuel-management-software/>, [Ziyaret Tarihi: 29.05.2020].

- OpenAirlines, 2020b, *MyFuelCoach, Pilot Engagement App*, <https://openairlines.com/pilot-engagement-app/>, [Ziyaret Tarihi: 29.05.2020].
- Özen, Z., 2016, *Kimlik Doğrulaması için Tuş Vuruş Dinamiklerine Dayalı Bir Güvenlik Sisteminin Yapay Sinir Ağları ile Geliştirilmesi*, Doktora Tezi, İstanbul Üniversitesi Fen Bilimleri Enstitüsü.
- Öztemel, E., 2012, *Yapay Sinir Ağları*, 3rd ed., Papatya Yayıncılık, İstanbul.
- Pirim, H., 2006, YAPAY ZEKÂ, *Journal of Yaşar University*, 1 (1), 81–93.
- ProMare, 2021, *Mayflower Autonomous Ship*, <https://mas400.com/story>, [Ziyaret Tarihi: 16.03.2021].
- R Development Core Team, 2021, *R: A Language and Environment for Statistical Computing*, <https://www.R-project.org/>, [Ziyaret Tarihi: 28.02.2021].
- Rahim Taleqani, A., Nygard, K., Bridgelall, R., Hough, J., 2018, Machine Learning Approach to Cyber Security in Aviation, *2018 IEEE International Conference on Electro/Information Technology (EIT)*, 03-05 May 2018 USA, IEEE, ISBN: 978-1-5386-5399-9, 0147-0152.
- Rao, V.K., 1999, Fuel price risk management using futures, *Journal of Air Transport Management*, 5 (1), 39–44.
- Raschka, S., 2018, *Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning*, <https://arxiv.org/abs/1811.12808>, [Ziyaret Tarihi: 26.03.2021].
- Rinker, T.W., Kurkiewicz, D., 2017, *pacman: Package Management for R*, <http://github.com/trinker/pacman>, [Ziyaret Tarihi: 26.03.2021].
- RStudio, 2021a, *RStudio: Integrated development environment for R*, <http://www.rstudio.com/>, [Ziyaret Tarihi: 28.02.2021].
- RStudio, 2021b, *Shiny*, <https://shiny.rstudio.com>, [Ziyaret Tarihi: 03.04.2021].
- RStudio, 2021c, *shinyapps.io*, <https://www.shinyapps.io>, [Ziyaret Tarihi: 03.04.2021].
- Russell, S.J., Norvig, P., 2010, *Artificial Intelligence : A Modern Approach*, 3rd ed., Pearson Education, USA, ISBN: 978-0-13-604259-4.
- Ryerson, M.S., Hansen, M., Bonn, J., 2011, Fuel consumption and operational performance, *Proceedings of the USA/FAA Air Traffic Management R&D Seminar 2011*, 14-17 June 2011 Germany.
- Salian, I., 2018, *NVIDIA Blog: Supervised Vs. Unsupervised Learning*, <https://blogs.nvidia.com/blog/2018/08/02/supervised-unsupervised-learning>, [Ziyaret Tarihi: 12.04.2020].
- Saravanan, R., Sujatha, P., 2018, A State of Art Techniques on Machine Learning Algorithms: A Perspective of Supervised Learning Approaches in Data Classification, *2018 Second*

International Conference on Intelligent Computing and Control Systems (ICICCS), 14-15 June 2018 India, IEEE, ISBN: 978-1-5386-2843-0, 945-949.

Sarkar, M., Leong, T. Y., 2000, Application of K-nearest neighbors algorithm on breast cancer diagnosis problem, *Proceedings of the American Medical Informatics Association (AMIA) Symposium*, 04-08 November 2000 Los Angeles, Hanley & Belfus, Inc, ISBN: 1-56053-480-X, 759-763.

SAS Institute Inc., 2017, *How do I interpret the Shapiro-Wilk test for normality in JMP?*, <http://www.jmp.com/support/notes/35/406.html>, [Ziyaret Tarihi: 28.06.2017].

Savaş, S., Topaloğlu, N., Yılmaz, M., 2012, Veri Madenciliği ve Türkiye'deki Uygulama Örnekleri, *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 21, 1-23.

Schiffmann, W., Joost, M., Werner, R., 1994, *Optimization of the Backpropagation Algorithm for Training Multilayer Perceptron*, <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.53.6869&rep=rep1&type=pdf>, [Ziyaret Tarihi: 26.03.2021].

Shearer, C., 2000, The CRISP-DM model: the new blueprint for data mining, *Journal of data Warehousing*, 5 (4), 13–22.

Shen, Y., 2010. *Artificial Intelligence*, https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-034-artificial-intelligence-fall-2010/tutorials/MIT6_034F10_tutor03.pdf, [Ziyaret Tarihi: 13.01.2020].

Shobha, G., Rangaswamy, S., Gudivada, V.N., Rao, C.R., 2018, Machine Learning, *Handbook of Statistics*, In: Arni S.R., Rao S., Rao C.R. (ed.), Chapter 8, Elsevier, ISBN: 978-0-444-64042-0, 197–228.

Shriram, S., Sivasankar, E., 2019, Anomaly Detection on Shuttle data using Unsupervised Learning Techniques, *2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)*, 11-12 December 2019 Dubai, IEEE, ISBN: 978-1-7281-3779-7, 221-225.

Silahtaroğlu, G., Canbolat, Z.N., 2020, An Early Prediction And Diagnosis Of Sepsis In Intensive Care Units: An Unsupervised Machine Learning Model, *Mugla Journal of Science and Technology*, 6 (1), 32-40.

Singh, V., Sharma, S.K., 2015, Fuel consumption optimization in air transport: a review, classification, critique, simple meta-analysis, and future research implications, *European Transport Research Review*, 7 (2), 1–24.

Singh, V., Sharma, S.K., Vaibhav, S., 2012, Modeling the civil aircraft operations for the optimization of fuel consumption in Indian air transport industry, *Industrial Engineering Letters*, 2 (7), 20–29.

SKYbrary, 2020, *Fuel - Flight Planning Definitions*, https://www.skybrary.aero/index.php/Fuel_-_Flight_Planning_Definitions, [Ziyaret Tarihi: 21.05.2020].

- Smeureanu, I., Ruxanda, G., Badea, L.M., 2013, Customer segmentation in private banking sector using machine learning Techniques, *Journal of Business Economics and Management*, 14 (5), 923–939.
- Sokolova, M., Lapalme, G., 2009, A systematic analysis of performance measures for classification tasks, *Information Processing & Management*, 45 (4), 427–437.
- Sun, S., Wang, C., Ding, H., Zou, Q., 2020, Machine learning and its applications in plant molecular studies, *Briefings in Functional Genomics*, 19 (1), 40–48.
- TAL Aviation Group, 2021, *Ethiopian Airlines Introduces Lido/Flight 4D into its System | News | TAL Aviation*, <https://www.talaviation.com/ethiopian-airlines-introduces-lido-flight-4d-into-its-system>, [Ziyaret Tarihi: 13.02.2021].
- Talabis, M.R.M., McPherson, R., Miyamoto, I., Martin, J.L., Kaye, D., 2015, *Analytics Defined*, Information Security Analytics, In: Talabis, M.R.M., McPherson, R., Miyamoto, I., Martin, J.L., Kaye, D. (ed.), Chapter 1, Syngress, Boston, ISBN: 978-0-12-800207-0, 1–12.
- T.C. Gümrük ve Ticaret Bakanlığı, 2009, *Gümrük Yönetmeliği* (7) (8), <https://kms.kaysis.gov.tr/Home/Goster/39273>, [Ziyaret Tarihi: 04.06.2020].
- Terzi, Ö., Küçüksille, E.U., Ergin, G., İlker, A., 2011, Veri Madenciliği Süreci Kullanılarak Güneş Işınımı Tahmini, *SDU International Technologic Science*, 3 (2), 29-37.
- Therneau, T., Atkinson, B., 2019, *rpart: Recursive Partitioning and Regression Trees*, <https://CRAN.R-project.org/package=rpart>, [Ziyaret Tarihi: 26.03.2021].
- Touzani, S., Granderson, J., Fernandes, S., 2018, Gradient boosting machine for modeling the energy consumption of commercial buildings, *Energy and Buildings*, 158, 1533–1543.
- TÜBİTAK, 2014, *Nöronlar*, http://www.bilimteknik.tubitak.gov.tr/sites/default/files/posterler/41x55_noronlar_poster.pdf, [Ziyaret Tarihi: 08.11.2019].
- Türk Dil Kurumu, 2019, *TÜRK DİL KURUMU*, http://tdk.gov.tr/index.php?option=com_gts&arama=gts&kelime=zek%C3%A2&uid=55814&guid=TDK.GTS.5901a6b63382e6.22318161, [Ziyaret Tarihi: 03.04.2019].
- Turkson, R.F., Yan, F., Ali, M.K.A., Hu, J., 2016, Artificial neural network applications in the calibration of spark-ignition engines: An overview, *Engineering Science and Technology, an International Journal*, 19 (3), 1346–1359.
- U.S. Energy Information Administration, 2020a, *Petroleum and other liquids CO2 emissions, World, Annual*, <https://www.eia.gov/opendata/qb.php?sdid=INTL.4006-8-WORLD-MMTCD.A>, [Ziyaret Tarihi: 06.03.2020].
- U.S. Energy Information Administration, 2020b, *Petroleum and other liquids CO2 emissions, Turkey, Annual*, <https://www.eia.gov/opendata/qb.php?sdid=INTL.4006-8-TUR-MMTCD.A>, [Ziyaret Tarihi: 06.03.2020].

- U.S. Energy Information Administration, 2019a, *U.S. Gulf Coast Kerosene-Type Jet Fuel Spot Price FOB (Dollars per Gallon)*, https://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=EER_EPIK_PF4_RGC_DPG&f=A, [Ziyaret Tarihi: 17.02.2019].
- U.S. Energy Information Administration, 2019b, *International Energy Statistics - World Regions*, <https://www.eia.gov/opendata/qb.php?sdid=INTL.63-2-WORL-TBPD.A>, [Ziyaret Tarihi: 02.02.2017].
- Uzbaş, B., 2017, *Sayısal Dental Modellerden Otomatik Cinsiyet Tespiti*, Doktora Tezi, Selçuk Üniversitesi Fen Bilimleri Enstitüsü.
- Wei, T., Simko, V., 2016, *corrplot: Visualization of a Correlation Matrix*, <https://CRAN.R-project.org/package=corrplot>, [Ziyaret Tarihi: 03.01.2020].
- Weihs, C., Ligges, U., Luebke, K., Raabe, N., 2005, *klaR Analyzing German Business Cycles, Data Analysis and Decision Support, Studies in Classification, Data Analysis, and Knowledge Organization*, In: Baier, D., Decker, R., Schmidt-Thieme, L. (ed.), Springer, Berlin, ISBN: 978-3-540-26007-3, 335–343.
- Wells, A.T., Wensveen, J.G., 2004, *Air Transportation, A Management Perspective*, Thompson Learning, Belmont: Brooks/Cole.
- Wensveen, J.G., Wells, A.T., 2011, *The Airline Industry*, Air Transportation: A Management Perspective, Ashgate Publishing, 184.
- Wickham, H., 2019, *plyr: Tools for Splitting, Applying and Combining Data*, <https://CRAN.R-project.org/package=plyr>, [Ziyaret Tarihi: 19.01.2020].
- Wigman, L., 2021, *Boeing: Pilotsuz savaş uçağının test uçuşu tamamlandı*, <https://tr.euronews.com/2021/03/02/boeing-ile-avustralya-ordusu-pilotsuz-savas-ucag-n-n-test-ucusunu-tamamlad>, [Ziyaret Tarihi: 02.03.2021].
- Williams, J.K., 2014, Using random forests to diagnose aviation turbulence, *Machine Learning*, 95, 51–70.
- Wirth, R., Hipp, J., 2000, CRISP-DM: Towards a standard process model for data mining, *Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining*, 11-13 April 2000 Manchester, Blackpool, Lancashire, Practical Application Company, ISBN: 1-902426-08-8, 29–40.
- Witten, I.H., Frank, E., Hall, M.A., Pal, C.J., 2017, *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, USA, ISBN: 978-0-12-804291-5.
- Wu, X., Zhu, X., Wu, G., Ding, W., 2014, Data mining with big data, *IEEE Transactions on Knowledge and Data Engineering*, 26 (1), 97–107.
- Yerlikaya-Özkurt, F., Askan, A., 2020, Prediction of potential seismic damage using classification and regression trees: a case study on earthquake damage databases from Turkey, *Natural Hazards*, 103 (3), 3163–3180.

- Yılmaz, İ.G., Kartal, E., Özen, Z., Gülseçen, S., 2021, Prediction of Fuel Tankering in Aviation Industry with Machine Learning Algorithms, *Journal of Aeronautics and Space Technologies*, 14 (1), 19–34.
- Yılmaz, İ.G., 2021, *Prediction Of Profitable Fuel Tankering Status With Machine Learning Techniques In Aviation*, <https://ilkerguvenyilmaz.shinyapps.io/fuel tankering/>, [Ziyaret Tarihi: 03.04.2021].
- Yüksek, A.G., Bircan, H., Zontul, M., Kaynar, O., 2007, Sivas İlinde Yapay Sinir Ağları ile Hava Kalitesi Modelinin Oluşturulması Üzerine Bir Uygulama, *Cumhuriyet Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 8 (1), 97–112.
- Zhou, Z.-H., 2012, *Ensemble Methods: Foundations and Algorithms*, CRC Press LLC, London, United Kingdom, ISBN: 978-1-4398-3005-5.
- Zhu, X., Li, L., 2020, *Improved Flight Time Predictions for Fuel Loading Decisions of Scheduled Flights with a Deep Learning Approach*, <https://arxiv.org/abs/2005.05684>, [Ziyaret Tarihi: 17.05.2020].

ÖZGEÇMİŞ

Kişisel Bilgiler	
Adı Soyadı	İlker Güven YILMAZ
Doğum Yeri	-
Doğum Tarihi	-
Uyruğu	☒ T.C. ☐ Diğer:
Telefon	-
E-Posta Adresi	-
Web Adresi	-

Eğitim Bilgileri	
Lisans	
Üniversite	Sakarya Üniversitesi
Fakülte	Fen-Edebiyat
Bölümü	Matematik
Mezuniyet Yılı	05.06.2012

Yüksek Lisans	
Üniversite	Sakarya Üniversitesi
Enstitü Adı	Fen Bilimleri
Anabilim Dalı	Matematik
Programı	Matematiğin Temelleri ve Matematik Lojik

Doktora	
Üniversite	İstanbul Üniversitesi
Enstitü Adı	Fen Bilimleri Enstitüsü
Anabilim Dalı	Enformatik Anabilim Dalı
Programı	Enformatik Programı

Makale ve Bildiriler	
Yılmaz, İ.G., Kartal, E., Özen, Z., Gülseçen, S., 2021. Prediction of Fuel Tankering in Aviation Industry with Machine Learning Algorithms, <i>Journal of Aeronautics and Space Technologies</i> , 14(1), 19-34.	
Aygün, D , Yılmaz, İ.G. , Gülseçen, S . (2019), Havayolları İçin Karar Vermenin Değeri: Check-in'lerde Yolcu Tercihlerinin Bir Analizi, <i>Journal of Transportation and Logistics</i> , 2 (1), 1-10.	
Yılmaz, İ.G, Aygün, D, Akadal, E, Gülseçen, S. (2019), Göz Hareketlerini İzleme Yöntemiyle Arama Motorlarının Otomatik Tamamlama Özelliğinin Kullanılabilirlik Açısından İncelenmesi, <i>Tasarım Enformatiği</i> , 1(1), 51-61.	

- Aygün D., Yılmaz İ.G., Satı Z., Industry 4.0 Awareness: A Research On Postgraduate Students, *6.Uluslararası Yönetim Bilişim Sistemleri Konferansı*, 09-12 Ekim 2019 İstanbul, Türkiye, 289-301.
- Yılmaz, İ. G., Aygün, D., Kartal, E. ve Özen, Z. (2018), *Otomotiv Sektöründe Twitter Veri Analizi: Pazarda Daha Önde Olabilmek İçin İpuçları*, Yönetim Bilişim Sistemlerinde Gündem, In: S. Gülseçen, Ç. Selçukcan Erol, Z. Ayvaz Reis ve M. Gezer (Ed.), Çağlayan Kitabevi, İstanbul, ISBN:978-975-436-153-7, 49-59.
- Yılmaz, İ.G., Aygün, D. and Tanrıkulu, Z. (2017), Social Media's Perspective on Industry 4.0: A Twitter Analysis, *Social Networking*, 6, 251-261.
- Yılmaz İ.G., Aygün D., Kartal E., Özen Z., Otomotiv Sektöründe Twitter Veri Analizi: Pazarda Daha Önde Olabilmek İçin İpuçları, *4.Uluslararası Yönetim Bilişim Sistemleri Konferansı*, 17-20 Ekim 2017 İstanbul, Türkiye, 57-57.