

**MAKİNE ÖĞRENMESİ ALGORİTMALARI İLE AVRUPA  
HAVALİMANLARI ANALİZİ**

**Pamukkale Üniversitesi  
Sosyal Bilimler Enstitüsü  
Yüksek Lisans Tezi  
İşletme Ana Bilim Dalı  
Sayısal Yöntemler Programı**

**Ahmet Arif ÇOLAKOĞLU**

**DANIŞMAN: Dr. Öğr. Üyesi Hüseyin KOÇAK**

**Haziran 2020  
DENİZLİ**

Bu tezin tasarımı, hazırlanması, yürütülmesi, arařtırmalarının yapılması ve bulgularının analizlerinde bilimsel etięe ve akademik kurallara özenle riayet edildiđini; bu çalışmanın doğrudan birincil ürünü olmayan bulguların, verilerin ve materyallerin bilimsel etięe uygun olarak kaynak gösterildiđini ve alıntı yapılan çalışmalara atıfta bulunulduđunu beyan ederim.

Ahmet Arif ÇOLAKOĐLU



## ÖNSÖZ

Bu tezin hazırlanmasında; İlk danışmanım ve hamilelik sebebi ile danışman değişikliği yapmak zorunda kaldığım fakat tezimin oluşmasında bana rehberlik anlamında çok ciddi katkıları olan Nilsen Kundakcı hocama teşekkür ederim.

Bölümdeki tüm hocalarımız kıymetlidir fakat danışmanınızdan memnunsanız danışman değişikliği tez aşamasındaki bir öğrencinin kabusudur. Bu süreçte Nilsen hocamızdan ayrılmak ne kadar üzücü olsa da bana her türlü kolaylığı sağlayan Hüseyin Koçak hocamızın danışmanım olması benim için büyük bir şans oldu. Kendisi bu süreçte bana danışmandan da öte akademik bir koç ve yol arkadaşı oldu. Ayrıca tezimi hazırlama sürecinde yaptığım tüm acemiliklerde kendisinden tüm akademik hayatımı şekillendirecek tecrübeler edindiğimi belirtmek istiyorum. Yaptığı tüm katkılardan ötürü kendisine teşekkür ederim.

Ayrıca Juri üyelerinden Utku Erdoğan hocama teşekkürü borç bilirim. Tezin geliştirilmesi ve gelecekte yapılabilecek çalışmalara ilişkin verdiği tavsiyeler benim için çok kıymetliydi.

Lisans eğitimimde dersini aldığım başta Oğuz Borat, Muammer Kalyon, Sıbgat Kaçtıoğlu, Berk Ayvaz ve Ertuğrul Çetinsoy olmak üzere tüm bölüm hocalarıma yetişmemde sarf ettikleri emek için teşekkür ederim.

İstanbul'da üniversiteyle birlikte tanıştığım meslek ahlakı aşılamayışı ve insan olmanın ne demek olduğunu öğretmeyi düstur edinen ve buna anlam katan ve bana maddi manevi her türlü desteği veren Hamilik Okulu Vakfı'na ve Lütfi Ertekin başta olmak üzere tüm Hamilik Okulu ailesine teşekkür ederim.

Yüksek Lisans eğitimim boyunca dersini aldığım tüm bölüm hocalarıma ve üniversite dışından akademik yayın araştırmaları, akademik dergileri tanıma ve makale hazırlama gibi akademik açıdan önem arz eden konularda çok ciddi deneyimler edindiğim İhsan Aytekin ve Abdullah Yıldızbaşı hocalarıma teşekkür ederim.

Arka planda her türlü kahrımı çeken, tüm stresli zamanlarımı ve zorlukları paylaştığım ve desteklerini benden esirgemeyen Annem ve Babam başta olmak üzere Kardeşlerim Fatma Sena ve Mehmet Fatih'e ve tüm geniş aileme, akademik kariyerimi

devam ettirmek için sarf ettiğim çabada bana çok anlamlı destekleri olan Ablam Belkıs Sultan Kuşaksız, Teyzem Ayşe Balcı ve Eniştelereim Zafer Kuşaksız ve İbrahim Balcı'ya özel teşekkür ederim.

Denizli'de her ne zaman bir ihtiyacım olsa bana ailemi aratmayan aile dostumuz Nezahat ve Ümit Belen çiftine, Ankara'da beni yalnız bırakmayan teyzem Şule ve eşi Bekir Başpınar çiftine, hemen her konuda fikir ve görüşlerini alabildiğim başta Cafer, Faruk, olmak üzere yakın arkadaşlarıma, bugünlere gelmemde üzerimde emeği olan başta Amcam Bahattin Çolakoğlu, Yengem Havva Çolakoğlu, Dayım Ömer Gürlekçe, Abim Selim Bostan ve yakın zamanda vefat eden Burhanettin Doğanay olmak üzere tüm akrabalarıma teşekkür ederim.

## ÖZET

### MAKİNE ÖĞRENMESİ ALGORİTMALARI İLE AVRUPA HAVALİMANLARI ANALİZİ

ÇOLAKOĞLU, Ahmet Arif  
Yüksek Lisans Tezi  
İşletme ABD  
Sayısal Yöntemler Programı  
Tez Yöneticisi: Dr. Öğr. Üyesi Hüseyin KOÇAK

Haziran 2020, X + 146 Sayfa

Bu çalışma, son yıllarda Türkiye’de ve Dünya’da artan havacılık faaliyetleri göz önüne alınarak her geçen gün sayıları katlanarak artan havalimanlarının daha iyi bir şekilde analiz edilmesi ihtiyacından doğmuştur. Türkiye ve Avrupa havalimanları baz alınarak oluşturulan veri seti kullanılarak makine öğrenmesi gözetmeli ve gözetmesiz öğrenme teknikleri ile Microsoft Azure Machine Learning (ML) Studio üzerinde iki farklı uygulama yapılmıştır. İlk olarak gözetmesiz öğrenme uygulamasında K-Ortalamalar algoritması ile benzer havalimanlarını tespit etmek amacıyla kümeleme yapılmıştır. Optimum küme sayısını belirlemek için Azure ML Studio ve Kaggle online platformlarında Python ve R programlama dilleri ile Simplified Silhouette, Davies Bouldin, Dunn, Average Deviation, Elbow grafik metodundan ve R kütüphanesinin barındırdığı NbClust fonksiyonunun sağladığı 30 farklı metriktten yararlanılmıştır. İkinci olarak gözetmeli öğrenme uygulamasında havalimanı taşınan yolcu sayıları tahmini gerçekleştirilmiştir. Burada Lineer Regresyon, Bayesyen Lineer Regresyon, Karar Ormanı Regresyonu, Artırılmış Karar Ağacı Regresyonu ve Poisson Regresyon algoritmalarından faydalanılmıştır. Ortaya çıkan sonuçlar Ortalama Mutlak Hata, Kök Ortalama Kare Hatası, Bağlı Mutlak Hata, Bağlı Kare Hatası ve Belirlilik Katsayısı metrikleri kullanılarak değerlendirilmiştir.

**Anahtar Kelimeler:** Azure ML Studio, Kaggle, K-Ortalamalar, Lineer Regresyon, Karar Ormanı Regresyonu, Artırılmış Karar Ağacı Regresyonu, Elbow, NbClust

**ABSTRACT**  
**EUROPEAN AIRPORTS ANALYSIS USING MACHINE LEARNING  
ALGORITHMS**

ÇOLAKOĞLU, Ahmet Arif  
Master Thesis  
Business Administration Department  
Quantitative Methods Programme  
Adviser of Thesis: Asst. Prof. Dr. Hüseyin KOÇAK

June 2020, X + 146 Pages

**This study has been arisen to analyse the airports, with incrementally increasing numbers, in a better way considering the constantly improving aviation operations in the world. Two different applications were implemented on the Azure ML Studio Platform employing supervised and unsupervised machine learning techniques as well as using the dataset based on the airports located in Turkey and Europe. Firstly, in the unsupervised learning application, clustering was performed with the K-Means algorithm to determine similar airports. In order to determine the optimum number of clusters, 30 different metrics provided by NbClust function found in R library, and Simplified Silhouette, Davies Bouldin, Dunn, Average Deviation and Elbow graphics method as well as Python and R programming languages on Azure ML Studio and Kaggle platforms were employed. Secondly, in the supervised learning application, the number of passengers carried at the airports were estimated. Therefore, linear Regression, Bayesian Linear Regression, Decision Forest Regression, Boosted Decision Tree Regression, and Poisson Regression were utilized. The results were evaluated using Mean Absolute Error, Root Mean Squared Error, Relative Absolute Error, Relative Squared Error, and Coefficient of Determination.**

**Key Words:** Azure ML Studio, Kaggle, K-Means, Linear Regression, Decision Forest Regression, Boosted Decision Tree Regression, Elbow, Nbclust

## İÇİNDEKİLER

ÖNSÖZ.....	i
ÖZET.....	iii
ABSTRACT.....	iv
İÇİNDEKİLER.....	v
ŞEKİLLER DİZİNİ.....	viii
TABLolar DİZİNİ.....	x
GİRİŞ.....	1

### BİRİNCİ BÖLÜM DİJİTAL DÖNÜŞÜM

1.1 Dijital Dönüşüm ve Bazı Güncel Kavramlar .....	2
1.1.1 Büyük Veri.....	2
1.1.2 Endüstride Dijital Dönüşüm .....	3
1.1.3 Yapay Zekâ .....	6
1.2 Dijital Dönüşümün Havacılık Üzerine Etkisi.....	8

### İKİNCİ BÖLÜM MAKİNE ÖĞRENMESİ VE BULUT BİLİŞİM TEKNOLOJİSİ

2.1 Makine Öğrenmesi .....	10
2.1.1 Makinelere Tam Olarak Nasıl Öğretiyoruz? .....	10
2.1.2 Makine Öğreniminde Kullanılan Adımlar Nelerdir?.....	11
2.1.3 Makine Öğrenmesi Çeşitleri .....	12
2.1.4 Makine Öğrenmesi Uygulamaları.....	14
2.1.5 Makine Öğrenmesi Algoritmaları.....	14
2.1.5.1 Regresyon Algoritmaları (Regression Algorithms) .....	14
2.1.5.2 Örnek Tabanlı Algoritmalar (Instance-based Algorithms) .....	15
2.1.5.3 Düzenleme Algoritmaları (Regularization Algorithms) .....	15
2.1.5.4 Karar Ağacı Algoritmaları (Decision Tree Algorithms).....	16
2.1.5.5 Bayesian Algoritmaları (Bayesian Algorithms).....	16
2.1.5.6 Kümeleme Algoritmaları (Clustering Algorithms).....	17
2.1.5.7 İlişkilendirme Kuralı Öğrenmesi Algoritmaları (Association Rule Learning Algorithms).....	17
2.1.5.8 Yapay Sinir Ağı Algoritmaları (Artificial Neural Network Algorithms) . .....	18
2.1.5.9 Derin Öğrenme Algoritmaları (Deep Learning Algorithms) .....	18
2.1.5.10 Boyutsallık Azaltma Algoritmaları (Dimensionality Reduction Algorithms) .....	19
2.1.5.11 Topluluk Algoritmaları (Ensemble Algorithms).....	19
2.1.5.12 Diğer Algoritmalar .....	20
2.1.5.13 Algoritmalara Genel Bakış.....	22
2.2 Bulut Bilişim Teknolojisi .....	23
2.2.1 Bulut Bilişim Temel Karakteristikleri .....	24
2.2.2 Bulut Bilişim Hizmet Sağlama Modelleri .....	25
2.2.3 Bulut Bilişim Konumlandırma Modelleri.....	26

2.2.3.1	Özel Bulut .....	26
2.2.3.2	Topluluk Bulutu .....	26
2.2.3.3	Genel Bulut .....	26
2.2.3.4	Hibrit Bulut .....	26
2.2.4	Tez Uygulamasında Kullanılan Platform ve Araçlar.....	26
2.2.4.1	Microsoft Azure Machine Learning Studio Platformu .....	26
2.2.4.2	Kaggle Platformu .....	27

## **ÜÇÜNCÜ BÖLÜM LİTERATÜR ARAŞTIRMASI**

3.1	Çeşitli Bilim Dallarında Yapay Zekâ ve Makine Öğrenmesi Araştırmalarının İncelenmesi .....	28
3.2	Havalimanları Kapsamında Yapılan Araştırmaların İncelenmesi.....	32
3.3	Makine Öğrenmesi & Azure ML Kapsamında Tahmine Dayalı Yapılan Araştırmaların İncelenmesi .....	35
3.4	Makine Öğrenmesi & Havalimanları Kapsamında Kümeleme Üzerine Yapılan Araştırmaların İncelenmesi .....	40

## **DÖRDÜNCÜ BÖLÜM METODOLOJİ**

4.1	Verilerin Toplanması ve Derlenmesi Süreci .....	44
4.2	Modelin Eğitimi için Kullanılacak Algoritmaların Seçimi .....	54
4.2.1	Havalimanı Yolcu Sayısı Tahmini için Seçilen Algoritmalar .....	54
4.2.1.1	Lineer Regresyon (Linear Regression) .....	54
4.2.1.2	Bayesyen Lineer Regresyon (Bayesian Linear Regression) .....	56
4.2.1.3	Karar Ormanı Regresyonu (Decision Forest Regression).....	56
4.2.1.4	Artırılmış Karar Ağacı Regresyonu (Boosted Decision Tree Regression) .....	57
4.2.1.5	Poisson Regresyon (Poisson Regression) .....	57
4.2.2	Havalimanları Kümeleme Uygulaması Algoritması.....	58
4.2.2.1	K-Ortalamalar Kümeleme ( K-Means Clustering).....	58
4.3	Modelin Değerlendirilmesinde Kullanılan Matematiksel Parametreler.....	58
4.3.1	Havalimanı Yolcu Sayısı Tahmini Değerlendirme Parametreleri .....	59
4.3.1.1	Ortalama Mutlak Hata (Mean Absolute Error (MAE)).....	59
4.3.1.2	Kök Ortalama Kare Hatası (Root Mean Squared Error (RMSE)) .....	59
4.3.1.3	Bağıl Mutlak Hata (Relative Absolute Error (RAE)) .....	59
4.3.1.4	Bağıl Kare Hatası (Relative Squared Error (RSE)).....	60
4.3.1.5	Belirlilik Katsayısı (Coefficient of Determination (CoD) .....	60
4.3.2	Havalimanları Kümeleme Uygulaması Değerlendirme Parametreleri .....	61



## BEŞİNCİ BÖLÜM UYGULAMA ÖRNEKLERİ

5.1 Benzer Havalimanlarını Tespit Etmeye Yönelik Kümeleme Uygulama Örneği	64
5.1.1 Kümeleme için Veri Seti Gösterimi.....	64
5.1.2 Kümeleme için Model Gösterimi .....	64
5.1.3 Kümeleme için Optimum Küme Sayısı Sonuçları.....	69
5.1.3.1 Azure ML Studio “Sweep Clustering” Modülü ile Optimum Küme Sayısı .....	73
5.1.3.2 Kaggle Platformu üzerinde Elbow Metodu ile Optimum Küme Sayısı	75
5.1.3.3 Kaggle Platformu üzerinde “NbClust” R Programlama Dili Paket İçeriğinde Yer Alan Yöntemler ile Optimum Küme Sayısı.....	80
5.1.4 Kümeleme için Sonuçların Gösterimi.....	94
5.1.5 Kümeleme için Sonuçların Değerlendirilmesi.....	107
5.2 Havalimanları İçin Taşınan Yolcu Sayısı Tahmini Uygulama Örneği .....	109
5.2.1 Taşınan Yolcu Sayısı Tahmini için Veri Seti Gösterimi .....	109
5.2.2 Taşınan Yolcu Sayısı Tahmini için Model Gösterimi .....	109
5.2.3 Taşınan Yolcu Sayısı Tahmini için Sonuçlar .....	116
5.2.4 Taşınan Yolcu Sayısı için Sonuçların Değerlendirilmesi.....	128
<b>SONUÇ.....</b>	<b>130</b>
<b>KAYNAKÇA .....</b>	<b>134</b>
<b>ÖZGEÇMİŞ.....</b>	<b>146</b>

## ŞEKİLLER DİZİNİ

Şekil 1: Endüstri Tarihinin Gelişimi .....	5
Şekil 2: Bir CPS’de İnsan - Makine Etkileşimi .....	5
Şekil 3: Akıllı Fabrikanın Görselleştirmesi.....	6
Şekil 4: Makine Öğretme Süreci .....	10
Şekil 5: Makine Öğrenmesi Çeşitleri .....	12
Şekil 6: Algoritmaların Görsel Açıklamaları .....	21
Şekil 7: Makine Öğrenmesi Algoritmaları Ana Tablo.....	22
Şekil 8: Linear Regresyon Matematiksel Formülü .....	55
Şekil 9: Linear Regresyon Grafikselsel Gösterimi .....	55
Şekil 10: Veri Setinin Harita Üzerinde Görselleştirilmesi .....	67
Şekil 11: Kümeleme Modeli .....	68
Şekil 12: Kümeleme Modeli Seçilen Sütunlar .....	70
Şekil 13: Kümeleme Uygulaması “Euclidean” Metodu Parametreleri .....	71
Şekil 14: Kümeleme Uygulaması “Cosine” Metodu Parametreleri.....	72
Şekil 15: Euclidean Simplified Silhouette .....	73
Şekil 16: Euclidean Davies – Bouldin .....	73
Şekil 17: Euclidean Dunn.....	74
Şekil 18: Euclidean Average Deviation .....	74
Şekil 19: Cosine Simplified Silhouette .....	74
Şekil 20: Cosine Dunn .....	74
Şekil 21: Cosine Average Deviation .....	74
Şekil 22: Cosine Davies – Bouldin .....	74
Şekil 23: Kaggle Veri Seti ilk 5 satır .....	76
Şekil 24: Kaggle Kategorik Değişkenin Çıkartıldığı Veri Seti.....	76
Şekil 25: Kaggle Veri İstatistikleri.....	77
Şekil 26: Kaggle Elbow (Dirsek) Metodu Grafiği 1 .....	79
Şekil 27: Kaggle Elbow (Dirsek) Metodu Grafiği 2 .....	79
Şekil 28: Kaggle Elbow (Dirsek) Metodu Grafiği 3 .....	80
Şekil 29: R Programlama ile Özet İstatistikler.....	81
Şekil 30: R Programlama ile Okunan Veri Seti .....	82
Şekil 31: R Programlama ile Temizlenen Veri Seti .....	82
Şekil 32: NbClust (Euclidean, ward.D2) Optimum Küme Sayısı (3).....	84
Şekil 33: NbClust (Euclidean, Single) Optimum Küme Sayısı (3).....	84
Şekil 34: NbClust (Euclidean, Complete) Optimum Küme Sayısı (3) .....	85
Şekil 35: NbClust (Euclidean, Average) Optimum Küme Sayısı (3) .....	86
Şekil 36: NbClust (Euclidean, Mcquitty) Optimum Küme Sayısı (3) .....	87
Şekil 37: NbClust (Euclidean, Median) Optimum Küme Sayısı (2).....	87
Şekil 38: NbClust (Euclidean, Centroid) Optimum Küme Sayısı (3).....	89
Şekil 39: NbClust (Euclidean, Kmeans) Optimum Küme Sayısı (2).....	89
Şekil 40: NbClust (Manhattan, Average) Optimum Küme Sayısı (4) .....	89
Şekil 41: NbClust (Minkowski, Single) Optimum Küme Sayısı (3) .....	91
Şekil 42: NbClust (Minkowski, Centroid) Optimum Küme Sayısı (3).....	91
Şekil 43: Hubert Endeksi R Programlama Grafikleri .....	92
Şekil 44: “D index” R Programlama Grafikleri .....	92
Şekil 45: Azure ML Studio (Euclidean) 3’lü Kümeleme.....	99

Şekil 46: Azure ML Studio (Euclidean) 5’li Kümeleme .....	100
Şekil 47: Azure ML Studio (Cosine) 3’lü Kümeleme .....	101
Şekil 48: Azure ML (Cosine) 5’li Kümeleme.....	102
Şekil 49: Tahminleme Modeli.....	112
Şekil 50: Tahminlemede Kullanılacak Veri Seti için Belirleyici Sütunların Seçimi ....	113
Şekil 51: Lineer Regresyon Algoritması için Teknik Parametreler .....	114
Şekil 52: Bayesyen Lineer Regresyon Algoritması için Teknik Parametreler .....	115
Şekil 53: Karar Ormanı Algoritması için Teknik Parametreler .....	117
Şekil 54: Artırılmış Karar Ağacı Regresyonu Algoritması için Teknik Parametreler ..	118
Şekil 55: Poisson Regresyon Algoritması için Teknik Parametreler .....	119
Şekil 56: Eğitim Veri Setinin Ayrılması .....	120
Şekil 57: Eğitim Modülüne Taşınan Yolcu Sayısı Sütununun Tanımlanması.....	121
Şekil 58: Azure ML Studio Lineer Regresyon Tahmin Skorları .....	122
Şekil 59: Lineer Regresyon Algoritması Sonuç Grafiği 1 .....	123
Şekil 60: Lineer Regresyon Algoritması Sonuç Grafiği 2 .....	123
Şekil 61: Bayesyen Lineer Regresyon Algoritması Sonuç Grafiği 1.....	124
Şekil 62: Bayesyen Lineer Regresyon Algoritması Sonuç Grafiği 2.....	124
Şekil 63: Karar Ormanı Algoritması Sonuç Grafiği 1 .....	125
Şekil 64: Karar Ormanı Algoritması Sonuç Grafiği 2 .....	125
Şekil 65: Artırılmış Karar Ağacı Algoritması Sonuç Grafiği 1 .....	126
Şekil 66: Artırılmış Karar Ağacı Algoritması Sonuç Grafiği 2 .....	126
Şekil 67: Poisson Regresyon Sonuç Grafiği 1 .....	127
Şekil 68: Poisson Regresyon Sonuç Grafiği 2 .....	127

## TABLOLAR DİZİNİ

Tablo 1: Literatür Taraması Bilim Dalları Bazında .....	31
Tablo 2: Literatür Taraması Havalimanları Bazında .....	35
Tablo 3: Literatür Taraması Makine Öğrenmesi & Azure ML Bazında.....	39
Tablo 4: Literatür Taraması Makine Öğrenmesi & Havalimanları Kümeleme Bazında	43
Tablo 5: Uygulama Örnekleri Veri Seti Özellikleri İncelemesi.....	52
Tablo 6: NbClust Optimum Küme Sayısı Bulma Yöntemleri / Endeksleri .....	63
Tablo 7: Kümeleme Ana Veri Seti Örneği (ilk 10 satır).....	65
Tablo 8: Kümeleme Veri Seti için Özet İstatistikler.....	66
Tablo 9: Seçilen Küme Sayısı Aralığında (2-20) Euclidean (Centroid) Metodu ile Tüm Endeksler için Hesaplanan Endeks Değerleri .....	93
Tablo 10: Euclidean (Centroid) Metodu ile Tüm Endeksler için Optimum Küme Sayısı ve Endeks Değerleri .....	93
Tablo 11: Ülke Bazında Kümelerdeki Havalimanları Sayısı (Euclidean) .....	96
Tablo 12: Ülke Bazında Kümelerdeki Havalimanları Sayısı (Cosine) .....	98
Tablo 13: Türkiye Havalimanları 3'lü Kümeleme .....	104
Tablo 14: Türkiye Havalimanları 5'li Kümeleme .....	106
Tablo 15: Tahminleme Ana Veri Seti Örneği (ilk 10 satır) .....	110
Tablo 16: Tahminleme Veri Seti Hakkında Özet İstatistikler.....	111
Tablo 17: Taşınan Yolcu Sayısı Sonuç Değerlendirme Metrikleri.....	129

## GİRİŞ

Havacılık var olduđu günden bugüne kadar dođası geređi hata kabul etmeyen bir sektör olmuştur. Bu durum, bizlerin havacılıđa bakış açısını her zaman farklı bir noktada tutmuştur. Havacılıđın, insan hayatında belki de en büyük sermaye olan zamanı daha verimli kullanmamızı sađlaması, onu vazgeçilmez kılmakla birlikte meydana gelebilecek en ufak bir problemde insan hayatını doğrudan etkilemesi sebebi ile gerek bilimsel gerek ticari açıdan üzerinde çok fazla düşünüp çıkarımlar yaptığımız bir alan olması, bizi de bu alanda çalışma yapmaya teşvik eden bir unsur olmuştur. Son 10 yılı baz aldığımızda havacılık sektöründeki gelişmeleri tetikleyen ana unsurları teknolojik gelişmeler, dünya nüfusundaki artış ve buna bađlı olarak ortaya çıkan ulaşım ihtiyacı olarak iki kategoriye indirgeyebiliriz. Artan dünya nüfusu ve buna bađlı olarak artan ihtiyaçlarımız herkesin malumu olmakla birlikte esas konumuza gelmeden evvel özellikle son yıllarda yaşadığımız teknolojik gelişmeler konusunu ele almanın yaptığımız bu çalışmaya bütüncül bir bakış açısı katacađı düşünölmektedir. Bu kapsamda öncelikle Dijital Dönüşüm kavramı ve bazı alt başlıkları daha sonra Dijital Dönüşüm'ün Havacılık üzerine etkileri ile bu tezi hazırlamaktaki asıl gaye olan havalimanları analizinde geçmiş çalışmaları, hangi yöntemler kullanıldığı, havalimanı analizlerinde veri seti içerisinde hangi parametrelerin ele alındığı, neden makine öğrenmesi tekniklerinin kullanıldığı açıklanmakta daha sonra makine öğrenmesi algoritmaları kullanarak yapılan analizler paylaşılmaktadır.

# 1 BİRİNCİ BÖLÜM

## DİJİTAL DÖNÜŞÜM

### 1.1 Dijital Dönüşüm ve Bazı Güncel Kavramlar

Geçmişten günümüze yapageldiğimiz hemen her iş kolunda yeni teknolojinin getirdiği yenilikler ile bir adaptasyon sürecinde bulunmaktayız. İnsanoğlu var oldu olalı bir değişim ve dönüşüm içerisindedir ve insanlık yazının icadından sanayi devrimine birçok yıkıcı olarak tabir edebileceğimiz akımlar yaşamıştır. Dijital Dönüşümün de son yüzyılın insanlık tarihine çok büyük değişiklikler getireceği bir akım olarak tanımlandığını söyleyebiliriz. Fakat bu kavramı sadece son yüzyıla sığdırmak da doğru olmayacaktır ve dijital dönüşümü teknoloji dönüşümü değil de iş yapma şeklinin dönüşümü olarak tanımlamak daha doğru olacaktır.

Dijital Dönüşüm perspektifinden bazı kavramların açıklamalarını ve bu çalışmanın gerçekleştirilmesindeki bakış açısı aşağıdaki alt başlıklar ile açıklanabilir.

#### 1.1.1 Büyük Veri

Hayatımızın akışı içerisinde; gelişen teknoloji ile tasarımdan üretime, seyahat planlamadan bankacılık işlemlerine, alışverişten ekonomiye, haberleşme ve iletişimde kısacası günlük hayatımızda hemen her alanda internetin icadı ve halka açılması, kişisel bilgisayarların ve akıllı telefonların yaygınlaşması, çevremizdeki elektronik aygıtların varlığı ile elle yapılan işlemlerin otomasyonu sayesinde yaşam tarzımızın artık çok daha fazla dijital ortamda şekillendiği tüm bu olaylar silsilesini, yekpare bir sistemin parçaları olarak düşündüğümüzde, adeta damarlarımızda akan kan gibi bir sistemin akışında olan tüm bilgi büyük veri olarak isimlendirilmektedir.

İnsanlık var oldu olalı veri vardır ve devletlerin yönetiminden bireysel kararlara kadar her zaman birtakım veriler insanoğlu tarafından bilinçli veya bilinçsiz bir şekilde üretilmiş ve kullanılmıştır. Bugün bu kavram bilgisayar ve internetin icadından sonra geçmişten farklı olarak teknoloji ile birlikte dijital ortamda değerlendirilmektedir. Özellikle son yıllarda dijital ortamda verinin sınıflandırılarak analiz edilmesi, anlamlı ve işlenebilir hale getirilmesinde depolama kabiliyetlerinin artması gibi gelişmeler

neticesinde elde edilen sonuçlar bu kavramı çok daha fazla ele almayı gerektirmektedir. Zira birçok bilimsel kaynağa göre büyük verinin hala çok az bir miktarı işlenebilmektedir. Büyük veri kavramı hakkında yazılan birçok kitap ve makale bulunmaktadır. Daha detaylı bilgi için Mustafa Acungil'in "24 Soruda Dijital Dönüşüm", Prof. Dr. Necmi Gürsakal'ın "Büyük Veri", Viktor Mayer-Schönberger ve Kenneth Cukier'in "Big Data: A Revolution That Will Transform How We Live, Work, and Think" kitapları incelenebilir (Acungil, 2018; Gürsakal, 2017; Mayer-Schönberger ve Cukier, 2013).

### **1.1.2 Endüstride Dijital Dönüşüm**

Endüstriyel üretimde; Endüstri 1.0, 2.0, 3.0 ve 4.0 olarak isimlendirilen süreçlerin tarihsel evrimi ve Endüstri 4.0'ın, girişimcilik sektörünün büyümesi, KOBİ sayılarının hızla artması, Ar-ge alanlarının genişlemesi, organize sanayi alanlarının artması, teknokent sayılarının artması, Home Office çalışan sayısının artması, kadınların iş dünyasında daha fazla yer alması, yeni meslekler ve istihdam gibi konularda nasıl etkili olduğuna ilişkin süreçler ele alınmaktadır.

18.yy. sonlarında ilk buhar makinesinin icadı ile üretimde makineleşme sürecinin yaşandığı dönem 1. Sanayi devrimi veya Endüstri 1.0 olarak isimlendirilmektedir. Elektriğin hayatımıza girmesiyle birlikte mekanik üretimde elektrik enerjisinden yararlanmaya başlanıldığı 20.yy başlangıcından itibaren ilk örneklerini Ford ve Toyota'da gördüğümüz seri üretim dönemi 2. Sanayi devrimi veya Endüstri 2.0 olarak, yarı iletken maddelerin gelişimi ve elektriğin imalatta kullanılması ile birlikte otomasyon sistemlerinin ve bilgi teknolojilerinin hayatımıza girdiği dönem (1970'lerin başı) 3. Sanayi devrimi veya Endüstri 3.0 dönemi ve bugün içinde bulunduğumuz internetin hayatımıza girmesiyle makinelerin dahi birbirleriyle iletişim kurduğu nesnelerin interneti gibi konuları konuştuğumuz teknoloji kavramlarını kolektif bir bütün olarak ele aldığımız siber – fiziksel sitemler dönemi de 4. Sanayi devrimi veya Endüstri 4.0 dönemi olarak isimlendirilmektedir (Kesayak, 2018). Endüstri tarihinin gelişimine ilişkin görsel Şekil 1'de yer almaktadır.

Endüstri 4.0 ifadesi ilk kez Alman hükümeti tarafından 2020 için yüksek teknoloji stratejisine yönelik bir girişimin sonucu olarak Kasım 2011 tarihli bir makalede yayımlandı. Dördüncü sanayi devrimi, endüstriyel üretimdeki ağın geliştirilmesi ve tüm

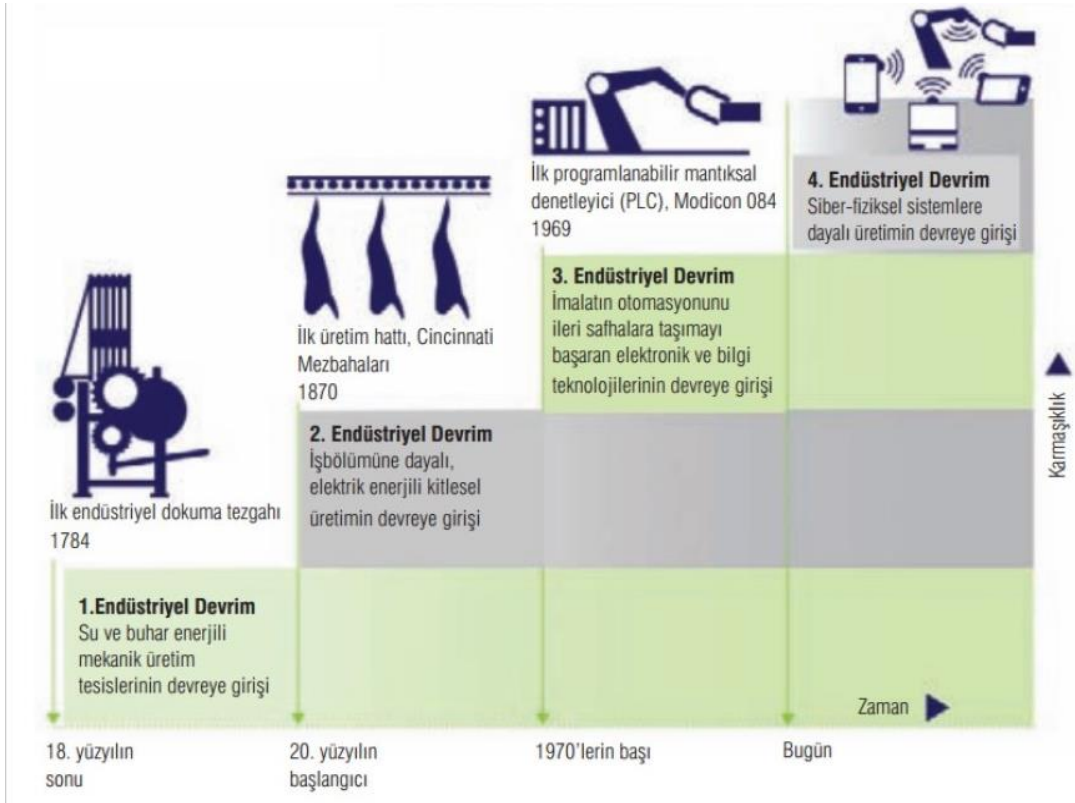
üretim alanlarının bilgisayarlaşmasından ortaya çıkan dijital bir devrim olarak kabul görmüştür (Zhou vd., 2015).

Endüstri 4.0 daha geniş bir perspektiften büyük veri analitiği, nesnelerin interneti, katmanlı üretim, sanal gerçeklik, bulut bilişim, robotik sistemler gibi yeni gelişen çeşitli teknolojilerde ekonomik, çevresel ve sosyal sürdürülebilir imalat sistemlerine öncülük ederek siber - fiziksel sistemler (CPS) ve insan – ekipman ara yüzünü etkin kullanmak için akıllı fabrikalar oluşturmayı amaçlar (Bahrin vd., 2016; Kamble vd., 2018; Lasi vd., 2014; I. Lee ve Lee, 2015; Schmidt vd., 2015). Bir CPS’de insan makine etkileşimi Şekil 2’de temsili olarak açıklanmaktadır.

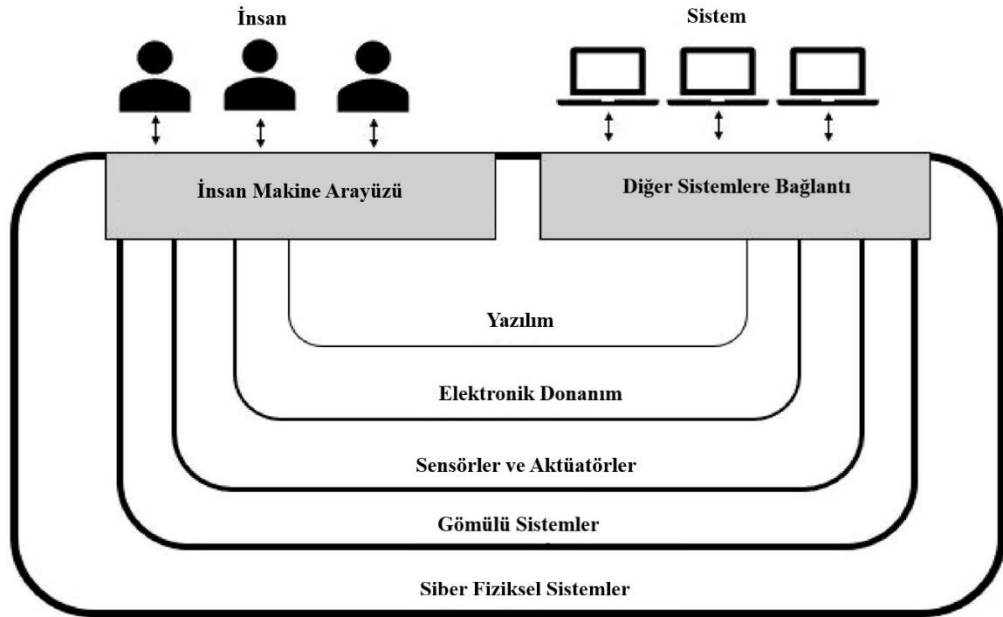
Endüstri 4.0’ın ana noktası, endüstriyel birimler arasında akıllı ağ oluşturma, süreçlerde hareketlilik, esnek bir şekilde birlikte çalışabilirlik, müşteriler ve tedarikçiler ile entegrasyon ve yenilikçi iş modellerinin benimsenmesi anlamına gelen dijital üretimin yani “akıllı fabrika” kavramının ortaya çıkışıdır. Burada ilişkilendirilen önemli husus, CPS’ye dayalı akıllı ağlardır. Endüstri 4.0, organizasyonlar arasında üretkenlik ve verimlilik artışı sağlayan CPS’yi kullanarak dijital ve fiziksel dünyayı bütünleştirir. Makineler, depolama ağları ve birbiriyle bilgi alışverişinde bulunabilecek, eylemleri tetikleyebilecek ve kontrol edebilecek üretim tesisleri CPS’i oluşturmaktadır (Barreto vd., 2017; Jazdi, 2014; Kamble vd., 2018; A. C. Pereira ve Romero, 2017; T. Pereira vd., 2017; Tjahjono vd., 2017). Akıllı fabrikaların yapısı Şekil 3’te yer almaktadır.

Endüstri 4.0, büyük veri, nesnelerin interneti (IoT), ve yapay zekanın (AI) kullanımını bir bütün olarak entegre eder ve organizasyonların ekonomik ve sosyal faydalar elde etmeleri için muazzam bir potansiyele sahiptir (Kagermann vd., 2013; Kamble vd., 2018).

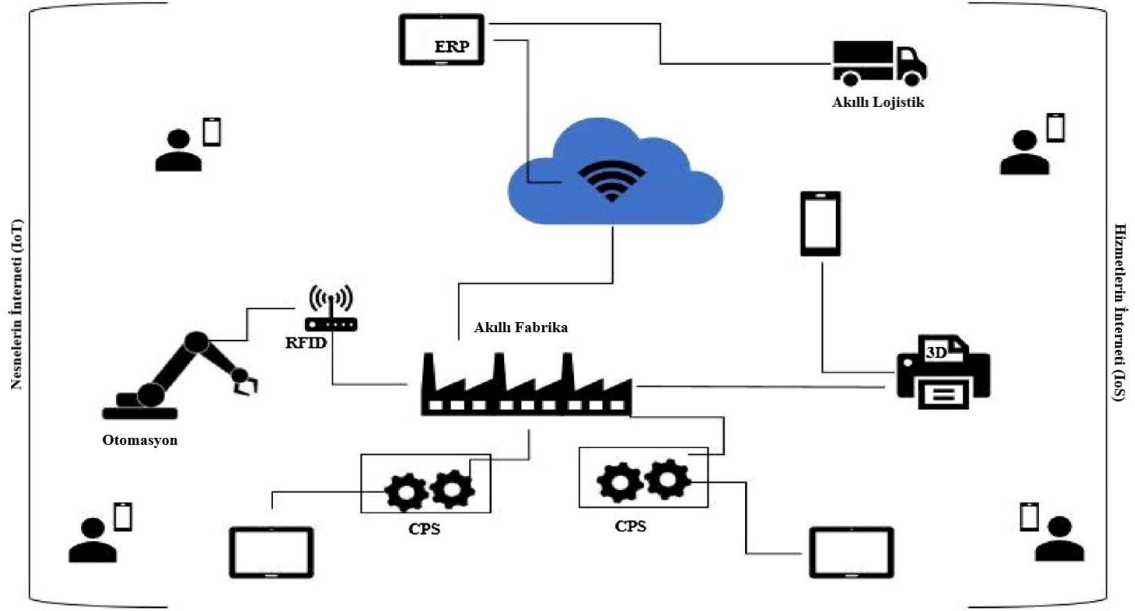




Şekil 1: Endüstri Tarihinin Gelişimi



Şekil 2: Bir CPS'de İnsan - Makine Etkileşimi



Şekil 3: Akıllı Fabrikanın Görselleştirilmesi

### 1.1.3 Yapay Zekâ

Bilimsel olarak modern bilim tarihini incelediğimizde yapay zekâ kavramının ilk kez ortaya çıktığı 1940 ve 1950 yılları, Fen bilimlerinden; matematik, mühendislik Sosyal bilimlerden; psikoloji, siyaset bilimi gibi farklı disiplinlerle çalışma yapan bilim adamlarının bir araya gelerek insan gibi düşünebilen, çalışabilen makineler yapma hayalinden yola çıkarak “yapay bir beyin oluşturma” hedefine yönelik ilk somut adımlara tanıklık etmiştir (Exastax, 2017). Alan Turing bir makinenin düşünüp düşünemediğini ortaya çıkarabilen bir başka deyişle makinelerin zeki olup olmadığını ayırt etmek amacı ile bir test geliştirmiştir. Bu test “Computing Machinery and Intelligence” adlı makalede yayımlanmış ve daha sonra Turing Testi diye anılmıştır. (Turing, 1950). Turing Testi, bir insanın karşısına fiziksel olarak görmeksizin etkileşimde bulunabileceği bir makine koyulduğu zaman bu makinenin insan olup olmadığı ayırt edilemiyorsa o makinenin düşünebilen zeki bir makine olduğu çıkarımını yapmamızı sağlayan bir testtir. Yapay Zekâ’nın akademik bir disiplin olarak kurulması ise 1956 yılında Alan Turing ve bu konuda gerçekleştirilen diğer çalışmalar hakkında John McCarthy tarafından düzenlenen akademik konferansta katılımcıların ortak görüşleri ile bu alandaki çalışmaların daha ileri bir seviyeye çıkması gerekliliği gündeme getirilmiştir. (Semih, 2018). Bu hareket de

Yapay Zekâ'nın 1956 yılında akademik bir disiplin olarak kurulmasına öncülük etmiştir (Köse, 2017).

Kurulduğu günden bugüne kadar yapay zeka hakkında bir çok tartışma ele alınmış ve farklı disiplinlerden bir çok bilim insanının katkıları ile yeni sentezler üretilmiştir. Günümüzde birçok üniversitede ders kitabı olarak okutulan kitaplarda yapay zeka 4 farklı yaklaşım üzerine oturtulmaktadır.

- İnsan gibi düşünebilen (Thinking Humanly),
- İnsan gibi davranabilen (Acting Humanly),
- Rasyonel düşünebilen (Thinking Rationally),
- Rasyonel davranabilen (Acting Rationally),

Bu kavramları ortaya koymak karar verme eyleminin gerçekleştirilmesinde duygusal tutarlılığın insan davranışına, rasyonel düşünce ve rasyonel davranış kavramlarına etkisini değerlendirmek açısından oldukça önemlidir. Daha detaylı bilgi için Stuart J. Russell ve Peter Norvig'in kaleme aldığı "Artificial Intelligence A Modern Approach" kitabı incelenebilir (Russell ve Norvig, 2010).

Yapay Zekâ; bir durumda yanıt vermek ve davranış göstermek için insanı taklit edecek bir sistem veya yazılım geliştiren bir bilimdir. Son derece geniş kapsamlı bir alan olan Yapay Zekâ, hedeflerini çoklu parçalara ayırmıştır. Daha sonra her bir parça kendi problemini çözmek için ayrı bir çalışma alanı haline gelmiştir (Sakthi, 2017; Staub vd., 2015).

Bu alanlar genel olarak aşağıda listelenmiştir.

1. Muhakeme
2. Bilgi Gösterimi / Temsili
3. Otomatik Planlama ve Çizelgeleme
4. Makine Öğrenmesi
5. Doğal Dil İşleme
6. Bilgisayarla Görme / Görüntü algılama
7. Robotik
8. Genel Zekâ veya Güçlü Yapay Zekâ

Listede yer alan Makine Öğrenmesi alanı, bir makine veya yazılımın karşılaşılabileceği sorunları çözme konusunda kendi kendine öğrenmesine yardımcı olacak bir yapay zeka hedefinden ortaya çıkmıştır (Sakthi, 2017).

Kendi kendine öğrenme algoritmalarının geliştirilmesini içeren bir bilim olan Makine öğrenmesi bu algoritmaları geliştirmek için istatistik biliminden faydalanır (Sakthi, 2017).

Makine öğrenmesi, Carnegie Mellon Üniversitesi'nde Profesör ve Eski Makine Öğrenmesi Bölüm Başkanı Tom M. Mitchell'in basitçe ortaya koyduğu sanatsal bir zekâ dalıdır. "Makine öğrenmesi, deneyimle otomatik olarak gelişen bilgisayar algoritmalarının incelenmesidir." (Mitchell, 2006). Yapay Zekâ'ya ulaşmayı umduğumuz yollardan birisi olan Makine Öğrenmesi, verileri genel kalıplarla inceleyerek, karşılaştırarak ve nüansları keşfederek büyük veri setleriyle çalışmaya dayanmaktadır (Roberto, 2018).

## **1.2 Dijital Dönüşümün Havacılık Üzerine Etkisi**

Havacılık ve Uzay her zaman güncel teknolojiyi dinamik olarak bünyesinde hayata geçiren ve ilkleri gördüğümüz bir alan olmuştur. Hemen her bilginin dijital ortamlarda kayıt altına alındığı bir çağda havacılık gibi önemli bir sektörde oluşan büyük verinin, analiz edilmesi ve yönetilmesi tam da dijital dönüşümün bir konusu olmaktadır.

Bugün de düşündüğümüz zaman hava taşımacılığı özelinde havalimanlarının teknik kabiliyetleri, yolcu kapasiteleri, havalimanlarının bulunduğu şehir nüfusu gibi birçok parametre büyük veriyi oluşturmaktadır. Örneğin geçmiş dönem bilimsel çalışmalarda kısıtlı parametreler ile kısıtlı bir coğrafyada belli sayıda havalimanının analizini yapabiliyorken bugün dijital dönüşümün bize sağladığı bulut bilişim sayesinde büyük veriye kişisel bilgisayarlarımızda depolamaksızın erişebilmekte, uygun parametreleri belirleyerek coğrafya farkı gözetmeksizin veri setleri oluşturabilmekte ve makine öğrenmesi gibi matematiksel teknikler ile bu verileri ister makro ölçekte ister mikro ölçekte çıkarımlar yapabilmek amacı ile kullanabilmekteyiz. Bu bize geçmişte hiç olmadığı kadar esneklik sağlamakla beraber birçok limitimizi de ortadan kaldırmaktadır. Ayrıca esneklik kabiliyetimizin artması ve birçok limitimizin ortadan kalkması ile gerçek

veriler üzerinde işlem yapma kabiliyetimiz artmıştır. Dolayısıyla gerçeğe daha yakın sonuçlar üretilmektedir.

Havacılık sektöründe ortaya çıkan maliyetler, iş gücü ve zaman kaybı açısından düşündüğümüzde Dijital Dönüşüm çerçevesinde büyük veri, makine öğrenmesi teknikleri ve diğer güncel teknoloji elemanları ile sağlanması amaçlanan dijital optimizasyon sadece havacılık sektöründe değil hayatın hemen her alanında birçok kolaylık sağlamaktadır.

## 2 İKİNCİ BÖLÜM

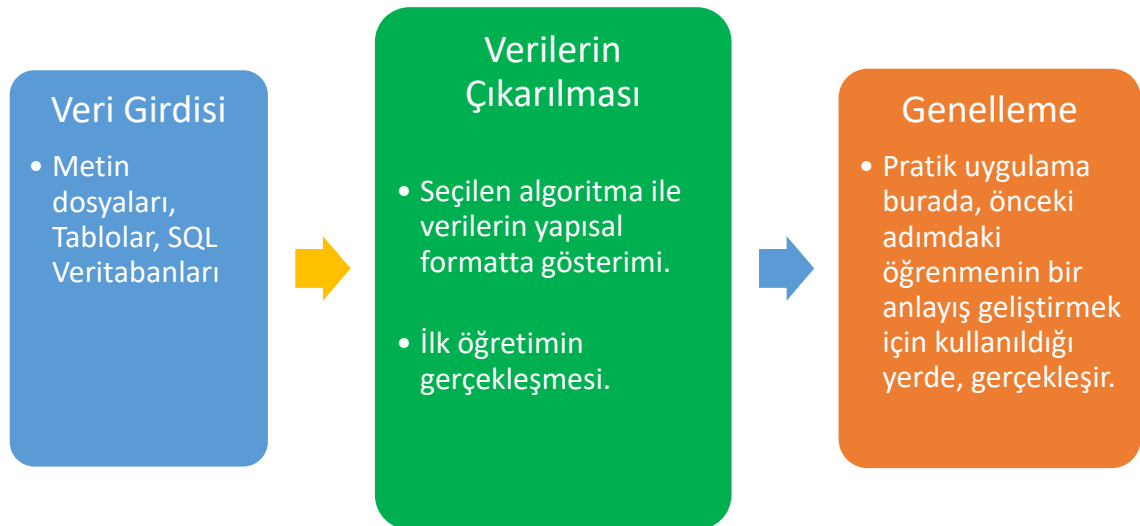
### MAKİNE ÖĞRENMESİ VE BULUT BİLİŞİM TEKNOLOJİSİ

#### 2.1 Makine Öğrenmesi

Bu bölümde son yıllarda üzerine en çok yayın yapılan alanlardan birisi olan aynı zamanda bu tezde analiz yöntemi olarak kullanılan Makine Öğrenmesi algoritmalarının sınıflandırılmasına ve bazı algoritmaların açıklamalarına yer verilecektir. Fakat bu sınıflandırmalara başlamadan önce konunun somut olarak anlaşılması için bir makineye öğretme adımları teknik detayları ile ele alınacaktır.

##### 2.1.1 Makinelere Tam Olarak Nasıl Öğretiyoruz?

Makinelere öğretme işi, her kademenin makinenin daha iyi bir versiyonunu oluşturduğu yapısal bir süreç içerir. Basitleştirme amacıyla, makine öğretme işlemi 3 parçaya ayrılabilir. 3 Parçadan oluşan bu süreç Şekil 4'te gösterilmektedir.



Şekil 4: Makine Öğretme Süreci

Bu 3 adım, makinenin bütünsel öğreniminin verilen görevi eşit derecede yerine getirmesini sağlar ve makinenin başarısı iki faktöre bağlıdır:

1. Çıkarılan verilerin genelleşmesinin ne kadar iyi gerçekleştiği.

2. Makinenin gelecekteki eylem sürecini tahmin etmek için öğrenmeyi pratik kullanıma sokabilmesi ne kadar iyi (Jain, 2015).

### 2.1.2 Makine Öğreniminde Kullanılan Adımlar Nelerdir?

Bir makine öğrenmesi görevi gerçekleştirmek için kullanılan 5 temel adım bulunmaktadır.

1. **Veri Toplama:** Gelecekteki öğrenmenin temelini Excel, Access, metin dosyaları vb. ile ham geçmiş verilerini toplamak gelecekteki öğrenme için temel unsurdur. Bu verilerin çeşitliliği ve miktarı ne kadar zengin olursa, makine öğrenmesi çıktıları da o kadar tutarlı olur (Jain, 2015).
2. **Verilerin Hazırlanması:** Herhangi bir analitik işlem, kullanılan verilerin kalitesine bağlıdır. Verilerin kalitesini belirlemek, eksik veriler ve aykırı değerlerin işlenmesi gibi sorunları çözmek ve sonrasında adımlar atmak için zaman harcanması gerekir. Keşif analizi verilerin nüanslarını incelemek ve içerik kalitesini arttırmak için bir yöntem olarak kullanılabilir (Jain, 2015; Sunil, 2016).
3. **Bir Modelin Eğitimi:** Bu adım, uygun algoritmanın seçilmesini ve verilerin model biçiminde gösterilmesini içerir. Temizlenen veriler iki bölüme ayrılmıştır: Eğitim- Test (ön koşullara bağlı oran), İlk bölüm (eğitim verileri) modeli geliştirmek için kullanılır. İkinci kısım (test verileri) referans olarak kullanılmıştır (Jain, 2015).
4. **Modeli Değerlendirme:** Doğruluğu test etmek için, verilerin ikinci kısmı (bekletme / test verileri) kullanılır. Bu adım, sonucu temel alan algoritma seçimindeki kesinliği belirler. Modelin doğruluğunu kontrol etmek için daha iyi bir test yapmak, model oluşturma sırasında hiç kullanılmamış veriler üzerindeki performansı görmek içindir (Jain, 2015).
5. **Performansı İyileştirme:** Elden edilen sonuçlara göre verimliliği arttırmak veya farklı bir model seçmek gerekebilir. Bu yüzden, veri toplama ve özellikle hazırlama işi, önemli miktarda zaman harcamayı gerektirir (Jain, 2015).

Herhangi bir modelde, bu 5 adım tekniği yapılandırmak için kullanılabilir.

### 2.1.3 Makine Öğrenmesi Çeşitleri

Makine Öğrenmesi algoritmaları üç kategori altında sınıflandırılmıştır. Bu sınıflandırma öğrenme stilleri baz alınarak yapılmıştır. Bunlar Gözetmeli / Denetimli Öğrenme (Supervised Learning), Gözetmesiz / Denetimsiz Öğrenme (Unsupervised Learning) ve Takviyeli / Kuvvetlendirmeli Öğrenme (Reinforcement Learning) yöntemleridir (Filiz, 2017; J. H. Lee vd., 2018). Bu üç kategori Şekil 5'te yer almaktadır.



Şekil 5: Makine Öğrenmesi Çeşitleri

**Gözetmeli Öğrenme (Supervised Learning / Predictive Models):** Sistem farklı girdilerin bir araya getirilmesi ile oluşturulan veri seti ve modelin değerlendirme yaparak ürettiği hedef sonuçları karşılaştırır. Hatalar sistem tarafından minimize edilerek optimumu yakalamak esastır (Jason, 2013). Tarihsel verilere dayanarak gelecekteki sonucu tahmin etmek amacı ile kullanılan tahmini modeller olarak da isimlendirilen gözetmeli öğrenmede genellikle en baştan ne öğrenilmesi gerektiği ve nasıl öğrenilmesi gerektiği konusunda net talimatlar verilmektedir (Jain, 2015).

Örneğin: Denetimli Öğrenme'yi bir pazarlama şirketi hangi müşterilerin zarar verebileceğini bulmaya çalışırken kullanır. Toplam Sigorta Değerini belirlemek amacıyla deprem, kasırga gibi tehlikelerin oluşma olasılığını tahmin etmek için de kullanabilir. Kullanılan bazı algoritma örnekleri: En yakın komşu, Naive Bayes, Karar Ağaçları, Regresyon vb. (Goodfellow vd., 2016).

**Gözetmesiz Öğrenme (Unsupervised Learning / Descriptive Models):** Sistemi oluşturan veri setinde hedef belirtmeksizin, modelin verilen parametrelere ait girdileri



kendi içinde değerlendirerek bir şablon oluşturması beklenmektedir. (Jason, 2013) Hedef belirlenmemiş ve tek bir özelliğin diğerinden önemli olmadığı açıklayıcı modelleri eğitmek için kullanılmaktadır (Jain, 2015).

Gözetmesiz öğrenme durumu şu şekilde olabilir: Bir perakendeci, ürünlerin kombinasyonunun ne olduğunu öğrenmek istediğinde, müşteriler daha sık satın alma eğilimindedir. Ayrıca, ilaç endüstrisinde, diyabetle birlikte hangi hastalıkların ortaya çıkabileceğini tahmin etmek için denetimsiz öğrenme kullanılabilir. Burada kullanılan algoritma örneği: K- Ortalama / Kümeleme Algoritması (Goodfellow vd., 2016).

**Güçlendirme / Takviyeli Öğrenme (Reinforcement Learning):** Temelinde gözetmeli öğrenmenin yer aldığı bu sistem, modelde hedef parametre çıktılarının ne derecede doğru olduğunu belirten yeni bir hedef parametresi oluşturma mantığına dayanmaktadır (Jason, 2013). Verimliliği (performansı) en üst düzeye çıkarmak amacıyla makinenin iş gereksinimine dayanarak belirli kararlar almak için tek bir sloganla eğitim aldığı bir makine öğrenmesi örneğidir. Güçlendirme öğrenmesinde yer alan fikir şudur: Makine / yazılım acentesi maruz kaldığı ortama bağlı olarak kendisini sürekli olarak eğitir ve işletme problemlerini çözmek için zenginleştirilmiş bilgilerini uygular. Bu sürekli öğrenme süreci, insan uzmanlığının daha az katılımını sağlayarak daha fazla zaman kazandıran bir süreçtir (Jain, 2015; J. H. Lee vd., 2018).

Güçlendirme Öğrenmesinde kullanılan bir algoritma örneği, Markov Karar Sürecidir.

Denetimli Öğrenme ve Takviyeli/Güçlendirme Öğrenmesi arasında ince bir fark vardır. Güçlendirme Öğrenmesi temelde bir çevre ile etkileşime girerek öğrenmeyi içerir. Bir Güçlendirme Öğrenmesi temsilcisi, harici bir gözetmenin/denetçinin örnekler sunduğu denetimli öğrenmeye karşı sürekli deneme ve hata öğrenme sürecinden ziyade geçmiş deneyimlerinden öğrenir (Jain, 2015).

Farkı anlamak için otonom araçlar iyi bir örnek olabilir. Kendi kendine sürüş gerçekleştiren bu arabalar sürekli kararlar vermek için güçlendirme öğrenmesini kullanmaktadır. Hangi rotaya gidiyor? Hangi hızda sürülmekte? gibi sorular çevre ile etkileşime girdikten sonra karar verilen sorulardan bazılarıdır. Denetimli öğrenme için

basit bir tezahür de bir yerden başka bir yere giden bir taksiden ücreti tahmin etmek olacaktır (Jain, 2015).

#### 2.1.4 Makine Öğrenmesi Uygulamaları

Google ve Facebook, ilgili reklamlarını hedef kullanıcılara iletmek için makine öğrenmesini yoğun olarak kullanmaktadır.

**Bankacılık ve Finansal Hizmetler:** Makine öğrenmesi kredi ödemekten veya kredi kartı faturalarından temerrüde düşmesi muhtemel müşterileri tahmin etmek için kullanılabilir. Makine öğrenmesi, bankalara kredi verilebilecek müşterileri ve kredi kartlarını tanımlamaları için yardımcı olacağından, bu çok önemlidir (Jain, 2015).

**Sağlık hizmeti:** Hasta semptomlarına dayanarak ölümcül hastalıkları (örneğin kanser) teşhis etmek ve bunlara benzer türden hastaların geçmiş verileriyle konuşmak için kullanılır (Jain, 2015).

**Perakendecilik:** Perakendecilerin rafa ne tür ürünler ekleyeceği veya çıkaracağı konusunda karar vermelerine yardımcı olan daha hızlı satan ürünleri (hızlı hareket eden) ve yavaş hareket eden ürünleri tanımlamak için kullanılır. Ayrıca, hangi iki/üç veya daha fazla ürünün birlikte sattığını bulmak için makine öğrenme algoritmaları kullanılabilir. Bu, perakendecilerin sadık müşterileri geliştirmelerine ve sürdürmelerine yardımcı olan müşteri sadakati girişimlerini tasarlamak için yapılır (Jain, 2015).

Bu örnekler buzdağının sadece görünen kısmıdır. Makine öğrenmesi pratikte her alanda kapsamlı uygulamalara sahiptir.

#### 2.1.5 Makine Öğrenmesi Algoritmaları

Burada algoritmalar benzerliklerine göre gruplandırılmıştır. Gruplandırılan algoritmalara ilişkin görsel anlatımlar Şekil 6'da ve tüm algoritma isimleri Şekil 7'de yer almaktadır.

##### 2.1.5.1 Regresyon Algoritmaları (Regression Algorithms)

Regresyon, model tarafından yapılan tahminlerde bir hata ölçüsü kullanılarak yinelemeli olarak rafine edilmiş değişkenler arasındaki ilişkinin modellenmesi ile ilgilidir (Jason, 2013).

Regresyon metotları bir istatistik çalışmasıdır ve istatistiksel makine öğrenimine dahil edilmiştir. Bu kafa karıştırıcı olabilir çünkü problem sınıfına ve algoritma sınıfına atıfta bulunmak için regresyon kullanılabilir. Gerçekten, regresyon bir süreçtir (Jason, 2013).

En popüler regresyon algoritmaları:

- Sıradan En Küçük Kareler Regresyonu (OLSR)
- Doğrusal Regresyon (Goodfellow vd., 2016)
- Lojistik regresyon (Goodfellow vd., 2016)
- Kademeli Regresyon
- Çok Değişkenli Uyarlanabilir Regresyon Eğrileri (MARS)
- Yerel Tahmini Dağılım Çizgisi Pürüzsüzleştirme (LOESS)

### **2.1.5.2 Örnek Tabanlı Algoritmalar (Instance-based Algorithms)**

Örnek temelli öğrenme modeli, model için gerekli görülen örneklem veya eğitim verilerinin örnekleri ile birlikte bir karar sorunudur (Jason, 2013).

Bu yöntemler tipik olarak bir örnek veri tabanı oluşturur ve en iyi eşleşmeyi bulmak ve bir tahmin yapmak için benzerlik ölçüsü kullanarak yeni verileri veri tabanı ile karşılaştırır. Bu sebeple, örnek tabanlı yöntemlere ayrıca kazanan yöntemlerle uygulama ve bellek tabanlı öğrenme de denir. Odak noktası depolanan örneklerin temsili ve örnekler arasında kullanılan benzerlik ölçümleridir (Jason, 2013; Wu vd., 2008).

En popüler örnek tabanlı algoritmalar:

- K - En Yakın Komşu (kNN) (Ansari vd., 2018)
- Vektör Ölçümü Öğrenmesi (LVQ)
- Kendi Kendini Düzenleyen Harita (SOM)
- Yerel Ağırlıklı Öğrenme (LWL)

### **2.1.5.3 Düzenleme Algoritmaları (Regularization Algorithms)**

Karmaşıklıklarına göre modelleri cezalandıran, genellikle genellemede daha iyi olan daha basit modelleri tercih eden başka bir yöntem bir uzantı oluşturur (tipik olarak regresyon yöntemleri) (Jason, 2013).

Düzenleme algoritmaları burada ayrıca listelendi, çünkü bunlar popüler, güçlü ve diğer yöntemlerde yapılan basit değişikliklerdir (Jason, 2013).

En popüler düzenleme algoritmaları:

- Ridge Regresyonu
- En Küçük Mutlak Çekme ve Seçme Operatörü (LASSO)
- Elastik Ağ
- En Küçük Açık Regresyonu (LARS)

#### **2.1.5.4 Karar Ağacı Algoritmaları (Decision Tree Algorithms)**

Karar ağacı yöntemleri, verilerdeki özniteliklerin gerçek değerlerine dayalı olarak alınan bir kararlar modeli oluşturur (Jason, 2013; Wu vd., 2008).

Ağaç yapılarında verilen kararlar, belirli bir kayıt için bir tahmin kararı verilene kadar geçerlidir. Karar ağaçları, sınıflandırma ve regresyon sorunları için veriler üzerinde eğitilir. Karar ağaçları genellikle hızlı ve kesindir ve makine öğrenmesinde büyük bir favoridir (Jason, 2013; Wu vd., 2008).

En popüler karar ağacı algoritmaları:

- Sınıflandırma ve Regresyon Ağacı (CART)
- Tekrarlı İkilikçi Ağacı (ID3)
- C4.5 ve C5.0 (güçlü bir yaklaşımın farklı sürümleri)
- Ki-kare Otomatik Etkileşim Tespiti (CHAID)
- Karar Kökü
- M5
- Koşullu Karar Ağaçları

#### **2.1.5.5 Bayesian Algoritmaları (Bayesian Algorithms)**

Bayesian yöntemleri, sınıflandırma ve regresyon gibi problemler için açıkça Bayes Teoremini uygulayan metotlardır (Jason, 2013; Wu vd., 2008).

En popüler Bayesian algoritmaları:

- Naive Bayes

- Gauss Naive Bayes
- Multinomial Naive Bayes
- Ortalama Tek Bağımlılık Tahmin Ediciler (AODE)
- Bayesian İnanç Ağı (BBN)
- Bayesian Ağı (BN) (Ansari vd., 2018)

#### **2.1.5.6 Kümeleme Algoritmaları (Clustering Algorithms)**

Kümeleme, regresyon gibi, problem sınıfını ve yöntem sınıfını tanımlar.

Kümeleme yöntemleri tipik olarak ağırlık merkezi tabanlı ve hiyerarşi gibi modelleme yaklaşımlarıyla düzenlenir. Tüm yöntemler, verileri en iyi ortaklığa sahip gruplar halinde en iyi şekilde düzenlemek için verilerdeki doğal yapıların kullanılmasıyla ilgilidir (Jason, 2013; Wu vd., 2008).

En popüler kümeleme algoritmaları:

- K- Ortalama
- K- Medyanlar
- Beklenti Maksimizasyonu (EM)
- Hiyerarşik Kümeleme

#### **2.1.5.7 İlişkilendirme Kuralı Öğrenmesi Algoritmaları (Association Rule Learning Algorithms)**

İlişkilendirme kuralı öğrenme yöntemleri, verilerdeki değişkenler arasındaki gözlenen ilişkileri en iyi açıklayan kuralları gösterir (Jason, 2013).

Bu kurallar, bir kuruluş tarafından yararlanılabilecek çok boyutlu büyük veri kümelerinde önemli ve ticari açıdan yararlı olan ilişkileri keşfedebilir (Jason, 2013; Wu vd., 2008).

En popüler ilişkilendirme kuralı öğrenme algoritmaları:

- Apriori algoritması
- Eclat algoritması

### **2.1.5.8 Yapay Sinir Ağı Algoritmaları (Artificial Neural Network Algorithms)**

Yapay Sinir Ağları, biyolojik sinir ağlarının yapısından ve/veya işlevinden ilham alan modellerdir (Jason, 2013).

Bunlar genellikle regresyon ve sınıflandırma problemleri için kullanılan bir desen eşleştirme sınıfıdır, ancak her türlü problem tipi için yüzlerce algoritma ve varyasyondan oluşan muazzam bir alt alandır (Jason, 2013).

Alandaki muazzam büyüme ve popülerlik nedeniyle Derin Öğrenmeyi sinir ağlarından ayrı tutarak burada daha klasik yöntemlerle ilgilenilmiştir (Jason, 2013).

En popüler yapay sinir ağı algoritmaları:

- Perceptron
- Geri Yayılım
- Hopfield Ağı
- Radyal Temel Fonksiyon Ağı (RBFN) (Sobie vd., 2018)

### **2.1.5.9 Derin Öğrenme Algoritmaları (Deep Learning Algorithms)**

Derin Öğrenme yöntemleri, bol miktarda kalitesiz hesaplamayı kullanan Yapay Sinir Ağları için modern bir güncellemedir (Jason, 2013).

Çok daha büyük ve daha karmaşık sinir ağları oluşturmakla ilgilendirler ve yukarıda da belirtildiği gibi, birçok metot, geniş veri kümelerinin çok az etiketli veri içerdiği yarı denetimli öğrenme problemleriyle ilgilidir (Jason, 2013; J. H. Lee vd., 2018).

En popüler derin öğrenme algoritmaları:

- Derin Boltzmann Makinesi (DBM) (Zhao vd., 2019)
- Derin İnanç Ağları (DBN) (Zhao vd., 2019)
- Dönüşümlü Sinir Ağı (CNN)
- Yığılmış Otomatik Kodlayıcılar

### **2.1.5.10 Boyutsallık Azaltma Algoritmaları (Dimensionality Reduction Algorithms)**

Kümeleme yöntemleri gibi, boyutsallığın azaltılması da verilerdeki doğal yapıyı araştırır ancak bu durumda denetimsiz bir şekilde veya sırayla daha az bilgi kullanarak verileri özetlemek veya açıklamak için kullanır (Jason, 2013).

Bu, boyutsal verileri görselleştirmek veya daha sonra denetimli bir öğrenme yönteminde kullanılabilir verileri basitleştirmek için faydalı olabilir. Bu yöntemlerin çoğu, sınıflandırma ve regresyonda kullanılmak için uyarlanabilir (Jason, 2013).

- Temel Bileşen Analizi (PCA)
- Temel Bileşen Regresyonu (PCR)
- Kısmi En Küçük Kareler Regresyonu (PLSR)
- Sammon Haritalama
- Çok Boyutlu Ölçeklendirme (YTH)
- Projeksiyon Takip
- Lineer Ayırt Edici Analiz (LDA)
- Karışım Ayırma Analizi (MDA)
- Karesel Ayırt Edici Analiz (QDA)
- Esnek Ayırt Edici Analiz (FDA)

### **2.1.5.11 Topluluk Algoritmaları (Ensemble Algorithms)**

Topluluk yöntemleri, bağımsız bir şekilde eğitilmiş ve öngörülerini bir şekilde genel tahminde bulunmak için bir araya getirilmiş olan daha zayıf modellerden oluşan modellerdir (Jason, 2013).

Öğrenmeyi zayıflatıcı etkenlerin bir araya getirilmesi ve bunların nasıl birleştirileceği konusunda çok çaba harcanmaktadır. Bunlar çok güçlü teknikler sınıfıdır ve bu nedenle çok popülerdir (Jason, 2013; Wu vd., 2008).

- Yükseltme
- Bootstrapped Toplama (Torbalama)
- AdaBoost
- Yiğilmiş Genelleme (harmanlama)

- Gradyan Artırma Makinaları (GBM)
- Gradyan Arttırılmış Regresyon Ağaçları (GBRT)
- Rastgele Orman (Sobie vd., 2018)

#### **2.1.5.12 Diğer Algoritmalar**

Birçok algoritma ele alınmamıştır. Yapay Zekâ ve Makine Öğrenmesi çok popüler ve gündemde birçok tartışmalar ile her gün yeni yayınların yapıldığı farklı disiplinlerde araştırmalar yapılan bir alandır (Feng vd., 2019). Gerek konu genişliği gerekse bilim dünyasında bu algoritmaların sınıflandırılmasında henüz tam anlamı ile fikir birliğine varılmamış olması bu konuda çalışma yapmayı güçleştirmektedir. Bu sebeple bazı algoritmalara bu çalışmada yer verilmemiştir. Gelecek çalışmalarda bu yöntemlere ilişkin araştırmalar genişletilebilir. Aşağıda kapsama alınmayan bazı algoritmalara ilişkin listeler yer almaktadır.

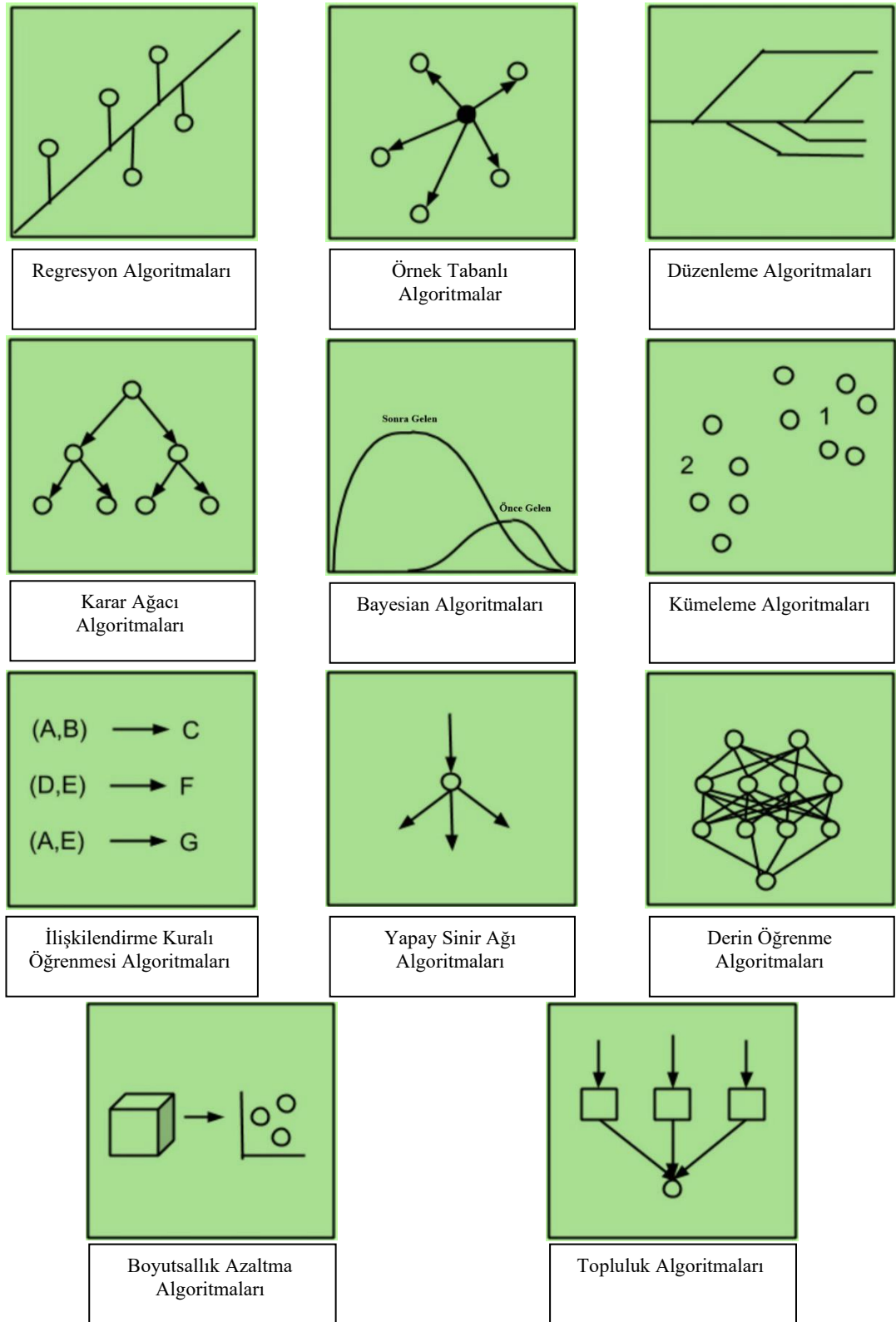
Makine öğrenmesi sürecinde uzmanlık görevleri içeren algoritmaları kapsama alınmamıştır.

- Özellik seçimi algoritmaları
- Algoritma doğruluğu değerlendirmesi
- Performans ölçüleri

Ayrıca, makine öğreniminin uzmanlık alt alanlarından gelen bazı algoritmalar da kapsama alınmamıştır.

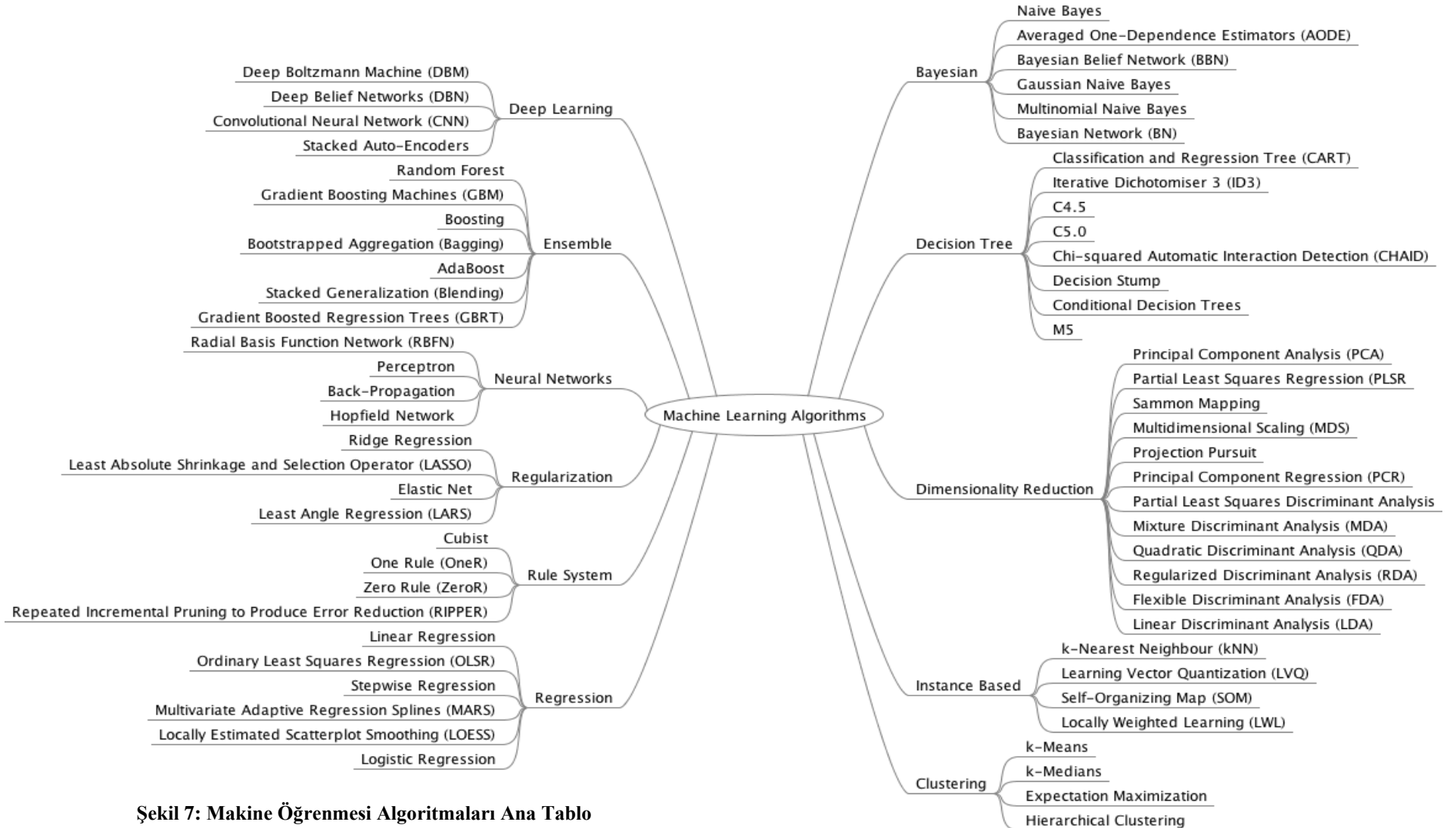
- Hesaplamalı zekâ (evrimsel algoritmalar, vb.)
- Bilgisayarla Görme (CV)
- Doğal Dil İşleme (NLP)
- Tavsiye Sistemleri
- Grafik Modeller





Şekil 6: Algoritmaların Görsel Açıklamaları

### 2.1.5.13 Algoritmaların Genel Bakışı



Şekil 7: Makine Öğrenmesi Algoritmaları Ana Tablo

## 2.2 Bulut Bilişim Teknolojisi

Bu bölümde tez uygulamasında ciddi kolaylıklar sağlayan Bulut Bilişim Teknolojisi hakkında özet sayılabilecek bazı temel bilgiler, tanımlar, kullanılan platform ve araçlar açıklanmaktadır.

İlk bilgisayarın icadı, bilgisayar ağları ve internetin ortaya çıkışı ile bilişim insan hayatına girmiş ve teknolojideki gelişme, önceleri sadece büyük kuruluşların edinebildiği sonraları hemen herkesin erişebildiği bilişim olanaklarına ivme kazandırmıştır. İnternet; içerik paylaşımı, yer ve servis sağlayıcılığı gibi yeni iş kolları oluşturmuş ve bir dönüşüm içerisine girmiştir. Bu dönüşüm kurum ve kuruluşların yeni sorumluluklar üstlenmesine yol açmış ve yaygın internet kullanımı çeşitli güvenlik sorunlarını da beraberinde getirmiştir. 2000'li yıllara geldiğimizde bilişim hizmetleri tedarikinde işletme, bakım ve enerji maliyetlerini azaltmak için bu hizmetlerin dış kaynak olarak temin edilebilmesi fikrini ortaya çıkarmıştır. Barındırma (Hosting), Dağıtım Bilişim (Grid Computing) ve Kamu Hizmeti Bilişimi (Utility Computing) vb. gelişen çeşitli hizmet alanları ihtiyaca cevap vermeye çalışmış fakat kendine hizmet (self-service) ve gerektiğinde kapasite artırma gibi fiziksel özellikleri sağlamada yetersiz kalmışlardır. Nihayetinde işlemci fiyatlarının ucuzlaması, geniş bant bilgisayar ağ erişimlerinin yaygınlaşması, sanallaştırma ve otomasyondaki hızlı gelişmeler, veri merkezlerinin endüstriyel ölçeklere ulaşması gibi bilgi teknolojilerindeki birçok eksponansiyel gelişme bulut bilişim kavramının doğuşunu tetiklemiştir (BTK, 2013).

NIST<sup>1</sup> tarafından oluşturulan bulut bilişim standartları yol haritası raporu; Amazon, Google ve Microsoft gibi büyük firmalar tarafından da kabul gören bir takım sınıflandırma ve standartları ortaya koymaktadır.

Bulut Bilişim; bilgisayar ağları, sunucular, veri tabanları uygulamalar ve servisler gibi yapılandırılabilir bilişim kaynaklarının paylaşılan bir havuzuna hizmet sağlayıcı etkileşimini hızlı bir şekilde temin edebilen ve minimum düzeyde yönetimsel çaba ile ihtiyari olarak dünyanın herhangi bir yerinden zaman kısıtı olmaksızın ağ erişimine

---

<sup>1</sup> NIST: İngilizcesi National Institute of Standards and Technology, Türkçesi Teknoloji ve Standartlar Ulusal Enstitüsü olan Amerika Birleşik Devletleri kurumudur.

olanak tanıyan bir modeldir (Hogan vd., 2011). Bu model beş karakteristik özellik ile üç hizmet sağlama modeli ve dört konumlandırma modeli üzerine inşa edilmiştir.

### 2.2.1 Bulut Bilişim Temel Karakteristikleri

- **Talep Dahilinde Kendi Kendine Hizmet:**

Tüketicilerin her bir hizmet için insan etkileşimi gerektirmeksizin ihtiyaç halinde otomatik olarak sunucu süresi ve ağ depolaması gibi bilgisayar özelliklerini tek taraflı olarak karşılayabildiği hizmetlerdir.

- **Geniş Ağ Erişimi:**

Üzerinde barındırdığı yetkinliklerin ağ üzerinde kullanılmasına ve yetkinliklere dizüstü bilgisayarlar, cep telefonları veya dijital kişisel asistanlar vasıtası ile erişime imkân veren altyapı hizmetidir.

- **Kaynak Havuzu:**

Tedarikçinin bilgi işlem kaynakları, dinamik olarak atanan farklı fiziksel ve sanal kaynaklar ile çoklu kiracı modeli kullanarak birçok tüketiciye hizmet vermek için bir havuzda toplanır ve tüketicilerin taleplerine göre tekrar atanır. Tüketicilerin genellikle sağlanan kaynakların tam yeri üzerinde hiçbir kontrol veya bilgiye sahip olmadığı ama ülke, eyalet ya da veri merkezi gibi daha yüksek bir çıkarım düzeyinde konum belirleyebildiği konum bağımsızlığı hissi söz konusudur. Bu kaynaklar işlem gücü depolama, bellek, ağ ve sanal makineler gibi fonksiyonları barındırmaktadır.

- **Hızlı Esneklik:**

Tüketicilerin sınırsız miktarda fonksiyonu herhangi bir zamanda ve miktarda hızlı bir şekilde satın alabildiği ve otomatik olarak ölçeklendirebildiği sistemler söz konusudur.

- **Ölçülebilir Hizmet:**

Bulut sistemleri işlem gücü, depolama, bant genişliği ve etkin kullanıcı hesapları gibi hizmet türlerine göre uygun düzeyde çıkarım yapabilen ölçümleme yeteneklerinden faydalanarak kaynak kullanımı otomatik olarak optimize edilmektedir. Kaynak kullanımı; kullanılan hizmetlerin hem tedarikçisi hem de tüketicisi için izlenebilen, kontrol edilebilen, raporlanabilen ölçülebilir şeffaf bir modeldir.

### 2.2.2 Bulut Bilişim Hizmet Sağlama Modelleri

- **Yazılım Hizmetleri (SaaS):**

Tüketicie sağlanan özellik, sağlayıcının bulut altyapısı üzerinde çalışan yazılım uygulamalarını kullanmaktır. Bu uygulamalara web tarayıcıları üzerinden isteğe bağlı olarak genellikle abonelik yöntemi ile erişilmektedir. Tüketici ağ, sunucular, işletim sistemleri, depolama ve bireysel uygulama yetenekleriyle birlikte temel bulut altyapısını, sınırlı kullanıcıya özgü uygulama yapılandırma ayarları olası istisnaları dışında yönetemez ya da kontrol edemez.

- **Platform Hizmetleri (PaaS):**

Tüketicie sağlanan özellik, sağlayıcı tarafından desteklenen programlama dilleri ve araçları kullanılarak oluşturulan tüketicinin oluşturduğu veya edindiği bulut altyapısı üzerindeki uygulamaları konuşlandırmaktır. Tüketici ağ, sunucular, işletim sistemleri ve depolamanın da yer aldığı temel bulut altyapısını kontrol etme yetkisine sahip değildir fakat konuşlandırılan uygulamalar ve bu uygulamaları barındırma ortamı uygulamaları üzerinde olasılık dahilinde denetime sahiptir.

- **Altyapı Hizmetleri (IaaS):**

Tüketicie sağlanan özellik, işletim sistemleri ve uygulamalarını içerebilen rastgele yazılımları konuşlandırılabilen ve çalıştırılabilen işlemci, depolama, ağlar ve diğer temel bilgi işlem kaynaklarını sağlamaktır. Tüketici temel bulut altyapısını kontrol etme yetkisine sahip değildir, fakat işletim sistemleri, depolama, konuşlandırılan uygulamalar ve belirli ağ bileşenlerinin (örneğin, ana bilgisayar ve güvenlik duvarları) sınırlı kontrolü üzerinde olasılık dahilinde denetime sahiptir.

### **2.2.3 Bulut Bilişim Konumlandırma Modelleri**

#### **2.2.3.1 Özel Bulut**

Bulut altyapısı yalnızca bir organizasyon için çalıştırılır. Organizasyonun bulut altyapısı kurum veya izin verilen başka bir tarafça yönetilebilir ve fiziksel olarak tesis içinde veya tesis dışında yer alabilir.

#### **2.2.3.2 Topluluk Bulutu**

Bulut altyapısı birkaç kuruluş tarafından paylaşılır ve görev, güvenlik gereksinimleri, gizlilik, politika, yargı ve uyumluluk hususları gibi ortak endişelere sahip belirli bir topluluğu destekler. Kuruluşlar veya izin verilen başka bir tarafça idaresi sağlanabilir ve benzer şekilde tesis içinde veya tesis dışında bulunabilir.

#### **2.2.3.3 Genel Bulut**

Bulut altyapısı kamusal kullanım veya büyük ölçekli endüstriler için hazır bulundurulur ve bulut hizmetleri satan bir kuruluşa aittir.

#### **2.2.3.4 Hibrit Bulut**

Bulut altyapısı, özel, topluluk ya da genel bulut gibi benzersiz iki ya da daha fazla bulutun veri ve uygulama taşınabilirliğine imkân veren standart veya patentli teknolojiyle birbirine bağlanmış bir birleşimdir.

### **2.2.4 Tez Uygulamasında Kullanılan Platform ve Araçlar**

#### **2.2.4.1 Microsoft Azure Machine Learning Studio Platformu**

Microsoft Azure, bulut sistemi içerisinde yer alan makine öğrenmesi konusunda özelleştirilmiş Python ve R programlama dilleri ile komut dosyası ekleri çalıştırılabilen internete bağlı olarak web tarayıcısı üzerinden belli limitler dahilinde ücretsiz abonelik ile erişim sağlanabilen bir platformdur. Microsoft Azure Machine Learning (ML) Studio Platformu herhangi bir kod yazmaksızın makine öğrenmesi modellerini sürükleyip bırak yöntemi ile kolayca oluşturmaya ve isteğe bağlı olarak modele Python veya R komut satırı ile eklemeler yapmaya olanak tanımaktadır. Kullanıcılar; son derece geniş bir dokümantasyon ile platform içerisinde yer alan araçlar hakkında bilgi edinebilir, içerisinde barındırdığı çeşitli veri setlerini veya uygun formatlara (csv, tsv, zip vb.)

dönüştürülmüş kendi veri setlerini sisteme yükleyerek kullanabilir, içerisinde birçok örneği barındıran galeriden yararlanabilir hatta burayı bir sosyal platform olarak ele alıp kendi oluşturduğu modelleri diğer kullanıcılarla paylaşabilir ve ortak çalışma gurupları oluşturabilir.

Microsoft Azure Bulut Platformu gibi Google Bulut platformu, AWS (Amazon Web Service) Bulut platformu ve benzeri diğer platformlar da bulut bilişim sektörüne; bilgi işlem, analiz, uygulama entegrasyonu, sanal ve artırılmış gerçeklik, yapay zekâ ve makine öğrenmesi, yazılım geliştirme, ağ iletişimi, nesnelerin interneti, blok zinciri, depolama, veri tabanları, kuantum teknolojiler ve daha birçok bilişsel alanda çok ciddi çözümler sunmaktadır. Fakat Azure ML Studio sağladığı kullanım kolaylığı açısından bu tezin konusu olan uygulamaların gerçekleştirildiği ana platform olarak seçilmiştir.

#### **2.2.4.2 Kaggle Platformu**

Veri Bilimi üzerine çalışan insanların projelerini gerçekleştirip paylaştığı bilgi alışverişinde bulunduğu ve kurum ve kuruluşların birtakım sorunların çözümüne yönelik ödüllü yarışmalar başlattığı bir veri bilimcinin sosyal medyası olarak tarif edebileceğimiz birkaç yıl önce de Google tarafından satın alınan bulut tabanlı platformdur.

Bir web tarayıcısı üzerinden ücretsiz abonelik ile erişime açılan bu platform üzerinde bilgisayarınıza herhangi bir kurulum yapmaksızın Python ve R dillerinde ve bu programlama dilleri için oluşturulmuş özel kütüphaneleri de kullanarak çeşitli veri setleri üzerine analizler ve geliştirmeler yapılabilmektedir. Bu tezin uygulamalarında bu platformdan da yararlanılmıştır.

### 3 ÜÇÜNCÜ BÖLÜM

#### LİTERATÜR ARAŞTIRMASI

##### 3.1 Çeşitli Bilim Dallarında Yapay Zekâ ve Makine Öğrenmesi Araştırmalarının İncelenmesi

Literatürde yapay zekâ ve makine öğrenmesi kendi sınırlarını aşarak farklı disiplinlerde kendisine yer bulmuş son zamanların belki de en popüler konuları arasında yer almaktadır. Dijital Dönüşüm ise bir araştırma alanı olmasının yanısıra teknolojiye liderlik yapan tüm şirketlerin kısa ve uzun vadeli stratejilerin yapılandırılması ve planlanması kapsamında gündemde baş köşede yer almaktadır. Dijital Dönüşümün gerçekleşmesine öncülük eden; akıllı telefonlar, otonom arabalar, insansız hava araçları, uzay çalışmaları, sağlık, biyoteknoloji gibi birçok örneği ele aldığımızda bu örneklerin arkasında temel olarak malzeme bilimi, matematik ve yazılım ana dalları ve bu dalların altındaki alt kümelerin içinde en önemli konulardan birisi olan yapay zekâ uygulamaları görülmektedir. (Burada malzeme bilimi olarak yarı iletkenlerin hayatımıza girmesi ile elektronik alanındaki gelişmeler ve son aşamada nanoteknoloji alanı kastedilmektedir.)

Popüler bilimsel araştırma konularına baktığımız zaman; biyoloji ve genetik, fizik ve astronomi, çevre ve ekoloji, nanoteknoloji, sağlık, robotik ve bilgisayar, psikoloji ve nörobiyoloji vb. yapay zekâ uygulamalarından yararlanılmıştır. Pavel Hamet ve Johanne Tremblay'in "Tıp'ta Yapay Zeka" makalesi (Hamet ve Tremblay, 2017), Kipp W. Johnson ve arkadaşlarının "Kardiyoloji'de Yapay Zeka" makalesi (K. W. Johnson vd., 2018), Nehmat Houssami ve arkadaşlarının "Göğüs kanseri taraması için yapay zeka: Fırsat mı, yutturmaca mı?" makalesi (Houssami vd., 2017), D. Douglas Miller ve Eric W. Brown'un "Tıbbi Uygulamada Yapay Zeka" makalesi (D. D. Miller ve Brown, 2018) gibi makaleler sağlık alanına, Paulius Cerka, Jurgita Grigiene ve Gintare Sirbikyte'nin "Yapay zeka yazılım sistemlerine tüzel kişilik vermek mümkün müdür?" makalesi (Čerka vd., 2017) hukuk alanına, Demis Hassabis ve arkadaşlarının "Sinirbilimden İlham Alan Yapay Zeka" makalesi (Hassabis vd., 2017) nörobilim alanına, Mehran Ali Azizi Oroumieh ve arkadaşlarının "Yapay zeka kullanarak uçak tasarım döngüsü zaman azaltma" makalesi (Azizi Oroumieh vd., 2013) havacılık ve uzay bilimi alanına,



Christopher Tack'ın "Yapay zeka ve makine öğrenmesi | kas iskelet sistemi fizyoterapisindeki uygulamalar" makalesi (Tack, 2019) fizyoterapi alanına, Tim Miller'in "Yapay zekada açıklama: Sosyal bilimlerden edinilen görüşler" makalesi (T. Miller, 2019) sosyal bilimler alanına, Chad Edwards ve arkadaşlarının "Yapay zeka öğretmeni sesinin değerlendirilmesi: İnsan-robot etkileşimlerinde sosyal kimlik teorisi" makalesi (Edwards vd., 2018) insan davranışlarında bilgisayar bilimi alanına, Larry D. Wall'un "Yapay zekanın bazı finansal düzenleyici etkileri" makalesi (Wall, 2018) finans alanına örnek verilebilir. Bunun yanı sıra Ian Goodfellow, Yoshua Bengio ve Aaron Courville'in hazırladığı "Derin Öğrenme" kitabı (Goodfellow vd., 2016), Xindong Wu ve arkadaşlarının "Veri madenciliğinde en iyi 10 algoritma" makalesi (Wu vd., 2008), Utku Köse'nin "Yapay Zekâ Tabanlı Optimizasyon Algoritmaları Geliştirilmesi" tezi (Köse, 2017) yapay zeka ve makine öğrenmesi alanında temel bilgilerden algoritmaların sınıflandırılmasına kadar geniş bir perspektiften bilgi sağlamaktadır. Bu değerlendirmede yazarlar ve konu başlıklarına ilişkin veriler Tablo 1'de yer almaktadır.

No	Yazar Adı	Başlık	Yıl	Kategori
1	Miller, D. Douglas Brown, Eric W.	Tıbbi Uygulamada Yapay Zekâ	2018	Tıp
2	Bianchini, Francesco	Yapay Zekâ ve Sentetik Biyoloji: Üç Geçici Bir Katkı	2016	Biyoloji
3	Edwards, Chad Edwards, Autumn Stoll, Brett Lin, Xialing Massey, Noelle Edwards, Chad	Yapay Zekâ Öğretmeninin Ses Değerlendirilmesi: İnsan-Robot Etkileşimlerinde Sosyal Kimlik Teorisi	2018	Bilgisayar ve Psikoloji
4	Wall, Larry D.	Yapay Zekâ'nın Bazı Finansal Düzenleyici Etkileri	2018	Ekonomi ve Finans
5	Miller, Tim	Yapay Zekâ'da Açıklama: Sosyal Bilimlerden Edinilen Görüşler	2019	Sosyal Bilimler
6	Tack, Christopher	Yapay Zekâ ve Makine Öğrenmesi Kas İskelet Sistemi Fizyoterapisindeki Uygulamalar	2018	Fizyoterapi

No	Yazar Adı	Başlık	Yıl	Kategori
7	Azizi Oroumieh, Mehran Ali Mohammad Bagher Malaek, S. Ashrafizaadeh, Mahmud Mahmoud Taheri, S.	Yapay Zekâ Kullanarak Uçak Tasarım Döngüsü Zaman Azaltma	2013	Havacılık ve Uzay
8	Hassabis, Demis Kumaran, Dharshan Summerfield, Christopher Botvinick, Matthew	Sinirbilimden İlham Alan Yapay Zekâ	2017	Sinir Bilimi
9	Čerka, Paulius Grigienė, Jurgita Sirbikytė, Gintarė	Yapay Zekâ Yazılım Sistemlerine Tüzel Kişilik Vermek Mümkün Müdür?	2017	Hukuk
10	Houssami, Nehmat Lee, Christoph I. Buist, Diana S.M. Tao, Dacheng	Göğüs Kanseri Taraması İçin Yapay Zekâ: Fırsat Mı, Yutturmaca Mı?	2017	Tıp
11	Johnson, Kipp W. Torres Soto, Jessica Glicksberg, Benjamin S. Shameer, Khader Miotto, Riccardo Ali, Mohsin Ashley, Euan Dudley, Joel T.	Kardiyoloji’de Yapay Zekâ	2018	Tıp
12	Hamet, Pavel Tremblay, Johanne	Tıp’ta Yapay Zekâ	2017	Tıp
13	Lovejoy, Christopher A. Buch, Varun Maruthappu, Mahiben	Teknoloji ve Zihin Sağlığı: Yapay Zekâ’nın Rolü	2019	Psikoloji
14	Dreyer, Keith Allen, Bibb	Sağlık Hizmetlerinde Yapay Zekâ: Cesur Yeni Dünya mı, Altın Fırsat mı?	2018	Sağlık
15	Martín, F. Matellán, V. Rodríguez, F. J. Ginés, J.	İç Mekân Robotları İçin RGB-D2 Verilerini Kullanarak, Octree3 Tabanlı Yerelleştirme	2019	Robotik
16	Kamble, Sachin S. Gunasekaran, Angappa Sharma, Rohit	Hindistan İmalat Endüstrisinde Endüstri 4.0’ı Benimseme Engellerinin İtici Güç ve Bağımlılık Gücünün Analizi	2018	Endüstri

<sup>2</sup> RGB-D: İngilizce’de Red, Green, Blue ve Dimension kelimelerinin baş harfleri ile oluşturulan Türkçesi Kırmızı, Yeşil, Mavi ve Boyut olan bir çeşit robotik renk sensörüdür.

<sup>3</sup> Octree: 3 Boyutlu veri görselleştirme aracı olarak kullanılabilen 8 düğüm noktasına sahip ağaç veri yapısı formunda çalışan metottur.

No	Yazar Adı	Başlık	Yıl	Kategori
17	Goodfellow, Ian Yoshua, Bengio Courville, Aaron	Derin Öğrenme	2016	Veri Bilimi
18	Wu, Xindong Kumar, Vipin Ross Quinlan, J. Ghosh, Joydeep Yang, Qiang Motoda, Hiroshi McLachlan, Geoffrey J. Ng, Angus Liu, Bing Yu, Philip S. Zhou, Zhi-Hua Steinbach, Michael Hand, David J. Steinberg, Dan	Veri Madenciliğinde En İyi 10 Algoritma	2008	Bilgi Sistemleri
19	Mitchell, Tom M	Makine Öğrenmesi Disiplini	2006	Bilgisayar Bilimi
20	Naili, Mohamed Bourahla, Mustapha Naili, Makhlof Tari, Abdel Kamel	Dinamik Veri Madenciliği İçin Kararlılık Tabanlı Dinamik Bayesian Ağ Yöntemi	2019	Mühendislik Uygulamaları
21	Rahman, Md. Mahmudur Ahmed, Chowdhury Farhan Leung, Carson Kai Sang	Belirsiz Veri tabanlarında Madencilik Ağırlıklı Sık Görülen Diziler	2019	Bilişim Bilimi
22	Sparks, Evan R. Talwalkar, Ameet Smith, Virginia Kottalam, Jey Pan, Xinghao Gonzalez, Joseph Franklin, Michael J. Jordan, Michael I. Kraska, Tim	MLI4: Dağıtılmış Makine Öğrenmesi için bir API5	2013	Veri Bilimi
23	Köse, Utku	Yapay Zekâ Tabanlı Optimizasyon Algoritmaları Geliştirilmesi	2017	Bilgisayar Bilimi

**Tablo 1: Literatür Taraması Bilim Dalları Bazında**

<sup>4</sup> MLI: Veri merkezli hesaplama dayalı dağılımlı bir ortamda Makine Öğrenimi algoritmaları oluşturmanın zorluklarını gidermek için tasarlanmış bir Uygulama Programlama Arayüzü'dür.

<sup>5</sup> API: (Application Programming Interface) Uygulama Programlama Arayüzü

### 3.2 Havalimanları Kapsamında Yapılan Araştırmaların İncelenmesi

Türkiye’de ve Dünya’da Havalimanları kapsamında yapılan analizleri, bir havalimanı analiz edilirken hangi kriterlerin baz alındığını hangi analiz yöntemlerinin kullanıldığını tespit etmek amacı ile literatür araştırması gerçekleştirilmiştir. Takip eden paragraflarda ayrıntılarına yer verilen bu araştırmalar Tablo 2’de toplanmıştır.

Mary Johnson ve Yue Gu’nun 2017 yılında yayınladıkları “Genel Havacılık Havalimanlarında FAA<sup>6</sup> NPIAS<sup>7</sup> Havalimanı Kategorileri Kullanarak Havalimanı Operasyonları Tahmini” adlı makalede Amerika Birleşik Devletleri’nde Genel havacılık otoriteleri tarafından yayınlanan kategorilere ayrılmış 5010 havalimanı ölçeğinde havalimanları operasyon tahmini yapılmıştır. Burada bahsi geçen kategoriler; Ulusal, Bölgesel, Yerel, Temel ve Sınıflandırılmamış olmak üzere Havalimanları; bir havalimanının yıllık operasyon sayısı, bir hava aracının yıllık operasyon sayısı, bir havalimanının toplam hava aracı sayısı gibi veriler kullanılarak analiz edilmiştir (M. Johnson ve Gu, 2017).

Joseph Sarkis’in 2000 yılında yayınladığı “Birleşik Devletlerde Ana Havalimanlarının Operasyonel Verimlilik Analizi” adlı makalede Veri Zarflama yöntemi kullanılarak Amerika Birleşik Devletleri’nde yer alan 44 havalimanı için operasyonel verimlilik analizi gerçekleştirilmiştir. Verimlilik ölçümleri; havalimanı operasyonel maliyetleri, havalimanı işçi sayıları, kapı sayıları ve pist sayıları olmak üzere 4 giriş ölçümü, operasyonel gelir, yolcu akışı, ticari ve genel havacılık hareketleri ve toplam kargo ulaşımı olmak üzere 5 çıkış ölçümü içermektedir (Sarkis, 2000).

M. Janic ve A. Reggiani’nin 2002 yılında yayınladığı “Havalimanı Yeni Aktarma Merkezi Seçiminde Çok Kriterli Karar Verme Analizi ile Bir Uygulama” adlı makalede Çok Kriterli Karar Verme yöntemleri kullanılarak Avrupa ölçeğinde yeni aktarma merkezi seçimine yönelik bir çalışma yapılmıştır. Yeni aktarma merkezi seçiminde; nüfus, gelir, havalimanı büyüklüğü, yolcu bazlı havalimanı erişim maliyeti, hizmet bazlı

---

<sup>6</sup> FAA: (Federal Aviation Administration) Federal Havacılık Kurulu Amerika’da bir devlet kuruluşudur.

<sup>7</sup> NPIAS: (National Plan of Integrated Airport Systems) Türkçesi Entegre Havalimanı Sistemleri Ulusal Planı olan Havalimanları kategorilerinin belirlendiği FAA tarafından yayınlanan rapordur.

ortalama havalimanı maliyeti, havalimanı kapasitesi, yerleşik pazar payı gibi kriterler baz alınmıştır (Janic ve Reggiani, 2002).

Yu-Chun Chang ve Ning Lee'nin "Düşük Maliyetli Operatörlerin Ağları için Çok Amaçlı Hedef Programlama Havaalanı Seçim Modeli" adlı 2010 yılında yayınlanan makalesinde Asya bölgesinde yer alan 12 havalimanı için en iyi merkezi havalimanını seçmek amacı ile uçuş sayısı, rota uzunluğu, birim maliyetler, talep, rakip sayısı, bilet ücretleri gibi bir çok parametre ele alınarak optimum sonucu bulmaya yönelik çalışma gerçekleştirilmiştir (Chang ve Lee, 2010).

Habip Koçak'ın "Veri Zarflama Yaklaşımı ile Türk Havalimanlarının Verimlilik İncelemesi" adlı 2011 yılında yayınlanan makalesinde Veri Zarflama yöntemi yardımıyla Türkiye'de yer alan 40 havalimanı ölçeğinde; operasyonel maliyetler, çalışan sayısı, uçuş sayısı ve yolcu sayısı giriş parametreleri kapsamında ve yolcu sayısı, uçuş sayısı, toplam yük ve operasyonel maliyetler çıkış parametreleri kapsamında dikkate alınarak verimlilik analizi gerçekleştirilmiştir (Koçak, 2011).

Bersam Bolat ve arkadaşlarının "Türkiye'deki Havalimanlarının Etkinlik Tahmini: Veri Zarflama Analizi ve Yapay Sinir Ağlarının Birlikte Kullanımı" adlı 2016 yılında yayınlanan makalesinde Veri Zarflama yöntemi kullanılarak Türkiye'de bulunan 41 havalimanı; yolcu biniş kapısı sayısı, check-in kontuarları sayısı, pist sayısı, bagaj konveyörleri sayısı, terminal büyüklüğü, personel sayısı, otopark araç kapasitesi sayısı giriş parametreleri kapsamında ve yolcu sayısı, yük sayısı, ticari uçuş sayısı çıkış parametreleri olarak değerlendirilerek verimlilik analizi yapılmış daha sonra havalimanı etkinlik tahmini yapan bir yapay sinir ağı modeli geliştirilmiştir.

Esra Öztürk ve Hasan Bal'ın "Veri Zarflama Analizi ve Kanonik Korelasyon Analizi ile Havaalanları Sıralaması" adlı 2017 yılında yayınlanan makalesinde Veri Zarflama ve Kanonik Korelasyon yöntemi kullanılarak Türkiye'de yer alan 23 havalimanı; terminal büyüklüğü, havalimanı büyüklüğü giriş verileri ve hava aracı hareketleri, yolcu trafiği, yük trafiği gibi çıkış verileri dikkate alınarak veriler türetilmiş ve türetilen veriler üzerinden kümeleme yapılmıştır.

Bahar Sennaroğlu ve Gülsay Varlık Çelebi'nin "PROMETHEE<sup>8</sup> , VIKOR<sup>9</sup> ve AHP<sup>10</sup> Yöntemleri Kullanarak Askeri Havalimanı Yer Seçimi" adlı 2018 yılında yayınlanan makalesinde Çok Kriterli Karar Verme yöntemleri kullanılarak askeri havalimanı için yer seçimi çalışmasında askeri kriterler, genişleme potansiyeli, maliyet, çevresel ve sosyal etkiler, iklim koşulları, altyapı özellikleri, saha, coğrafi özellikler ve ihtiyaçlar gibi ana kriterler altında belirlenen alt kriterler ile değerlendirme yapılmıştır.

Eva Stichhauerova ve Natalie Pelloneova'nın "Veri Zarflama Modeli Kullanarak Seçilen Alman Havalimanlarının Verimlilik Değerlendirmesi" adlı 2019 yılında yayınlanan makalesinde Veri Zarflama yöntemi kullanılarak Almanya'da yer alan 27 büyük havalimanı; personel sayısı, pist sayısı ve havalimanı büyüklüğü olmak üzere 3 adet giriş verisi, hava aracı hareketi ve kargo trafiği olmak üzere 2 adet çıkış verisi kullanılarak analiz edilmiştir (Stichhauerova ve Pelloneova, 2019).

No	Yazar Adı	Başlık	Yıl	Kategori
1	Johnson, Mary Gu, Yue	Genel Havacılık Havalimanlarında FAA NPIAS Havalimanı Kategorileri Kullanarak Havalimanı Operasyonları Tahmini	2017	Havacılık, Operasyon Tahmini
2	Sarkis, Joseph	Birleşik Devletlerde Ana Havalimanlarının Operasyonel Verimlilik Analizi	2000	Havacılık, Verimlilik Analizi
3	Janic, M. Reggiani, A.	Havalimanı Yeni Aktarma Merkezi Seçiminde Çok Kriterli Karar Verme Analizi ile Bir Uygulama	2002	Havacılık, Havalimanı Seçimi
4	Chang, Yu-Chun Lee, Ning	Düşük Maliyetli Operatörlerin Ağları için Çok Amaçlı Hedef Programlama Havaalanı Seçim Modeli	2010	Havacılık, Havalimanı Seçimi
5	Kocak, Habip	Veri Zarflama Yaklaşımı ile Türk Havalimanlarının Verimlilik İncelemesi	2011	Havacılık, Verimlilik Analizi

<sup>8</sup> PROMETHEE: (Preference Ranking Organization Method for Enrichment Evaluation) Türkçesi Değerlendirmelerin Zenginleştirilmesi için Tercih Sıralaması Organizasyon Yöntemi olan Çok Kriterli Karar Verme yöntemleri içerisinde ele alınan sayısal yöntemlerden biridir (Sennaroglu ve Varlık Celebi, 2018).

<sup>9</sup> VIKOR: (VİseKriterijumska Optimizacija I Kompromisno Resenje) Çok Kriterli Karar Verme yöntemleri içerisinde ele alınan sayısal yöntemlerden biridir (Sennaroglu ve Varlık Celebi, 2018).

<sup>10</sup> AHP: (Analytic Hierarchy Process) Çok Kriterli Karar Verme yöntemleri içerisinde en yaygın kullanılan sayısal yöntemdir (Sennaroglu ve Varlık Celebi, 2018).

No	Yazar Adı	Başlık	Yıl	Kategori
6	Bolat, Bersam Temur, Gül T. Gürler, Haktan	Türkiye'deki Havalimanlarının Etkinlik Tahmini: Veri Zarflama Analizi ve Yapay Sinir Ağlarının Birlikte Kullanımı	2016	Havacılık, Verimlilik Analizi
7	Öztürk, Esra Bal, Hasan	Veri Zarflama Analizi ve Kanonik Korelasyon Analizi ile Havaalanları Sıralaması	2017	Havacılık, Havalimanı Seçimi
8	Sennaroglu, Bahar Varlık Celebi, Gulsay	PROMETHEE, VIKOR ve AHP Yöntemleri Kullanarak Askeri Havalimanı Yer Seçimi	2018	Havacılık, Havalimanı Lokasyon Seçimi
9	Stichhauerova, Eva Pelloneova, Natalie	Veri Zarflama Modeli Kullanarak Seçilen Alman Havalimanlarının Verimlilik Değerlendirmesi	2019	Havacılık, Verimlilik Analizi

**Tablo 2: Literatür Taraması Havalimanları Bazında**

### 3.3 Makine Öğrenmesi & Azure ML Kapsamında Tahmine Dayalı

#### Yapılan Araştırmaların İncelenmesi

Makine Öğrenmesi alanında tahmine dayalı gerçekleştirilen uygulamaların çeşitli örnekleri ve bu tez içerisinde yer alan iki adet uygulama örneğini gerçekleştirmek için ana platform olarak kullanılan Microsoft Azure ML platformu üzerinde çalışılmış bazı bilimsel araştırmalar;

- Tahmine dayalı ne çeşit uygulamalar yapılmıştır?
- Tahmine dayalı uygulama yapılırken izlenen adımlar ve yöntemler nelerdir?
- Elde edilen sonuçların doğruluğu hangi matematiksel parametreler ile ortaya konulmuştur?

gibi sorulara cevap bulmak amacı ile incelenmiştir. Yapılan incelemeler Tablo 3'te listelenmiştir.

Jamilu Awwalu ve arkadaşlarının “Veri Madenciliği Algoritmalarının Performans Karşılaştırmaları: Otomobil Değerlendirme Veri Seti Üzerinde bir Vaka Analizi” adlı 2014 yılında yayınlanan makalesi Otomobil Değerlendirme Veri Seti kullanılarak Veri Madenciliği kapsamında üretim ve iş süreçlerine katkı yapabilecek bir tahmin veya

sınıflandırma geliştirmek ve 3 farklı veri madenciliği algoritmasının performanslarını karşılaştırmak amacı ile ortaya konmuştur (Awwalu vd., 2014).

Shin Hayakawa ve Hitoshi Hayashi'nin “Kapalı Alanlarda Konum Tahmini için Azure ML Kullanma” adlı 2017 yılında yayınlanan makalesi, kapalı alanlarda istikrarsız sinyaller yüzünden doğru konumlama yapmanın zor olması sebebi ile konum tahmini yapmaya yönelik bir araştırmadır. IEEE<sup>11</sup> ICDM<sup>12</sup> veri setinden alınan kapalı alanlarda ölçümleri yapılmış gerçek wireless verileri kullanılarak Azure Machine Learning Studio platformunda makine öğrenmesi teknikleri ile konum tahminlemesine yönelik yapılan uygulamada sadece 505 eğitim verisi ile 247 alan arasından doğru alanı tanımlamada %69'dan daha fazla bir oranda kesinlik elde edilmiştir(Hayakawa ve Hayashi, 2017).

Rahul Nigam ve K. Govinda 'nın “Lojistik Regresyon kullanarak Bulut Tabanlı Uçuş Gecikmesi Tahmini” adlı 2017 yılında yayınladığı makalesi ticari havayolu şirketlerinin süreç planlamasında uçuş gecikmesinin hayati bir rol oynaması sebebi ile daha doğru uçuş gecikmesi tahmini yapılması ihtiyacına binaen hazırlanmıştır. Uçuş gecikmesi tahmini yapmaya yönelik birçok metot bulunmakla beraber bu makalede bulut tabanlı Azure ML Studio platformu üzerinde makine öğrenmesi tekniklerinden gözetmeli öğrenme kategorisi altında lojistik regresyon metodu kullanılmıştır. Amerika Birleşik Devletleri'nde yer alan en yoğun 70 havalimanına ait veriler, hava araçlarının gecikme verileri ve daha doğru sonuçlar üretmek amacı ile hava şartlarını içeren veriler ile bir veri seti oluşturulmuş ve %80 doğruluk oranında sonuçlar elde edilmiştir (Nigam ve Govinda, 2017).

Alexei Botchkarev “Azure ML Studio kullanarak Hastane Maliyet Tahmin Modeli Vaka Değerlendirmesi” adlı 2018 yılında yayınladığı makalesinde sağlık hizmetlerinde finansal yönetim ve bütçe planlamasının kritik bir önem taşıması sebebi ile maliyet tahminine yönelik bir çalışma gerçekleştirmiştir. Bulut tabanlı Azure ML Studio platformu üzerinde hastane bilgilerini taklit etme amaçlı simule edilmiş bir veri seti

---

<sup>11</sup> IEEE: İngilizcesi: The Institute of Electrical and Electronics Engineers, Türkçesi Elektrik ve Elektronik Mühendisleri Enstitüsü olan kelimelerin kısaltmasıdır.

<sup>12</sup> ICDM: İngilizcesi International Conference on Data Mining, Türkçesi Veri Madenciliği üzerine Uluslararası Konferans olan kelimelerin kısaltmasıdır.



kullanarak 14 farklı makine öğrenmesi algoritmasının performans değerlendirmesi MAE<sup>13</sup>, RMSE<sup>14</sup>, RAE<sup>15</sup>, RSE<sup>16</sup>, CoD<sup>17</sup> vb. gibi matematiksel parametreler ile ortaya konulmaktadır (Botchkarev, 2018a).

Byung Wan Jo ve Rana Muhammad Asad Khan'ın "Azure ML Tabanlı Yeraltı Madenleri Hava Kalitesi Kirlilik Tahmini için Nesnelerin İnterneti Sistemi" adlı 2018 yılında yayınladığı makalesi, yeraltı kömür madenlerinde güvenliği artırmak amacı ile oluşturulan geçmişte sınırlı izleme ve birkaç olayı raporlama kabiliyetine sahip olan kablosuz sensör ağlarına, değerlendirme ve kirlilik tahmini özelliği eklenerek Arduino<sup>18</sup> tabanlı sensör modülleri, iletişim protokolleri ve üzerinde Azure ML Studio çalıştıran bir baz istasyonu gibi farklı modülleri bünyesinde barındıran güvenilir, verimli ve uygun maliyetli bütünsel bir nesnelerin interneti sistemine dönüştürülmesine dayalı bir çalışmadır. Sistem tarafından toplanan veriler makine öğrenmesi teknikleri ile maden hava kalitesini maden ortamı endeksi (MEI<sup>19</sup>) açısından hızlı bir şekilde değerlendirmekte ve maden ortam güvenliğinin artırılmasına katkıda bulunmaktadır (Jo ve Khan, 2018).

Xiaojia Guo ve arkadaşlarının "Gerçek Zamanlı Veri ve Makine Öğrenmesi Kullanarak Havalimanı Yolcu Transferi Akışı Tahminlemesi" adlı 2018 yılında yayınlanan makalesinde Londra Heathrow havalimanında gerçek zamanlı veri ile makine öğrenmesi tekniklerinin kullanıldığı transfer yolcuların bağlantı zamanlarının dağılımlı tahminini üreten bir öngörü sistemi üzerine çalışılmıştır. Araştırma sonuçları olarak yolcuları 16 segmente bölen ve yolcuların bağlantı zamanlarını etkileyen 7 anahtar faktör tanımlanmaktadır. Ayrıca aynı uçuş üzerinde gelen yolcuların bağlantı zamanları arasına

---

<sup>13</sup> MAE: İngilizcesi Mean Absolute Error, Türkçesi Ortalama Mutlak Hata olan tahmin edilen verilerin gerçek verilerden ne kadar uzaklaştığını anlamamızı sağlayan matematiksel parametrenin kısaltmasıdır.

<sup>14</sup> RMSE: İngilizcesi Root Mean Squared Error, Türkçesi Kök Ortalama Hatalar Karesi olan matematiksel parametrenin kısaltmasıdır.

<sup>15</sup> RAE: İngilizcesi Relative Absolute Error, Türkçesi Bağıl Mutlak Hata olan matematiksel parametrenin kısaltmasıdır.

<sup>16</sup> RSE: İngilizcesi Relative Squared Error, Türkçesi Bağıl Kareler Hatası olan matematiksel parametrenin kısaltmasıdır.

<sup>17</sup> CoD: İngilizcesi Coefficient of Determination, Türkçesi Belirlilik Katsayısı olan matematiksel parametrenin kısaltmasıdır.

<sup>18</sup> Arduino: İtalyan mühendisler tarafından geliştirilen açık kaynak kodlu geliştirme ortamına sahip fiziksel programlama yapılabilen bir bilgi işlem platformudur.

<sup>19</sup> MEI: İngilizcesi Mine Environment Index, Türkçesi Maden Ortam Endeksi olan kelimelerin kısaltmasıdır.

eklenen korelasyonların göç ve güvenlik alanlarına varış tahminlerini iyileştirebileceği saptanmaktadır (Guo vd., 2018).

İsmail Koç ve Emel Arslan'ın “Yapay Sinir Ağları Kullanarak Türkiye'deki İç Hatlar Taşımacılığı için Talep Tahmini” adlı 2018 yılında yayınlanan makalesi, 2007-2015 yıllarını içeren 8 yıllık bir veri seti ile Azure ML, R Script<sup>20</sup> ve MATLAB<sup>21</sup>'den yararlanılan ve yapay sinir ağları metodu kullanılan bir uygulamalı araştırmadır. Sonuç değerlendirmesi için performans kriteri olarak Pearson Korelasyon Katsayısı<sup>22</sup> kullanılmış ve önerilen modelden elde edilen sonuçların 0.79 – 0.93 arasında kabul edilebilir bir seviyede olduğu gözlemlenmiştir (Koc ve Arslan, 2018).

Navoneel Chakrabarty ve Sanket Biswas'ın “Yetişkin Nüfus Sayımı Gelir Düzeyi Tahminine İstatiksel Bir Yaklaşım” adlı 2018 yılında yayınlanan makalesi gelir eşitliği probleminde bir çözüm sağlamayı makine öğrenmesi ve veri madenciliği teknikleri kullanarak göstermeyi amaçlayan bir çalışmadır. UCI<sup>23</sup> yetişkin veri seti kullanılarak bir kişinin Amerika Birleşik Devletleri'ndeki yıllık gelirini belirli özelliklere dayanarak 50 bin dolardan daha büyük veya en az 50 bin dolara eşit olan gelir kategorisine düşüp düşmediğini tahmin etmek için sınıflandırma yapılmıştır. Gradient Boosting Classifier<sup>24</sup> %88.16 ile en yüksek doğruluk sonucunu ortaya koyan yöntem olmuştur (Chakrabarty ve Biswas, 2018).

Navoneel Chakrabarty'nin “Amerikan Havayolları için Uçuş Varış Gecikmesi Tahminine bir Veri Madenciliği Yaklaşımı” adlı 2019 yılında yayınlanan makalesi Amerikan Havayolları firması tarafından işletilen Amerika Birleşik Devletleri'ndeki en yoğun 5 havalimanını kapsayan iç hat uçuşlarının uçuş bilgilerinin Veri Madenciliği ve

---

<sup>20</sup> R Script: İstatistik konusunda özelleşmiş açık kaynaklı programlama dilidir.

<sup>21</sup> MATLAB: Sayısal analiz ve görselleştirme amacı ile kullanılan programlama dilidir. Matrix Laboratory kelimelerinin kısaltmasıdır.

<sup>22</sup> Pearson Korelasyon Katsayısı: Korelasyon kelime anlamı ilişki demektir. Korelasyon katsayısı ise iki değişkene ait değerler ile yapılan hesaplama sonucu ortaya çıkan -1 ile +1 arasında değerler alan katsayıdır. Bu katsayı Karl Pearson tarafından geliştirildiği için bu ismi almıştır.

<sup>23</sup> UCI: Makine Öğrenmesi için bir kaynak oluşturmak amacı ile California Üniversitesi Bilgisayar Bilimleri Bölümü tarafından oluşturulan herkesin kullanımına açık veri seti kaynağıdır. University of California Irvine Kelimelerinin kısaltmasıdır.

<sup>24</sup> Gradient Boosting Clasifier: Makine Öğrenmesi Gözetmeli Öğrenme tekniklerinden bir sınıflandırma algoritmasıdır.

Makine Öğrenmesi yaklaşımı kullanarak uçuşların olası varış gecikmesini tahmin etmeyi amaçlayan bir çalışmadır. Çalışmada Gradient Boosting Classifier modeli kullanılmış ve %85.73 doğrulukta sınıflandırma kategorisinde tahmin sonuçları elde edilmiştir (Chakrabarty, 2019).

No	Yazar Adı	Başlık	Yıl	Kategori
1	Awwalu, Jamilu Ghazvini, Anahita Abu Bakar, Azuraliza	Veri Madenciliği Algoritmalarının Performans Karşılaştırmaları: Otomobil Değerlendirme Veri Seti Üzerinde bir Vaka Analizi	2014	Veri Madenciliği
2	Hayakawa, Shin Hayashi, Hitoshi	Kapalı Alanlarda Konum Tahmini için Azure ML Kullanma	2017	Makine Öğrenmesi, Tahmin
3	Nigam, Rahul Govinda, K.	Lojistik Regresyon kullanarak Bulut Tabanlı Uçuş Gecikmesi Tahmini	2017	Makine Öğrenmesi, Tahmin
4	Botchkarev, Alexei	Azure ML Studio kullanarak Hastane Maliyet Tahmin Modeli Vaka Değerlendirmesi	2018	Makine Öğrenmesi, Tahmin
5	Jo, Byung Wan Khan, Rana Muhammad Asad	Azure ML Tabanlı Yeraltı Madenleri Hava Kalitesi Kirlilik Tahmini için Nesnelerin İnterneti Sistemi	2018	Makine Öğrenmesi, Tahmin
6	Guo, Xiaojia Grushka-Cockayne, Yael De Reyck, Bert	Gerçek Zamanlı Veri ve Makine Öğrenmesi Kullanarak Havalimanı Yolcu Transferi Akışı Tahminlemesi	2018	Makine Öğrenmesi, Tahmin
7	Koc, Ismail Arslan, Emel	Yapay Sinir Ağları Kullanarak Türkiye'deki İç Hatlar Taşımacılığı için Talep Tahmini	2018	Yapay Sinir Ağları, Tahmin
8	Chakrabarty, Navoneel Biswas, Sanket	Yetişkin Nüfus Sayımı Gelir Düzeyi Tahminine İstatiksel Bir Yaklaşım	2018	Makine Öğrenmesi, Veri Madenciliği
9	Chakrabarty, Navoneel	Amerikan Havayolları için Uçuş Varış Gecikmesi Tahminine bir Veri Madenciliği Yaklaşımı	2019	Veri Madenciliği

**Tablo 3: Literatür Taraması Makine Öğrenmesi & Azure ML Bazında**

### 3.4 Makine Öğrenmesi & Havalimanları Kapsamında Kümeleme Üzerine Yapılan Araştırmaların İncelenmesi

Makine Öğrenmesi kümeleme konusunda havalimanları özelinde gerçekleştirilen uygulamaların çeşitli örneklerine ilişkin çalışılmış bazı bilimsel araştırmalar;

- Kümeleme tekniği ile ne çeşit uygulamalar yapılmıştır?
- Kümeleme tekniği ile uygulama yapılırken izlenen adımlar ve yöntemler nelerdir?
- Elde edilen sonuçların doğruluğu hangi matematiksel parametreler ile ortaya konulmuştur?

Gibi sorulara cevap bulmak amacı ile incelenmiştir. Bu araştırmalar Tablo 4’te toplanmıştır.

Joseph Sarkis ve Srinivas Talluri’nin “Amerika Birleşik Devletleri Havalimanlarının Karşılaştırması için Performansa Dayalı Kümeleme” adlı 2004 yılında yayınlanan makalesi Amerika Birleşik Devletleri’nde yer alan 44 havalimanı için operasyonel verimliliği değerlendirmek amacı ile Veri Zarflama Analizi ve Kümeleme metodu kullanılarak havalimanlarının karşılaştırılmasına yönelik bir çalışmayı kapsamaktadır. Verimlilik ölçümleri için girdi verileri; havalimanı operasyonel maliyetleri, havalimanı çalışan sayıları, havalimanı yolcu giriş kapısı sayısı ve havalimanı pist sayısı olmak üzere 4 parametreyi, çıktı verileri; operasyonel gelir, yolcu akışı, ticari ve genel havacılık hareketleri ve toplam kargo taşımacılığı olmak üzere 5 parametreyi içermektedir. Veriler direk olarak havalimanlarının kendisinden veya ACI<sup>25</sup>’dan toplanmıştır (Sarkis ve Talluri, 2004).

Paolo Malighetti ve arkadaşlarının “Avrupa Ağı içerisinde Havalimanı Sınıflandırması ve İşlevselliği” adlı 2009 yayınlanan makalesi havalimanlarını sınıflandırmak ve stratejik grupları tanımlamak amacı ile havalimanlarının hem özelliklerini hem de Avrupa ağı içerisindeki konumlarını değerlendiren bir çalışmadır. 2007 sonbaharında en az bir tarifeli uçuşa sahip 467 adet havalimanı geleneksel

---

<sup>25</sup> ACI: İngilizcesi Airports Council International, Türkçesi Uluslararası Havalimanları Konseyi olan 1991 yılında İsviçre’de kurulan günümüzde merkezi Kanada da bulunan uluslararası havacılık otoritesidir.

kümeleme ve Simulated Annealing<sup>26</sup> metodu kullanılarak analiz edilmiştir. Araştırma kapsamında optimum küme sayısını bulmak için farklı teknikler incelenmiş ve Dendogram<sup>27</sup> yapısı tarafından önerilen ve Duda / Hart <sup>28</sup>endeksi tarafından da teyit edilen 8, optimum küme sayısı olarak kabul edilmiştir (Malighetti vd., 2009).

Zambochova Marta'nın "Dünya Havaalanlarının Ele Alınan Yolcu Sayısı Bazında Küme Analizi (Önemli Olayların Etkisinin İncelenmesi Vaka Çalışması)" adlı 2017 yılında yayınlanan makalesi K-Ortalamlar kümeleme metodu ile küme analizini ele almaktadır. Optimum küme sayısını belirlemek amacı ile Bayesian Information Criterion (BIC) ve <sup>29</sup>Akaike Information Criterion (AIC) <sup>30</sup>olarak isimlendirilen 2 farklı matematiksel metot kullanılmıştır. Her iki metot için de optimal küme sayısı 3 bulunmuş ve oluşturulan 3 kümenin kalitesi Silhouette Coefficient <sup>31</sup>metodu ile değerlendirilmiştir (Marta, 2017).

Ertuğrul Ayyıldız ve Selin Yalçın'ın "Havalimanlarının Kümeleme Yöntemleri ile İncelenmesi: Türkiye Örneği" adlı 2018 yılında yayınlanan makalesi Türkiye'de işletilen 55 havalimanının kümeleme metodu kullanılarak benzer özelliklerine göre analiz edilmesi hakkındadır. T.C. Devlet Hava Meydanları İşletmesi 2017 Aralık ayı verilerinin

---

<sup>26</sup> Simulated Annealing: Türkçesi Benzetilmiş Tavlama olan simülasyon tabanlı optimizasyon tekniğidir.

<sup>27</sup> Dendogram: Hiyerarşik Kümeleme kapsamında diyagram şeklinde kümelerin dizilişini gösteren metottur.

<sup>28</sup> Duda/Hart endeksi: Kümeleme yaparken optimum küme sayısını bulmada nerede durulması gerektiğini araştıran durdurma kuralı olarak da isimlendirilen yöntemlerden bir tanesidir. Detaylı bilgi için "Cluster Analysis Stopping Rules in Stata" (Halpin, 2016) makalesine ve "Duda, R. O., Hart, P. E. and Stork, D. G. (2000) Pattern Classification, 2 edn. New York, John Wiley." kitabına bakabilirsiniz.

<sup>29</sup> Bayesian Information Criterion (BIC): 1978'de Gideon E. Schwarz tarafında yayınlanan makalede formüle edilen optimum küme sayısını bulmaya yönelik minimum değeri belirlemeyi sağlayan yöntemlerden biridir. AIC yöntemi ile doğrudan ilişkilidir.

<sup>30</sup> Akaike Information Criterion (AIC): 1970'lerin başında Hirotugu Akaike tarafından formüle edilen optimum küme sayısını bulmaya yönelik ortalama değeri belirlemeyi sağlayan yöntemlerden biridir.

<sup>31</sup> Silhouette Coefficient: Veri kümelerinin ne kadar iyi sınıflandırıldığı hakkında fikir geliştirmeyi sağlayan bir küme elemanının kendi kümesine ne kadar benzediğinin ölçüsünü ortaya koyan tutarlılığı yorumlama yöntemidir.

kullanıldığı çalışmada Elbow<sup>32</sup> ve Ward<sup>33</sup>'s metotları kullanılarak ideal küme sayısı 6 olarak belirlenmiştir (Ayyildiz ve Yalcin, 2018).

Qingjun Xia ve arkadaşlarının “Geliştirilmiş K-Ortalamlar Kümeleme Algoritmasına Dayalı Havaalanı Rol Oryantasyonu” adlı 2019 yılında yayınlanan makalesi geliştirilmiş K-Ortalamlar algoritması kullanılarak Çin’de yer alan farklı tipteki 213 havalimanının rolleri hakkında bir fikir vermeyi amaçlamaktadır. MATLAB üzerinde simule edilen algoritma geleneksel K-Ortalamlar algoritmasında kümelerin kalitesinin büyük oranda küme merkezinin seçimine bağlı olması sebebi ile başlangıç küme merkezini optimize eden bir algoritma ile geliştirilerek sonuçlar değerlendirilmiştir (Xia vd., 2019).

D L Trenggonowati ve arkadaşlarının “K-Ortalamlar Kümeleme Algoritması Kullanarak Havalimanları Kümeleme Organizasyonu” adlı 2019 yılında yayınlanan makalesi Endonezya’da havalimanlarının işletmesinden sorumlu “PT. Angkasa Pura II” olarak isimlendirilen devlet kuruluşunun 21 havalimanının projelendirilmesi kapsamında yapılan bir incelemedir. Bu çalışmada şirketin 5 küme oluşturmak istediği bilindiği için Bölümleyici Kümeleme metodu (Partition Clustering Method<sup>34</sup>) kullanılmıştır (Trenggonowati vd., 2019).

---

<sup>32</sup> Elbow Metodu: Bir veri setinde optimum küme sayısını bulmak amacı ile kullanılan bir metottur. Oluşturulan küme fonksiyonun açıklanan varyans yüzdeleri ilk kümeler için çok fazla varyansı açıkladığı için büyük bir değer almaktadır ancak bir noktadan sonra bu değerler belirgin şekilde düşmeye başlar. Bu durumu bir grafik olarak incelediğimizde bahsedilen belirgin nokta gözlemlenebilmektedir. Bu da bize Elbow kriterini ya da başka bir deyişle optimum küme sayısını veren ifadedir. Türkçe’de dirsek metodu olarak da anılmaktadır.

<sup>33</sup> Ward Metodu: Joe H. Ward tarafından formüle edilen Hiyerarşik Kümeleme analizinde kullanılan minimum varyans metodu olarak da bilinen metottur.

<sup>34</sup> Partition Clustering Method: Türkçesi bölümlenmeli kümeleme metottur. Hiyerarşik, Merkezi Uzaklık kümeleme yöntemleri gibi bir başka yöntemdir.

No	Yazar Adı	Başlık	Yıl	Kategori
1	Sarkis, Joseph Talluri, Srinivas	Amerika Birleşik Devletleri Havalimanlarının Karşılaştırması için Performansa Dayalı Kümeleme	2004	Veri Zarflama, Kümeleme
2	Malighetti, Paolo Paleari, Stefano Redondi, Renato	Avrupa Ağı içerisinde Havalimanı Sınıflandırması ve İşlevselliği	2009	Kümeleme
3	Marta, Zambochova	Dünya Havaalanlarının Ele Alınan Yolcu Sayısı Bazında Küme Analizi (Önemli Olayların Etkisinin İncelenmesi Vaka Çalışması)	2017	Makine Öğrenmesi, Kümeleme
4	Ayyıldız, Ertugrul Yalcin, Selin	Havalimanlarının Kümeleme Yöntemleri ile İncelenmesi: Türkiye Örneği	2018	Makine Öğrenmesi, Kümeleme
5	Xia, Qingjun Zhang, Zhaoyue Zhang, Baochen	Geliştirilmiş K-Ortalamlar Kümeleme Algoritmasına Dayalı Havaalanı Rol Oryantasyonu	2019	Makine Öğrenmesi, Kümeleme
6	Trenggonowati, D L Ulfah, M Ekawati, R Yusuf, V A	K-Ortalamlar Kümeleme Algoritması Kullanarak Havalimanları Kümeleme Organizasyonu	2019	Makine Öğrenmesi, Kümeleme

**Tablo 4: Literatür Taraması Makine Öğrenmesi & Havalimanları Kümeleme Bazında**

## 4 DÖRDÜNCÜ BÖLÜM

### METODOLOJİ

Tez çalışmasına ilk 3 bölümde yapılan araştırmalar ele alınarak başlanmıştır daha sonra literatür taraması kapsamında uygulamaya yönelik makaleler de hangi tür veri setleri ile hangi tür metotları kullanarak ne çeşit analizler yapılmıştır sorularına cevap arayarak havalimanı analizi için bir yol haritası belirlemenin ilk adımı atılmıştır. Tablo 5 hem havalimanı analizine yönelik hem de genel anlamda makine öğrenmesi uygulamalarında kullanılan veri setlerinin özelliklerini incelemek amacı ile hazırlanmıştır.

#### 4.1 Verilerin Toplanması ve Derlenmesi Süreci

Havalimanları analizinde teknik veriler, yolcu sayılarına ilişkin veriler ve bölge nüfusuna ilişkin veriler havalimanları analizlerinde uygulama örneklerinde kullanılan veri seti özellikleri baz alınarak 3 farklı kategoride toplanmıştır. İlk başta sadece Türkiye havalimanlarını içeren bir veri seti oluşturulmuş ve ön değerlendirmeler dikkate alınarak havalimanı sayısının 54 ile sınırlı kalmasının nispeten daha kısır sonuçlar ürettiği saptanmıştır. Örneklendirecek olursak; Türkiye de yer alan İstanbul Sabiha Gökçen Havalimanı, İstanbul Havalimanı, İstanbul Atatürk Havalimanı, Ankara Esenboğa Havalimanı, Antalya Havalimanı ve İzmir Adnan Menderes Havalimanı gibi büyük havalimanlarının içerdiği veriler matematiksel anlamda diğer havalimanlarına göre çok farklı bir boyutta olabilmektedir. Nitekim Veri zarflama vb. yöntemler kullanılarak yapılan çalışmalarda bu havalimanlarının kapsam dışı bırakıldığı birçok örnekte gözlemlenmiştir. Ayrıca Makine Öğrenmesi yaklaşımı ile yapılan uygulama örneklerinde örneklem sayısı ne kadar fazla ile o kadar gerçeğe yakın sonuçların elde edildiği bilinen bir durumdur. Bu sebeple veri seti Türkiye'ye en yakın ve Türkiye'nin gerek ekonomik gerekse kültürel anlamda en çok etkileşimde bulunduğu Avrupa bölgesi havalimanlarını kapsayacak şekilde genişletilmiştir.



Kaynak	Başlık	Veri Seti Örnekleme	Veri Seti Parametreleri	Model / Hedef	Algoritmalar	Değerlendirme Sonuçları
<b>Awwalu, Jamilu Ghazvini, Anahita Abu Bakar, Azuraliza 2014</b>	Veri Madenciliği Algoritmalarının Performans Karşılaştırması; Araba Değerlendirme Veri Seti Üzerinde Vaka Analizi	Araba Modelleri (Satır Sayısı: 1728)	Satın Alma, Bakım, Kişi Sayısı, Kapı Sayısı, Bagaj Kapasitesi, Güvenlik	Sınıflandırma, Kümeleme, Algoritma Performanslarını karşılaştırma	Karar Ağacı (Decision Tree) (Sınıflandırma)	93.22
					Naive Bayesian (Sınıflandırma)	93.51
					Yapay Sinir Ağları (Artificial Neural Network) (Sınıflandırma)	93.51
					Karar Ağacı (Decision Tree) (Kümeleme)	100
					Naive Bayesian (Kümeleme)	100
					10 Kat Cross Validation (Kümeleme)	100

Kaynak	Başlık	Veri Seti Örnekleme	Veri Seti Parametreleri	Model / Hedef	Algoritmalar	Değerlendirme Sonuçları
Botchkarev, Alexei 2018	Azure ML Studio Kullanarak maliyet tahmin modeli ile hastane değerlendirmesi	Teşhis ya da Müdahale (Satır Sayısı: 7000)	Cinsiyet, Yaş Grubu, Kalma Uzunluğu, Maliyet	Maliyet Tahmini, Algoritma Performanslarını karşılaştırma	Lineer Regresyon	0.55 - 0.75
					Bayesyen Lineer Regresyon	0.71 - 0.75
					Sinir Ağları (Neural Network) Regresyonu	(-0.01) – (-0.09)
					Karar Ormanı (Decision Forest) Regresyonu	0.57 - 0.72
					Poisson Regresyon	0.23 - 0.17
					Artırılmış Karar Ağacı (Boosted Decision Tree) Regresyonu	0.66 - 0.71
Hayakawa, Shin Hayashi, Hitoshi 2017	Kapalı Mekân Konum Tahmini için Azure ML Kullanma	Çeşitli Kablosuz Erişim noktalarından alınan sinyaller (Satır Sayısı: 3196)	Konum Etiketleri, Alan Sayısı, Alınan Sinyal Gücü	Alınan Sinyal Gücü Sayılarından Alan Tahmini	Çok Sınıflı Karar Ormanı (Multiclass Decision Forest) (Alt Kriter: Bagging)	0.6967

Kaynak	Başlık	Veri Seti Örnekleme	Veri Seti Parametreleri	Model / Hedef	Algoritmalar	Değerlendirme Sonuçları
Bihis, Matthew Roychowdhury, Sohini 2015	Çoklu Sınıf ve İkili Sınıflandırma görevleri için bir geliştirme; Azure ML Yaklaşımı	1 Wisconsin Göğüs Kanseri - İkili kategorik içerikte veri Seti (Satır sayısı: 683)	1 Wisconsin Göğüs Kanseri Veri Seti (Sütun Sayısı: 10)	Medikal görüntüler üzerinden patoloji tespiti	1 Wisconsin Breast Cancer - Binary Support Vector Machine	1 Wisconsin Breast Cancer 99.7
		2 Almanya Kredi Kartı İkili Sınıflandırma Veri Seti (Satır Sayısı: 1000)	2 Almanya Kredi Kartı Veri Seti (Sütun Sayısı: 20),		2 German Credit Card - Binary Neural Network	2 German Credit Card 77.1
		3 Teleskop ikili sınıflandırma veri seti (Satır sayısı: 19020)	3 Teleskop Veri Seti (Sütun Sayısı: 10)		3 Telescope Data - Binary Boosted Decision Tree	3 Telescope Data 87.3
		4 DIARETDB1 - 89 iç organ görüntüleri içeren veri seti (1500 x 1152) piksel	Kırmızı, Yeşil, Mavi renklerinin her bir piksel için oran değeri		4 DIARETDB1 - One vs All Boosted Decision Tree - Binary Boosted Decision Tree - Binary Decision Forest - Binary Support Vector Machine	4 DIARETDB1 - 0.82 - 0.71

Kaynak	Başlık	Veri Seti Örnekleme	Veri Seti Parametreleri	Model / Hedef	Algoritmalar	Değerlendirme Sonuçları
Koc, Ismail Arslan, Emel 2018	Yapay Sinir Ağları kullanarak Türkiye'de İç Hatlar Hava Ulaşımı için Talep Tahmini	01.01.2007 - 01.11.2015 tarihleri arasında aylık periyotlarla elde edilen 107 gözlemden oluşan veri seti	Yolcu Sayısı, Nüfus Sayısı Yurt dışına çıkan vatandaş sayısı, Gelen Turist Sayısı, Kişi Başına Düşen Milli Gelir Tüketici fiyat endeksi Petrol Fiyatı, İş Gücü Turist Konaklaması için Yatak Kapasitesi, Havayolu şirketi Koltuk kapasitesi, Bilet Fiyatı	Türkiye'de iç hatlar yolcu taşımacılığı için talep tahmini ve algoritmaların performans karşılaştırması	Multiple Linear Regression	Azure ML Neural Network Tune Regression: 0.87 to 0.91
					Back-propagation Neural Network	Azure ML Neural Network Train Regression: 0.88 to 0.91
					Gravity	R Script Neural Network 0.91
					Time series	MATLAB Time Delay 0.79 to 0.89

Kaynak	Başlık	Veri Seti Örnekleme	Veri Seti Parametreleri	Model / Hedef	Algoritmalar	Değerlendirme Sonuçları
Chakrabarty, Navoneel 2019	Amerikan Havayolları için Uçuş Varış Gecikmesi Tahminine Bir Veri Madenciliği Yaklaşımı	97360 Örneklem	Uçuş Numarası, Havalimanı Kodu, Dünya alan Kodu, Varış Havalimanı Kodu, Varış Dünya Alan Kodu, Gidiş Zamanı, Benzersiz Havayolu Şirketi Kodu, Geliş Zamanı, İkili etiketleme (gecikeme varsa 1, yoksa 0)	Varış Gecikmesi Tahmini	Gradient Boosting Classifier	0.76
					Gradient Boosting Classifier (Grid Search)	0.77

Kaynak	Başlık	Veri Seti Örneklemi	Veri Seti Parametreleri	Model / Hedef	Algoritmalar	Değerlendirme Sonuçları
Guo, Xiaojia Grushka-Cockayne, Yael De Reyck, Bert 2018	Gerçek Zamanlı Veri ve Makine Öğrenmesi Kullanarak Havalimanı Yolcu Akışı Tahmini	Yolcu bazlı bireysel veriler içeren veri seti (sattır sayısı: 3,762,690)	30 değişken bulunmaktadır. Zaman, Hava aracı Gövdesi, Hava aracı Tipi, Uçuş Kapasitesi, Toplam Yolcu Kapasitesi, Toplam Taşınan Yolcu Sayısı, Pist Numarası, Tarife Zamanı, Stant Numarası, Havalimanına giriş yapan hava aracı zamanı, Uçuş Numarası, Yolcu Seyahat Sınıfı Varış Terminali, Yerel Saat, Bölge Numarası, (varış için geliş bölgesi numarası) Varış Uçuşu için Dakiklik ...	Yolcu Bağlantı Zamanı, Yolcu Akışı, Yolcu Gecikmeleri gibi 19 farklı tahmin incelemesi	Naive Model	MAE (Mean Absolute Error);10.20, Pinball Losses Average;3.38
					Lineer Regresyon	MAE (Mean Absolute Error);8.52, Pinball Losses Average;3.01
					Quantile Regression	MAE (Mean Absolute Error);8.18, Pinball Losses Average;2.74
					Quantile Regression Forest	MAE (Mean Absolute Error);8.24, Pinball Losses Average;2.77
					Gradient Boosting Machine	MAE (Mean Absolute Error);8.38, Pinball Losses Average;2.81
					Regression Tree	MAE (Mean Absolute Error);8.16, Pinball Losses Average;2.73

Kaynak	Başlık	Veri Seti Örnekleme	Veri Seti Parametreleri	Model / Hedef	Algoritmalar	Değerlendirme Sonuçları
Nigam, Rahul Govinda, K. 2017	Lojistik Regresyon kullanarak Bulut Tabanlı Uçuş Gecikmesi Tahmini	Amerika Birleşik Devletlerindeki en yoğun 70 havalimanı bilgilerini içeren veri seti	Gecikmeleri içeren Veri Setinden; Havalimanı Numarası, Destinasyon Sayısı, Geliş Zamanı, Tarife ve Gerçek Geliş zamanı arasındaki fark, 15 dakikadan daha fazla olan Geliş gecikmeleri, Varış Zamanı, Tarife ve gerçek varış zamanı arasındaki fark, İptal edilen uçuşlar. Hava koşullarını içeren Veri setinden; Havalimanı Numarası, Yıl, Ay, Gün, Saat, Zaman Dilimi, Gökyüzü Durumu, Görünürlük, Kuru termometre Sıcaklığı, Hava Tipi, Bağlı Nem, Rüzgar Hızı, Rüzgar Yönü, Basınç, Basınç Trendi, Basınç Değişimi, Deniz Seviyesi Basıncı, İrtifa.	Uçuş Gecikmesi Tahmini	Lojistik Regresyon	Accuracy:80.6

Kaynak	Başlık	Veri Seti Örnekleme	Veri Seti Parametreleri	Model / Hedef	Algoritmalar	Değerlendirme Sonuçları
Chakrabarty, Navoneel Biswas, Sanket 2018	Yetişkin Nüfus Sayımı Gelir Düzeyi Tahminine İstatistiksel Bir Yaklaşım	Nüfus Sayımı Veri seti (Satır sayısı: 48,842)	Yaş, Çalışma Sınıfı, Eğitim Seviyesi, Uyuşu, Medeni Hal, İş, Akrabalık, Irk, Cinsiyet, Ana Kazanç, Ana Harcama, Haftalık Çalışma Saati, Ülke	Bir kişinin yıllık 50 bin dolar veya verilen başka bir değerden daha fazla kazanıp kazanmadığının tahmini	Gradient Boosting Classifier	Accuracy:88.16
Chakrabarty, Navoneel 2019	Zehirlilik Sınıflandırmasına Yorum Yapmak için Bir Makine Öğrenmesi Yaklaşımı	159571 Örneklem	Yorum Metni, Zehir, Şiddetli Zehir, Tiksindirme, Korkutma, Saldırma, Nefret	Binominal Sütunlar için zehirlilik tahmini	Linear Support Vector Machine, Decision Tree Classifier	96.01 - 99.66
Qasem, Mohammed Thulasiram, Ruppa Thulasiram, Parimala 2015	Borsalar için Makine Öğrenmesi Teknikleri kullanarak Twitter Duyarlılık Sınıflandırması	Tweet veri seti (Satır Sayısı: 42 000)	Kullanıcı Adı, Web Bağlantıları, Hashtagleri, Cashtagleri, Harf tekrarları, İnkâr etmeler, Ünlem ve Soru işaretleri, Emojiler	Borsa Fiyat Tahmini	Multiclass Neural Network	0.68 - 0.72
					Multiclass Logistic Regression	0.70 - 0.72

**Tablo 5: Uygulama Örnekleri Veri Seti Özellikleri İncelemesi**



Verilerin toplanmasında ACI<sup>35</sup>, EUROSTAT<sup>36</sup>, OurAirports (DAFIF<sup>37</sup>), ICAO<sup>38</sup>, IATA<sup>39</sup>, TÜİK<sup>40</sup>, DHMİ<sup>41</sup> ve bazı ülkeler için yerel resmi sitelerden yararlanılmıştır.

Verilerin toplanmasında yukarıda ismi geçen kurumların web sitelerinde detaylı inceleme yapıldıktan sonra Türkiye’de yer alan 51 havalimanına ait 2017, 2018 ve 2019 dönemini kapsayan uçak trafiği, taşınan yolcu ve yük sayıları gibi veriler (DHMİ, 2019), Türkiye il bazlı nüfus verileri (TÜİK, 2019) öncelikli olarak temin edilmiştir. Daha sonra incelemeler genişletilerek dünya havalimanlarına ait pist uzunlukları, rakım, konum, kapasite gibi teknik veriler (OurAirports, 2019) uluslararası benzersiz havacılık kodları (IATA, 2019; ICAO, 2019) ile alınmıştır. Aynı şekilde Avrupa kapsamında havalimanı trafik verileri ve nüfus verileri (Eurostat, 2019) toplanmıştır. Bazı eksik veriler manuel olarak devletlerin yerel resmi sitelerinden ve Uluslararası Havalimanları Konseyi’nin (ACI, 2019) de faydalandığı ve sitesinde paylaştığı güvenilir sitelerden elde edilmiştir.

Nihayetinde farklı kaynaklardan toplanan veriler IATA ve ICAO benzersiz alan isimleri baz alınarak birleştirilmiştir. Başlangıçta 828 havalimanı için 24 farklı parametre içeren bir veri seti oluşturulmuştur. Tamamlanması mümkün olmayan eksik veri içeren 268 havalimanı çıkartılarak geriye kalan 560 havalimanı ile veri seti analiz için hazır hale getirilmiştir.

---

<sup>35</sup> ACI (Airports Council International): 1991 yılında dünya havalimanlarının ticari anlamda küresel olarak temsil edilmesi, havalimanları için standartlar, politikalar ve faydalı uygulamalar geliştirilmesi amacı ile kurulan uluslararası bir kurumdur.

<sup>36</sup> EUROSTAT: Lüksemburg’da konumlanmış Avrupa Birliği İstatistik resmi ofisidir.

<sup>37</sup> DAFIF (Digital Aeronautical Flight Information File): 2006 yılında halka açık erişime kapatılan Amerika merkezli resmi havacılık veri kaynağıdır. 2006 yılı sonrası bu ihtiyacı karşılamak amacı ile ourairports.com sitesinde David Megginson tarafından günümüze dek veriler halka açık biçimde güncellenmektedir. Daha fazla bilgi için : <https://ourairports.com/>

<sup>38</sup> ICAO (International Civil Aviation Organization): 1994 yılında Uluslararası Sivil Havacılık Organizasyonu olarak Uluslararası Sivil Havacılık sözleşmesinin (Chicago Sözleşmesi) yönetimini gerçekleştirmek için kurulmuş Birleşmiş Milletler uzman kuruluşudur. 193 üye devlet ve endüstri grupları birlikte çalışmaktadır.

<sup>39</sup> IATA (International Air Transport Association): 290 havayolu şirketi ve dünya hava trafiğinin %82’sini temsil eden ticari birliktir. 1945’te Küba Havana’da kurulmuştur.

<sup>40</sup> TÜİK: Türkiye İstatistik Kurumudur. Geçmiş Osmanlı Devleti’ne kadar dayanan Türkiye’nin birçok farklı konuda verilerini toplayan, düzenleyen ve halka açık paylaşımını yapan resmi kuruluştur.

<sup>41</sup> DHMİ: Devlet Hava Meydanları İşletmesi Türkiye hava sahası hava trafiği kontrolü ve düzenlenmesi görevini üstlenen resmi kuruluştur.

İstanbul Atatürk havalimanının kapatılması sonrası faaliyete geçen İstanbul Havalimanı ve Avrupa’da yer alan 12 havalimanı sadece havalimanı yolcu sayısı tahmini uygulamasında eksik ya da tutarsız veri içerdiği için hariç tutulmuştur. Böylelikle Yolcu Sayısı Tahmini uygulaması için 548 satırlık, Kümeleme Uygulaması için ise 560 satırlık 2 veri seti derlenmiştir.

Her iki uygulama için de derlenen veriler sisteme yüklendikten sonra ön işleme tabi tutulmaktadır. Ön işlemeden kasıt, modele dahil edilecek sütunların (feature) seçimi, normalizasyon ve eğitimde kullanılacak sütunların seçilmesidir. Modele dahil edilecek sütunlar, veri setinin derlenmesi sürecinde farklı veri setlerinden toplanan verileri eşleştirmek amacı ile oluşturulan ve anahtar olarak tutulan bazı demografik özellikteki sütunların elimine edilmesi ayrıca kategorik özellikteki sütunların numerik değerlerinin seçilmesi süreçleri göz önünde bulundurularak seçilmiştir.

Verilerin normalizasyonu Azure ML üzerinde “Z-score” ve “Log-Normal” metotları kullanılarak gerçekleştirilmiştir. Z-score bir değer eğer tüm değerlerin ortalamasına eşitse 0’a, ortalamasından büyükse pozitif bir değere ve küçük ise negatif bir değere normalleştirme işlemi gerçekleştirirken Log-Normal doğal logaritma fonksiyonu ile normalleştirme işlemini gerçekleştirmektedir.

## **4.2 Modelin Eğitimi için Kullanılacak Algoritmaların Seçimi**

### **4.2.1 Havalimanı Yolcu Sayısı Tahmini için Seçilen Algoritmalar**

#### **4.2.1.1 Lineer Regresyon (Linear Regression)**

Lineer regresyon bir diğer adıyla doğrusal regresyon bir ya da birden fazla bağımsız değişken ile bağımlı değişken veya sayısal bir sonuç arasında doğrusal bir ilişki kurmayı amaçlamaktadır. En basit anlamda sayısal bir hedefin tahmin edilmesini ifade eden lineer regresyon karmaşık olmayan çok boyutlu seyrek veri kümelerinde hızlı ve güvenilir sonuçlar üretme eğilimindedir.

Azure Machine Learning Studio içerisinde yer alan Lineer Regresyon modülü hatayı ölçmek ve regresyon çizgisine uymak için En Küçük Kareler Yöntemi ve Gradyan

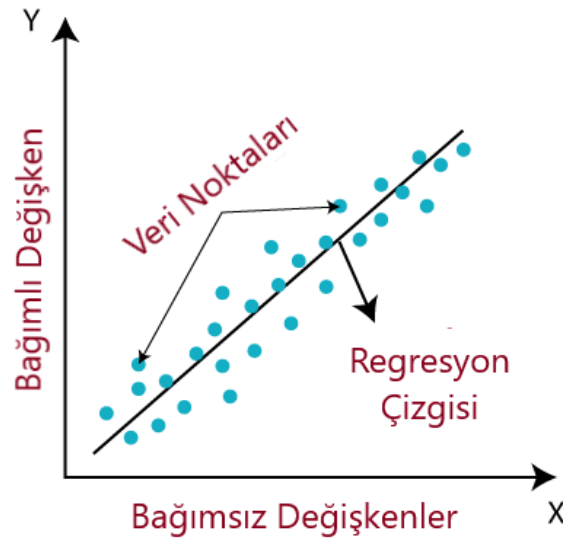
İniş (Gradient Descent)<sup>42</sup> yöntemini desteklemektedir. En Küçük Kareler Yöntemi lineer regresyonda en sık kullanılan tekniklerden biri olup hatanın daha açıklayıcı bir ifadeyle gerçek değerden tahmin edilen çizgiye olan uzaklığın karesi toplamı alınarak hesaplanan değer in karekökünü en aza indirmeyi hedefleyen kayıp fonksiyonunu ifade etmektedir. Gradyan İniş yöntemi ise bir modelde eğitim sürecinin her bir adımında hata miktarını en aza indirmeyi amaçlayan farklı öğrenme problemleri için ayarlanabilir çeşitli parametrelere sahip bir yöntemdir (Microsoft, 2019). Şekil 8’de Lineer Regresyon matematiksel formülü ve Şekil 9’da grafiksel gösterimi yer almaktadır.

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

Sabit Terim
Eğim
Bağımsız Değişken
Rastgele Hata Terimi

Bağımlı Değişken
İndis
Doğrusallık Bileşeni
Rastgele Hata Bileşeni

Şekil 8: Lineer Regresyon Matematiksel Formülü



Şekil 9: Lineer Regresyon Grafiksel Gösterimi

<sup>42</sup> Gradyan İniş (gradient descent): Bir kayıp fonksiyonunu optimize eden bir dizi parametreyi bulmaya yarayan bir algoritmadır.

#### 4.2.1.2 Bayesyen Lineer Regresyon (Bayesian Linear Regression)

Bayesyen yaklaşım, önsel olasılık dağılımı (tecrübeye dayalı olasılık olarak da isimlendirilir.) formunda ek bilgilerle desteklenen lineer regresyon modelini kullanmaktadır. Parametreler hakkında önsel bilgiler parametreler için tahminler üretmek üzere bir olasılık fonksiyonu ile birleştirilir (Ekici, 2005; Microsoft, 2019).

Azure ML Studio üzerinde yer alan Bayesyen Lineer Regresyon modülü, Aşırı Uymayı (Overfitting)<sup>43</sup> önlemek amacı ile düzenleme ağırlığı (Regularization Weight) ve bilinmeyen değerleri gruplamak için bilinmeyen kategorik düzeylere izin ver (allow unknown categorical levels) olmak üzere iki araç ile küçük ayarlar yapılmasına olanak sağlamaktadır (Bishop ve Tipping, 2003).

#### 4.2.1.3 Karar Ormanı Regresyonu (Decision Forest Regression)

Karar ormanı olarak isimlendirilen bu algoritma bir grup karar ağacına dayanan bir regresyon modeli oluşturmak amacı ile kullanılmaktadır. Karar ağaçları her bir örnek için bir dizi basit test gerçekleştiren parametrik olmayan, bir yaprak düğüme (karar) ulaşılan kadar ikili ağaç veri yapısını uçtan uca geçen modellerdir.

Doğrusal olmayan karar sınırlarının temsili, eğitim ve tahmin sırasında hem hesaplamada hem de bellek kullanımında randımanlı çalışması gibi avantajlara sahiptir.

Bir karar ormanı regresyonundaki her bir ağaç tahmin yoluyla Gauss dağılımına uygun çıktılar üretmektedir. Modeldeki tüm ağaç yapıları için kombine dağılıma en yakın Gauss dağılımını bulmak amacı ile ağaç grubu üzerinde bir kümeleme yapılmaktadır. Bu kümeleme (aggregation) için Azure ML üzerinde Karar Ormanı modülü içerisinde Torbalama (Bagging or Bootstrap Aggregating) ve Çoğaltma (Replicate) olmak üzere iki farklı yöntem bulunmaktadır. Torbalama (Bagging or Bootstrap Aggregating) tek tek ağaçların döndürdüğü tüm dağılımları birleştirerek verilen Gauss Dağılımlarının karışım döngülerinin eşleştiği Gauss dağılımının ilk iki döngüsünü bulmaktır. Çoğaltma (Replicate) ise her bir ağacın tam olarak aynı giriş verisi ile eğitildiği ve her bir ağaç

---

<sup>43</sup> Aşırı Uyma (Overfitting): Bir modelde analizi gerçekleştirilen veri kümesinin analiz sonuçları ile aşırı oranda uyum sağlaması dolayısıyla modelin farklı bir veri seti ile karşılaştığında bu yeni verilere uyum sağlayamaması durumudur.

düğümü için hangi ayrık ifadenin kullanıldığının tespit edilmesi rastgele olan bir yöntemdir. Rastgele olması ağaç yapısı için çeşitliliği sağlamaktadır (Microsoft, 2019).

#### **4.2.1.4 Artırılmış Karar Ağacı Regresyonu (Boosted Decision Tree Regression)**

Artırma (boosting) her bir ağacın önceki ağaçlara bağlı olduğu anlamında kullanılmıştır. Artırma (boosting); torbalama (bagging), rastgele ormanlar (random forests) ve benzerleri ile topluluk modelleri oluşturmak için kullanılan birkaç klasik yöntemden biridir.

Azure ML Studio'da Artırılmış Karar Ağaçları Regresyonu MART<sup>44</sup> Gradyan Artırma (gradient boosting)<sup>45</sup> algoritmasının etkili bir uygulamasını kullanmaktadır. Bu teknik her bir adımdaki hatayı ölçmek ve bir sonraki adımda düzeltmek için önceden tanımlanmış bir kayıp fonksiyonu kullanarak her regresyon ağacını adım adım oluşturmaktadır. Böylece tahmin modeli gerçekte daha zayıf olan tahmin modellerinin toplanmasından oluşmaktadır (Microsoft, 2019).

#### **4.2.1.5 Poisson Regresyon (Poisson Regression)**

Poisson regresyonu tipik olarak sayılabilen (pozitif) sayısal değerlere sahip verileri tahmin etmek için kullanılan özel bir regresyon analizidir. Poisson regresyon modelinde tahmin etmeye çalışılan değerler için yanıt değişkenleri Poisson dağılıma<sup>46</sup> sahip olmalıdır. Bir etkinlik sırasında acil servis çağrı sayısını tahmin etme, bir promosyondan sonraki müşteri sorularının sayısını projelendirme, acil durum tabloları oluşturma, uçak uçuşları ile ilgili soğuk algınlığı sayısını tahmin etme vb. senaryolar için bu yöntem kullanılabilir (Microsoft, 2019).

---

<sup>44</sup> MART : (Multiple Additive Regression Trees) veri madenciliği tahmini için kullanılan yöntemlerden biridir. Daha fazla bilgi için; Friedman, J. H. "Tutorial: Getting Started with MART in R." (Friedman, 2002).

<sup>45</sup> Gradyan Artırma (gradient boosting); Regresyon problemleri için bir makine öğrenmesi tekniğidir.

<sup>46</sup> Poisson Dağılımı: Belirli bir zaman veya mekân aralığında meydana gelen belirli sayıda olayın olasılığını ifade eden ayrık bir olasılık dağılımıdır.

## 4.2.2 Havalimanları Kümeleme Uygulaması Algoritması

### 4.2.2.1 K-Ortalamlar Kümeleme ( K-Means Clustering)

Kümeleme bir veri setindeki vakaları benzer özelliklere sahip kümeler halinde gruplamak, yinelemeli teknikler kullanarak verileri keşfetmek, verilerdeki anormallikleri gözlemlemek ve çıkarımlar yapmak için elverişlidir. K-Ortalamlar algoritması her bir veri noktası için küme içi kareler toplamını en aza indirerek kümelere birine atama yapmaktadır. Küme içi kareler toplamı her bir kümenin merkezini temsil eden küme merkezi (centroid) olarak isimlendirilen bir nokta baz alınarak hesaplanmaktadır. Bir veri setinde küme merkezleri (centroids) sayısı isteğe bağlı olarak belirleneceği gibi optimum küme sayısını bulmaya yönelik çeşitli yöntemler de kullanılabilir. Bu yöntemlere değerlendirme parametreleri başlığı altında yer verilmektedir.

Eğitim verileri işlenirken K-Ortalamlar kümeleme algoritması rastgele seçilmiş bir başlangıç merkezi seti ile başlamaktadır. Bu merkezler kümeler için başlangıç noktası görevi yerine getirerek konumlarını yinelemeli olarak iyileştirirler. İterasyonlar küme merkezleri stabilize olduğunda veya her bir nokta için küme atamaları artık değişmez hale geldiğinde son bularak bir çözüm üzerinde birleşmektedir. Bu sürecin gerçekleştirilmesinde Azure ML Studio Lloyd<sup>47</sup>'un algoritmasından faydalanmaktadır (Microsoft, 2019).

## 4.3 Modelin Değerlendirilmesinde Kullanılan Matematiksel

### Parametreler

Bir modelde gözlenen ve tahmin edilen değerler arasındaki fark ne kadar küçük olursa tahmin edilen değerlerin gözlenen değerlere o kadar iyi uyum sağladığı bilinmektedir. Bu sebeple aşağıdaki başlıklarda oluşturulan modellerin analiz sonuçlarının doğruluğunu değerlendirmek amacıyla bir takım değerlendirme metriklerinden yararlanılmıştır.

---

<sup>47</sup> Lloyd Algoritması: Voronoi iterasyonu veya gevşemesi olarak da bilinen K-Ortalamlar algoritması ile yakından ilişkili olup değerlerin küme merkezine olan mesafesini en aza indirmek için Öklid mesafesi yöntemi ile yinelemeli ölçümler olarak küme merkezini tespit eden bir algoritmadır.

### 4.3.1 Havalimanı Yolcu Sayısı Tahmini Değerlendirme Parametreleri

Havalimanı yolcu sayısı için tahmin edilen değerler 5 farklı metrik kullanılarak değerlendirilmiştir. Bu metriklerin formüllerinde yer alan semboller aşağıda yer almaktadır.

$Y_i = \text{Gerçek Değerler}$

$\bar{Y} = \text{Gerçek Değerlerin Ortalaması}$

$\lambda(X_i) = \text{Tahmin Edilen Değerler}$

$n = \text{Veri Sayısı}$

#### 4.3.1.1 Ortalama Mutlak Hata (Mean Absolute Error (MAE))

Her bir veri noktası ile tahmin doğrusu arasındaki mesafelerin (mutlak değerlerin) ortalamasıdır. Tahmin sonuçlarının gerçek sonuçlara ne kadar yakın olduğunu ölçmeye olanak tanıyan bu metrik hata puanı ne kadar düşük ise o kadar iyi sonuçlar üretildiğinin bir göstergesidir (*Encycl. Mach. Learn.*, 2010; Fürnkranz vd., 2011; Microsoft, 2019).

$$MAE = \frac{\sum_{i=1}^n |Y_i - \lambda(X_i)|}{n}$$

#### 4.3.1.2 Kök Ortalama Kare Hatası (Root Mean Squared Error (RMSE))

Tahmin ve gerçek gözlem arasındaki ortalama kare farklarının karekökünün alınması ile elde edilen hata metriğidir. Ortalama mutlak hata gibi sıfır ve sonsuz arası değerler alan bu metrikte de daha iyi sonuçlar için hata puanının düşük olması beklenmektedir (Fürnkranz vd., 2011; Microsoft, 2019; Willmott ve Matsuura, 2005).

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Y_i - \lambda(X_i))^2}{n}}$$

#### 4.3.1.3 Bağıl Mutlak Hata (Relative Absolute Error (RAE))

Tahmin edilen değerler ile gerçek değerler arasındaki bağıl mutlak farktır. Ortalama fark aritmetik ortalamaya bölüldüğü için bağıl olmaktadır. Hatanın yeterince küçük olup olmadığını anlamayı mümkün kılmaktadır (Botchkarev, 2018b, 2019; Microsoft, 2019; Sammut ve Webb, 2010).

$$RAE = \frac{\sum_{i=1}^n |(Y_i - \lambda(X_i))|}{\sum_{i=1}^n |(Y_i - \bar{Y})|}$$

#### 4.3.1.4 Bağlı Kare Hatası (Relative Squared Error (RSE))

Gerçek değerlerin toplam kare hatasına bölünerek tahmini değerlerin toplam kare hatalarını normalleştiren bir diğer benzer metriktir (Botchkarev, 2018b; Microsoft, 2019; Sammut ve Webb, 2010).

$$RSE = \sqrt{\frac{\sum_{i=1}^n (Y_i - \lambda(X_i))^2}{\sum_{i=1}^n |(Y_i - \bar{Y})|}}$$

#### 4.3.1.5 Belirlilik Katsayısı (Coefficient of Determination (CoD))

Belirlilik katsayısı 0 ile 1 arasında değerler alabilen modelde tahmin edilen değerlerin gücünü temsil eden ve genellikle  $R^2$  olarak isimlendirilen bir katsayıdır. 0 modelin rastgele olduğu anlamına gelirken 1 mükemmel bir uyum olduğu anlamına gelmektedir. Ancak 1'e yakın değerlerin iyi sonuçlar ürettiği yorumu sadece belirlilik katsayısına bakarak değil hata değerlerini de göz önünde bulundurularak yapılmaktadır (Botchkarev, 2018b, 2019; Fürnkranz vd., 2011; Microsoft, 2019; Sammut ve Webb, 2010).

$$CoD = 1 - \frac{\sum_{i=1}^n (Y_i - \lambda(X_i))^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}$$

Daha detaylı bilgi için “Encyclopedia of Machine Learning, Springer 2010, Sammut, Claude Webb, Geoffrey I.” kitabına ve “Botchkarev, A. (2018b). Evaluating Performance of Regression Machine Learning Models Using Multiple Error Metrics in Azure Machine Learning Studio. SSRN Electronic Journal” makalesine bakılabilir.



### 4.3.2 Havalimanları Kümeleme Uygulaması Değerlendirme Parametreleri

Azure ML Studio üzerinde yer alan ‘‘Sweep Clustering<sup>48</sup>’’ parametre taraması (a parameter sweep) kullanarak bir modeli eğitmek için kullanılan bir modüldür. Parametre taraması bir veri seti verildiğinde düzenlenen model için en iyi hiperparametreleri bulmanın bir yoludur. Modül en iyi kümeleme setini bulana kadar farklı parametrelerle modeller oluşturup test eder ve belirlenen parametreler üzerinde iterasyonlar gerçekleştirerek test edilen modelleri açıklayan bir metrik seti oluşturur ve en iyi modeli temel alan küme atamalarını yapar (Microsoft, 2019). Optimum küme sayısı tespitine yönelik ‘‘Sweep Clustering’’ modülü altında Simplified Silhouette, Davies Bouldin, Dunn, Average Deviation olmak üzere 4 parametre bulunmaktadır. Kaggle üzerinde Python ve R paketleri kapsamında bu 4 parametreye ek olarak 27 parametre daha bulunmaktadır. Azure ML üzerinde bulunan 3 endeks (Simplified Silhouette, Davies Bouldin, Dunn) NbClust R paketinde de bulunmaktadır. Bu parametreler Tablo 6’da yer almaktadır.

NbClust paketinde küme merkezlerinin optimizasyonun da ‘‘Euclidean’’, ‘‘Maximum’’, ‘‘Manhattan’’, ‘‘Canberra’’, ‘‘Binary’’, ‘‘Minkowski’’ veya ‘‘Null’’ gibi mesafe hesaplama yöntemleri yer almaktadır. Varsayılan olarak ‘‘Euclidean’’ metod kullanılmaktadır (Charrad vd., 2014).

No	Endeksin Kısa Adı	Endeksin Uzun Adı / Açıklaması	Endeksin Optimum Küme Sayısı	Platform Adı
1	Elbow	Dirsek Metodu	Grafik yöntem	Kaggle / Python
2	Silhouette	Silhouette	Maksimum değer	Azure ML Studio
3	Davies Bouldin	Davies Bouldin	Minimum değer	Azure ML Studio
4	Dunn	Dunn	Maksimum değer	Azure ML Studio
5	Average Deviation	Average Deviation	En iyi başlangıç kaynağını bulmak için kullanılmaktadır.	Azure ML Studio
6	KL	Krzanowski ve Lai 1988	Maksimum değer	Kaggle / NbClust R Paketi

<sup>48</sup> Sweep Clustering: ‘‘Sweep’’ Kümeleme uygulamalarında bir tarama yöntemi olup birçok farklı varyasyona sahiptir. Daha ayrıntılı bilgi için Krista Rizman ve Žalik Borut Žalik’in ‘‘A sweep-line algorithm for spatial clustering’’ adlı 2009’da ‘‘Advances in Engineering Software’’ dergisinde yayınlanan makalesi incelenebilir.

No	Endeksin Kısa Adı	Endeksin Uzun Adı / Açıklaması	Endeksin Optimum Küme Sayısı	Platform Adı
7	CH	Calinski ve Harabasz 1974	Maksimum değer	Kaggle / NbClust R Paketi
8	Hartigan	Hartigan 1975	Hiyerarşi seviyeleri arasındaki maksimum fark	Kaggle / NbClust R Paketi
9	CCC	(Sarle 1983)	Maksimum değer	Kaggle / NbClust R Paketi
10	Scott	(Scott ve Symons 1971)	Hiyerarşi seviyeleri arasındaki maksimum fark	Kaggle / NbClust R Paketi
11	Marriot	(Marriot 1971)	Seviyeler arasındaki ikinci farkların maksimum değeri	Kaggle / NbClust R Paketi
12	TRCOVW	(Milligan ve Cooper 1985)	Hiyerarşi seviyeleri arasındaki maksimum fark	Kaggle / NbClust R Paketi
13	TRACEW	(Milligan ve Cooper 1985)	Seviyeler arasındaki ikinci farkların mutlak maksimum değeri	Kaggle / NbClust R Paketi
14	Friedman	(Friedman ve Rubin 1967)	Hiyerarşi seviyeleri arasındaki maksimum fark	Kaggle / NbClust R Paketi
15	Rubin	(Friedman ve Rubin 1967)	Seviyeler arasındaki ikinci farkların minimum değeri	Kaggle / NbClust R Paketi
16	CINDEX	(Hubert ve Levin 1976)	Minimum değer	Kaggle / NbClust R Paketi
17	DB	(Davies ve Bouldin 1979)	Minimum değer	Kaggle / NbClust R Paketi
18	Silhouette	(Rousseeuw 1987)	Maksimum değer	Kaggle / NbClust R Paketi
19	DUDA	(Duda ve Hart 1973)	Endeksin kritik değerden büyük olduğu en küçük değer	Kaggle / NbClust R Paketi
20	PSEUDOT2	(Duda ve Hart 1973)	Endeksin kritik değerden küçük olduğu en küçük değer	Kaggle / NbClust R Paketi
21	Beale	(Beale 1969)	Kritik değer in alfaya eşit veya büyük olduğu değer	Kaggle / NbClust R Paketi
22	Ratkowsky	(Ratkowsky ve Lance 1978)	Maksimum değer	Kaggle / NbClust R Paketi
23	Ball	(Ball ve Hall 1965)	Hiyerarşi seviyeleri arasındaki maksimum fark	Kaggle / NbClust R Paketi
24	PTBISERIAL	(Milligan 1980, 1981)	Maksimum değer	Kaggle / NbClust R Paketi
25	GAP	(Tibshirani vd. 2001)	Endeksin sıfırdan büyük veya sıfıra eşit olduğu en küçük değer	Kaggle / NbClust R Paketi
26	Frey	(Frey ve Van Groenewoud 1972)	Endeks değerinin 1.00 'den küçük olduğu küme düzeyi değeri	Kaggle / NbClust R Paketi
27	McClain	(McClain ve Rao 1975)	Minimum değer	Kaggle / NbClust R Paketi

No	Endeksin Kısa Adı	Endeksin Uzun Adı / Açıklaması	Endeksin Optimum Küme Sayısı	Platform Adı
28	Gamma	(Baker ve Hubert 1975)	Maksimum değer	Kaggle / NbClust R Paketi
29	GPLUS	(Rohlf 1974) (Milligan 1981)	Minimum değer	Kaggle / NbClust R Paketi
30	TAU	(Rohlf 1974) (Milligan 1981)	Maksimum değer	Kaggle / NbClust R Paketi
31	Dunn	(Dunn 1974)	Maksimum değer	Kaggle / NbClust R Paketi
32	Hubert	(Hubert ve Arabie 1985)	Grafik yöntem	Kaggle / NbClust R Paketi
33	SDINDEX	(Halkidi vd. 2000)	Minimum değer	Kaggle / NbClust R Paketi
34	DINDEX	(Lebart vd. 2000)	Grafik yöntem	Kaggle / NbClust R Paketi
35	SDBW	(Halkidi ve Vazirgiannis 2001)	Minimum değer	Kaggle / NbClust R Paketi

**Tablo 6: NbClust Optimum Küme Sayısı Bulma Yöntemleri / Endeksleri**

## 5 BEŞİNCİ BÖLÜM

### UYGULAMA ÖRNEKLERİ

#### 5.1 Benzer Havalimanlarını Tespit Etmeye Yönelik Kümeleme Uygulama Örneği

##### 5.1.1 Kümeleme için Veri Seti Gösterimi

560 farklı havalimanına ilişkin 23 sütundan oluşan veri setinde ICAO kodu benzersiz tanımlama için kullanılmış olup Kıta Kodu, Ülke Kodu, Ülke Adı ve Şehir demografik bilgiler olarak yer almaktadır. Havalimanı Kategorisi ve Pist Yüzeyi sütunları kategorik veriler içerdiğinden bu bilgiler numerik formatta yeni bir sütun olarak eklenmiştir. Ana veri seti ilk 10 satırı Tablo 6’da gösterilmiştir.

Ana veri setine ilişkin istatistiksel özet Tablo 7’de yer almaktadır. Tabloda her bir sütuna (feature) ait satır sayısı, benzersiz değer sayısı (unique value count), kayıp değer sayısı (missing value count), her bir sütundaki minimum değer (min) ve maksimum değer (max), ortalama değer (mean), ortalama sapma (mean deviation), birinci çeyrek (1st Quartile), medyan (median), üçüncü çeyrek (3rd Quartile), mod (mode), açıklık değeri (range), varyans (variance) ve standart sapma (standart deviation) bilgilerini içeren değerler bulunmaktadır.

Veri seti taşınan yolcu sayıları baz alınarak farklı ülkelerin farklı renklerle gösterildiği harita üzerinde görselleştirilmiştir. Şekil 10’da görselleştirilen harita veri setinde yer alan bazı ülkeler için daire boyutları taşınan yolcu sayısının fazla olduğu havalimanlarını temsil edecek şekilde havalimanlarının coğrafi konumunu işaret etmektedir.

##### 5.1.2 Kümeleme için Model Gösterimi

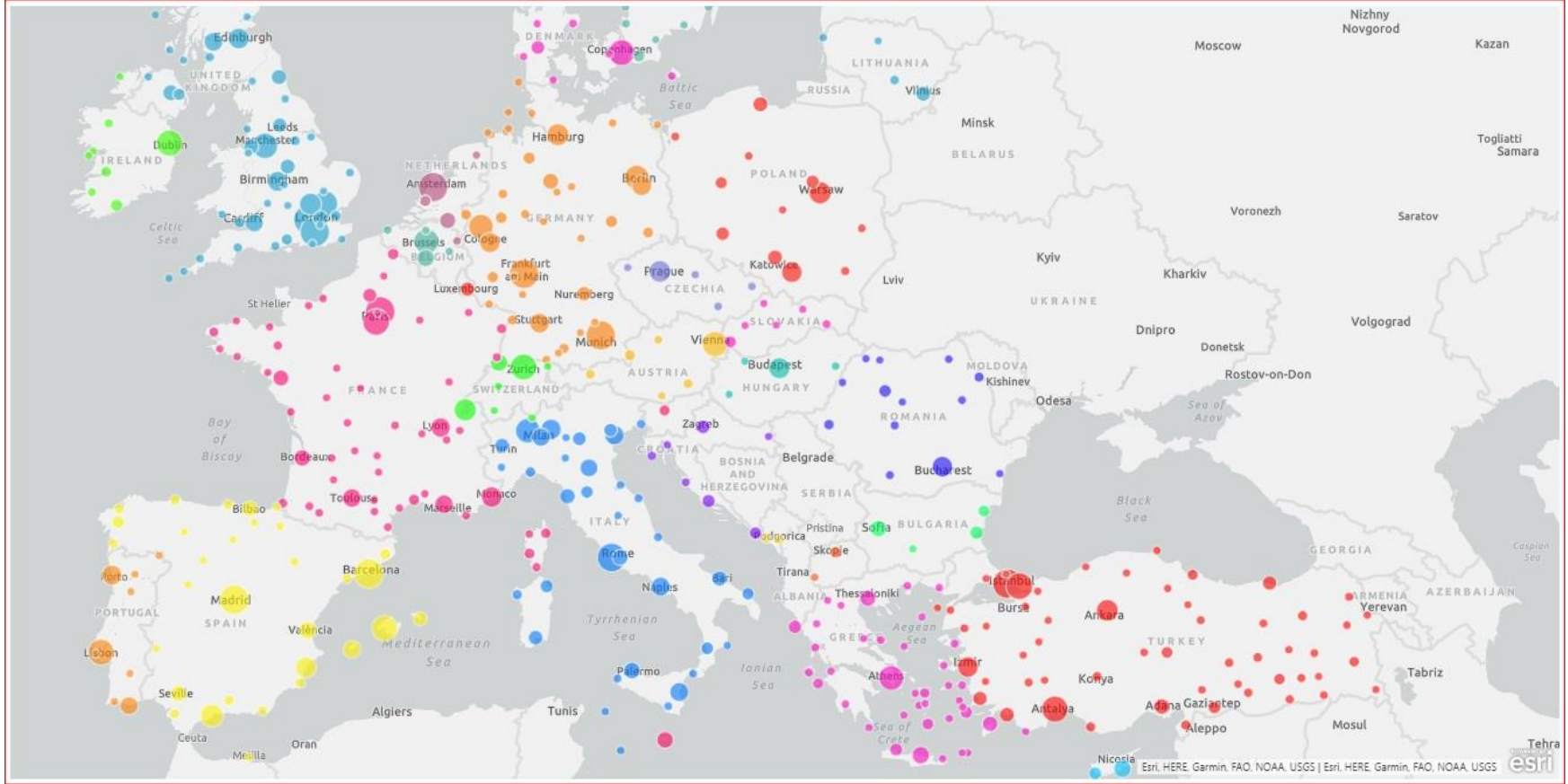
Kümeleme uygulaması için Azure ML Studio programına yüklenen ve analize hazır hale getirilen ham veri seti çalışma alanı üzerinde ilgili sütunların seçilmesi, seçilen sütunların normalizasyonu, son değişikliklerin yapılması, algoritmanın tanımlanması ve sonuçların değerlendirilmesi süreçleri şematik olarak Şekil 11’de yer almaktadır.

ICAO Kodu	Havalimani Adı	Kita Kodu	Ulke Kodu (iso)	Ulke Adı	Sehir	Ulke Nufusu	Sehir Nufusu	Sehirlerin Nufus Oranı	Havalimani Kategorisi	Havalimani Kategorisi (Num)	Rakim Ft	Tarifeli Ucus Bilgisi	Pist Uzunlugu Ft	Pist Genisligi Ft	Pist Yuzeyi	Pist Yuzeyi Num	Rota Sayisi	Yuk Posta 2018	Direk Gecis 2018	Tum Ucak Hareketleri 2018	Yolcu Uçagi Hareketleri 2018	Tasinan Yolcu 2018
LOWG	Graz Airport	EU	AT	Austria	Graz	8955102	222326	40.279149	medium_ai	2	1115	1	9842	148	Asphalt	5	11	226	7917	41416	7941	1026686
LOWI	Innsbruck Airport	EU	AT	Austria	Innsbruck	8955102	132493	67.589246	medium_ai	2	1907	1	6562	148	Asphalt	5	22	73	417	50902	7397	1124389
LOWK	Klagenfurt Airport	EU	AT	Austria	Klagenfurt	8955102	90610	98.831277	medium_ai	2	1472	1	8924	148	Concrete	3	4	0	50	20550	1432	229655
LOWL	Linz Hoersching Airport	EU	AT	Austria	Linz	8955102	204846	43.716265	medium_ai	2	980	1	9843	197	Asphalt	5	5	7702	8128	27414	1703	458868
LOWS	Salzburg Airport	EU	AT	Austria	Salzburg	8955102	153377	58.386212	medium_ai	2	1411	1	9022	148	Concrete	3	50	185	3342	58271	10220	1854326
LOWW	Vienna International Airport	EU	AT	Austria	Vienna	8955102	1691468	5.2942781	large_airpo	3	600	1	11811	148	Asphalt	5	211	229607	85911	256640	177763	27024787
EBAW	Antwerp International Airport (Deur	EU	BE	Belgium	Antwerpen	11539328	1106422	10.429409	medium_ai	2	39	1	4954	148	Asphalt	5	14	2449	7	39467	6070	282126
EBBR	Brussels Airport	EU	BE	Belgium	Brussels	11539328	1019022	11.323924	large_airpo	3	184	1	11936	148	Asphalt	5	190	539082	38855	223577	136323	25622344
EBCI	Brussels South Charleroi Airport	EU	BE	Belgium	Brussels	11539328	1019022	11.323924	large_airpo	3	614	1	8366	148	Asphalt	5	113	438	1154	53948	47911	8022370
EBLG	Liege Airport	EU	BE	Belgium	Liege	11539328	750318	15.37925	large_airpo	3	659	1	12106	148	Asphalt	5	9	869201	2744	31196	1195	169108

**Tablo 7: Kümeleme Ana Veri Seti Örneği (ilk 10 satır)**

Feature	Count	Unique Value Count	Missing Value Count	Min	Max	Mean	Mean Deviation	1st Quartile	Median	3rd Quartile	Mode	Range	Sample Variance	Sample Standard Deviation
ICAO Kodu	560	560	0											
Havalimani Adi	560	560	0											
Kita Kodu	560	5	0											
Ulke Kodu (iso)	560	40	0											
Ulke Adi	560	40	0											
Sehir	560	531	0											
Ulke Nufusu	560	40	0	77741	83517045	38820781.34	29757194.17	8591365	37887768	67530172	{67530172,83429615}	83439304	9.90583E+14	31473520.65
Sehir Nufusu	560	531	0	259	15067724	642099.9143	779841.5981	23815.25	143260	541869	12346478	15067465	3.22445E+12	1795676.586
Sehirlerin Nufus Orani	560	531	0	1	322459.6332	2531.47626	3886.839727	52.12928273	174.9488292	782.9736691	5.469589951	322458.6332	250631618.9	15831.34924
Havalimani Kategorisi	560	3	0											
Havalimani Kategorisi (Num)	560	3	0	1	3	2.175	0.430178571	2	2	3	2	2	0.316368515	0.562466457
Rakim Ft	560	394	0	-15	6400	578.0892857	597.2531888	51.75	235.5	642.5	7	6415	833602.4214	913.0183029
Tarifeli Ucus Bilgisi	560	2	0	0	1	0.916071429	0.153769133	1	1	1	1	1	0.077022106	0.277528568
Pist Uzunlugu Ft	560	366	0	1574	13829	7884.6375	2058.417946	6377	8062	9842	8202	12255	6731884.307	2594.587502
Pist Genisligi Ft	560	41	0	49	262	141.0232143	21.32929847	144.75	148	148	148	213	908.4091202	30.13982615
Pist Yuzeyi	560	7	0											
Pist Yuzeyi Num	560	5	0	1	5	4.585714286	0.649540816	5	5	5	5	4	0.654536162	0.809034092
Rota Sayisi	560	128	0	0	357	31.59821429	36.24885204	2	7	38.25	0	357	2763.06726	52.56488619
Yuk Posta 2018	560	304	0	0	2175747	37815.9875	64743.18795	0	60.5	2079.75	0	2175747	39678025912	199193.4384
Direk Gecis 2018	560	400	0	0	250101	8898.844643	11373.56364	0.75	862.5	8174.5	0	250101	441689963.3	21016.42128
Tum Ucak Hareketleri 2018	560	559	0	26	510958	37887.12321	39959.68619	4009.75	12471.5	38699	2232	510932	4972587165	70516.57369
Yolcu Ucagi Hareketleri 2018	560	482	0	0	460777	20056.96429	27130.19247	650	2355.5	11835.75	0	460777	2653054460	51507.8097
Tasinan Yolcu 2018	560	550	0	0	80016117	3752146.304	5084296.807	75825	373150.5	2366663.75	0	80016117	9.50036E+13	9746979.947

**Tablo 8: Kümeleme Veri Seti için Özet İstatistikler**

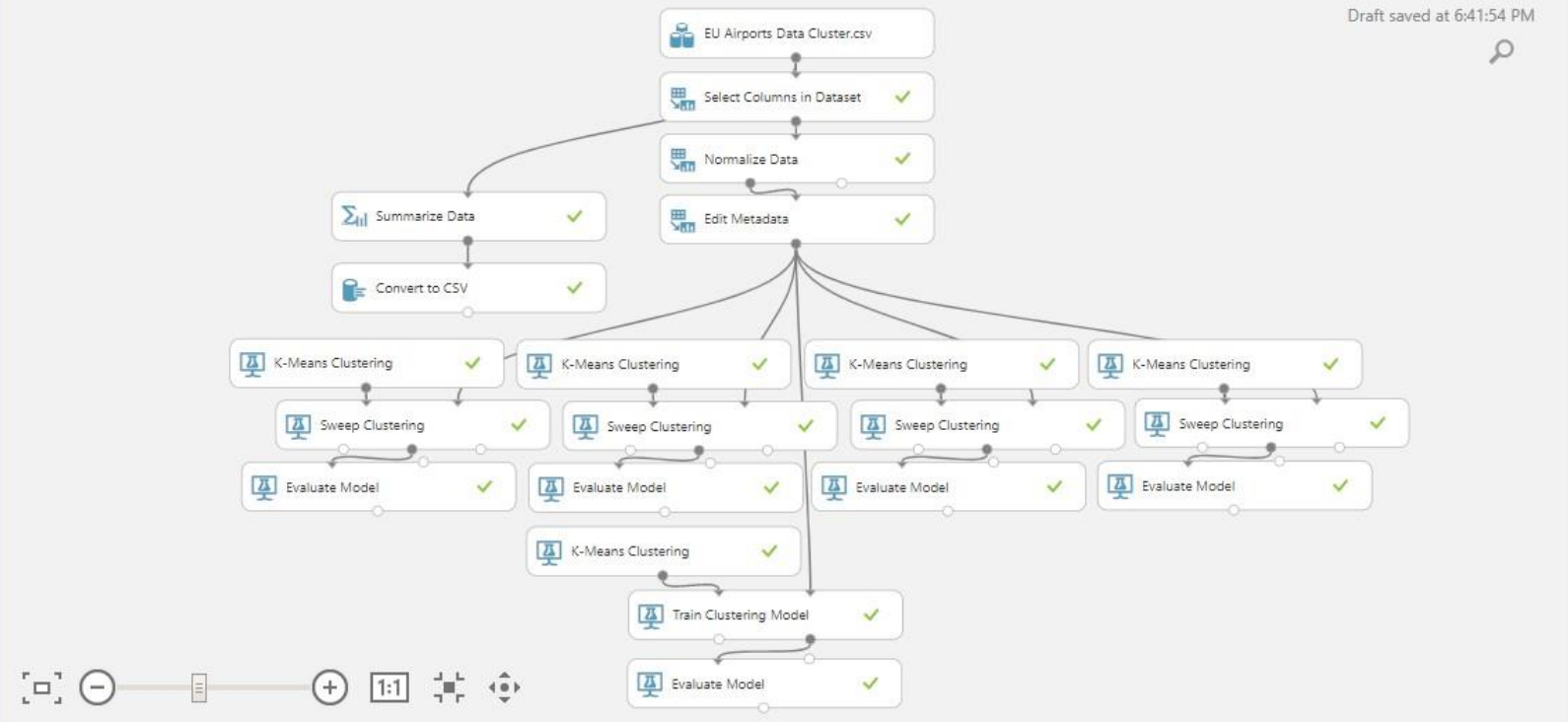


Şekil 10: Veri Setinin Harita Üzerinde Görselleştirilmesi

# Sweep Testing / Clustering

Finished running ✓

Draft saved at 6:41:54 PM



Şekil 11: Kümeleme Modeli



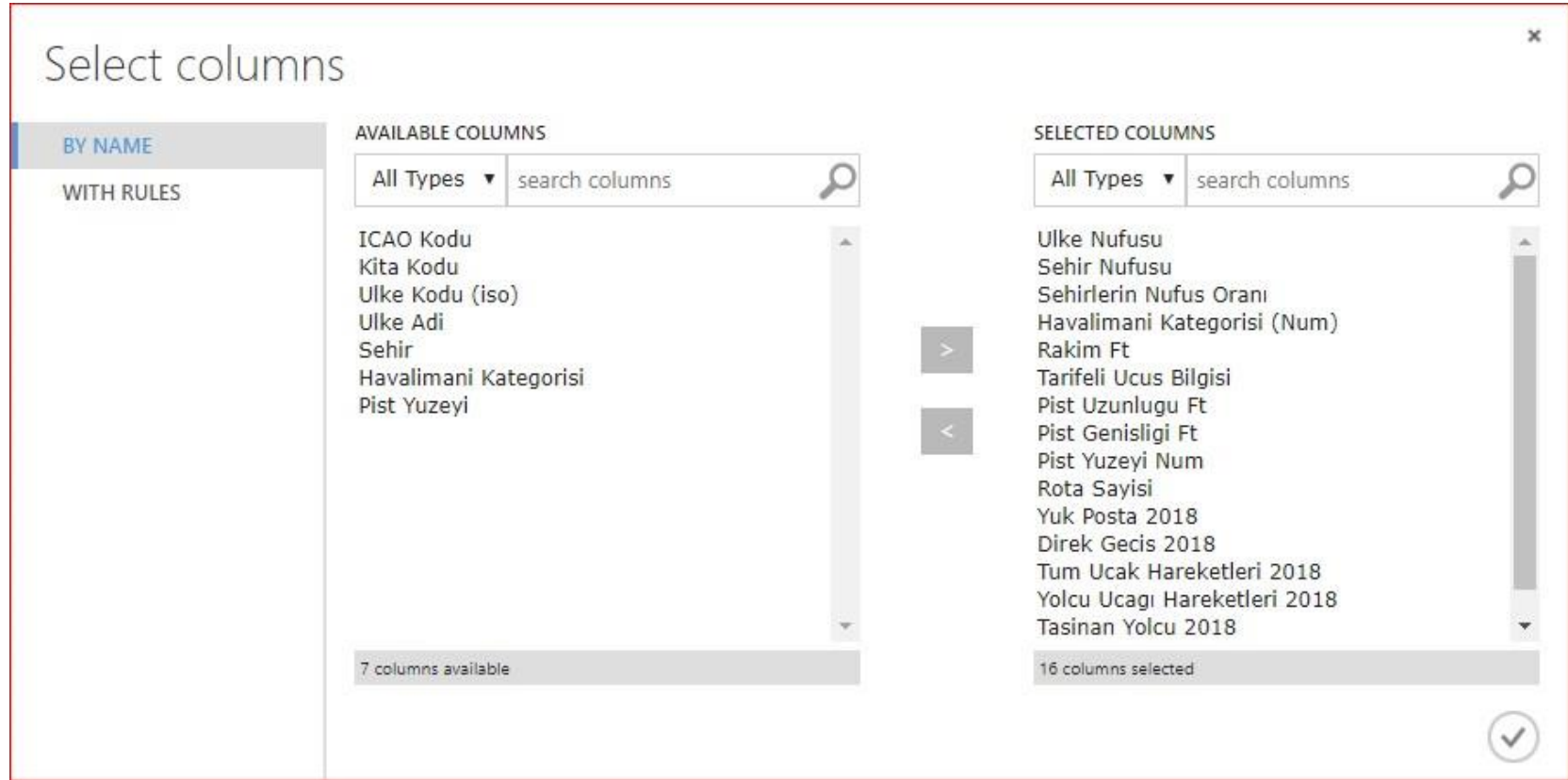
Makine öğrenmesi kapsamında eğitim modülüne hazırlık açısından eğitimde kullanılacak özelliklerin seçimi kritiktir. Veri setinde seçilen özellikler birçok akademik çalışma incelenerek belirlenmiştir. Azure ML Stüdyo çalışma alanında yapılan seçim demografik bilgilerin ayrıştırılması ve sadece numerik değerlerin seçilmesi amacı ile gerçekleştirilmiştir. Seçilen sütunlar Şekil 12’de yer almaktadır. “Havalimanı Adı” sütunu sonuçların değerlendirilmesinde demografik olarak gerekli olduğu için seçilen 16 sütunun içerisinde yer almaktadır. Bir başka önemli aşama olan normalizasyon farklı birim veya kategorilerdeki verileri belirli bir aralıktaki ölçüğe indirgeyerek birbirleri ile anlamlı olarak kıyaslamaya imkân tanımaktadır. Kümeleme uygulaması için “Z score” yöntemi kullanılmıştır. Azure ML üzerinde “Z-score” haricinde “Log-Normal”, “Min-Max”, “Logistic” ve “Tanh” metotları bulunmaktadır. Bazı normalizasyon yöntemleri farklı alanlarda kullanıma uygun olmakla birlikte “Z score” daha yaygın olarak kullanılan bir metot olması sebebi ile seçilmiştir.

Verilerin eğitim modülüne akışından hemen bir önceki adımda “havalimanı adı” sütunu içerik bilgilerinin eğitime etkisini sıfırlamak için “Edit Metadata” fonksiyonunda yer alan “clear feature” işlevinden yararlanılmıştır. K-Ortalamalar algoritması için yer alan parametreler Şekil 13’te yer almakta olup küme merkezlerinin belirlenmesine yönelik ilk olarak “Euclidean” tekniği kullanılmıştır. Algoritma özellik olarak rastgele besleme ile çalışmaktadır ancak sonuçların karşılaştırılabilir olması açısından “12345” rastgele besleme sayısı olarak belirlenmiştir. Optimum küme sayısının belirlenmesi için ise 3 ila 20 arasında bir skala seçilmiştir.

Şekil 14’te ise aynı model üzerinde K-Ortalamalar algoritmasının küme merkezlerini belirlemeye yönelik olarak kullandığı bir diğer teknik olan “Cosine” tekniği ile oluşturulan modelin görüntüsü yer almaktadır. Diğer tüm parametreler kıyaslama yapabilmek amacı ile “Euclidean” tekniğinde seçilen parametreler ile aynı tutulmuştur.

### 5.1.3 Kümeleme için Optimum Küme Sayısı Sonuçları

Kümeleme uygulamasında sonuçlar için öncelikle optimum küme sayısının tespit edilmesi kritik bir öneme sahiptir. Optimum küme sayısını belirlemek üzere Azure ML Studio haricinde



Şekil 12: Kümeleme Modeli Seçilen Sütunlar

## Sweep Testing / Clustering Euclidean

Finished running ✓

**Workflow Steps:**

- EU Airports Data Cluster.csv
- Select Columns in Dataset ✓
- Normalize Data Zscore ✓
- Edit Metadata Havalimani Adi ✓
- Summarize Data ✓
- Convert to CSV ✓
- K-Means Clustering Euclidean ✓
- Sweep Clustering Simplified Silhouette ✓
- Sweep Clustering Davies - Bouldin ✓
- Sweep Clustering Dunn ✓
- Sweep Clustering Average Deviation ✓
- Evaluate Model ✓
- Evaluate Model ✓
- Evaluate Model ✓
- Evaluate Model ✓
- Evaluate Model ✓
- Train Clustering Model Except (Havalimani Adi) 15 Columns ✓
- Evaluate Model ✓

**Properties Panel: K-Means Clustering**

- Create trainer mode: Parameter Range
- Range for Number of Centroids: 3-20
- Initialization for sweep: K-Means++
- Random number seed: 12345
- Number of seeds to sweep: 1
- Metric: Euclidean
- Iterations: 100
- Assign Label Mode: Ignore label column

Şekil 13: Kümeleme Uygulaması “Euclidean” Metodu Parametreleri

### Sweep Testing / Clustering Cosine

Finished running ✓

**Properties Project**

- ▲ K-Means Clustering
  - Create trainer mode: Parameter Range
  - Range for Number of Centroids: 3-20
  - Initialization for sweep: K-Means++
  - Random number seed: 12345
  - Number of seeds to sweep: 1
  - Metric: Cosine
  - Iterations: 100
  - Assign Label Mode: Ignore label column

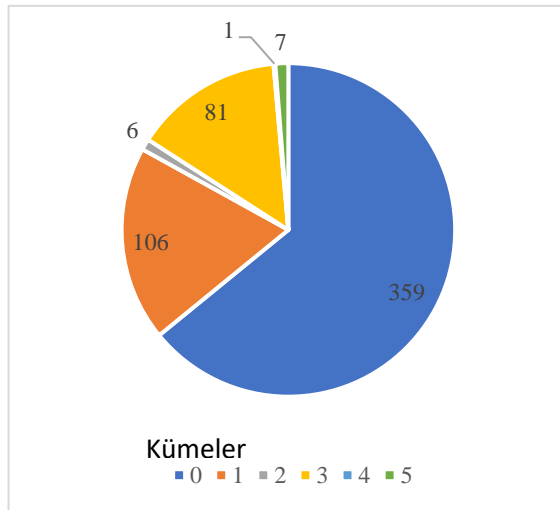
Quick Help

Şekil 14: Kümeleme Uygulaması “Cosine” Metodu Parametreleri

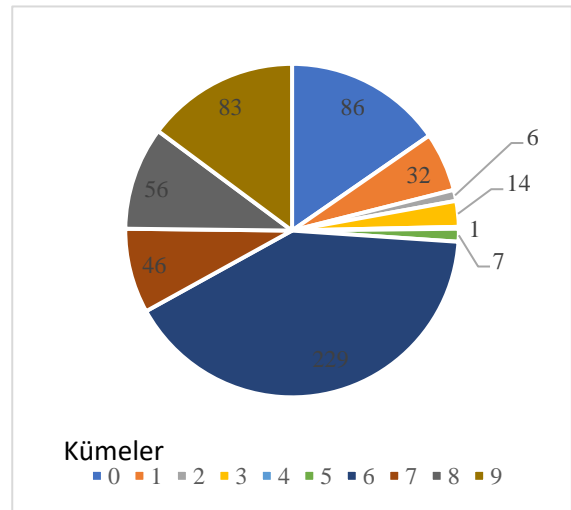
Kaggle üzerinde python ve R programlama dilleri ile de birtakım çalışmalar gerçekleştirilmiştir. Bu çalışmalara ilişkin optimum küme sayısını belirlemeye yönelik sonuçlar aşağıdaki başlıklarda ele alınmaktadır.

### 5.1.3.1 Azure ML Studio “Sweep Clustering” Modülü ile Optimum Küme Sayısı

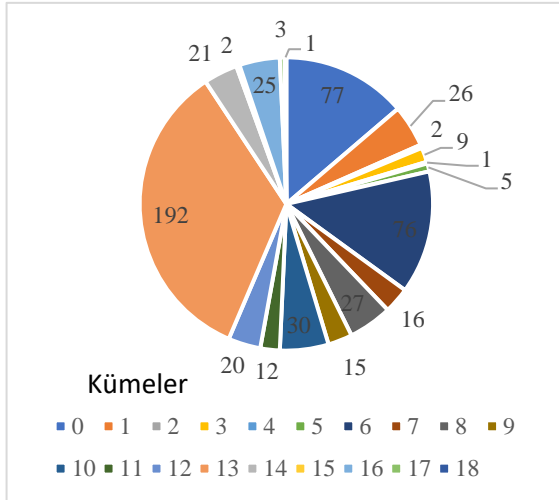
Sweep Clustering Modülü Simplified Silhouette, Davies – Bouldin, Dunn, Average Deviation olmak üzere 4 farklı yöntem ile optimum küme sayısı belirlemede yardımcı bir rol alabilmektedir. Sadece belirtilen 4 farklı parametreden bir tanesini seçerek oluşturulan modül en iyi kümeleme dizisini bulana kadar kendi içerisinde iterasyonlar döndürerek kümeleme modellerini test eder ve test edilen modelleri açıklayan bir dizi metrik oluşturarak en iyi modeli temel alan küme atamalarını gerçekleştirir. Aşağıdaki grafiklerde (Şekil 15-22) “Sweep Clustering” metotları ile belirlenen optimum küme sayılarını ve her bir kümede yer alan havalimanı sayılarını gösteren sonuçlar yer almaktadır.



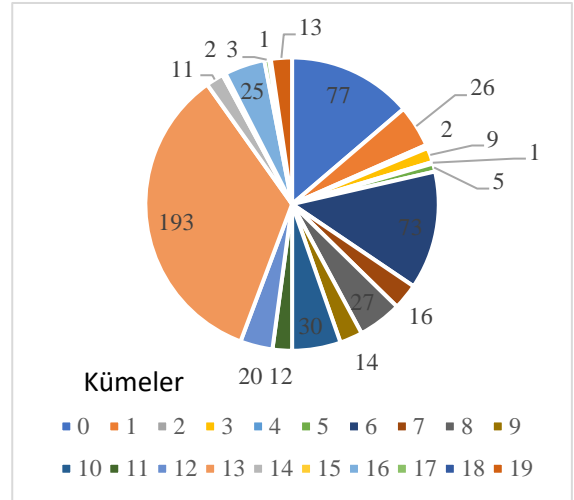
Şekil 15: Euclidean Simplified Silhouette



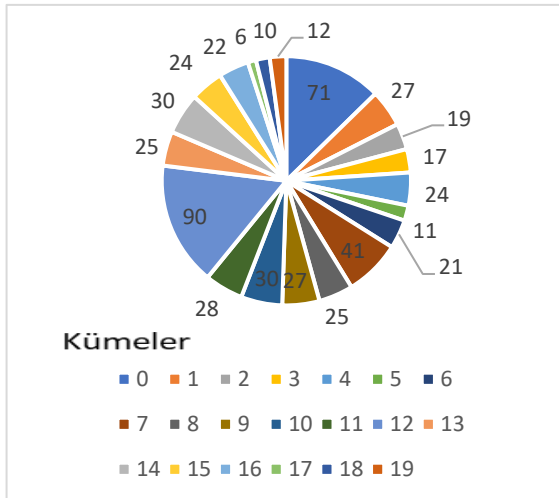
Şekil 16: Euclidean Davies – Bouldin



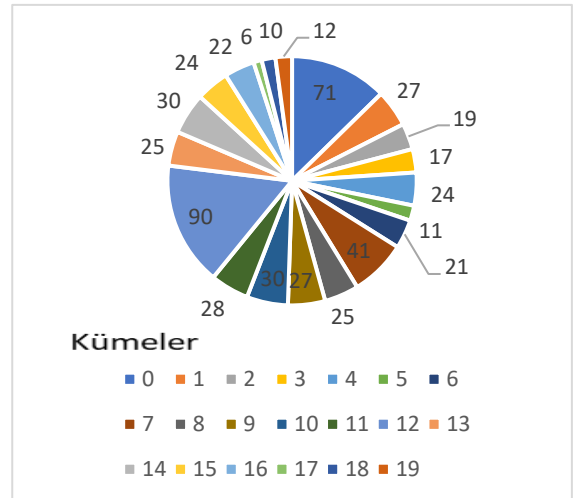
Şekil 17: Euclidean Dunn



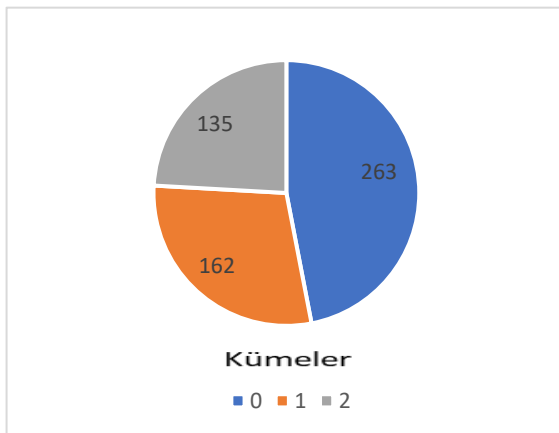
Şekil 18: Euclidean Average Deviation



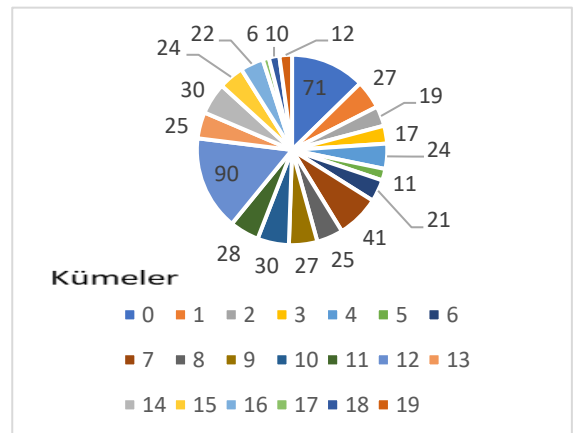
Şekil 19: Cosine Simplified Silhouette



Şekil 22: Cosine Davies – Bouldin



Şekil 20: Cosine Dunn



Şekil 21: Cosine Average Deviation

### 5.1.3.2 Kaggle Platformu üzerinde Elbow Metodu ile Optimum Küme Sayısı

Optimum küme sayısını tespit etmek amacı ile Azure ML Studio üzerinde yer alan Sweep Clustering modülünün ürettiği sonuçlar önemli bir çıkarım yapmaya olanak tanımaktadır fakat bazı kümelere sadece bir eleman atanması ve ortak bir optimum küme sayısı üretilmemiş olması sebebi ile farklı platformlarda benzer yöntemler araştırılmıştır.

Araştırmalar çerçevesinde basit fakat yaygın bir kullanıma sahip olan Elbow (Dirsek) metodu ile optimum küme sayısı değerlendirilmiştir.

Azure ML Studio platformundan normalize edilmiş veriler CSV<sup>49</sup> formatında indirildikten sonra Kaggle platformu üzerinde oluşturulan python programlama dili ile komut dosyasına yüklenmiştir. Komut dizinine ilişkin satırlar ve çalıştırdıktan sonra oluşturulan görseller aşağıda yer almaktadır.

```
#import libraries / Gerekli kütüphanelerin içe aktarılması
import pandas as pd
import numpy as np
import random as rd
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
```

```
#Platforma yüklenmiş olan verinin okunması ve ilk 5 satırının
görüntülenmesi
data = pd.read_csv("../input/eu-airports-data-cluster-n/EU Airports
Data Cluster N1.csv")
data.head()
```

Sonuç çıktısı Şekil 23'te yer almaktadır.

```
#"Havalimanı Adı" sütununun kategorik değişken içerdiği için veri
setinden çıkartılması
del data['Havalimanı Adı']
data.head()
```

Sonuç çıktısı Şekil 24'te yer almaktadır.

```
# statistics of the data / Veri istatistikleri
data.describe()
```

Sonuç çıktısı Şekil 25'te yer almaktadır.

---

<sup>49</sup> CSV (Comma Separated Values): Virgülle ayrılmış şekilde formatlanmış metin dosyalarıdır.

	Havalimani Adi	Ulke Nufusu	Sehir Nufusu	Sehirlerin Nufus Orani	Havalimani Kategorisi (Num)	Rakim Ft	Tarifeli Ucus Bilgisi	Pist Uzunlugu Ft	Pist Genisligi Ft	Pist Yuzeyi Num	Rota Sayisi	Yuk Posta 2018	Direk Gecis 2018	Tum Ucak Hareketleri 2018	Yolcu Ucagi Hareketleri 2018	Tasinan Yolcu 2018
0	Graz Airport	-0.949763	-0.233978	-0.157499	-0.311408	0.588587	0.302685	0.755077	0.231688	0.512532	-0.392213	-0.188880	-0.046760	0.050088	-0.235436	-0.279871
1	Innsbruck Airport	-0.949763	-0.284050	-0.155773	-0.311408	1.456815	0.302685	-0.510224	0.231688	0.512532	-0.182761	-0.189648	-0.403943	0.184730	-0.246007	-0.269838
2	Klagenfurt Airport	-0.949763	-0.307396	-0.153797	-0.311408	0.979947	0.302685	0.400947	0.231688	-1.961762	-0.525501	-0.190015	-0.421421	-0.246079	-0.361918	-0.361716
3	Linz Hoersching Airport	-0.949763	-0.243721	-0.157282	-0.311408	0.440594	0.302685	0.755462	1.858897	0.512532	-0.506460	-0.151315	-0.036711	-0.148653	-0.356652	-0.338179
4	Salzburg Airport	-0.949763	-0.272410	-0.156354	-0.311408	0.913076	0.302685	0.438752	0.231688	-1.961762	0.350391	-0.189086	-0.264641	0.289323	-0.191151	-0.194883

Şekil 23: Kaggle Veri Seti ilk 5 satır

	Ulke Nufusu	Sehir Nufusu	Sehirlerin Nufus Orani	Havalimani Kategorisi (Num)	Rakim Ft	Tarifeli Ucus Bilgisi	Pist Uzunlugu Ft	Pist Genisligi Ft	Pist Yuzeyi Num	Rota Sayisi	Yuk Posta 2018	Direk Gecis 2018	Tum Ucak Hareketleri 2018	Yolcu Ucagi Hareketleri 2018	Tasinan Yolcu 2018
0	-0.949763	-0.233978	-0.157499	-0.311408	0.588587	0.302685	0.755077	0.231688	0.512532	-0.392213	-0.188880	-0.046760	0.050088	-0.235436	-0.279871
1	-0.949763	-0.284050	-0.155773	-0.311408	1.456815	0.302685	-0.510224	0.231688	0.512532	-0.182761	-0.189648	-0.403943	0.184730	-0.246007	-0.269838
2	-0.949763	-0.307396	-0.153797	-0.311408	0.979947	0.302685	0.400947	0.231688	-1.961762	-0.525501	-0.190015	-0.421421	-0.246079	-0.361918	-0.361716
3	-0.949763	-0.243721	-0.157282	-0.311408	0.440594	0.302685	0.755462	1.858897	0.512532	-0.506460	-0.151315	-0.036711	-0.148653	-0.356652	-0.338179
4	-0.949763	-0.272410	-0.156354	-0.311408	0.913076	0.302685	0.438752	0.231688	-1.961762	0.350391	-0.189086	-0.264641	0.289323	-0.191151	-0.194883

Şekil 24: Kaggle Kategorik Değişkenin Çıkarıldığı Veri Seti



	Ulke Nufusu	Sehir Nufusu	Sehirlerin Nufus Orani	Havalimani Kategorisi (Num)	Rakim Ft	Tarifeli Ucus Bilgisi	Pist Uzunlugu Ft	Pist Genisligi Ft	Pist Yuzeyi Num	Rota Sayisi	Yuk Posta 2018	Direk Gecis 2018	Tum Ucak Hareketleri 2018	Yolu Ucagi Hareketleri 2018	Tasinan Yolcu 2018
count	5.600000e+02	5.600000e+02	5.600000e+02	5.600000e+02	5.600000e+02	5.600000e+02	5.600000e+02	5.600000e+02	5.600000e+02	5.600000e+02	5.600000e+02	5.600000e+02	5.600000e+02	5.600000e+02	5.600000e+02
mean	1.078899e-15	1.546382e-17	-1.408595e-16	6.879418e-17	-7.862521e-17	8.201773e-16	-5.570941e-17	-1.387134e-15	1.054712e-15	6.839767e-17	-2.953986e-17	3.171075e-16	3.588399e-17	-3.816392e-17	-2.180795e-17
std	1.000894e+00	1.000894e+00	1.000894e+00	1.000894e+00	1.000894e+00	1.000894e+00	1.000894e+00	1.000894e+00	1.000894e+00	1.000894e+00	1.000894e+00	1.000894e+00	1.000894e+00	1.000894e+00	1.000894e+00
min	-1.232073e+00	-3.577564e-01	-1.599925e-01	-2.090881e+00	-8.501727e-01	-3.303770e+00	-2.434406e+00	-3.055940e+00	-4.436055e+00	-6.016852e-01	-1.900153e-01	-4.238020e-01	-5.373910e-01	-3.897447e-01	-3.852989e-01
25%	-9.613301e-01	-3.446283e-01	-1.567500e-01	-3.114078e-01	-5.769981e-01	3.026845e-01	-5.815897e-01	1.237604e-01	5.125323e-01	-5.835830e-01	-1.900153e-01	-4.237682e-01	-4.808467e-01	-3.771140e-01	-3.775126e-01
50%	-2.967089e-02	-2.780489e-01	-1.489850e-01	-3.114078e-01	-3.755627e-01	3.026845e-01	6.841977e-02	2.316876e-01	5.125323e-01	-4.683774e-01	-1.897113e-01	-3.827259e-01	-3.607428e-01	-3.439729e-01	-3.469810e-01
75%	9.129916e-01	-5.586781e-02	-1.105443e-01	1.468065e+00	7.061009e-02	3.026845e-01	7.550767e-01	2.316876e-01	5.125323e-01	1.266574e-01	-1.795651e-01	-3.449647e-02	1.152357e-02	-1.597537e-01	-1.422719e-01
max	1.421392e+00	8.040714e+00	2.022659e+01	1.468065e+00	8.382255e+00	3.026845e-01	2.293111e+00	4.017440e+00	5.125323e-01	6.198013e+00	1.074253e+01	1.148710e+01	6.714646e+00	8.564023e+00	7.831365e+00

Şekil 25: Kaggle Veri İstatistikleri

```

# Boş bir listede değerlerin muhafaza edilmesi ve K-Ortalamlar
# algoritmasına yerleştirilmesi
SSE = []
for cluster in range(1,20):
    kmeans = KMeans(n_jobs = -1, n_clusters = cluster, init='k-
means++')
    kmeans.fit(data)
    SSE.append(kmeans.inertia_)
# Sonuçların dataframe formatına dönüştürülmesi ve görselleştirilmesi
frame = pd.DataFrame({'Cluster':range(1,20), 'SSE':SSE})
plt.figure(figsize=(12,6))
plt.plot(frame['Cluster'], frame['SSE'], marker='o')
plt.title('The Elbow Method showing the Optimum Number of Cluster')
plt.xlabel('Number of Clusters')
plt.ylabel('Inertia')
plt.show()

```

Sonuç çıktısı Şekil 26’da yer almaktadır.

Elbow metot ile optimum küme sayısını belirleyen yukarıdaki komut dizini haricinde daha kesin bir görselleştirme sunmak üzere oluşturulan kütüphanelerde bulunmaktadır. “yellow brick” kütüphanesi de kendi görselleştirme seçenekleri ile aşağıdaki komut dizinleri kullanılarak Elbow metodunu çalıştırabilen bir kütüphanedir (Bengfort vd., 2020).

```

from sklearn.cluster import KMeans
from yellowbrick.cluster import KElbowVisualizer
model = KElbowVisualizer(KMeans(), k=(1,20))
model.fit(data)
model.show()

```

Sonuç çıktısı Şekil 27’de yer almaktadır.

```

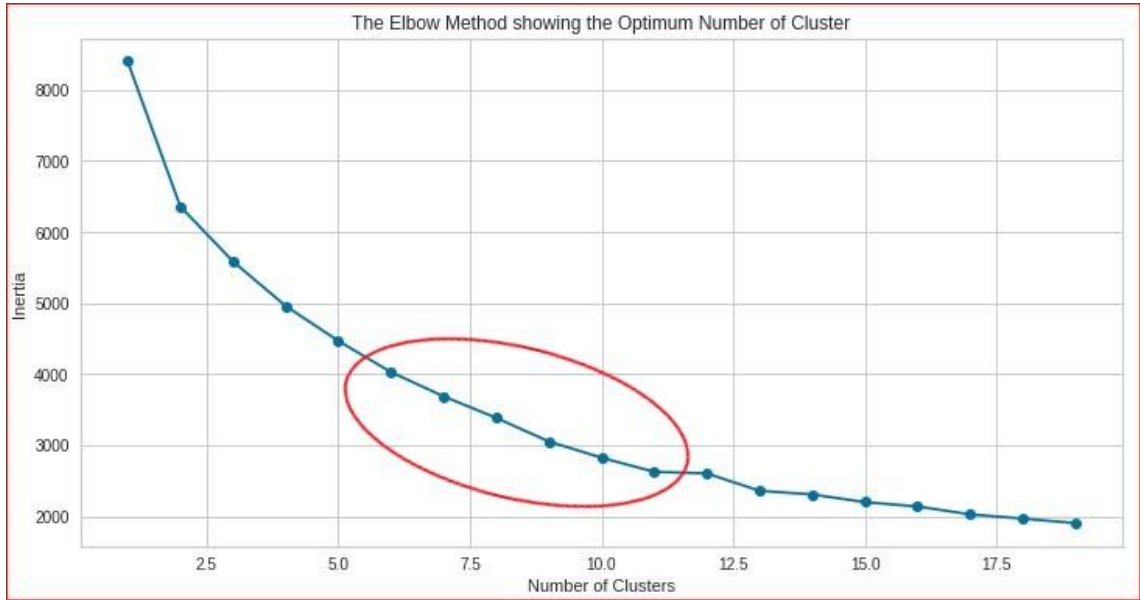
from sklearn.cluster import KMeans
from yellowbrick.cluster import KElbowVisualizer

# Instantiate the clustering model and visualizer
model = KMeans()
visualizer = KElbowVisualizer(model, k=(1,20))

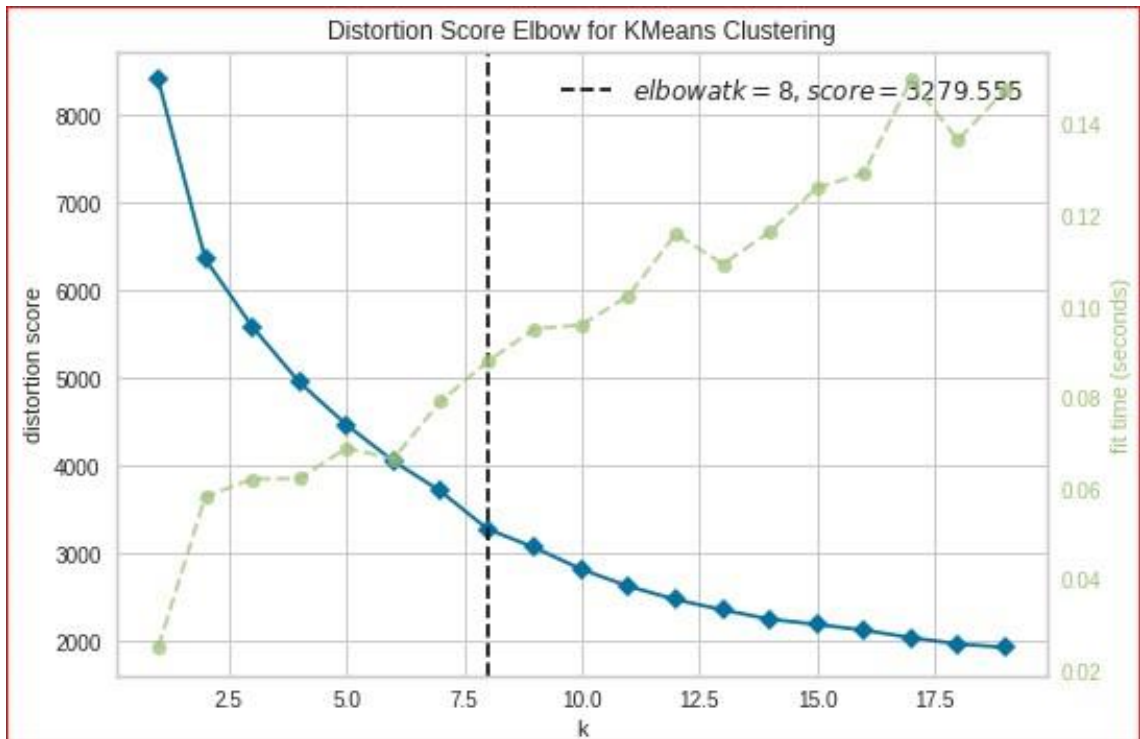
visualizer.fit(data)          # Fit the data to the visualizer
visualizer.show()            # Finalize and render the figure

```

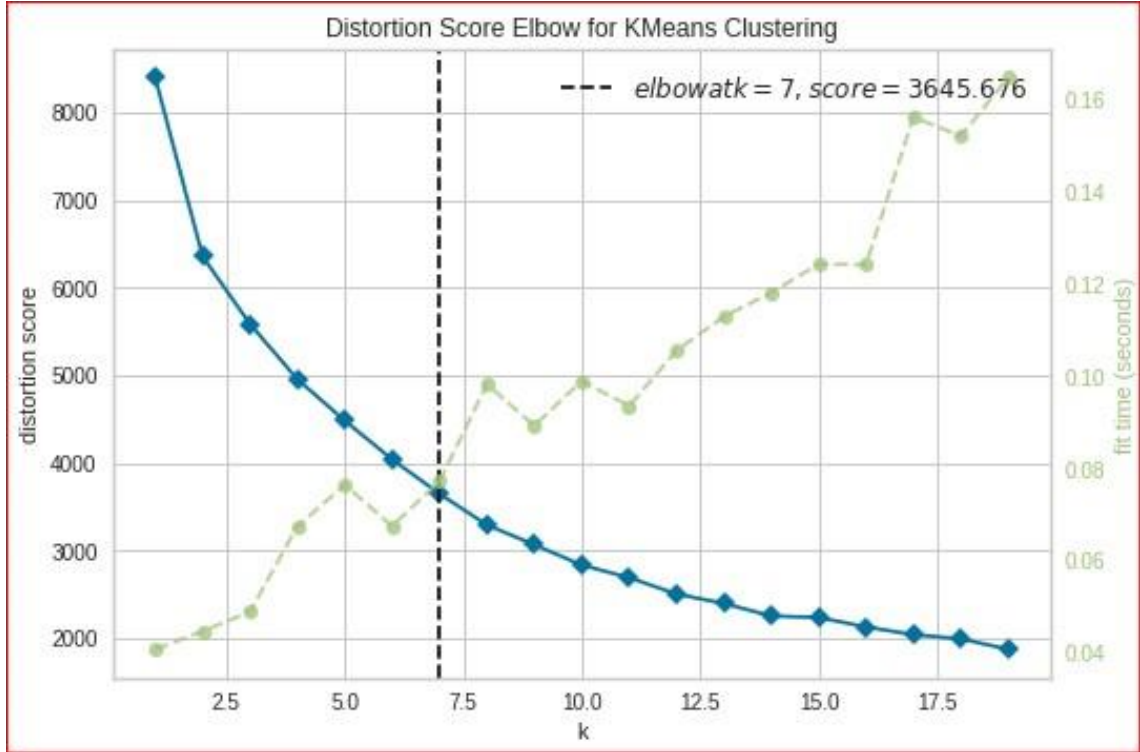
Sonuç çıktısı Şekil 28’de yer almaktadır.



Şekil 26: Kaggle Elbow (Dirsek) Metodu Grafiği 1



Şekil 27: Kaggle Elbow (Dirsek) Metodu Grafiği 2



Şekil 28:Kaggle Elbow (Dirsek) Metodu Grafiği 3

### 5.1.3.3 Kaggle Platformu üzerinde “NbClust” R Programlama Dili Paket İçeriğinde Yer Alan Yöntemler ile Optimum Küme Sayısı

Araştırmalar çerçevesinde optimum küme sayısını tespit etmeye yönelik daha gelişmiş çözümler sunan “NbClust” R programlama dili paketi değerlendirilmiştir. NbClust paketi 30 farklı metot ile eş zamanlı olarak küme sayılarını hesaplamaktadır ve metotların ürettiği sonuçlardan en çok tekrar eden sonuçları sıralayarak optimum küme sayısını tespit etmektedir (Charrad vd., 2014).

Azure ML Studio platformundan normalize edilmiş veriler CSV<sup>50</sup> formatında indirildikten sonra Kaggle platformu üzerinde oluşturulan R programlama dili ile komut dosyasına yüklenmiştir. Komut dizinine ilişkin satırlar ve çalıştırdıktan sonra oluşturulan görseller aşağıda yer almaktadır.

```
#import libraries / Gerekli kütüphanelerin içe aktarılması
```

<sup>50</sup> CSV (Comma Separated Values): Virgülle ayrılmış şekilde formatlanmış metin dosyalarıdır.

```
library(tidyverse) # Birçok yardımcı işlev için kullanılan bir
pakettir.
library(NbClust)
```

```
#Platforma yüklenmiş olan verinin okunması ve ilk 6 satırının
görüntülenmesi
x <- read_csv("../input/eu-airports-data-cluster-n/EU Airports Data
Cluster N1.csv")
head(x)
```

Sonuç çıktısı Şekil 30'da yer almaktadır.

```
# Veri setinden kategorik özelliğe sahip olan "havalimanı adı"
sütununun çıkartılması
data<-x[,-c(1)]
head(data) # Temizlenen veri setinin ilk 6 satırının görüntülenmesi
Sonuç çıktısı Şekil 31'de yer almaktadır.
```

```
# Veri setinin özet istatistiklerinin görüntülenmesi
summary(x)
Sonuç çıktısı Şekil 29'da yer almaktadır.
```

```
Havalimani Adi      Ulke Nufusu      Sehir Nufusu
Length:560         Min. :-1.23207   Min. :-0.35776
Class :character   1st Qu.: -0.96133 1st Qu.: -0.34463
Mode  :character   Median : -0.02967 Median : -0.27805
                Mean  :  0.00000 Mean  :  0.00000
                3rd Qu.:  0.91299 3rd Qu.: -0.05587
                Max.  :  1.42139 Max.  :  8.04071

Sehirlerin Nufus Oranı Havalimani Kategorisi (Num)   Rakim Ft
Min. :-0.1600         Min. :-2.0909         Min. :-0.65017
1st Qu.: -0.1568      1st Qu.: -0.3114      1st Qu.: -0.57700
Median : -0.1490      Median : -0.3114      Median : -0.37556
Mean  :  0.0000         Mean  :  0.0000         Mean  :  0.00000
3rd Qu.: -0.1105      3rd Qu.:  1.4681      3rd Qu.:  0.07061
Max.  : 20.2266        Max.  :  1.4681        Max.  :  6.38225

Tarifeli Ucus Bilgisi Pist Uzunlugu Ft   Pist Genisligi Ft   Pist Yuzeyi Num
Min. :-3.3038         Min. :-2.43441       Min. :-3.0559       Min. :-4.4361
1st Qu.:  0.3027      1st Qu.: -0.58159     1st Qu.:  0.1238     1st Qu.:  0.5125
Median :  0.3027      Median :  0.06842     Median :  0.2317     Median :  0.5125
Mean  :  0.0000         Mean  :  0.00000      Mean  :  0.0000      Mean  :  0.0000
3rd Qu.:  0.3027      3rd Qu.:  0.75508     3rd Qu.:  0.2317     3rd Qu.:  0.5125
Max.  :  0.3027        Max.  :  2.29311       Max.  :  4.0174       Max.  :  0.5125

Rota Sayisi        Yuk Posta 2018      Direk Gecis 2018
Min. :-0.6017       Min. :-0.1900       Min. :-0.4238
1st Qu.: -0.5636    1st Qu.: -0.1900    1st Qu.: -0.4238
Median : -0.4684    Median : -0.1897    Median : -0.3827
Mean  :  0.0000     Mean  :  0.0000     Mean  :  0.0000
3rd Qu.:  0.1267    3rd Qu.: -0.1796    3rd Qu.: -0.0345
Max.  :  6.1960     Max.  : 10.7425     Max.  : 11.4871

Tum Ucak Hareketleri 2018 Yolcu Ucagi Hareketleri 2018 Tasinan Yolcu 2018
Min. :-0.53739      Min. :-0.3897       Min. :-0.3853
1st Qu.: -0.48085   1st Qu.: -0.3771    1st Qu.: -0.3775
Median : -0.36074   Median : -0.3440    Median : -0.3470
Mean  :  0.00000    Mean  :  0.0000     Mean  :  0.0000
3rd Qu.:  0.01152   3rd Qu.: -0.1598    3rd Qu.: -0.1423
Max.  :  6.71465    Max.  :  8.5640     Max.  :  7.8314
```

Şekil 29: R Programlama ile Özet İstatistikler

Havalimani Adi	Ulke Nufusu	Sehir Nufusu	Sehirlerin Nufus Orani	Havalimani Kategorisi (Num)	Rakim Ft	Tarifeli Ucus Bilgisi	Pist Uzunlugu Ft	Pist Genisligi Ft	Pist Yuzeyi Num	Rota Sayisi	Yuk Posta 2018	Direk Gecis 2018	Tum Ucak Hareketleri 2018	Yolcu Ucagi Hareketleri 2018	Tasinan Yolcu 2018
<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
Graz Airport	-0.9497629	-0.2339782	-0.1574992	-0.3114078	0.58858704	0.3026845	0.7550767	0.2316876	0.5125323	-0.3922130	-0.1888797	-0.04675974	0.05008797	-0.2354361	-0.2798710
Innsbruck Airport	-0.9497629	-0.2840503	-0.1557726	-0.3114078	1.45681508	0.3026845	-0.5102237	0.2316876	0.5125323	-0.1827607	-0.1896485	-0.40394260	0.18472980	-0.2460070	-0.2698381
Klagenfurt Airport	-0.9497629	-0.3073955	-0.1537974	-0.3114078	0.97994741	0.3026845	0.4009469	0.2316876	-1.9617616	-0.5255008	-0.1900153	-0.42142075	-0.24607865	-0.3619182	-0.3617162
Linz Hoersching Airport	-0.9497629	-0.2437214	-0.1572819	-0.3114078	0.44059363	0.3026845	0.7554624	1.8588970	0.5125323	-0.5064597	-0.1513148	-0.03671100	-0.14865281	-0.3566522	-0.3381789
Salzburg Airport	-0.9497629	-0.2724098	-0.1563544	-0.3114078	0.91307631	0.3026845	0.4387516	0.2316876	-1.9617616	0.3503905	-0.1890857	-0.26464129	0.28932349	-0.1911508	-0.1948826
Vienna International Airport	-0.9497629	0.5849084	-0.1597110	1.4680655	0.02401957	0.3026845	1.5146426	0.2316876	0.5125323	3.4160101	0.9636988	3.66765623	3.10492190	3.0645262	2.3898118

Şekil 30: R Programlama ile Okunan Veri Seti

Ulke Nufusu	Sehir Nufusu	Sehirlerin Nufus Orani	Havalimani Kategorisi (Num)	Rakim Ft	Tarifeli Ucus Bilgisi	Pist Uzunlugu Ft	Pist Genisligi Ft	Pist Yuzeyi Num	Rota Sayisi	Yuk Posta 2018	Direk Gecis 2018	Tum Ucak Hareketleri 2018	Yolcu Ucagi Hareketleri 2018	Tasinan Yolcu 2018
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
-0.9497629	-0.2339782	-0.1574992	-0.3114078	0.58858704	0.3026845	0.7550767	0.2316876	0.5125323	-0.3922130	-0.1888797	-0.04675974	0.05008797	-0.2354361	-0.2798710
-0.9497629	-0.2840503	-0.1557726	-0.3114078	1.45681508	0.3026845	-0.5102237	0.2316876	0.5125323	-0.1827607	-0.1896485	-0.40394260	0.18472980	-0.2460070	-0.2698381
-0.9497629	-0.3073955	-0.1537974	-0.3114078	0.97994741	0.3026845	0.4009469	0.2316876	-1.9617616	-0.5255008	-0.1900153	-0.42142075	-0.24607865	-0.3619182	-0.3617162
-0.9497629	-0.2437214	-0.1572819	-0.3114078	0.44059363	0.3026845	0.7554624	1.8588970	0.5125323	-0.5064597	-0.1513148	-0.03671100	-0.14865281	-0.3566522	-0.3381789
-0.9497629	-0.2724098	-0.1563544	-0.3114078	0.91307631	0.3026845	0.4387516	0.2316876	-1.9617616	0.3503905	-0.1890857	-0.26464129	0.28932349	-0.1911508	-0.1948826
-0.9497629	0.5849084	-0.1597110	1.4680655	0.02401957	0.3026845	1.5146426	0.2316876	0.5125323	3.4160101	0.9636988	3.66765623	3.10492190	3.0645262	2.3898118

Şekil 31: R Programlama ile Temizlenen Veri Seti



```
# NbClust Paket içeriğinde yer alan parametreler ile optimum küme
sayısı tespiti 1. yol
res<-NbClust(data, diss=NULL, distance = "euclidean", min.nc=2,
max.nc=20, method = "ward.D2", index = "all")
res$All.index
res$Best.nc
```

Yukarıda yer alan kod NbClust içerisinde yer alan 30 farklı endekste (metot) öklidyen mesafe hesaplaması baz alınarak yine paket içeriğinde yer alan “ward.D2” metodu üzerine inşa edilmiştir.

Şekil 32 “ward.D2” metodu ile tüm endeksler için tespit edilen en iyi küme sayılarından en çok tekrar eden sonuçları sıralamaktadır. Tüm endekslerden 8 tanesi en iyi küme sayısını 3 olarak belirlediği için çoğunluk kuralı gereği optimum küme sayısı 3 olmuştur.

```
# NbClust Paket içeriğinde yer alan parametreler ile optimum küme
sayısı tespiti 2. yol
res<-NbClust(data, diss=NULL, distance = "euclidean", min.nc=2,
max.nc=20, method = "single", index = "all")
res$All.index
res$Best.nc
```

Yukarıda yer alan kod NbClust içerisinde yer alan 30 farklı endekste (metot) öklidyen mesafe hesaplaması baz alınarak yine paket içeriğinde yer alan “single” metodu üzerine inşa edilmiştir.

Şekil 33 “single” metodu ile tüm endeksler için tespit edilen en iyi küme sayılarından en çok tekrar eden sonuçları sıralamaktadır. Tüm endekslerden 14 tanesi en iyi küme sayısını 3 olarak belirlediği için çoğunluk kuralı gereği optimum küme sayısı 3 olmuştur.

```
# NbClust Paket içeriğinde yer alan parametreler ile optimum küme
sayısı tespiti 3. yol
res<-NbClust(data, diss=NULL, distance = "euclidean", min.nc=2,
max.nc=20, method = "complete", index = "all")
res$All.index
res$Best.nc
```

Yukarıda yer alan kod NbClust içerisinde yer alan 30 farklı endekste (metot) öklidyen mesafe hesaplaması baz alınarak yine paket içeriğinde yer alan “complete” metodu üzerine inşa edilmiştir.

Şekil 34 “complete” metodu ile tüm endeksler için tespit edilen en iyi küme sayılarından en çok tekrar eden sonuçları sıralamaktadır. Tüm endekslerden 8 tanesi en iyi küme sayısını 3 olarak belirlediği için çoğunluk kuralı gereği optimum küme sayısı 3 olmuştur.

```
*****
* Among all indices:
* 2 proposed 2 as the best number of clusters
* 8 proposed 3 as the best number of clusters
* 3 proposed 4 as the best number of clusters
* 1 proposed 5 as the best number of clusters
* 4 proposed 6 as the best number of clusters
* 1 proposed 7 as the best number of clusters
* 4 proposed 20 as the best number of clusters

***** Conclusion *****

* According to the majority rule, the best number of clusters is 3
```

Şekil 32: NbClust (Euclidean, ward.D2) Optimum Küme Sayısı (3)

```
*****
* Among all indices:
* 6 proposed 2 as the best number of clusters
* 14 proposed 3 as the best number of clusters
* 1 proposed 11 as the best number of clusters
* 2 proposed 19 as the best number of clusters
* 1 proposed 20 as the best number of clusters

***** Conclusion *****

* According to the majority rule, the best number of clusters is 3
```

Şekil 33: NbClust (Euclidean, Single) Optimum Küme Sayısı (3)



```

*****
* Among all indices:
* 7 proposed 2 as the best number of clusters
* 8 proposed 3 as the best number of clusters
* 2 proposed 4 as the best number of clusters
* 2 proposed 5 as the best number of clusters
* 3 proposed 8 as the best number of clusters
* 1 proposed 12 as the best number of clusters
* 1 proposed 20 as the best number of clusters

***** Conclusion *****

* According to the majority rule, the best number of clusters is 3

```

Şekil 34: NbClust (Euclidean, Complete) Optimum Küme Sayısı (3)

```

# NbClust Paket içeriğinde yer alan parametreler ile optimum küme
sayısı tespiti 4. yol
res<-NbClust(data, diss=NULL, distance = "euclidean", min.nc=2,
max.nc=20, method = "average", index = "all")
res$All.index
res$Best.nc

```

Yukarıda yer alan kod NbClust içerisinde yer alan 30 farklı endekste (metot) öklidyen mesafe hesaplaması baz alınarak paket içeriğinde yer alan “average” metodu üzerine inşa edilmiştir.

Şekil 35 “average” metodu ile tüm endeksler için tespit edilen en iyi küme sayılarından en çok tekrar eden sonuçları sıralamaktadır. Tüm endekslerden 11 tanesi en iyi küme sayısını 3 olarak belirlediği için çoğunluk kuralı gereği optimum küme sayısı 3 olmuştur.

```

# NbClust Paket içeriğinde yer alan parametreler ile optimum küme
sayısı tespiti 5. yol
res<-NbClust(data, diss=NULL, distance = "euclidean", min.nc=2,
max.nc=20, method = "mcquitty", index = "all")
res$All.index
res$Best.nc

```

Yukarıda yer alan kod NbClust içerisinde yer alan 30 farklı endekste (metot) öklidyen mesafe hesaplaması baz alınarak paket içeriğinde yer alan “mcquitty” metodu üzerine inşa edilmiştir.

Şekil 36 “mcquitty” metodu ile tüm endeksler için tespit edilen en iyi küme sayılarından en çok tekrar eden sonuçları sıralamaktadır. Tüm endekslerden 8 tanesi en iyi küme sayısını 3 olarak belirlediği için çoğunluk kuralı gereği optimum küme sayısı 3 olmuştur.

```
# NbClust Paket içeriğinde yer alan parametreler ile optimum küme
sayısı tespiti 6. yol
res<-NbClust(data, diss=NULL, distance = "euclidean", min.nc=2,
max.nc=20, method = "median", index = "all")
res$All.index
res$Best.nc
```

Yukarıda yer alan kod NbClust içerisinde yer alan 30 farklı endekste (metot) öklidyen mesafe hesaplaması baz alınarak yine paket içeriğinde yer alan “median” metodu üzerine inşa edilmiştir.

Şekil 37 “median” metodu ile tüm endeksler için tespit edilen en iyi küme sayılarından en çok tekrar eden sonuçları sıralamaktadır. Tüm endekslerden 9 tanesi en iyi küme sayısını 2 olarak belirlediği için çoğunluk kuralı gereği optimum küme sayısı 2 olmuştur.

```
*****
* Among all indices:
* 6 proposed 2 as the best number of clusters
* 11 proposed 3 as the best number of clusters
* 1 proposed 7 as the best number of clusters
* 1 proposed 15 as the best number of clusters
* 3 proposed 17 as the best number of clusters
* 1 proposed 20 as the best number of clusters

***** Conclusion *****
* According to the majority rule, the best number of clusters is 3
```

Şekil 35: NbClust (Euclidean, Average) Optimum Küme Sayısı (3)

```

*****
* Among all indices:
* 6 proposed 2 as the best number of clusters
* 8 proposed 3 as the best number of clusters
* 3 proposed 6 as the best number of clusters
* 1 proposed 7 as the best number of clusters
* 1 proposed 13 as the best number of clusters
* 2 proposed 14 as the best number of clusters
* 2 proposed 19 as the best number of clusters
* 1 proposed 20 as the best number of clusters

***** Conclusion *****

* According to the majority rule, the best number of clusters is 3

```

Şekil 36: NbClust (Euclidean, Mcquitty) Optimum Küme Sayısı (3)

```

*****
* Among all indices:
* 9 proposed 2 as the best number of clusters
* 3 proposed 3 as the best number of clusters
* 9 proposed 4 as the best number of clusters
* 2 proposed 12 as the best number of clusters
* 1 proposed 20 as the best number of clusters

***** Conclusion *****

* According to the majority rule, the best number of clusters is 2

```

Şekil 37: NbClust (Euclidean, Median) Optimum Küme Sayısı (2)

```

# NbClust Paket içeriğinde yer alan parametreler ile optimum küme
sayısı tespiti 7. yol
res<-NbClust(data, diss=NULL, distance = "euclidean", min.nc=2,
max.nc=20, method = "centroid", index = "all")
res$All.index
res$Best.nc

```

Yukarıda yer alan kod NbClust içerisinde yer alan 30 farklı endekste (metot) öklidyen mesafe hesaplaması baz alınarak paket içeriğinde yer alan “centroid” metodu üzerine inşa edilmiştir.

Şekil 38 “centroid” metodu ile tüm endeksler için tespit edilen en iyi küme sayılarından en çok tekrar eden sonuçları sıralamaktadır. Tüm endekslerden 15 tanesi en iyi küme sayısını 3 olarak belirlediği için çoğunluk kuralı gereği optimum küme sayısı 3 olmuştur.

```
# NbClust Paket içeriğinde yer alan parametreler ile optimum küme
sayısı tespiti 8. yol
res<-NbClust(data, diss=NULL, distance = "euclidean", min.nc=2,
max.nc=20, method = "kmeans", index = "all")
res$All.index
res$Best.nc
```

Yukarıda yer alan kod NbClust içerisinde yer alan 30 farklı endekste (metot) öklidyen mesafe hesaplaması baz alınarak yine paket içeriğinde yer alan “kmeans” metodu üzerine inşa edilmiştir.

Şekil 39 “kmeans” metodu ile tüm endeksler için tespit edilen en iyi küme sayılarından en çok tekrar eden sonuçları sıralamaktadır. Tüm endekslerden 8 tanesi en iyi küme sayısını 2 olarak belirlediği için çoğunluk kuralı gereği optimum küme sayısı 2 olmuştur.

```
# NbClust Paket içeriğinde yer alan parametreler ile optimum küme
sayısı tespiti 9. yol
res<-NbClust(data, diss=NULL, distance = "manhattan", min.nc=2,
max.nc=20, method = "average", index = "all")
res$All.index
res$Best.nc
```

Yukarıda yer alan kod NbClust içerisinde yer alan 30 farklı endekste (metot) “manhattan” mesafe hesaplaması baz alınarak yine paket içeriğinde yer alan “average” metodu üzerine inşa edilmiştir.

Şekil 40 “average” metodu ile tüm endeksler için tespit edilen en iyi küme sayılarından en çok tekrar eden sonuçları sıralamaktadır. Tüm endekslerden 9 tanesi en iyi küme sayısını 4 olarak belirlediği için çoğunluk kuralı gereği optimum küme sayısı 4 olmuştur.

```

*****
* Among all indices:
* 6 proposed 2 as the best number of clusters
* 15 proposed 3 as the best number of clusters
* 1 proposed 10 as the best number of clusters
* 1 proposed 16 as the best number of clusters
* 1 proposed 20 as the best number of clusters

***** Conclusion *****

* According to the majority rule, the best number of clusters is 3

```

Şekil 38: NbClust (Euclidean, Centroid) Optimum Küme Sayısı (3)

```

*****
* Among all indices:
* 8 proposed 2 as the best number of clusters
* 2 proposed 3 as the best number of clusters
* 3 proposed 4 as the best number of clusters
* 1 proposed 10 as the best number of clusters
* 1 proposed 11 as the best number of clusters
* 4 proposed 12 as the best number of clusters
* 2 proposed 15 as the best number of clusters
* 1 proposed 16 as the best number of clusters
* 2 proposed 20 as the best number of clusters

***** Conclusion *****

* According to the majority rule, the best number of clusters is 2

```

Şekil 39: NbClust (Euclidean, Kmeans) Optimum Küme Sayısı (2)

```

*****
* Among all indices:
* 8 proposed 2 as the best number of clusters
* 3 proposed 3 as the best number of clusters
* 9 proposed 4 as the best number of clusters
* 1 proposed 6 as the best number of clusters
* 1 proposed 11 as the best number of clusters
* 1 proposed 20 as the best number of clusters

***** Conclusion *****

* According to the majority rule, the best number of clusters is 4

```

Şekil 40: NbClust (Manhattan, Average) Optimum Küme Sayısı (4)

```
# NbClust parametreleri ile optimum küme sayısı tespiti 10. yol
res<-NbClust(data, diss=NULL, distance = "minkowski", min.nc=2,
max.nc=20, method = "single", index = "all")
res$All.index
res$Best.nc
```

Yukarıda yer alan kod NbClust içerisinde yer alan 30 farklı endekste “minkowski” mesafe hesaplaması baz alınarak paket içeriğinde yer alan “single” metodu üzerine inşa edilmiştir.

Şekil 41 “single” metodu ile tüm endeksler için tespit edilen en iyi küme sayılarından en çok tekrar eden sonuçları sıralamaktadır. Tüm endekslerden 14 tanesi en iyi küme sayısını 3 olarak belirlediği için çoğunluk kuralı gereği optimum küme sayısı 3 olmuştur.

```
# NbClust parametreleri ile optimum küme sayısı tespiti 11. yol
res<-NbClust(data, diss=NULL, distance = "minkowski", min.nc=2,
max.nc=20, method = "centroid", index = "all")
res$All.index
res$Best.nc
```

Yukarıda yer alan kod NbClust içerisinde yer alan 30 farklı endekste “minkowski” mesafe hesaplaması baz alınarak paket içeriğinde yer alan “centroid” metodu üzerine inşa edilmiştir.

Şekil 42 “centroid” metodu ile tüm endeksler için tespit edilen en iyi küme sayılarından en çok tekrar eden sonuçları sıralamaktadır. Tüm endekslerden 15 tanesi en iyi küme sayısını 3 olarak belirlediği için çoğunluk kuralı gereği optimum küme sayısı 3 olmuştur.

Hubert endeksi elbow metodu gibi küme sayısını grafiksel olarak belirleyen bir metottur. Hubert endeksi grafiğinde ölçülen değerlerin önemli bir artışa tekabül ettiği değişim noktası araştırılmaktadır. Hubert Endeks değerleri ve ikincil farklarına ilişkin grafikler Şekil 43’te yer almaktadır.

NbClust’ta küme sayısını grafiksel olarak belirleyen bir diğer endeks “D index” tir ve Hubert endeksi gibi ölçülen değerlerin önemli bir artışa tekabül ettiği değişim noktalarını araştırarak oluşturulan değerler ve ikincil farklar grafikleri Şekil 44’te yer almaktadır.



Seçilen Küme Sayısı Aralığında (2-20) Euclidean (Centroid) Metodu tüm metotlar arasında optimum küme sayısını 30 endeksten 15'i ile en çok aynı sonucu bulan metot olmuştur. Bu sebeple Euclidean (Centroid) metodu nihai sonuç olarak kabul edilmiş ve Tablo 9 ve Tablo 10'da Metodun içerdiği tüm endeksler için hesaplanan endeks değerlerine yer verilmiştir.

```

*****
* Among all indices:
* 6 proposed 2 as the best number of clusters
* 14 proposed 3 as the best number of clusters
* 1 proposed 11 as the best number of clusters
* 2 proposed 19 as the best number of clusters
* 1 proposed 20 as the best number of clusters

***** Conclusion *****

* According to the majority rule, the best number of clusters is 3

*****

```

Şekil 41: NbClust (Minkowski, Single) Optimum Küme Sayısı (3)

```

*****
* Among all indices:
* 6 proposed 2 as the best number of clusters
* 15 proposed 3 as the best number of clusters
* 1 proposed 10 as the best number of clusters
* 1 proposed 16 as the best number of clusters
* 1 proposed 20 as the best number of clusters

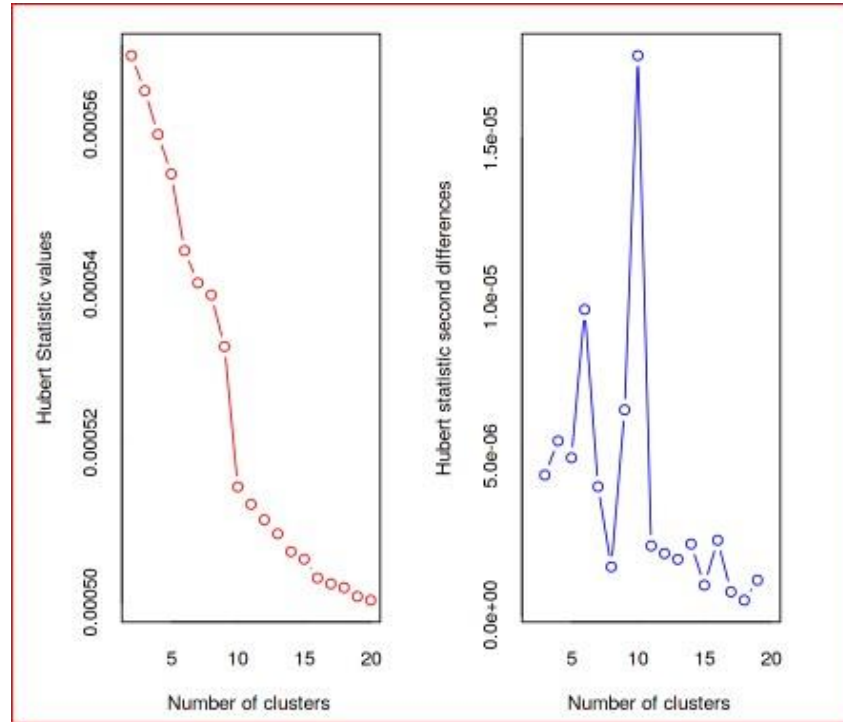
***** Conclusion *****

* According to the majority rule, the best number of clusters is 3

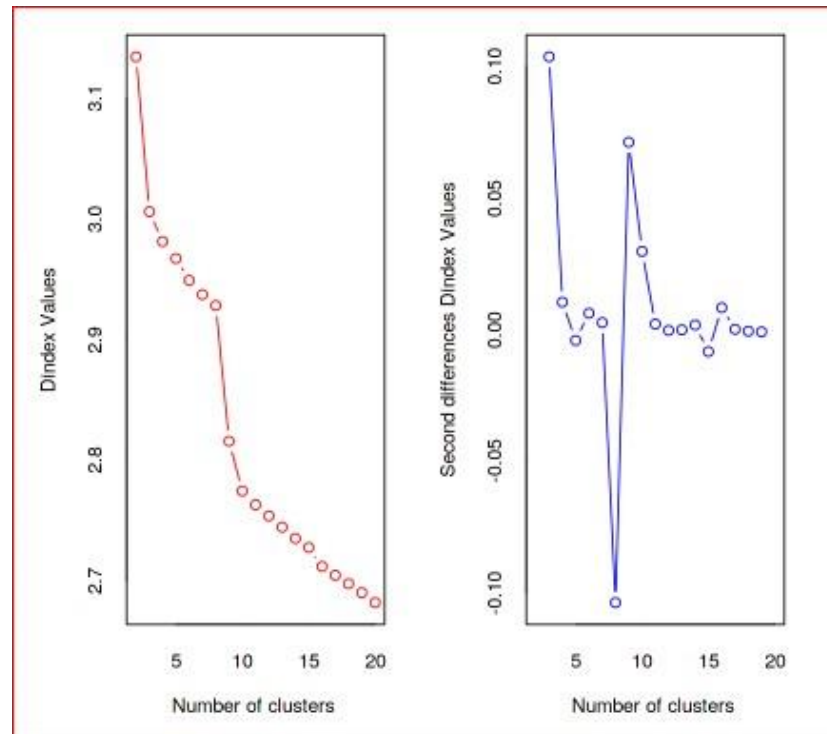
*****

```

Şekil 42: NbClust (Minkowski, Centroid) Optimum Küme Sayısı (3)



Şekil 43: Hubert Endeksi R Programlama Grafikleri



Şekil 44: "D index" R Programlama Grafikleri



	KL	CH	Hartigan	CCC	Scott	Marriot	TrCovW	TraceW	Friedman	Rubin	...	Ratkowsky	Ball	Ptbiserial	Frey	McClain	Dunn	Hubert	SDindex	Dindex	SDbw
2	0.3446	28.9635	110.4391	-21.4235	805.9624	3.704115e+36	445205.4	7985.504	3.3057	1.0519	...	0.0539	3992.7522	0.3292	12.3908	0.0008	0.7022	6e-04	0.3831	3.1348	0.4849
3	9.6635	72.4375	13.8132	-17.0795	1676.8257	1.759908e+36	291586.9	6666.144	37.3099	1.2601	...	0.1999	2222.0481	0.6669	21.2637	0.0053	0.5023	6e-04	0.8873	3.0063	1.3101
4	0.6477	53.9966	4.8704	-22.5640	1922.1039	2.019053e+36	280464.3	6504.829	40.0980	1.2913	...	0.1886	1626.2073	0.6846	79.0445	0.0064	0.4493	6e-04	0.7216	2.9816	0.9781
5	1.7331	41.9941	6.9095	-27.7654	2026.3410	2.618962e+36	276747.1	6448.343	40.7375	1.3027	...	0.1755	1289.6686	0.6678	54.9302	0.0078	0.3257	6e-04	0.7155	2.9674	0.7814
6	0.8459	35.3317	3.6811	-31.9100	2203.5386	2.748340e+36	271773.0	6369.051	42.0407	1.3189	...	0.1647	1061.5085	0.6677	91.8594	0.0078	0.3257	5e-04	0.4912	2.9493	0.4081
7	0.8691	30.1976	1.1132	-36.1258	2296.7286	3.167326e+36	269520.2	6327.010	42.5213	1.3276	...	0.1570	903.8586	0.6486	407.7232	0.0094	0.2553	5e-04	0.5058	2.9375	0.3494
8	0.1418	26.0476	78.4003	-40.4347	2378.9799	3.571812e+36	268966.8	6314.300	44.2729	1.3303	...	0.1476	789.2875	0.6485	11.2222	0.0094	0.2553	5e-04	0.5081	2.9285	0.2138
9	5.2438	35.7639	21.3552	-30.4407	3041.9284	1.383757e+36	212976.9	5529.016	66.5768	1.5193	...	0.1667	614.3351	0.7300	5.9663	0.0277	0.2493	5e-04	0.7173	2.8158	0.5293
10	4.1273	35.3308	3.7781	-30.0209	3762.7289	4.716019e+35	203521.7	5322.722	76.0300	1.5781	...	0.1641	532.2722	0.7305	22.9628	0.0278	0.2642	5e-04	0.6642	2.7745	0.4173
11	0.8860	32.3350	2.7662	-32.9828	3863.3071	4.768260e+35	202534.2	5286.409	77.9352	1.5890	...	0.1591	480.5826	0.7249	26.2649	0.0297	0.2642	5e-04	0.6574	2.7631	0.3791
12	1.0863	29.7408	2.5581	-33.8953	4138.3617	3.472362e+35	201799.9	5259.906	85.7049	1.5970	...	0.1532	438.3255	0.7248	35.2974	0.0297	0.2642	5e-04	0.6321	2.7539	0.3262
13	1.2842	27.5525	2.8639	-34.7294	4211.9956	3.573092e+35	200447.8	5235.466	90.9774	1.6044	...	0.1482	402.7282	0.7248	24.6750	0.0297	0.2642	5e-04	0.6174	2.7445	0.2819
14	0.7227	25.7394	1.5244	-35.4047	4301.9499	3.529001e+35	199497.6	5208.198	92.3480	1.6128	...	0.1435	372.0141	0.7247	73.7150	0.0297	0.2642	5e-04	0.6006	2.7352	0.2406
15	1.1060	24.0324	7.9047	-36.2573	4461.5577	3.046479e+35	199000.7	5193.697	97.4423	1.6173	...	0.1390	346.2465	0.7246	16.4257	0.0297	0.2642	5e-04	0.6091	2.7277	0.2116
16	0.8955	23.2398	0.8789	-35.8271	4771.8416	1.991693e+35	194652.0	5119.444	100.8950	1.6408	...	0.1366	319.9653	0.7182	1040.1291	0.0334	0.2642	5e-04	0.6214	2.7120	0.2344
17	1.0513	21.8372	0.8187	-36.6786	4809.3487	2.102774e+35	194416.0	5111.187	101.3378	1.6435	...	0.1329	300.6581	0.7182	-1173.7515	0.0334	0.2642	5e-04	0.6648	2.7047	0.1865
18	1.6159	20.5938	1.8696	-37.4825	4844.3293	2.214682e+35	194209.0	5103.493	101.8893	1.6459	...	0.1293	283.5274	0.7181	40.2126	0.0334	0.2642	5e-04	0.6449	2.6977	0.1366
19	1.3037	19.5845	2.1469	-38.0408	4896.1177	2.249626e+35	193729.8	5085.949	103.3169	1.6516	...	0.1265	267.6815	0.7181	35.3183	0.0334	0.2642	5e-04	0.6510	2.6902	0.1280
20	0.7801	18.7057	1.5289	-38.5070	4983.6345	2.132022e+35	193138.6	5065.846	105.5060	1.6582	...	0.1238	253.2923	0.7180	60.6277	0.0334	0.2642	5e-04	0.6532	2.6820	0.1113

**Tablo 9: Seçilen Küme Sayısı Aralığında (2-20) Euclidean (Centroid) Metodu ile Tüm Endeksler için Hesaplanan Endeks Değerleri**

	KL	CH	Hartigan	CCC	Scott	Marriot	TrCovW	TraceW	Friedman	Rubin	...	Ratkowsky	Ball	Ptbiserial	Frey	McClain	Dunn	Hubert	SDindex	Dindex	SDbw
Number_clusters	3.0000	3.0000	3.0000	3.0000	3.0000	3.000000e+00	3.0	3.000	3.0000	3.0000	...	3.0000	3.000	10.0000	16.000	2e+00	2.0000	0	2.0000	0	20.0000
Value_index	9.6635	72.4375	96.6259	-17.0795	870.8833	2.203352e+36	153618.5	1158.045	34.0041	-0.1789	...	0.1999	1770.704	0.7305	1040.129	8e-04	0.7022	0	0.3831	0	0.1113

**Tablo 10: Euclidean (Centroid) Metodu ile Tüm Endeksler için Optimum Küme Sayısı ve Endeks Değerleri**

#### 5.1.4 Kümeleme için Sonuçların Gösterimi

Optimum küme sayısını belirleme yöntemleri, verileri çoğunlukla 3 kümeye ayırmıştır. Bu bağlamda optimum sayı 3 olarak belirlenmiştir. Karşılaştırma yapabilmek amacı ile 3'lü ve 5'li küme sonuçları tablo ve harita görselleştirmelerinde sunulmaktadır.

Veri seti bir tabloda gösterilemeyecek kadar çok satır içermesi sebebi ile ülke bazında 3'lü ve 5'li kümelere atanan havalimanı sayıları, "euclidean" ve "cosine" metotlarına göre sonuçları gösteren iki farklı özet tablo oluşturulmuştur. "Euclidean" metodu ile bulunan 3'lü ve 5'li kümeleme sonuçları Tablo 11'de ve "Cosine" metodu ile bulunan sonuçlar Tablo 12'de yer almaktadır. Tablolar incelendiğinde boş olan hücreler, bazı ülkeler için her bir kümeye atanan eleman (havalimanı) olmadığını göstermektedir.

Ayrıca sonuçlar harita üzerinde de görselleştirilmiştir. Bu görselleştirme Microsoft ürünü olan "Power BI" iş analizi platformunda yapılmıştır. Azure ML Stüdyo üzerinde kümeleme sonuçları veri setine dönüştürüldükten sonra ".csv" formatında bilgisayara indirilmiştir. İndirilen veriler daha önce ham veri setinde yer alan havalimanı adı, ülke adı, şehir adı, enlem ve boylam bilgileri gibi demografik bilgileri de içeren veri seti ile eşleştirildikten ve atanan kümeler bilgisini içeren sütun Euclidean yönteminde 3'lü kümeler için "A", "B", "C" 5'li kümeler için "A", "B", "C", "D", "E" ve Cosine yönteminde 3'lü kümeler için "K", "L", "M" 5'li kümeler için "K", "L", "M", "N", "O" olmak üzere harflerle temsil edilen kategorik veriye dönüştürülen yeni bir sütun eklendikten sonra "Power BI" platformuna yüklenmiştir. Haritalarda havalimanı konum bilgileri kullanılarak bir daire ile temsil edilen veriler, aynı renkteki daireler aynı kümede yer alan havalimanlarını gösterecek şekilde organize edilerek görselleştirilmiştir.

"Euclidean" tekniği ile elde edilen 3'lü kümeleme sonuçları Şekil 45'te, "Euclidean" tekniği ile elde edilen 5'li kümeleme sonuçları Şekil 46'da "Cosine" tekniği ile elde edilen 3'lü kümeleme sonuçları Şekil 47'de ve "Cosine" tekniği ile elde edilen 5'li kümeleme sonuçları Şekil 48'de harita üzerinde görselleştirilmiştir. Son olarak tek bir ülkeye ait sonuçların da değerlendirilebileceğini göstermek açısından ve bu tezin Türkiye'de yapılmış olmasından Türkiye havalimanları örneği ele alınmıştır. Türkiye'de yer alan 55 havalimanı için bulunan sonuçlar tüm veri setinden filtrelenerek Tablo 13 ve 14'te paylaşılmıştır.

No	Ülke Adı	Kümelerdeki Havalimanları Sayısı (Euclidean 5)					Kümelerdeki Havalimanları Sayısı (Euclidean 3)			Toplam
		A Kümesi	B Kümesi	C Kümesi	D Kümesi	E Kümesi	A Kümesi	B Kümesi	C Kümesi	
1	Avusturya	3	2		1		5	1		6
2	Belçika	3	1		1		2	3		5
3	Bulgaristan		3		1		1	3		4
4	Hırvatistan	7					6	1		7
5	Kıbrıs Rum Kesimi	1			1		1	1		2
6	Çekya	1	3		1		4	1		5
7	Danimarka	9			1		8	2		10
8	Estonya	1					1			1
9	Finlandiya	20			1		20	1		21
10	Fransa	36	10	1	7		46	7	1	54
11	Fransız Guyanası	2					2			2
12	Almanya	19	11	1	8	1	28	10	2	40
13	Yunanistan	38			1		35	4		39
14	Guadeloupe (Fransız Bölgesi)	1						1		1
15	Macaristan	1	2		1		3	1		4
16	İzlanda	6					5	1		6
17	İrlanda	7			1		7	1		8
18	İtalya	27			9		21	14	1	36
19	Letonya				1			1		1
20	Litvanya	3	1				3	1		4
21	Lüksemburg				1			1		1
22	Makedonya	2					2			2
23	Malta				1			1		1
24	Martinik (Fransız Bölgesi)	1					1			1
25	Mayotte (Fransız Bölgesi)	1					1			1
26	Karadağ	2					2			2

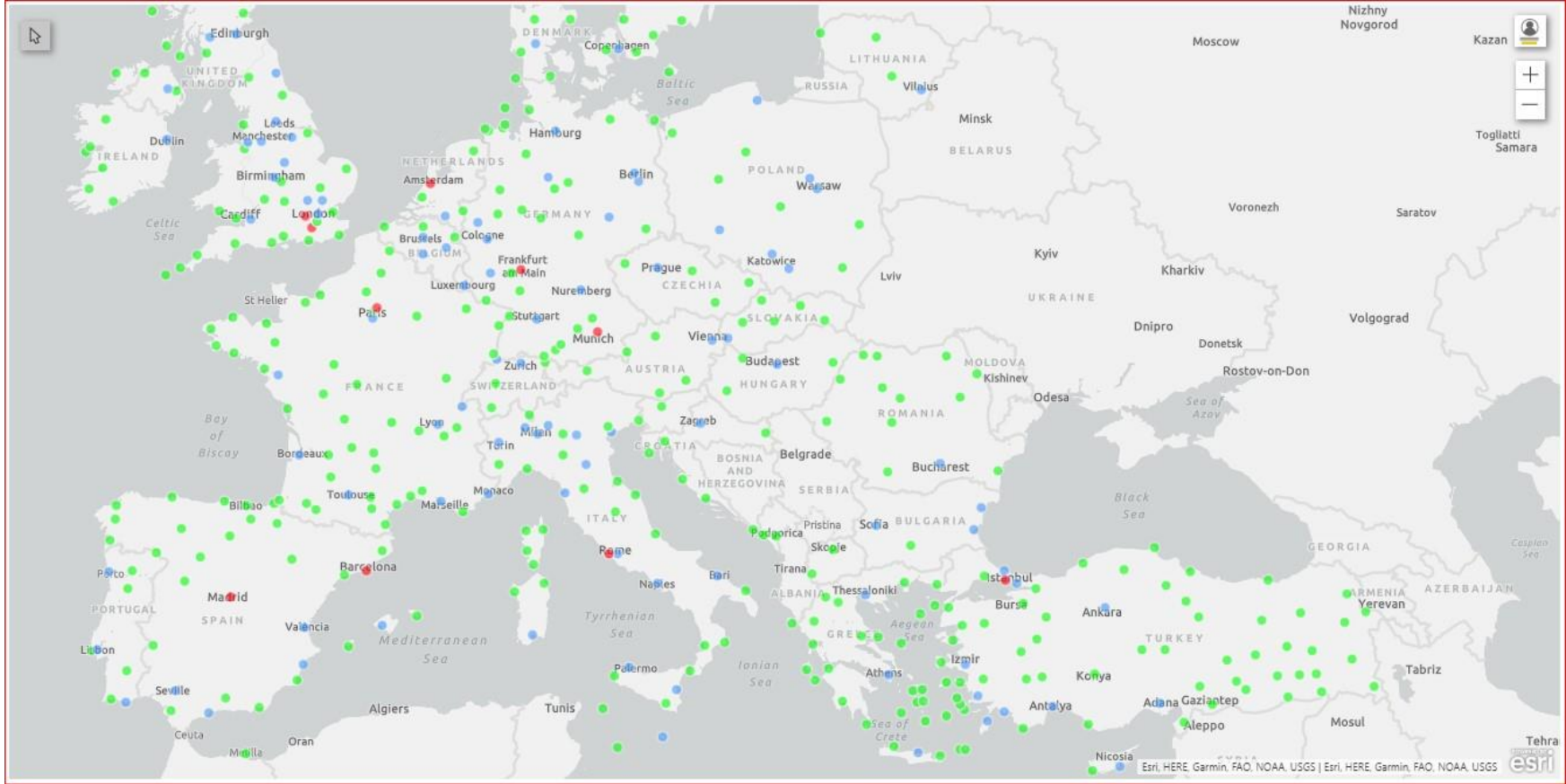
No	Ülke Adı	Kümelerdeki Havalimanları Sayısı (Euclidean 5)					Kümelerdeki Havalimanları Sayısı (Euclidean 3)			Toplam
		A Kümesi	B Kümesi	C Kümesi	D Kümesi	E Kümesi	A Kümesi	B Kümesi	C Kümesi	
27	Hollanda	4		1			3	1	1	5
28	Norveç	46			3		43	6		49
29	Polonya	6	4		2		6	6		12
30	Portekiz	14	3		3		16	4		20
31	Reunion (Fransız Bölgesi)	2					2			2
32	Romanya	1	11		1		12	1		13
33	Saint Martin (Fransız Bölgesi)	1					1			1
34	Slovakya	2	4				5	1		6
35	Slovenya	1					1			1
36	İspanya	27	5	1	6		27	10	2	39
37	İsveç	29			2		29	2		31
38	İsviçre	4			3		4	3		7
39	Türkiye	7	42	1	5		46	8	1	55
40	Birleşik Krallık	38	8	1	8		39	14	2	55
	Toplam	373	110	6	70	1	438	112	10	560

**Tablo 11: Ülke Bazında Kümelerdeki Havalimanları Sayısı (Euclidean)**

No	Ülke Adı	Kümelerdeki Havalimanları Sayısı (Cosine 5)					Kümelerdeki Havalimanları Sayısı (Cosine 3)			Toplam
		K Kümesi	L Kümesi	M Kümesi	N Kümesi	O Kümesi	K Kümesi	L Kümesi	M Kümesi	
1	Avusturya	1	1	2	2		3	1	2	6
2	Belçika	1	3	1			1	3	1	5
3	Bulgaristan		1	3				3	1	4
4	Hırvatistan	3	2	1		1	5	2		7
5	Kıbrıs Rum Kesimi		2					2		2
6	Çekya	1	1	3			1	1	3	5
7	Danimarka	4	2	2		2	7	3		10
8	Estonya		1					1		1
9	Finlandiya	18	1			2	19	1	1	21
10	Fransa	16	10	5	16	7	19	11	24	54
11	Fransız Guyanası	2					2			2
12	Almanya	2	16	6	7	9	6	16	18	40
13	Yunanistan	32	5		1	1	34	5		39
14	Guadeloupe (Fransız Bölgesi)		1					1		1
15	Macaristan	1	1	2			1	1	2	4
16	İzlanda	4	1			1	5	1		6
17	İrlanda	5	3				5	3		8
18	İtalya	11	17		7	1	15	18	3	36
19	Letonya		1					1		1
20	Litvanya	2	1			1	2	1	1	4
21	Lüksemburg		1					1		1
22	Makedonya		1		1		1	1		2
23	Malta		1					1		1
24	Martinik (Fransız Bölgesi)		1					1		1
25	Mayotte (Fransız Bölgesi)	1					1			1
26	Karadağ	2					1	1		2

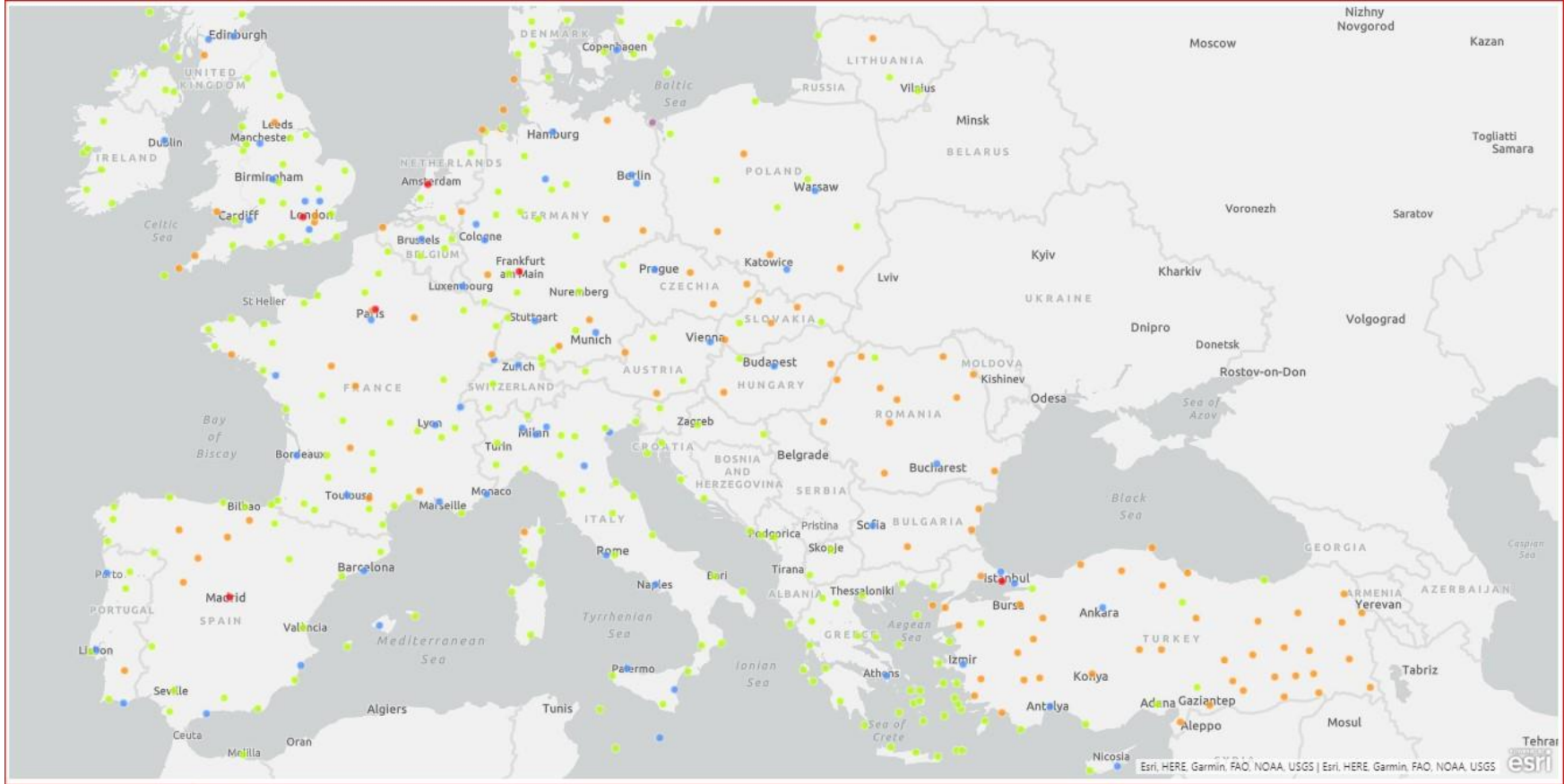
No	Ülke Adı	Kümelerdeki Havalimanları Sayısı (Cosine 5)					Kümelerdeki Havalimanları Sayısı (Cosine 3)			Toplam
		K Kümesi	L Kümesi	M Kümesi	N Kümesi	O Kümesi	K Kümesi	L Kümesi	M Kümesi	
27	Hollanda	3	2				3	2		5
28	Norveç	41	6		1	1	43	6		49
29	Polonya	2	5	5			2	7	3	12
30	Portekiz	11	5	3		1	12	5	3	20
31	Reunion (Fransız Bölgesi)	1	1				1	1		2
32	Romanya	1	1	10		1	3	1	9	13
33	Saint Martin (Fransız Bölgesi)	1					1			1
34	Slovakya	1		4		1	3	1	2	6
35	Slovenya		1					1		1
36	İspanya	13	16	1	7	2	14	16	9	39
37	İsveç	27	4				27	4		31
38	İsviçre	3	3			1	4	3		7
39	Türkiye		8	9	32	6		10	45	55
40	Birleşik Krallık	18	23	4	1	9	22	25	8	55
	Toplam	228	149	61	75	47	263	162	135	560

**Tablo 12: Ülke Bazında Kümelerdeki Havalimanları Sayısı (Cosine)**



Şekil 45: Azure ML Studio (Euclidean) 3'lü Kümeleme

A kümesi: ● B Kümesi: ● C Kümesi: ●



Şekil 46: Azure ML Studio (Euclidean) 5'li Kümeleme

A kümesi: ●

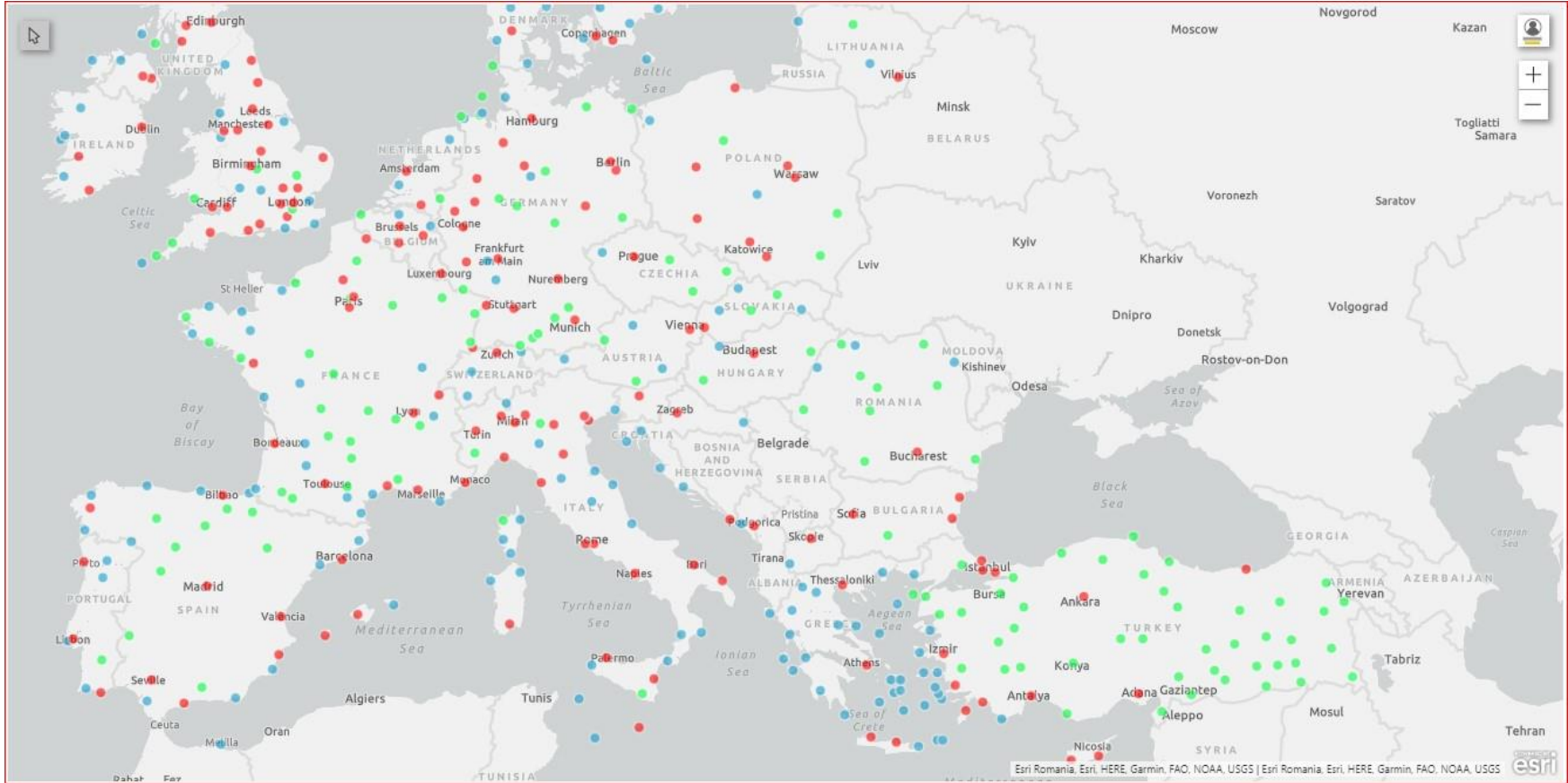
B Kümesi: ●

C Kümesi: ●

D Kümesi: ●

E Kümesi: ●



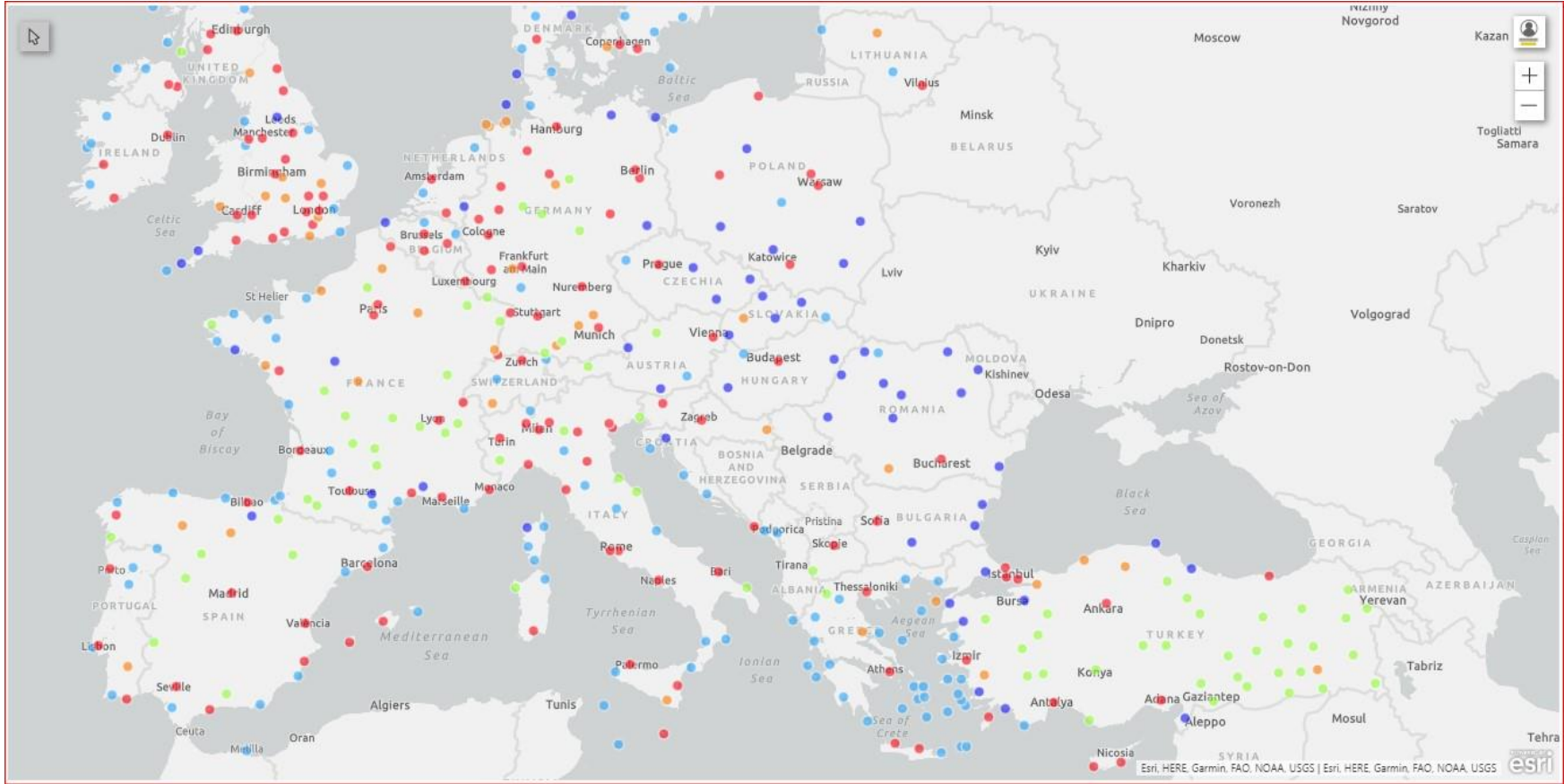


Şekil 47: Azure ML Studio (Cosine) 3'lü Kümeleme

K kümesi: ●

L Kümesi: ●

M Kümesi: ●



Şekil 48: Azure ML (Cosine) 5'li Kümeleme

K kümesi: ●

L Kümesi: ●

M Kümesi: ●

N Kümesi: ●

O Kümesi: ●

Euclidean			Cosine		
A Kümesi	B Kümesi	C Kümesi	K Kümesi	L Kümesi	M Kümesi
Adiyaman Airport	Adana Airport	Atatürk Airport		Adana Airport	Adiyaman Airport
Ağrı Airport	Adnan Menderes Airport			Adnan Menderes Airport	Ağrı Airport
Amasya Merzifon Airport	Antalya Airport			Antalya Airport	Amasya Merzifon Airport
Anadolu Airport	Dalaman Airport			Atatürk Airport	Anadolu Airport
Balıkesir Körfez Airport	Esenboğa Airport			Dalaman Airport	Balıkesir Körfez Airport
Balıkesir Merkez Airport	İstanbul Airport			Esenboğa Airport	Balıkesir Merkez Airport
Batman Airport	Milas Bodrum Airport			İstanbul Airport	Batman Airport
Bingöl Çeltiksuyu Airport	Sabiha Gökçen Airport			Milas Bodrum Airport	Bingöl Çeltiksuyu Airport
Bursa Yenişehir Airport				Sabiha Gökçen Airport	Bursa Yenişehir Airport
Çanakkale Airport				Trabzon Airport	Çanakkale Airport
Çardak Airport					Çardak Airport
Cengiz Topel Airport					Cengiz Topel Airport
Çıldır Airport					Çıldır Airport
Diyarbakır Airport					Diyarbakır Airport
Elâzığ Airport					Elâzığ Airport
Erzincan Airport					Erzincan Airport
Erzurum Airport					Erzurum Airport
Gaziantep Airport					Gaziantep Airport
Gazipaşa Airport					Gazipaşa Airport
Gökçeada Airport					Gökçeada Airport
Hakkâri Yüksekova Airport					Hakkâri Yüksekova Airport
Hatay Airport					Hatay Airport
İğdır Airport					İğdır Airport
Kahramanmaraş Airport					Kahramanmaraş Airport
Kars Airport					Kars Airport
Kastamonu Airport					Kastamonu Airport
Kayseri Erkilet Airport					Kayseri Erkilet Airport

Euclidean			Cosine		
A Kümesi	B Kümesi	C Kümesi	K Kümesi	L Kümesi	M Kümesi
Konya Airport					Konya Airport
Malatya Erhac Airport					Malatya Erhac Airport
Mardin Airport					Mardin Airport
Muş Airport					Muş Airport
Nevşehir Kapadokya Airport					Nevşehir Kapadokya Airport
Samsun Çarşamba Airport					Samsun Çarşamba Airport
Şanlıurfa GAP Airport					Şanlıurfa GAP Airport
Siirt Airport					Siirt Airport
Sinop Airport					Sinop Airport
Şırnak Şerafettin Elçi Airport					Şırnak Şerafettin Elçi Airport
Sivas Nuri Demirağ Airport					Sivas Nuri Demirağ Airport
Süleyman Demirel Airport					Süleyman Demirel Airport
Tekirdağ Çorlu Airport					Tekirdağ Çorlu Airport
Tokat Airport					Tokat Airport
Trabzon Airport					Uşak Airport
Uşak Airport					Van Ferit Melen Airport
Van Ferit Melen Airport					Zafer Airport
Zafer Airport					Zonguldak Airport
Zonguldak Airport					

**Tablo 13: Türkiye Havalimanları 3'lü Kümeleme**

Euclidean					Cosine				
A Kümesi	B Kümesi	C Kümesi	D Kümesi	E Kümesi	K Kümesi	L Kümesi	M Kümesi	N Kümesi	O Kümesi
Adana Airport	Adıyaman Airport	Atatürk Airport	Adnan Menderes Airport			Adana Airport	Balıkesir Körfez Airport	Adıyaman Airport	Cengiz Topel Airport
Balıkesir Merkez Airport	Ağrı Airport		Antalya Airport			Adnan Menderes Airport	Bursa Yenişehir Airport	Ağrı Airport	Çıldır Airport
Cengiz Topel Airport	Amasya Merzifon Airport		Esenboğa Airport			Antalya Airport	Çanakkale Airport	Amasya Merzifon Airport	Gökçeada Airport
Gazipaşa Airport	Anadolu Airport		İstanbul Airport			Atatürk Airport	Dalaman Airport	Anadolu Airport	Kastamonu Airport
Kahramanmaraş Airport	Balıkesir Körfez Airport		Sabiha Gökçen Airport			Esenboğa Airport	Hatay Airport	Balıkesir Merkez Airport	Siirt Airport
Tokat Airport	Batman Airport					İstanbul Airport	Milas Bodrum Airport	Batman Airport	Zonguldak Airport
Trabzon Airport	Bingöl Çeltiksuyu Airport					Sabiha Gökçen Airport	Samsun Çarşamba Airport	Bingöl Çeltiksuyu Airport	
	Bursa Yenişehir Airport					Trabzon Airport	Sinop Airport	Çardak Airport	
	Çanakkale Airport						Tekirdağ Çorlu Airport	Diyarbakır Airport	
	Çardak Airport							Elâzığ Airport	
	Çıldır Airport							Erzincan Airport	
	Dalaman Airport							Erzurum Airport	
	Diyarbakır Airport							Gaziantep Airport	
	Elâzığ Airport							Gazipaşa Airport	
	Erzincan Airport							Hakkâri Yüksekova Airport	
	Erzurum Airport							İğdır Airport	
	Gaziantep Airport							Kahramanmaraş Airport	
	Gökçeada Airport							Kars Airport	
	Hakkâri Yüksekova Airport							Kayseri Erkilet Airport	
	Hatay Airport							Konya Airport	
	İğdır Airport							Malatya Erhac Airport	
	Kars Airport							Mardin Airport	
	Kastamonu Airport							Muş Airport	

Euclidean					Cosine				
A Kümesi	B Kümesi	C Kümesi	D Kümesi	E Kümesi	K Kümesi	L Kümesi	M Kümesi	N Kümesi	O Kümesi
	Kayseri Erkilet Airport Konya Airport Malatya Erhac Airport Mardin Airport Milas Bodrum Airport Muş Airport Nevşehir Kapadokya Airport Samsun Çarşamba Airport Şanlıurfa GAP Airport Siirt Airport Sinop Airport Şırnak Şerafettin Elçi Airport Sivas Nuri Demirağ Airport Süleyman Demirel Airport Tekirdağ Çorlu Airport Uşak Airport Van Ferit Melen Airport Zafer Airport Zonguldak Airport							Nevşehir Kapadokya Airport Şanlıurfa GAP Airport Şırnak Şerafettin Elçi Airport Sivas Nuri Demirağ Airport Süleyman Demirel Airport Tokat Airport Uşak Airport Van Ferit Melen Airport Zafer Airport	

Tablo 14: Türkiye Havalimanları 5'li Kümeleme

### 5.1.5 Kümeleme için Sonuçların Değerlendirilmesi

40 Avrupa ülkesine ait 560 satırdan oluşan veri setinin tamamı kullanılarak yapılan kümeleme uygulamasında Euclidean ve Cosine metotlarına göre sonuçlar üretilmiştir.

Euclidean metodunda 3'lü Kümelemede A kümesi 438, B kümesi 112, C kümesi 10 havalimanına ve 5'li kümelemede A kümesi 373, B kümesi 110, C kümesi 6, D kümesi 70, E kümesi 1 havalimanına sahiptir. Kendi içerisinde değerlendirdiğimizde E kümesinde 1 havalimanı olması Euclidean tekniği için 5 farklı kümeye ihtiyaç olup olmadığını sorgulamaktadır. 3'lü kümelemede ise Şekil 45'teki harita görseli incelendiğinde küçük ölçekli havalimanlarının çoğunun yeşil renkli A kümesinde, orta büyüklükte havalimanlarının mavi renkli B kümesinde ve nihayet kırmızı noktalar olarak işaretlenen İstanbul, Roma, Barcelona, Londra, Frankfurt gibi mega büyüklükte havalimanlarının ise C kümesinde toplandığı yorumu kolaylıkla yapılabilmektedir. Bu da Euclidean tekniği için çıkan sonuçlarda havalimanlarının kümelere mantıklı bir yolla atandığını göstermektedir. Şekil 46'da 5'li kümeleme sonuçları için atanan kümelerin mantıklı yolla gerçekleştiği; kırmızı ve mavi noktalarda çok büyük ve büyük ölçekli havalimanlarının, turuncu noktalarda orta ölçekli havalimanlarının ve diğer renkler için küçük ölçekli havalimanlarının atanmış olmasından anlaşılabilir. Fakat veriler incelendiğinde mor renk ile temsil edilen E kümesine atanan Heringsdorf havalimanı kendine has özelliklere sahip olsa da diğer havalimanlarından farklı bir küme oluşturacak kadar farklı özellikler içermediği kanaatine varılmıştır bu sebeple Euclidean tekniği için veri setini 3 kümede yapılandırmak daha doğru bir çıkarım olacaktır.

Cosine metodunda 3'lü kümelemede K kümesi 263, L kümesi 162, M kümesi 135, havalimanına ve 5'li kümelemede K kümesi 228, L kümesi 149, M kümesi 61, N kümesi 75, O kümesi 45 havalimanına sahiptir. Şekil 47 incelendiğinde 3'lü kümeleme için Kırmızı noktalar olarak belirtilen havalimanlarının İstanbul, Roma, Ankara, Milano, Muğla, Monaco, Zürih, Edinburgh, Glasgow, Hamburg, Hannover, Frankfurt, Londra, Liverpool, Bremen gibi yoğun trafiğe sahip havalimanlarından oluştuğu, Yeşil noktaların normal yoğunluğa Mavi noktaların ise daha az yoğunluğa sahip havalimanlarını temsil ettiği yorumu yapılabilmektedir. 5'li kümeleme için Şekil 48 incelendiğinde; kırmızı

renklerin yine büyük veya yoğun diyebileceğimiz havalimanlarını temsil ettiğini, koyu mavi renklerin nispeten yoğun havalimanlarını, yeşil renklerin normal trafiğe sahip havalimanlarını, açık mavi renklerin yoğun olmayan fakat belli düzeyde hareket içeren havalimanlarından oluştuğunu, turuncu renklerin ise Aydın, Çanakkale Gökçeada, Kocaeli, Siirt, Zonguldak, Cambridge gibi diğer havalimanlarına göre çok az hareket içeren havalimanlarını temsil ettiğini görebiliriz. 5'li kümeleme sonuçları Euclidean tekniğine göre atanan küme sayılarının dağılımı açısından Cosine tekniğinde daha iyi sonuçlar üretmiş olmasına rağmen ülke bazlı incelemede birkaç ülkede bazı kümelere atanan eleman bulunmamaktadır. Bu da ülke bazlı değerlendirmelerde sonuçların değerlendirilmesini nispeten zorlaştırabilmektedir. 3'lü kümeleme için de aynı şekilde Türkiye örneğinde olduğu gibi bazı kümelere atanan eleman bulunmamaktadır. Fakat bu 5'li kümelemedeki kadar çok ülkeyi kapsamamaktadır. Bu Cosine metodu için de 3'lü küme yapılandırmasında daha iyi sonuçlar üretildiğini göstermektedir.

Euclidean ve Cosine metotlarından her ikisi için de 3'lü küme yapılandırması daha iyi sonuçlar vermektedir. Bu iki tekniğin ürettiği sonuçlardan hangisinin daha elverişli olduğu, veri setini nasıl yorumlamak istediğinize bağlı olarak değişebilmektedir. Euclidean tekniği büyük ve yoğun havalimanları ayrımını çok keskin şekilde ayırmıştır. Cosine de ise bu ayrım o kadar keskin gerçekleşmemiştir. Euclidean tekniğinde kırmızı renkli C kümesi büyük ve yoğun ölçekli olan; İstanbul Atatürk Havalimanı, Londra Heathrow, Paris, Frankfurt, Münih, Roma, Amsterdam, Madrid, Barcelona havalimanlarını kapsarken İstanbul Sabiha Gökçen, Ankara Esenboğa, Berlin, Köln, Hannover gibi büyük fakat nispeten daha az hareketliliğe sahip olan havalimanları bu kümede yer almamaktadır. Fakat Cosine tekniğinde bu havalimanlarının hepsi büyük ve yoğun havalimanlarını kapsayan kırmızı renkli L kümesinde yer almaktadır. Havacılık açısından değerlendirilirse çok büyük havalimanlarında var olan hareketliliğin diğer büyük havalimanlarından dahi belirgin olarak ayrıştığı bilinmektedir. Dolayısıyla Euclidean tekniği çok büyük havalimanlarını ayrı bir kümede toplamayı başararak sonuçların değerlendirilmesi açısından daha elverişli bir çıkarım sunmaktadır.



## 5.2 Havalimanları İçin Taşınan Yolcu Sayısı Tahmini Uygulama Örneği

### 5.2.1 Taşınan Yolcu Sayısı Tahmini için Veri Seti Gösterimi

Kümeleme uygulamasında kullanılan veri setinde yer alan tüm sütunlar (feature) aynı kalmıştır ancak eksik veriler içeren İstanbul Havalimanı gibi yeni açılan havalimanlarının ve bir takım tutarsız verilere sahip havalimanlarının (aktif olarak tarifeli uçuşa sahip olmayan bazı havalimanları) yapılacak olan tahmine zarar vermesini önlemek amacı ile taşınan yolcu sayısı tahmini uygulamasında kapsam dışı bırakılmıştır. Örnek veri setine ilişkin ilk 10 satır Tablo 15’te yer almaktadır.

Veri setinde kümeleme uygulamasından farklı olarak 2018 yılına ait taşınan yolcu sayısı “0” olarak belirtilen ve yeni açılması sebebi ile verileri eksik olan 12 adet havalimanı, tahmini olması gereken değerden sapacağı için bu veri setinden çıkartılmıştır. Geriye kalan 548 adet havalimanı için özet istatistikler Tablo 16’da gösterilmiştir.

### 5.2.2 Taşınan Yolcu Sayısı Tahmini için Model Gösterimi

Tahmin modeli ön işleme, eğitim ve değerlendirme olmak üzere 3 ana yapıdan oluşmaktadır. Model Şekil 49’da gösterilmektedir. Ön işleme olarak isimlendirilen bölümde kümeleme uygulamasında seçilen 16 sütun tahminleme uygulamasında da seçilmiş olup Poisson Regresyon algoritması negatif değerleri kabul etmediği için “Log-Normal” metodu ile diğer 4 algoritma ise “Z-Score” normalizasyon işlemine tabi tutulmuştur. “Havalimanı Adı” kümelemede olduğu gibi eğitim modelinden çıkartılmıştır. Veri seti için belirleyici sütunların seçimi Şekil 50 görselinde yer almaktadır.

Lineer Regresyon algoritması varsayılan ayarları benimsenmiş olup çözüm metodu olarak en küçük kareler metodu kullanılmıştır. Düzenleme ağırlığı: 0.001 ve deney sonuçlarını sistematik bir biçimde karşılaştırabilmek amacı ile rastgele besleme sayısı “12345” seçilmiştir. Lineer Regresyon teknik parametrelerine ilişkin görsel Şekil 51’de yer almaktadır.

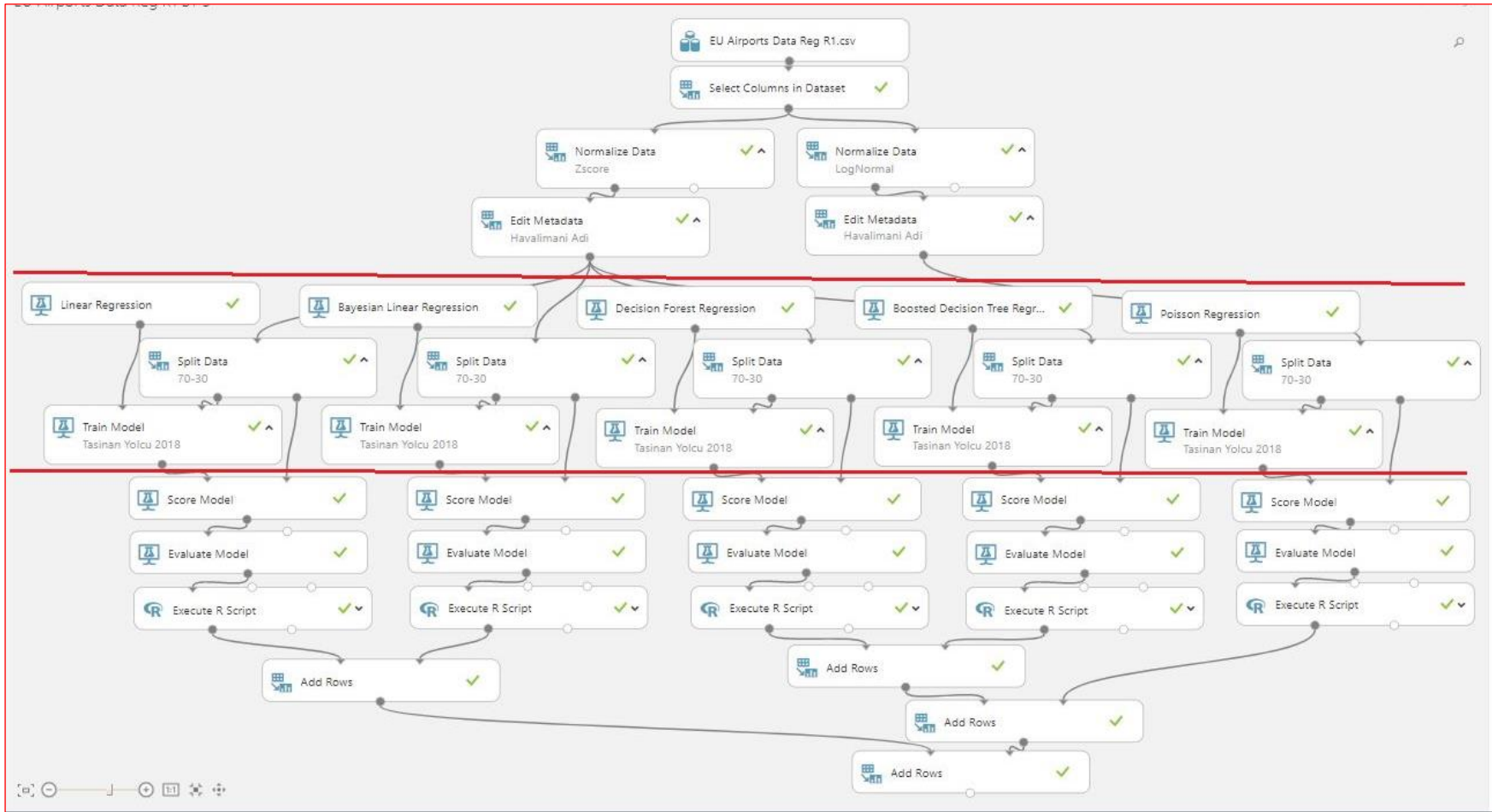
Bayesyen Regresyon algoritması Azure ML Studio varsayılan ayarları seçilmiştir. Platformdan alınan görsel Şekil 52’de yer almaktadır.

ICAO Kodu	Havalimani Adı	Kita Kodu	Ulke Kodu (iso)	Ulke Adı	Sehir	Ulke Nufusu	Sehir Nufusu	Sehirlerin Nufus Oranı	Havalimani Kategorisi	Havalimani Kategorisi (Num)	Rakim Ft	Tarifeli Ucus Bilgisi	Pist Uzunlugu Ft	Pist Genisligi Ft	Pist Yuzeyi	Pist Yuzeyi Num	Rota Sayisi	Yuk Posta 2018	Direk Gecis 2018	Tum Ucak Hareketleri 2018	Yolcu Uçak Hareketleri 2018	Tasinan Yolcu 2018
LOWG	Graz Airport	EU	AT	Austria	Graz	8955102	222326	40.279149	medium_air	2	1115	1	9842	148	Asphalt	5	11	226	7917	41416	7941	1026686
LOWI	Innsbruck Airport	EU	AT	Austria	Innsbruck	8955102	132493	67.589246	medium_air	2	1907	1	6562	148	Asphalt	5	22	73	417	50902	7397	1124389
LOWK	Klagenfurt Airport	EU	AT	Austria	Klagenfurt	8955102	90610	98.831277	medium_air	2	1472	1	8924	148	Concrete	3	4	0	50	20550	1432	229655
LOWL	Linz Hoersching Airport	EU	AT	Austria	Linz	8955102	204846	43.716265	medium_air	2	980	1	9843	197	Asphalt	5	5	7702	8128	27414	1703	458868
LOWS	Salzburg Airport	EU	AT	Austria	Salzburg	8955102	153377	58.386212	medium_air	2	1411	1	9022	148	Concrete	3	50	185	3342	58271	10220	1854326
LOWW	Vienna International Airport	EU	AT	Austria	Vienna	8955102	1691468	5.2942781	large_airport	3	600	1	11811	148	Asphalt	5	211	229607	85911	256640	177763	27024787
EBAW	Antwerp International Airport (Deur	EU	BE	Belgium	Antwerpen	11539328	1106422	10.429409	medium_air	2	39	1	4954	148	Asphalt	5	14	2449	7	39467	6070	282126
EBBR	Brussels Airport	EU	BE	Belgium	Brussels	11539328	1019022	11.323924	large_airport	3	184	1	11936	148	Asphalt	5	190	539082	38855	223577	136323	25622344
EBCI	Brussels South Charleroi Airport	EU	BE	Belgium	Brussels	11539328	1019022	11.323924	large_airport	3	614	1	8366	148	Asphalt	5	113	438	1154	53948	47911	8022370
EBLG	Liege Airport	EU	BE	Belgium	Liege	11539328	750318	15.37925	large_airport	3	659	1	12106	148	Asphalt	5	9	869201	2744	31196	1195	169108

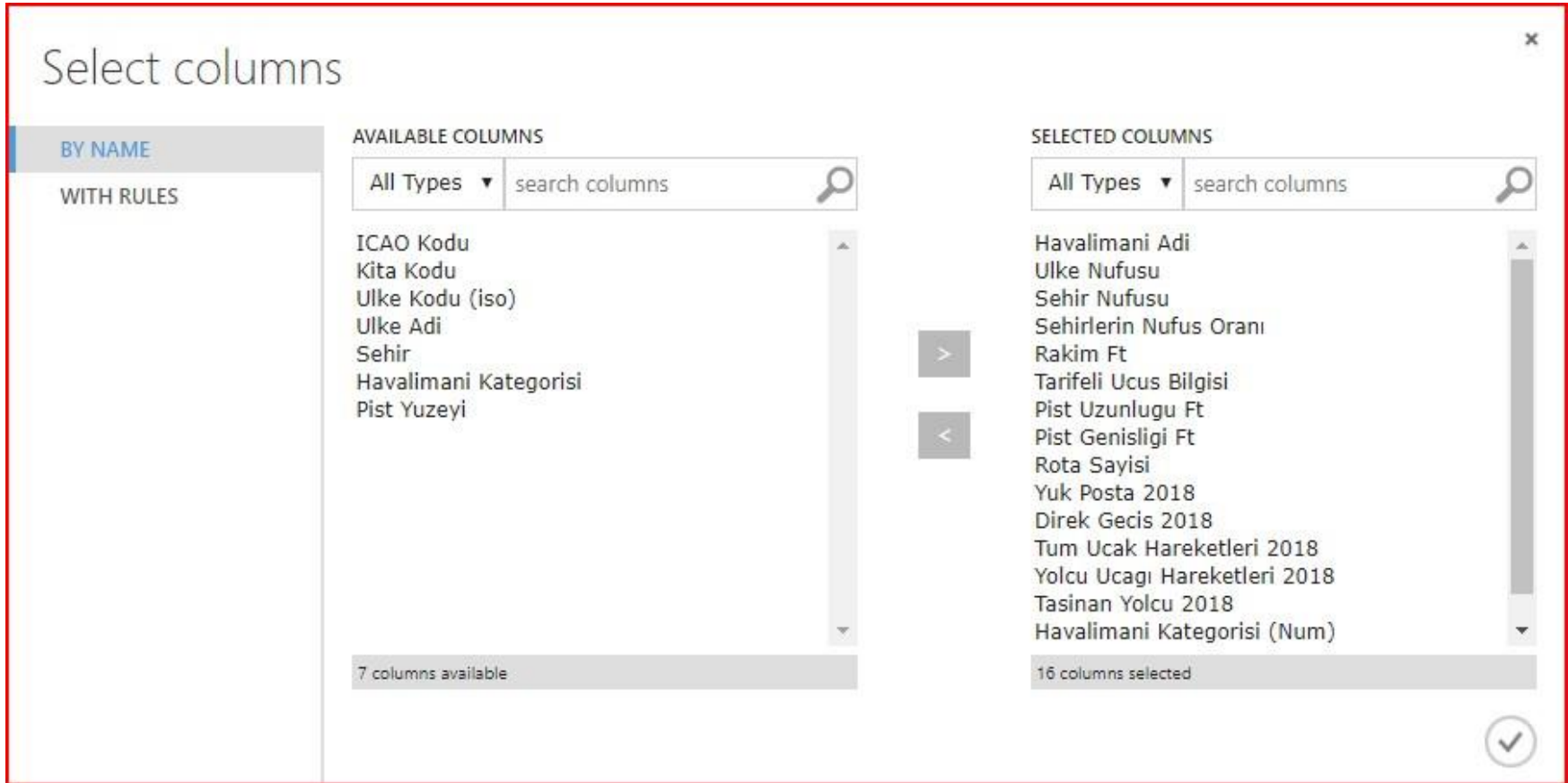
Tablo 15: Tahminleme Ana Veri Seti Örneği (ilk 10 satır)

Feature	Count	Unique Value Count	Missing Value Count	Min	Max	Mean	Mean Deviation	1st Quartile	Median	3rd Quartile	Mode	Range	Sample Variance	Sample Standard Deviation
ICAO Kodu	548	548	0											
Havalimani Adi	548	548	0											
Kita Kodu	548	5	0											
Ulke Kodu (iso)	548	40	0											
Ulke Adi	548	40	0											
Sehir	548	520	0											
Ulke Nufusu	548	40	0	77741	83517045	38189469	29641624.08	8591365	37887768	67530172	65129728	83439304	9.8477E+14	31381052.87
Sehir Nufusu	548	520	0	259	15067724	621450.9	751469.1504	23232.5	139928.5	540918	12346478	15067465	2.91064E+12	1706058.535
Sehirlerin Nufus Orani	548	520	0	1	322459.6	2516.47	3853.486095	51.65923	174.7801	785.5034	5.46959	322458.6	254002603.1	15937.45912
Havalimani Kategorisi	548	3	0											
Havalimani Kategorisi (Num)	548	3	0	1	3	2.178832	0.434559646	2	2	3	2	2	0.318966092	0.564770832
Rakim Ft	548	387	0	-15	6400	580.3869	599.2578454	51.75	235.5	644.25	7	6415	842115.4698	917.6684967
Tarifeli Ucus Bilgisi	548	2	0	0	1	0.930657	0.12906921	1	1	1	1	1	0.064652584	0.254268724
Pist Uzunlugu Ft	548	358	0	1574	13829	7932.325	2031.966527	6562	8166.5	9842	8202	12255	6627095.116	2574.314494
Pist Genisligi Ft	548	39	0	49	262	141.4799	20.7659505	147	148	148	148	213	887.6357938	29.79321724
Pist Yuzeyi	548	7	0											
Pist Yuzeyi Num	548	5	0	1	5	4.587591	0.647210826	5	5	5	5	4	0.652277185	0.807636791
Rota Sayisi	548	127	0	0	357	31.7573	36.02415552	2	7.5	39	2	357	2685.928192	51.8259413
Yuk Posta 2018	548	304	0	0	2175747	38644.07	66009.81116	0	73.5	2577.5	0	2175747	40516417036	201286.9023
Direk Gecis 2018	548	400	0	0	250101	9093.71	11527.9912	6.5	995	8395.5	0	250101	449604401	21203.87703
Tum Ucak Hareketleri 2018	548	547	0	54	510958	38320.76	40458.88125	4022.5	12471.5	38863.5	2232	510904	5062445526	71150.86455
Yolcu Ucagi Hareketleri 2018	548	478	0	0	460777	20233.19	27209.29381	710.5	2477	12000.75	0	460777	2676034063	51730.39787
Tasinan Yolcu 2018	548	548	0	6	80016117	3834136	5163193.15	85487.25	401182	2477330	{6,7,10,15,	80016111	9.67735E+13	9837351.943

**Tablo 16: Tahminleme Veri Seti Hakkında Özet İstatistikler**



Şekil 49: Tahminleme Modeli



Şekil 50: Tahminlemede Kullanılacak Veri Seti için Belirleyici Sütunların Seçimi

EU Airports Data Reg R1 B1-5 Finished running ✓

EU Airports Data Reg R1.csv Draft saved at 9:26:16 PM

Select Columns in Dataset ✓

Normalize Data ✓

Normalize Data ✓

Edit Metadata ✓

Edit Metadata ✓

Linear Regression ✓

Bayesian Linear Regression ✓

Decision Forest Regression ✓

Boosted Decision Tree Regr... ✓

Split Data ✓

Split Data ✓

Split Data ✓

Split Data ✓

Train Model ✓

Train Model ✓

Train Model ✓

Train Model ✓

Score Model ✓

Score Model ✓

Score Model ✓

Score Model ✓

Evaluate Model ✓

Evaluate Model ✓

Evaluate Model ✓

Evaluate Model ✓

Execute R Script ✓

Execute R Script ✓

Execute R Script ✓

Execute R Script ✓

Add Rows ✓

Add Rows ✓

Add Rows ✓

Add Rows ✓

Properties Project

Linear Regression

Solution method  
Ordinary Least Squares

L2 regularization weight  
0.001

Include intercept term

Random number seed  
12345

Allow unknown categorical levels

START TIME 2/27/2020 9:00:52 PM

END TIME 2/27/2020 9:00:52 PM

ELAPSED TIME 0:00:00.000

STATUS CODE Finished

STATUS DETAILS Task output was present in output cache

Quick Help

Şekil 51: Linear Regresyon Algoritması için Teknik Parametreler

EU Airports Data Reg R1 B1-5 In draft

Draft saved at 1:53:39 PM

Properties Project

Bayesian Linear Regression

Regularization weight

Allow unknown categorical levels

START TIME 3/4/2020 1:42:56 PM

END TIME 3/4/2020 1:42:56 PM

ELAPSED TIME 0:00:00.000

STATUS CODE Finished

STATUS DETAILS Task output was present in output cache

Quick Help

Şekil 52: Bayesyen Lineer Regresyon Algoritması için Teknik Parametreler

Karar Ormanı algoritması Azure ML Studio varsayılan ayarları uygulanmıştır. Platformdan alınan görsel Şekil 53'te yer almaktadır.

Artırılmış Karar Ağacı Algoritması Azure ML Studio varsayılan ayarları seçilmiştir. Rastgele sayı besleme seçeneği verileri her defasında aynı şekilde bölmek ve ard arda yapılan deney sonuçlarını karşılaştırabilmek amacı ile "12345" seçilmiştir. Platformdan alınan görsel Şekil 54'te yer almaktadır.

Poisson Regresyon Algoritması Azure ML Studio varsayılan ayarları seçilmiş olup yine aynı şekilde "12345" rastgele sayısı seçilmiştir. Platformdan alınan görsel Şekil 55'te yer almaktadır.

Veri setinin %70'i modeli eğitmek için kullanılırken kalan %30'u tahmin skorları üretmek ve ortaya çıkan sonuçların doğruluğunu test etmek amacı ile ayrılmaktadır. Bu oran (80-20), (60-40) gibi farklı oranlarda yapılabilmektedir. 5 farklı algoritmanın ürettiği tahmin skorlarını aynı havalimanları ile karşılaştırabilmek ve her seferinde aynı bölümlenmeyi elde edebilmek amacı ile rastgele besleme sayısı "12345" seçilmiştir. Eğitim veri setinin ayrılmasına ilişkin görsel Şekil 56'da yer almaktadır.

Eğitim Modülüne Taşınan Yolcu Sayısı Sütununun Tanımlanmasına ilişkin görsel ise Şekil 57'de yer almaktadır.

### **5.2.3 Taşınan Yolcu Sayısı Tahmini için Sonuçlar**

Veri setinin %70'i kullanılarak eğitilen model, kalan %30'luk işlem yapılmayan veri seti üzerinde test edilmiştir. Veri setinin %30'luk kısmı 164 satıra yani 164 havalimanına tekabül etmektedir. Bu 164 havalimanına ait tahmin skorlarını da içeren ilk birkaç satırı gösteren sonuçların Azure ML Studio Ekran görüntüsü (Lineer Regresyon Algoritması ile oluşturulan sonuçlar) Şekil 58'de yer almaktadır.

Şekil 59-68 görsellerinde yer alan sonuçlar tüm algoritmalar için ayrı ayrı üretilen sonuçların grafikleridir.



EU Airports Data Reg R1 B1-5 Finished running ✓

Draft saved at 9:26:16 PM

Properties Project

Decision Forest Regression

Resampling method: Bagging

Create trainer mode: Single Parameter

Number of decision trees: 8

Maximum depth of the decision trees: 32

Number of random splits per node: 128

Minimum number of samples per leaf node: 1

Allow unknown values for categorical features

START TIME: 2/27/2020 9:00:52 PM

END TIME: 2/27/2020 9:00:52 PM

ELAPSED TIME: 0:00:00.000

STATUS CODE: Finished

Quick Help

Şekil 53: Karar Ormanı Algoritması için Teknik Parametreler

EU Airports Data Reg R1 B1-5 Finished running ✓

Draft saved at 9:26:16 PM

**Properties Project**

**Boosted Decision Tree Regression**

Create trainer mode  
Single Parameter

Maximum number of leaves per tree  
20

Minimum number of samples per leaf node  
10

Learning rate  
0.2

Total number of trees constructed  
100

Random number seed  
12345

Allow unknown categorical levels

START TIME 2/27/2020 9:00:53 PM  
END TIME 2/27/2020 9:00:53 PM  
ELAPSED TIME 0:00:00.000  
STATUS CODE Finished

Quick Help

Şekil 54: Artırılmış Karar Ağacı Regresyonu Algoritması için Teknik Parametreler

EU Airports Data Reg R1 B1-5 Finished running ✓

Draft saved at 9:26:16 PM

Properties Project

▲ Poisson Regression

Create trainer mode  
Single Parameter

Optimization tolerance  
1E-07

L1 regularization weight  
1

L2 regularization weight  
1

Memory size for L-BFGS  
20

Random number seed  
12345

Allow unknown categorical levels

START TIME 2/27/2020 9:00:53 PM  
END TIME 2/27/2020 9:00:53 PM  
ELAPSED TIME 0:00:00.000  
STATUS CODE Finished

Quick Help

Şekil 55: Poisson Regresyon Algoritması için Teknik Parametreler

EU Airports Data Reg R1 B1-5 Finished running ✓

Draft saved at 9:26:16 PM

**Properties Project**

Split Data

Splitting mode  
Split Rows

Fraction of rows in the first output dataset  
0.7

Randomized split

Random seed  
12345

Stratified split  
False

START TIME 2/27/2020 9:00:56 PM  
END TIME 2/27/2020 9:00:56 PM  
ELAPSED TIME 0:00:00.000  
STATUS CODE Finished  
STATUS DETAILS Task output was present in output cache

Quick Help

Şekil 56: Eğitim Veri Setinin Ayrılması

EU Airports Data Reg R1 B1-5 Finished running ✓

Properties Project

Train Model

Label column

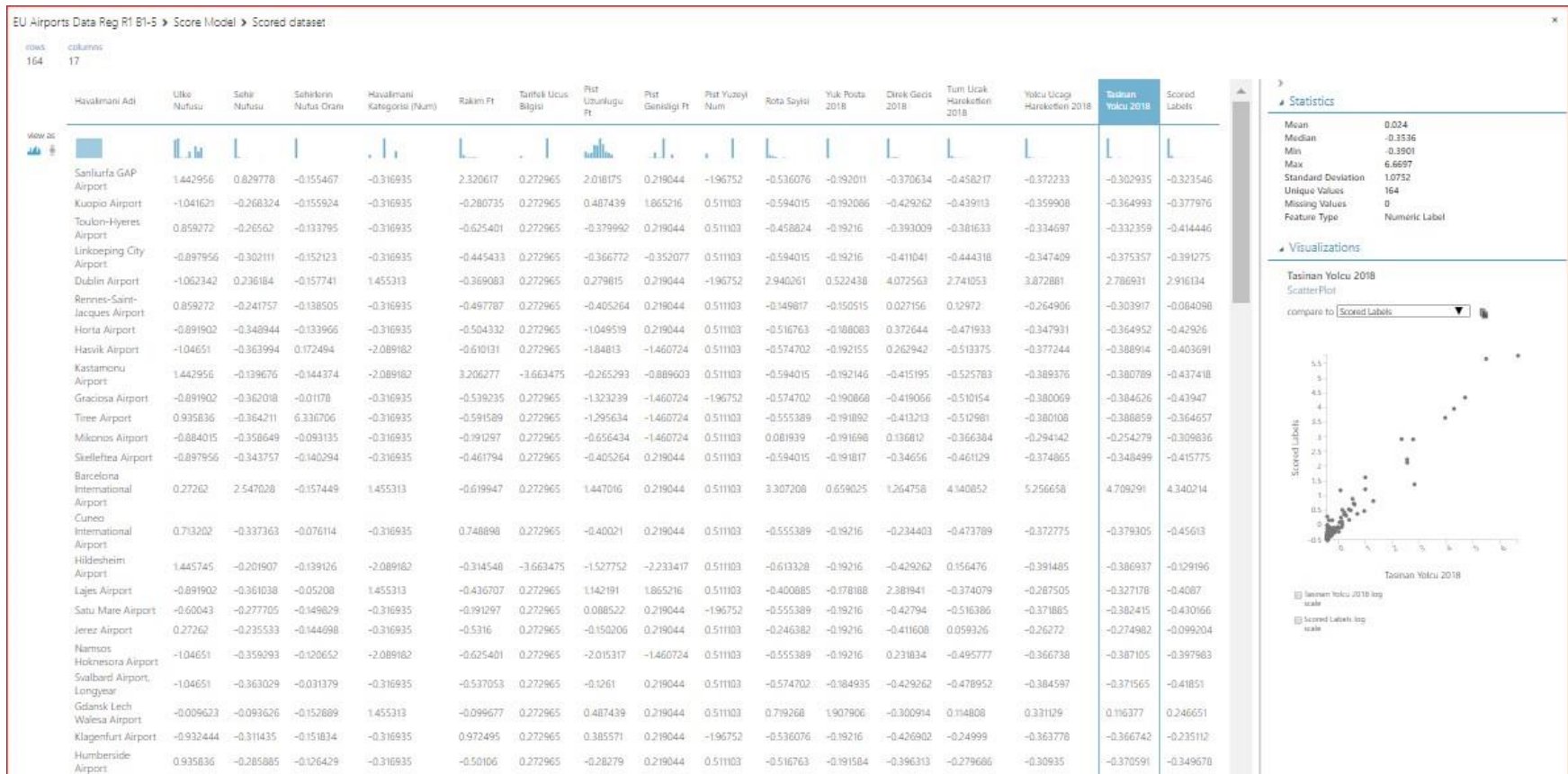
**Selected columns:**  
Column names: Tasinan Yolcu  
2018

Launch column selector

START TIME 2/27/2020 9:00:58 PM  
END TIME 2/27/2020 9:00:58 PM  
ELAPSED TIME 0:00:00.000  
STATUS CODE Finished  
STATUS DETAILS Task output was present in output cache

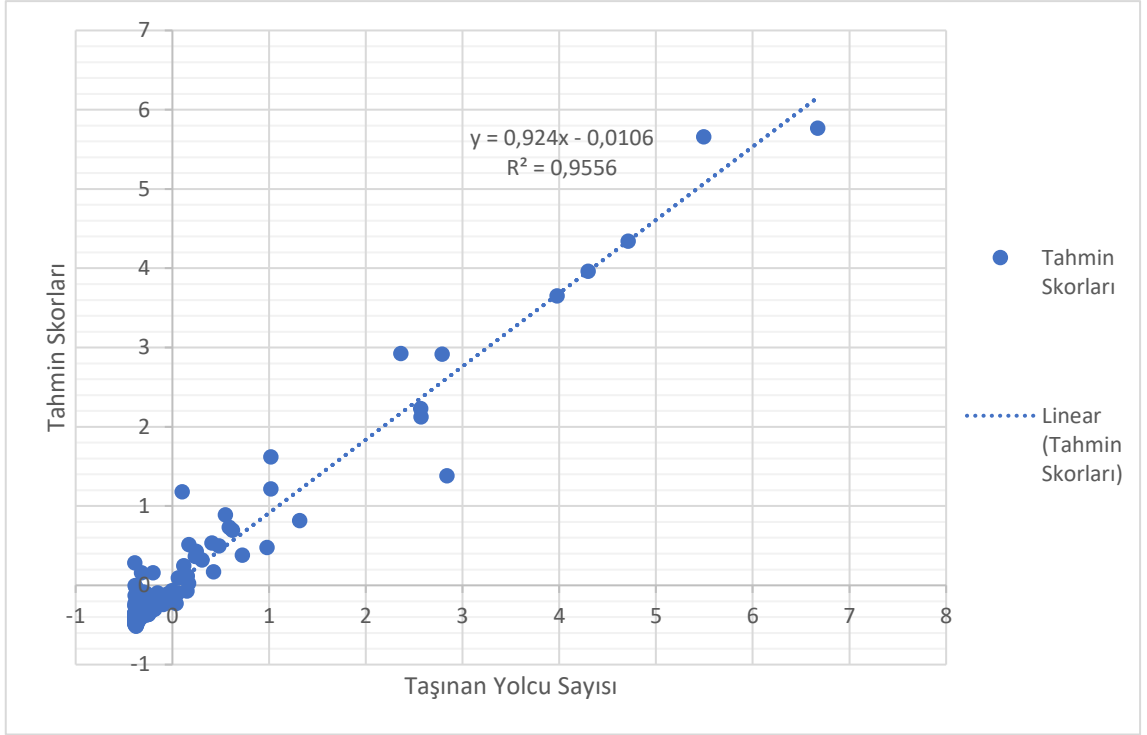
Quick Help

Şekil 57: Eğitim Modülüne Taşınan Yolcu Sayısı Sütununun Tanımlanması

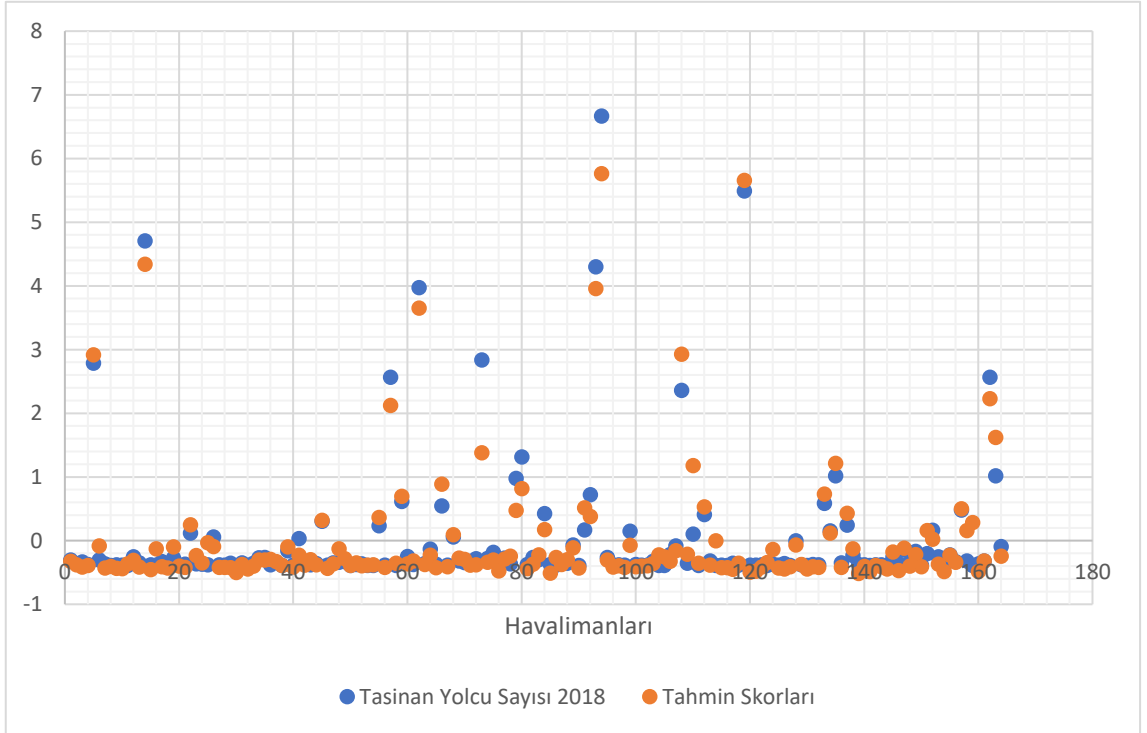


Şekil 58: Azure ML Studio Lineer Regresyon Tahmin Skorları

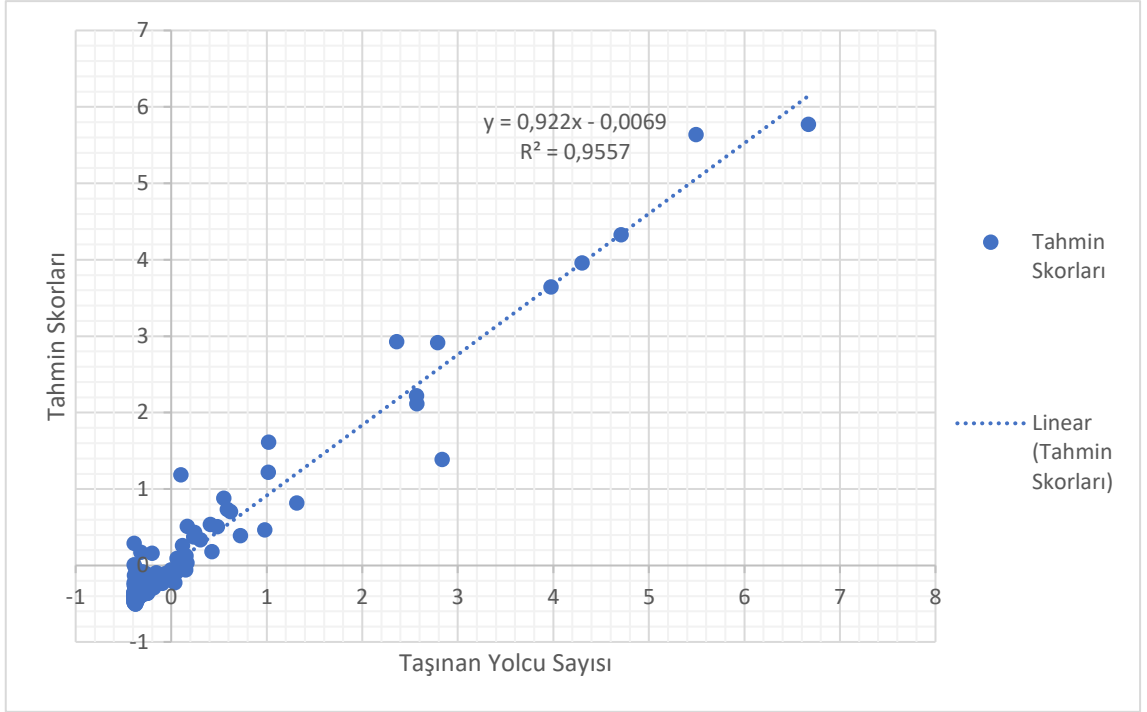




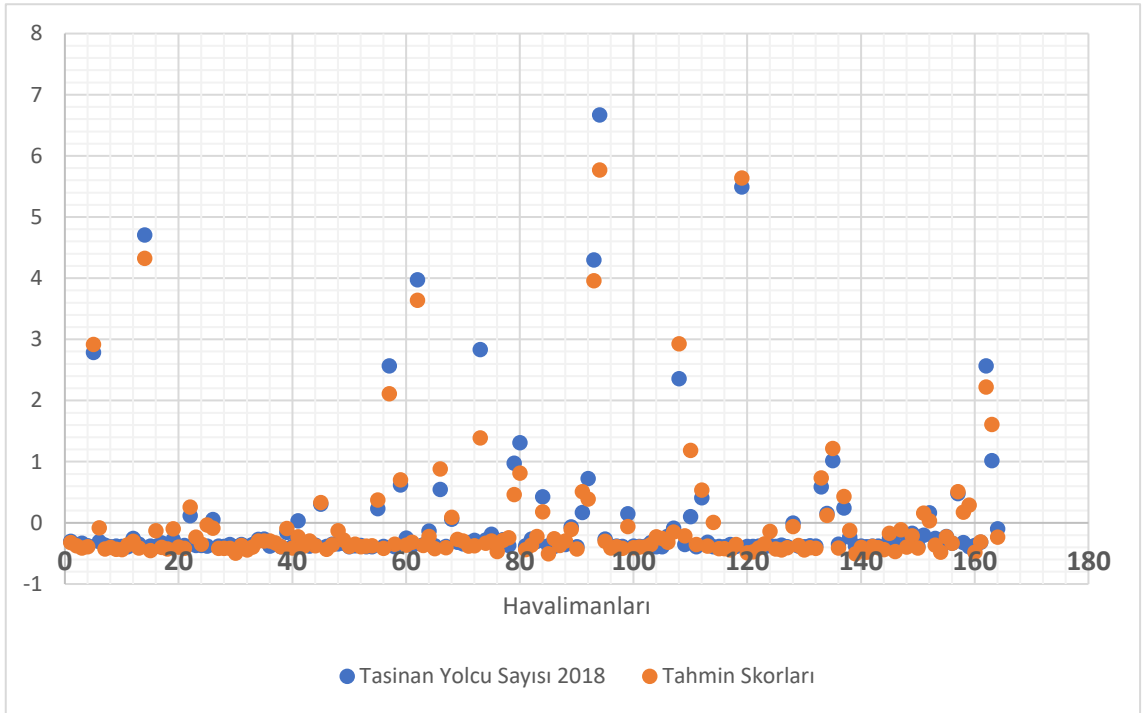
Şekil 59: Lineer Regresyon Algoritması Sonuç Grafiği 1



Şekil 60: Lineer Regresyon Algoritması Sonuç Grafiği 2

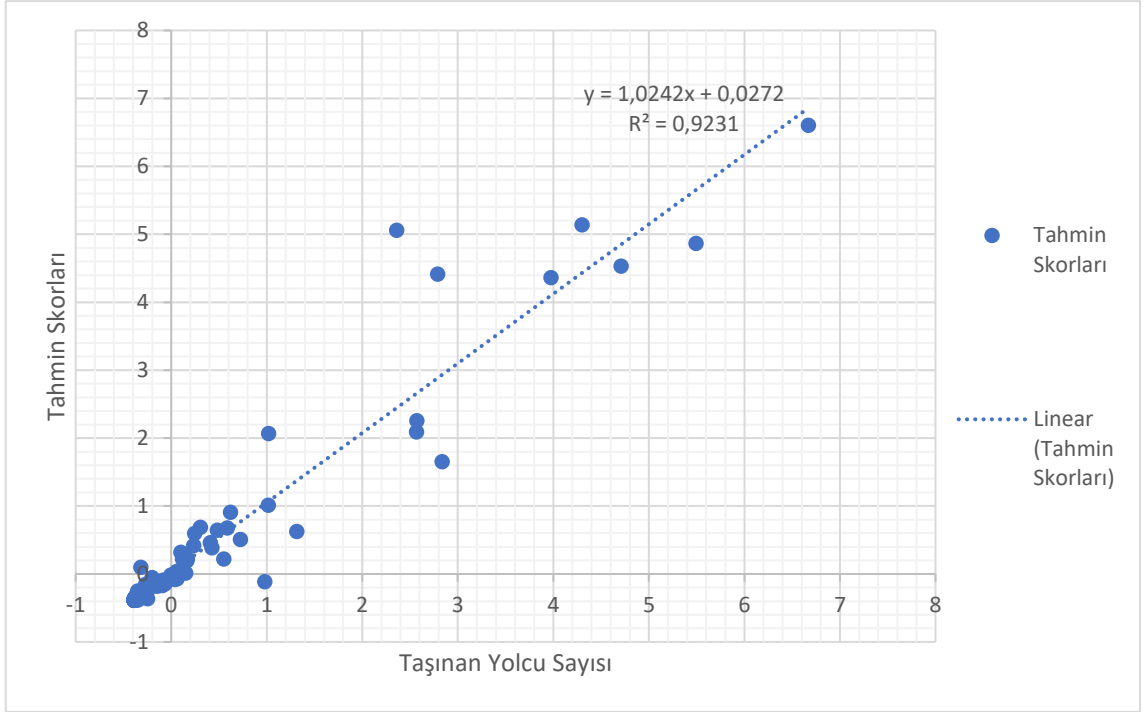


Şekil 61: Bayesyen Linear Regresyon Algoritması Sonuç Grafiği 1

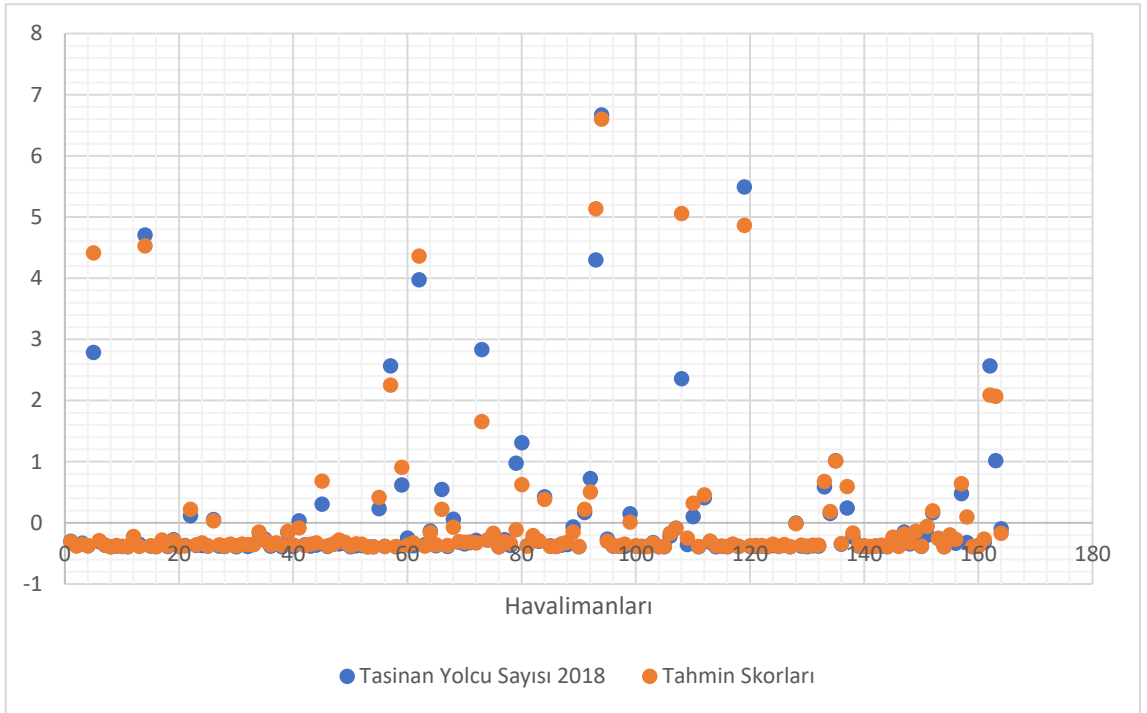


Şekil 62: Bayesyen Linear Regresyon Algoritması Sonuç Grafiği 2

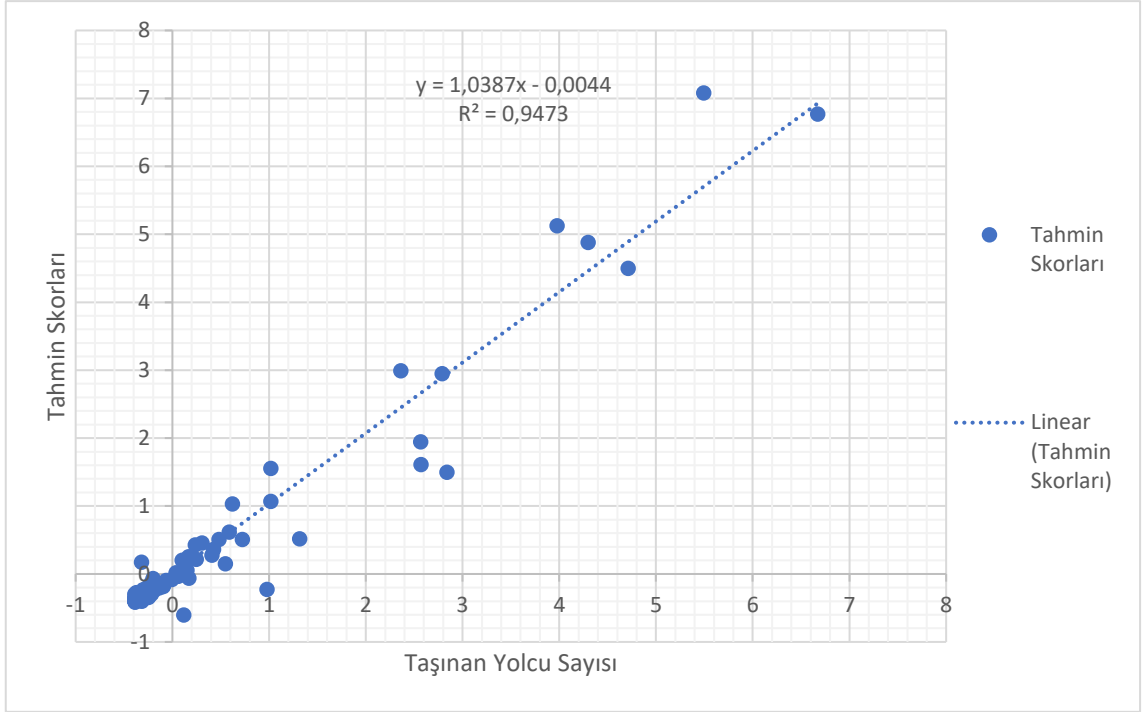




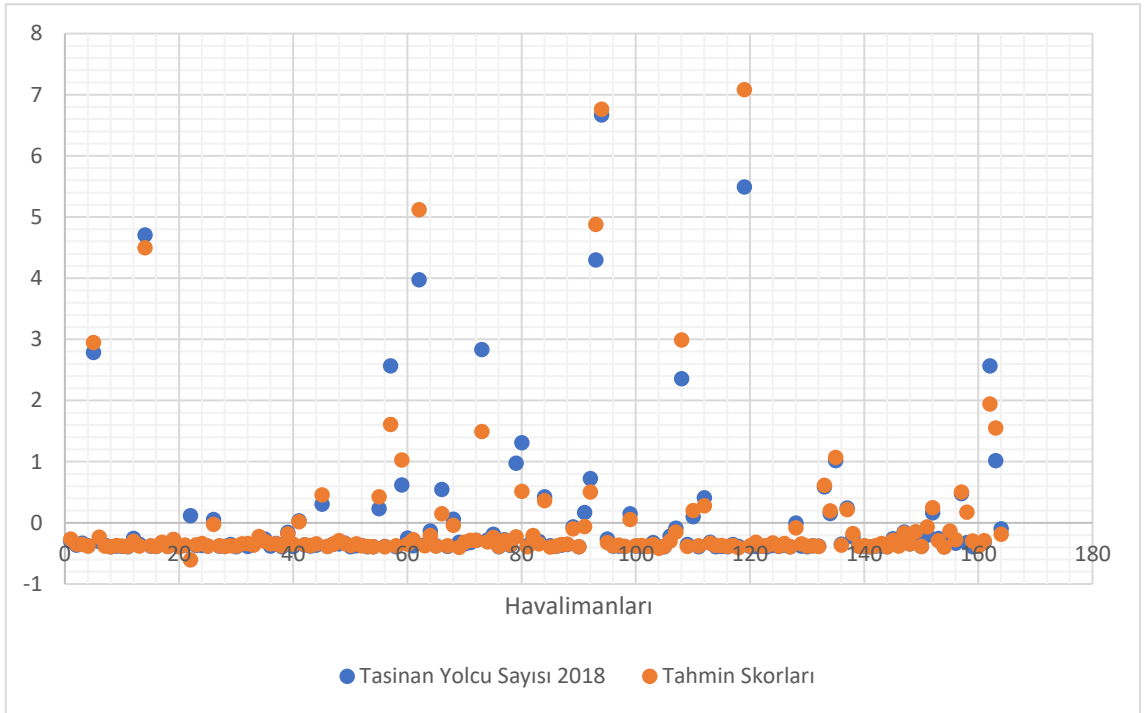
Şekil 63: Karar Ormanı Algoritması Sonuç Grafiği 1



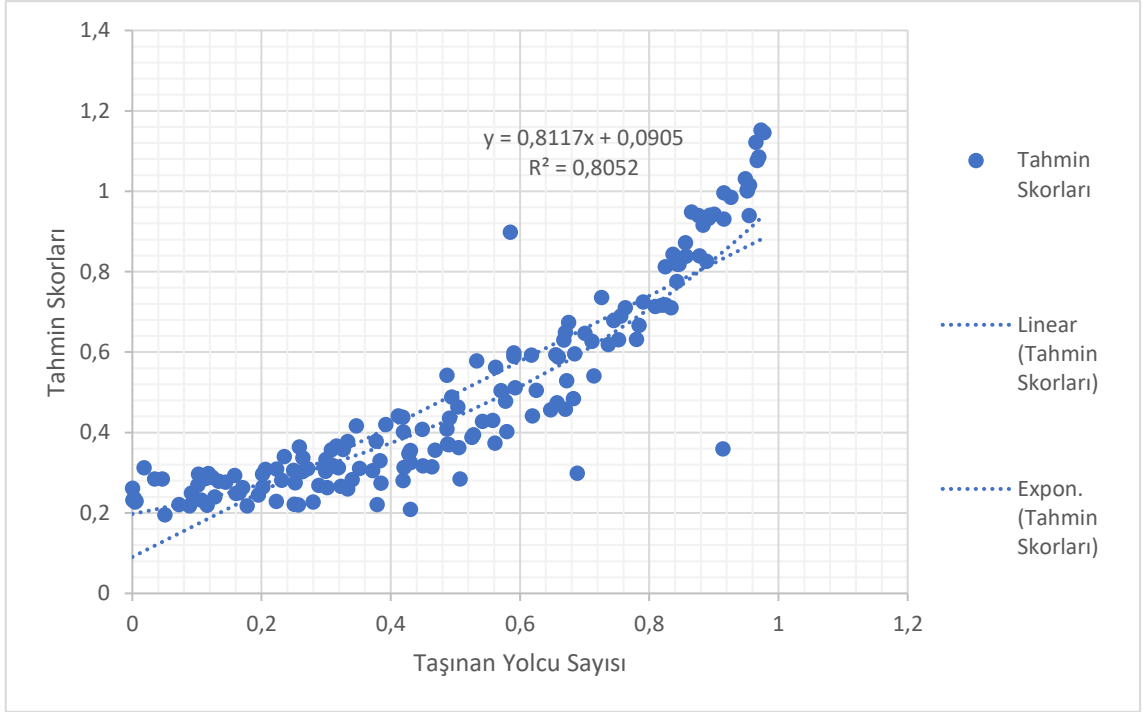
Şekil 64: Karar Ormanı Algoritması Sonuç Grafiği 2



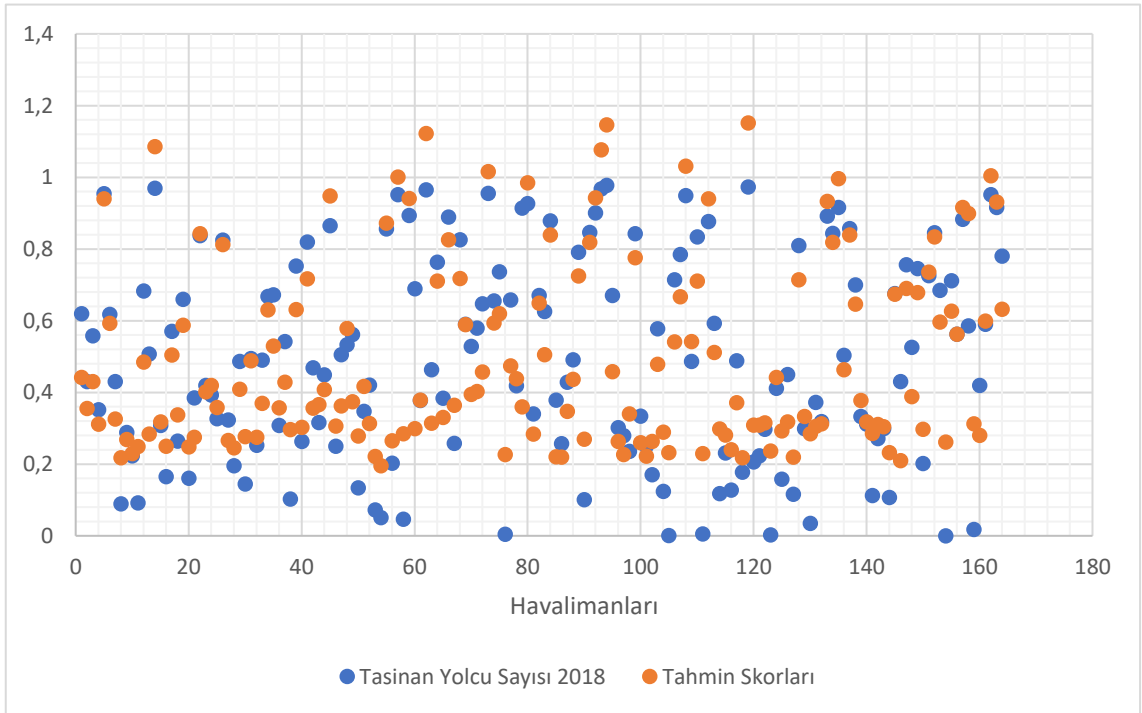
Şekil 65: Artırılmış Karar Ağacı Algoritması Sonuç Grafiği 1



Şekil 66: Artırılmış Karar Ağacı Algoritması Sonuç Grafiği 2



Şekil 67: Poisson Regresyon Sonuç Grafiği 1



Şekil 68: Poisson Regresyon Sonuç Grafiği 2

#### 5.2.4 Taşınan Yolcu Sayısı için Sonuçların Değerlendirilmesi

560 satırlık veri setinden İstanbul havalimanı gibi yeni açılması sebebi ile verileri tam olmayan veya çeşitli nedenlerle eksik ya da hatalı veriler içeren 12 havalimanının çıkartılması ile 548 satırlık veri seti derlenmiştir. Bu veri seti de %70'i eğitim verisi ve %30'u test verisi olarak ikiye ayrılmıştır. 5 farklı makine öğrenmesi algoritması eğitim verisi üzerinde çalıştırılmış ve 5 farklı makine öğrenmesi modeli ortaya konulmuştur. Modellerin her biri 548 satırlık ana veri setinden ayrılan aynı satırları içeren test veri seti üzerinde tahmin skorları üretmiştir. Lineer Regresyon Algoritması için “score labels” etiketi ile tahmin sonuçlarını ve “taşınan yolcu sayısı 2018” etiketi ile gerçek verileri de kapsayan test veri setinin bir kısmı Azure ML Studio platformundan alınan Şekil 58'deki ekran görüntüsünde paylaşılmıştır. Üretilen skorlar için 4 hata metriği ve belirlilik katsayısı hesaplanmıştır. Hesaplama sonuçları Tablo 17'de gösterilmiştir.

4 farklı hata metriği için hesaplanan değerler ne kadar düşük ise ve belirlilik katsayısı 1'e ne kadar yakın ise sonuçların doğruluğu o kadar artmaktadır. Bu 5 farklı metrik hesaplanan değerler itibari ile tek başına yeterli olmayıp birlikte değerlendirildiği zaman anlamlı olmaktadır. Lineer Regresyon ve Bayesyen Lineer Regresyon algoritmaları 0.95 ile en iyi belirlilik katsayısına Karar Ormanı ve Artırılmış Karar Ağacı Algoritmaları ise en iyi hata değerlerine sahip algoritmalar olarak öne çıkmaktadır.

Sonuçların daha iyi değerlendirilebilmesi açısından her bir algoritmanın ürettiği tahmin skorları ve gerçek değerler iki farklı grafik ile bir önceki tahmin sonuçları başlığı altında paylaşılmıştı. Lineer Regresyon ve Bayesyen Lineer Regresyon sonuçları birbirlerine neredeyse eşit değerler olarak karşımıza çıkmaktadır. Diğer algoritmalar arasından da en iyi skorları Artırılmış Karar Ağacı Regresyonu sunmaktadır. Bu algoritmaya ait Şekil 65'te yer alan grafik, tahmin skorlarının regresyon eğrisine Lineer Regresyon ve Bayesyen Lineer Regresyon algoritmaları hariç diğer algoritmalarından daha yakın sonuçlar ürettiğini gözler önüne sermektedir. Bu grafiğin, x eksenini doğrultusunda -1,2 aralığında bakıldığında Şekil 59 ve 61 Lineer Regresyon ve Bayesyen Lineer Regresyon grafiklerinden 3 veri noktası hariç tutulursa daha iyi sonuçlar ürettiği görülebilir. Fakat 2,5 aralığı incelendiğinde Lineer Regresyon ve Bayesyen Lineer Regresyon sonuçlarında regresyon eğrisi ile doğrudan çakışan noktalar bulunmaktadır.

Ayrıca Lineer Regresyon ve Bayesyen Lineer Regresyon sonuçları 2 nokta hariç regresyon eğrisine daha yakın oturan sonuçlar sergilemektedir. Ortalama Mutlak Hata metrikleri açısından Lineer Regresyon 0,128 puana, Bayesyen Lineer Regresyon 0,126 puana ve Artırılmış Karar Ağacı Regresyonu 0,101 puana sahiptir. Artırılmış Karar Ağacı 0,027 puan daha iyi bir değere sahiptir. Fakat Lineer Regresyon ve Bayesyen Lineer Regresyon algoritmaları açısından bağıl kare hatası 0,046 puan ile diğer algoritmalarından daha başarılıdır. Değerlendirilen tüm parametreler Bayesyen Lineer Regresyon algoritmasının daha iyi sonuçlar ürettiğini göstermektedir.

Sonuçlar değerlendirilirken ortaya çıkan en büyük zorluk çok büyük ve çok küçük havalimanlarını içeren bir veri setinde var olan sayısal değerlerin de birbirinden çok farklı olmasıdır. Örnek verecek olursak Akdeniz’de küçük bir adada yer alan bir havalimanı için taşınan yolcu sayısı yıllık 1000 rakamını bile bulmazken İstanbul, Frankfurt, Londra gibi şehirlerde bu rakam milyonlarla ifade edilmektedir. Öte yandan bu havalimanlarına ait pist uzunluğu ya da taşınan kargo sayısı gibi bazı özelliklerde birbirleri ile aynı veya çok yakın değerler alabilmektedir. Bu durum farklı birimleri aynı ölçekte değerlendirebilmek açısından gerçekleştirilen normalizasyon işlemlerinde sayısal sapmalara neden olabilmektedir. Yolcu sayısı tahmini uygulamasında bu durumun göz önünde bulundurulması önemlidir.

Algoritmalar	Ortalama Mutlak Hata (MAE)	Kök Ortalama Kare Hatası (RMSE)	Bağıl Mutlak Hata (RAE)	Bağıl Kare Hatası (RSE)	Belirlilik Katsayısı (CoF)
Lineer Regresyon	0.128	0.229	0.221	0.046	0.954
Bayesyen Lineer Regresyon	0.126	0.229	0.218	0.046	0.954
Karar Ormanı Regresyonu	0.105	0.319	0.181	0.089	0.911
Artırılmış Karar Ağacı Regresyonu	0.101	0.266	0.175	0.062	0.938
Poisson Regresyon	0.096	0.125	0.396	0.195	0.805

**Tablo 17: Taşınan Yolcu Sayısı Sonuç Değerlendirme Metrikleri**

## SONUÇ

Yapay Zekâ alanındaki gelişmeler son yıllarda çok hızlı bir biçimde gerçekleşmiştir. Öyle ki tek başına bir bilim dalı haline gelmiş ve dallara ayrılmıştır. Bu hızlı gelişmenin etkisi ile veya bu kadar popüler bir konuda birçok farklı görüşün hararetli tartışmalara ve son derece yenilikçi gelişmelere sahne olması sebebi ile olsa gerek yapay zekâ bileşenlerinin taksonomisi ile ilgili önemli çalışmalar yapılmış fakat henüz bir fikir birliğine varılmamış olduğu görülmektedir. Bu durumu; yapay zekâ alanına her geçen gün yeni bir konunun girmesi, sınıflandırmada sürekli bir değişim ve dönüşüm gerçekleşmesi oldukça zorlaştırmaktadır. Bu sebeple bu çalışmada dijital dönüşüm ve yapay zekâ konularında bütüncül bir bakış açısı sergileyerek yapay zekâ'nın makine öğrenmesi algoritmaları üzerine araştırma yapılmış ve havalimanları analizinde bu algoritmalar ile 2 farklı sınıfta (Gözetmesiz ve Gözetmeli Öğrenme) makine öğrenmesi tekniği kullanılarak yapılan uygulamalarda literatüre katkı sağlamak amaçlanmıştır.

Kümeleme uygulaması benzer havalimanlarını tespit etmek amacı ile gerçekleştirilmiştir. Bu sayede artan havacılık faaliyetleri çerçevesinde gelecekte yapılacak her türlü havacılık organizasyonuna bir çıkarım ve öngörü sağlanabilir. Nitekim hem Avrupa havalimanlarını hem de münferit bir ülkeyi kapsayan birçok benzer çalışma bulunmaktadır.

Robert Mayer 2016 yılında hiyerarşik kümeleme yöntemini kullanarak dünya genelinde kargo karakteristikleri baz alınan 114 havalimanından oluşan veri seti üzerinde çalışmıştır. Bu çalışmada, Kuzey Amerika ve Avrupa gibi bölge havalimanları benzersiz özelliklerine göre karakterize edilmiştir. Kargoculuk yapan havayolları ağlarında, diğer havalimanları yolcu operasyonlarının önemli bir kapasitesinin sonucu olarak yüksek kargo hacmi üretilirken hava taşımacılığına yüksek oranda bağımlı havalimanlarının merkezi bir lokasyondan yarar sağlamaya eğilimli olduğu saptanmıştır (Mayer, 2016).

H. Rodriguez ve A. Voltes 2014'te havalimanları verimlilik karşılaştırması için sınır temelli hiyerarşik kümeleme metodu kullanarak dünya çapında 106 havalimanını analiz etmiştir. Çalışmada 17 farklı küme tespit edilmiştir. Büyük uçakların kullanımı ve düşük maliyetli taşımacıların hakimiyeti gibi faktörlerin havalimanı endüstrisindeki maliyet performansını artırdığı gösterilmiştir (Rodríguez-Déniz ve Voltes-Dorta, 2014).

Paolo Malighetti ve arkadaşları tarafından 2009 yılında en az bir tarifeli yolcu taşımacılığına sahip 467 Avrupa havalimanının baz alındığı benzetilmiş tavlama ve kümeleme metotları kullanılarak havalimanları sınıflandırılması üzerine bir çalışma gerçekleştirilmiştir. Kümeleme metodu ile 8 küme ve benzetilmiş tavlama metodu ile 13 modül tespit edilmiştir. Bu çalışmada Avrupa havalimanlarının modüller halinde sınıflandırılması ve kümelenmesi, tanımlanan gruplar ve özellikler açısından aynı gruba ait benzer havalimanları ve farklı gruplara ait benzer havalimanları arasında rekabetin daha ağır olup olmadığını test etmek için bir temel sağlamaktadır (Malighetti vd., 2009).

Bu tezin konusu olan kümeleme uygulamasında ise 40 Avrupa ülkesine ait 560 havalimanı Euclidean ve Cosine olmak üzere iki farklı metot kullanılarak kümelendi. Doğru bir sınıflandırma yapmak ve bu sayede benzer havalimanlarını tespit edebilmek için optimum küme sayısı araştırılmıştır. Python ve R programlama dillerinin kütüphanelerinde tanımlı 30 farklı teknik ile optimum küme sayıları tespit edilmiştir. Optimum küme sayısı 3 olarak saptanmış ve kıyaslama yapabilmek açısından her iki teknik için de 3'lü ve 5'li kümeleme sonuçları paylaşılmıştır. Kümeleme sonuçları farklı ülkelerde birbiri ile benzer karakteristikteki havalimanlarını gözler önüne sermektedir. Kümeleme sonuçları aynı kümelerde gruplanan havalimanlarının taşınan yolcu sayısı, taşınan yük miktarı, şehir nüfusu ve destinasyon sayısı gibi parametrelerde birbirine benzer özellikler sergilediğini göstermektedir. Özellikle taşınan yolcu sayısının, havalimanlarının aynı kümeye atanmasında diğer parametrelerden daha etkili olduğu görülmüştür. Bir diğer yandan Avrupa ülkeleri havalimanları sınıflandırmasında Türkiye'deki havalimanlarının benzer sonuçlar ürettiği saptanmıştır. Türkiye, İspanya, Polonya, İtalya, Almanya, Amerika Birleşik Devletleri, Çin, Endonezya gibi ülke bazlı çalışmalar literatür taramasında incelenmiştir. Bu çalışmalarda Veri Zarflama Metodu, Analitik Hiyerarşi Prosesi, Kümeleme, Temel Bileşen Analizi, Yapay Sinir Ağları Bulanık Yaklaşım gibi birçok metot havalimanı seçimi, verimlilik analizi ve kıyaslama konularının araştırılmasında kullanılmıştır.

Eva Stichhauerova ve Natalie Pelloneova'nın 2019 yılında Almanya'nın en önemli 27 havalimanının performansını değerlendirdiği çalışmada veri zarflama metodu kullanılmıştır. Operasyon yönetiminde en iyi uygulamaları ve uygun süreçleri işledikleri için verilen girdileri verimli çıktılara dönüştürebilen 13 havalimanı bulunmuştur. 5

havalimanı da optimal ve en üretken büyüklüğe ulaşabilen tesisler olarak tanımlanmıştır (Stichhauerova ve Pelloneova, 2019). Ayrıca bu çalışmada diğer ülkeler için yapılan çalışmalara ilişkin son derece kullanışlı bir özet istatistik ve karşılaştırma tablosu paylaşılmıştır.

Selin Yalçın ve Ertuğrul Ayyıldız 2018 yılında Türkiye’de yer alan 55 havalimanına ilişkin kümeleme metodu ile bir vaka çalışması gerçekleştirmişlerdir. Çalışma da kümeleme yöntemi ile havalimanları 6 kümeye atanmıştır. Kümelerin 2 tanesinde sadece 1 eleman bulunmaktadır. Belirlenen kriterlerde yüksek değerlere sahip olan 5 büyük havalimanı (İstanbul Atatürk, İstanbul Sabiha Gökçen, Ankara Esenboğa, İzmir Adnan Menderes, Antalya) hariç tutulmuştur (Ayyildiz ve Yalcin, 2018).

Tez’de kümeleme uygulaması sonunda diğer ülkelerden farklı olarak Türkiye havalimanları ayrıca değerlendirilmiştir. Türkiye de yer alan 55 havalimanı tüm Avrupa havalimanlarıyla beraber değerlendirilmiş ve 3’lü ve 5’li kümelere atanan elemanlar Tablo 13 ve 14’te paylaşılmıştır. Kümeler incelendiğinde havalimanlarının Selin Yalçın ve Ertuğrul Ayyıldız’ın çalışmasına benzer şekilde kümelendiği görülebilir. Fakat Avrupa havalimanları ile yeterince büyük bir veri setine sahip olmanın avantajı kullanılarak tüm havalimanları kümeleme uygulamasına dahil edilebilmiştir. Bu şekilde Türkiye’de yer alan İstanbul Atatürk, İstanbul Sabiha Gökçen, Ankara Esenboğa, İzmir Adnan Menderes, Antalya gibi havalimanları hem kendi içinde hem de Avrupa’da yer alan Londra Heathrow, Frankfurt, Madrid, Barcelona gibi büyük havalimanları ile mukayese edilebilmiştir. Sonuçlar Avrupa çapında havalimanları sınıflandırmasında Türkiye havalimanlarının birbirine yakın değerlerde olduğunu göstermektedir.

Taşınan yolcu sayısı tahmini uygulaması, havacılık kuruluşlarının ve havayolları şirketlerinin son derece ciddi gelecek planları hazırlamak zorunda olduğu bir alanda taşınan yolcu sayısındaki trendi görebilmenin bir o kadar önem arz etmesi sebebi ile gerçekleştirilmiştir. Bu uygulamada lineer regresyon ve regresyona dayalı ağaç yapısı algoritmaları kullanılmıştır.

Martin Dziedzic ve arkadaşları 2020 yılında 21 Avrupa ülkesini kapsayan 146 küçük bölgesel havalimanını çoklu lineer regresyon algoritması ve korelasyon analizi yöntemlerini kullanarak hava trafik hacmi ve yapısının belirlenmesi üzerine çalışmıştır.



Avrupa’da yer alan küçük bölgesel havalimanlarının yolcu sayısındaki kapasite sınırlamaları ve havalimanı ücretleri konusunda hava ulaşımı pazarının kompleks bir parçasını oluşturduğu ve nüfus büyüklüğünün trafik hacmini doğrudan etkilediği bulunmuştur (Dziedzic vd., 2020).

Bu tez de yer alan taşınan yolcu sayısı uygulamasında ise; İstanbul havalimanı gibi yeni açılan havalimanları ve tarifeli uçuşa sahip olmadığı için veri tabanlarında taşınan yolcu sayısının “0” olarak belirtildiği 12 havalimanı hesaplamaları bozmaması için hariç tutulmuştur. Böylece 548 havalimanı için 5 farklı makine öğrenmesi algoritması kullanılarak taşınan yolcu sayısı tahmin edilmesi ve en doğru trend çizgisinin yakalanması hedeflenmiştir. Ayrıca belirtilen 5 farklı makine öğrenmesi algoritmasına ilişkin bir performans değerlendirmesi öngörülmüştür. Sonuçlar incelendiğinde 5 algoritmadan 3’ünün (Lineer Regresyon, Bayesyen Lineer Regresyon ve Artırılmış Karar Ağacı Regresyonu) hata metrikleri açısından tatmin edici değerler ortaya koyduğu gözlemlenmiştir. Tüm değerler göz önüne alındığında Bayesyen Lineer Regresyon Algoritmasının daha etkili bir performans sergilediği görülmüştür. Tüm kısıtlara rağmen ve bir zaman çizelgesine bağlı olmaksızın Taşınan Yük Sayısı, Rota Sayısı, Şehir ve Ülke Nüfusu, Havalimanı Kapasitesi ve diğer teknik özellikler kullanılarak taşınan yolcu sayısına ilişkin trendi görebilmek açısından bir öngörü sağlanabileceği ortaya konulmuştur.

Tüm bilimsel çalışmalarda olduğu gibi farklı yaklaşımlardan sentez çıkarabilmek için bu çalışmaya da alternatif çalışmalar hazırlanarak yorumlanması ve desteklenmesi olağan bir süreçtir. Bu kapsamda literatüre yeni eklenen algoritmaların incelenmesi, havalimanları analizinde kullanılabilecek yeni özellikler içeren veri setlerinin güncellenmesi ve veri setinin Avrupa ölçeğinden dünyadaki tüm havalimanlarını kapsayacak şekilde genişletilmesi ve bir uygulama üzerinde çalışma süreçlerinin ve ortaya çıkan sonuçların yorumlanması sonraki çalışmalar için öneri olabilir. Ayrıca Derin Öğrenme konusu ile gelişen “Doğal Dil Anlama”, “Doğal Dil İşleme” ve “Yapay Sinir Ağları” gibi konularda kapsamlı araştırmalar yapılabilir.

## KAYNAKÇA

- ACI. (2019). Data Centre - Airports Council International (ACI) World. <https://aci.aero/data-centre/>. 14 Şubat 2020 tarihinde <https://aci.aero/data-centre/> adresinden erişildi.
- Acungil, M. (2018). *24 Soruda Dijital Dönüşüm*. İstanbul: Tuti Kitap. 8 Ocak 2019 tarihinde <https://tuti.com.tr/kitap/24-soruda-dijital-dönüşüm/> adresinden erişildi.
- Ansari, F., Erol, S., ve Sihn, W. (2018). Rethinking Human-Machine Learning in Industry 4.0: How Does the Paradigm Shift Treat the Role of Human Learning? *Procedia Manufacturing*, 23(2017), ss. 117–122. Elsevier B.V.
- Awwalu, J., Ghazvini, A., ve Abu Bakar, A. (2014). Performance Comparison of Data Mining Algorithms: A Case Study on Car Evaluation Dataset. *International Journal of Computer Trends and Technology*, 13(2), ss. 78–82. 6 Kasım 2019 tarihinde <http://www.ijcttjournal.org> adresinden erişildi.
- Ayyildiz, E., ve Yalcin, S. (2018). Analysis of airports using clustering methods: case study in Turkey. *Pressacademia*, 5(3), ss. 194–205. 17 Aralık 2019 tarihinde <http://doi.org/10.17261/Pressacademia.2018.963> adresinden erişildi.
- Azizi Oroumieh, M. A., Mohammad Bagher Malaek, S., Ashrafizaadeh, M., ve Mahmoud Taheri, S. (2013). Aircraft design cycle time reduction using artificial intelligence. *Aerospace Science and Technology*, 26(1), ss. 244–258. Elsevier Masson SAS.
- Bahrin, M., Othman, M., Azli, N., Teknologi, M. T.-J., ve 2016, U. (2016). Industry 4.0: A review on industrial automation and robotic. *researchgate.net*. 11 Ocak 2019 tarihinde [https://www.researchgate.net/profile/Fauzi\\_Othman/publication/304614356\\_Industry\\_4\\_0\\_A\\_review\\_on\\_industrial\\_automation\\_and\\_robotic/links/57ac15aa08ae3765c3b7bab8.pdf](https://www.researchgate.net/profile/Fauzi_Othman/publication/304614356_Industry_4_0_A_review_on_industrial_automation_and_robotic/links/57ac15aa08ae3765c3b7bab8.pdf) adresinden erişildi.
- Barreto, L., Amaral, A., ve Pereira, T. (2017). Industry 4.0 implications in logistics: an overview. *Procedia Manufacturing*, 13, ss. 1245–1252. Elsevier. 11 Ocak 2019 tarihinde

<https://www.sciencedirect.com/science/article/pii/S2351978917306807?via%3Dihub> adresinden erişildi.

Bengfort, B., Bilbro, R., McIntyre, K., Gray, L., Roman, P., Morris, A., Sharma, S., vd. (2020). Yellowbrick v1.1. 16 Mart 2020 tarihinde <http://www.scikit-yb.org/en/stable/> adresinden erişildi.

Bishop, C. M., ve Tipping, M. E. (2003). *Bayesian Regression and Classification. Nato Science Series sub Series III Computer And Systems Sciences* (C. 190). IOS Press.

Botchkarev, A. (2018a). Evaluating Hospital Case Cost Prediction Models Using Azure Machine Learning Studio. 15 Ekim 2019 tarihinde <http://arxiv.org/abs/1804.01825> adresinden erişildi.

Botchkarev, A. (2018b). Evaluating Performance of Regression Machine Learning Models Using Multiple Error Metrics in Azure Machine Learning Studio. *SSRN Electronic Journal*. Elsevier BV. 15 Nisan 2020 tarihinde <https://www.ssrn.com/abstract=3177507> adresinden erişildi.

Botchkarev, A. (2019). A New Typology Design of Performance Metrics to Measure Errors in Machine Learning Regression Algorithms. *Interdisciplinary Journal of Information, Knowledge, and Management*, 14, ss. 045–076. Informing Science Institute. 15 Nisan 2020 tarihinde <http://arxiv.org/abs/1809.03006> adresinden erişildi.

BTK. (2013). *Bilgi Teknolojileri ve İletişim Kurumu Raporu*. Ankara.

Čerka, P., Grigienė, J., ve Sirbikytė, G. (2017). Is it possible to grant legal personality to artificial intelligence software systems? *Computer Law and Security Review*, 33(5), ss. 685–699.

Chakrabarty, N. (2019). A Data Mining Approach to Flight Arrival Delay Prediction for American Airlines. 22 Kasım 2019 tarihinde <http://arxiv.org/abs/1903.06740> adresinden erişildi.

Chakrabarty, N., ve Biswas, S. (2018). A Statistical Approach to Adult Census Income Level Prediction. *2018 International Conference on Advances in Computing*,

*Communication Control and Networking (ICACCCN)* içinde (ss. 207–212). IEEE. 22 Kasım 2019 tarihinde <https://ieeexplore.ieee.org/document/8748528/> adresinden erişildi.

Chang, Y. C., ve Lee, N. (2010). A Multi-Objective Goal Programming airport selection model for low-cost carriers' networks. *Transportation Research Part E: Logistics and Transportation Review*.

Charrad, M., Ghazzali, N., Boiteau, V., ve Niknafs, A. (2014). NbClust : An R Package for Determining the Relevant Number of Clusters in a Data Set. *Journal of Statistical Software*, 61(6), ss. 1–36. American Statistical Association. 16 Şubat 2020 tarihinde <http://www.jstatsoft.org/v61/i06/> adresinden erişildi.

DHMI. (2019). Devlet Hava Meydanları İşletmesi Genel Müdürlüğü. <https://www.dhmi.gov.tr/Sayfalar/default.aspx>. 14 Şubat 2020 tarihinde <https://www.dhmi.gov.tr/Sayfalar/default.aspx> adresinden erişildi.

Dziedzic, M., Njoya, E. T., Warnock-Smith, D., ve Hubbard, N. (2020). Determinants of air traffic volumes and structure at small European airports. *Research in Transportation Economics*, 79, s. 100749. JAI Press. 20 Nisan 2020 tarihinde <https://linkinghub.elsevier.com/retrieve/pii/S0739885919302616> adresinden erişildi.

Edwards, C., Edwards, A., Stoll, B., Lin, X., Massey, N., ve Edwards, C. (2018). Evaluations of an Artificial Intelligence Instructor's Voice: Social Identity Theory in Human-Robot Interactions. *Computers in Human Behavior*, 90(May 2018), ss. 357–362. Elsevier.

Ekici, O. (2005). *Bayesyen Regresyon ve WinBUGS ile Bir Uygulama*. İstanbul Üniversitesi.

*Encyclopedia of Machine Learning*. (2010). *Encyclopedia of Machine Learning*. Springer US.

Eurostat. (2019). Directorate-General (DG) of the European Commission. <https://ec.europa.eu/eurostat/web/main/home>. 12 Şubat 2020 tarihinde

<https://ec.europa.eu/eurostat/web/main/home> adresinden erişildi.

Exastax. (2017). Yapay Zeka ve Makine Öğrenimi Arasındaki Fark Nedir? 12 Ocak 2019 tarihinde <https://www.exastax.com.tr/makine-ogrenimi/yapay-zeka-ve-makine-ogrenimi-arasindaki-fark-nedir/> adresinden erişildi.

Feng, Z., Hou, Q., Zheng, Y., Ren, W., Ge, J. Y., Li, T., Cheng, C., vd. (2019). Method of artificial intelligence algorithm to improve the automation level of Rietveld refinement. *Computational Materials Science*, 156(October 2018), ss. 310–314.

Filiz, F. (2017). 4.1 Artificial Intelligent Algorithms.

Friedman, J. H. (2002). *Tutorial: Getting Started with MART in R*. 13 Nisan 2020 tarihinde <http://www-stat.stanford.edu/> adresinden erişildi.

Fürnkranz, J., Chan, P. K., Craw, S., Sammut, C., Uther, W., Ratnaparkhi, A., Jin, X., vd. (2011). Mean Squared Error. *Encyclopedia of Machine Learning* (ss. 653–653). Boston, MA: Springer US. 15 Nisan 2020 tarihinde [http://link.springer.com/10.1007/978-0-387-30164-8\\_528](http://link.springer.com/10.1007/978-0-387-30164-8_528) adresinden erişildi.

Goodfellow, I., Yoshua, B., ve Courville, A. (2016). *Deep Learning*. MIT Press. 12 Ocak 2019 tarihinde <http://www.deeplearningbook.org/> adresinden erişildi.

Guo, X., Grushka-Cockayne, Y., ve De Reyck, B. (2018). Forecasting Airport Transfer Passenger Flow Using Real-Time Data and Machine Learning. *SSRN Electronic Journal*. Elsevier BV. 21 Kasım 2019 tarihinde <https://www.ssrn.com/abstract=3245609> adresinden erişildi.

Gürsakal, N. (2017). *Büyük Veri*. Dora Yayınları.

Halpin, B. (2016). *Cluster Analysis Stopping Rules in Stata*. 26 Ocak 2020 tarihinde <http://ulsites.ul.ie/sociology/sites/default/files/wp2016-01.pdf> adresinden erişildi.

Hamet, P., ve Tremblay, J. (2017). Artificial intelligence in medicine. *Metabolism: Clinical and Experimental*, 69, ss. S36–S40. Elsevier Inc.

Hassabis, D., Kumaran, D., Summerfield, C., ve Botvinick, M. (2017). Neuroscience-Inspired Artificial Intelligence. *Neuron*, 95(2), ss. 245–258. Elsevier Inc.

- Hayakawa, S., ve Hayashi, H. (2017). Using Azure Machine Learning for Estimating Indoor Locations. *2017 International Conference on Platform Technology and Service (PlatCon)* içinde (ss. 1–4). IEEE. 15 Ekim 2019 tarihinde <http://ieeexplore.ieee.org/document/7883736/> adresinden erişildi.
- Hogan, M., Liu, F., Sokol, A., ve Tong, J. (2011). *NIST Cloud Computing Standards Roadmap*.
- Houssami, N., Lee, C. I., Buist, D. S. M., ve Tao, D. (2017). Artificial intelligence for breast cancer screening: Opportunity or hype? *Breast*, 36, ss. 31–33. Elsevier Ltd.
- IATA. (2019). IATA - Home. <https://www.iata.org/>. 14 Şubat 2020 tarihinde <https://www.iata.org/> adresinden erişildi.
- ICAO. (2019). International Civil Aviation Organization (ICAO). <https://www.icao.int/Pages/default.aspx>. 14 Şubat 2020 tarihinde <https://www.icao.int/Pages/default.aspx> adresinden erişildi.
- Jain, K. (2015). Machine Learning basics for a newbie. *Analytics Vidhya*. 7 Ocak 2019 tarihinde <https://www.analyticsvidhya.com/blog/2015/06/machine-learning-basics/> adresinden erişildi.
- Janic, M., ve Reggiani, A. (2002). An Application of the Multiple Criteria Decision Making (MCDM) Analysis to the Selection of a New Hub Airport. *European Journal of Transport and Infrastructure Research*, 2, ss. 113–xx.
- Jason, B. (2013). A Tour of Machine Learning Algorithms. 12 Ocak 2019 tarihinde <https://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/> adresinden erişildi.
- Jazdi, N. (2014). Cyber physical systems in the context of Industry 4.0. *2014 IEEE International Conference on Automation, Quality and Testing, Robotics* içinde (ss. 1–4). IEEE. 11 Ocak 2019 tarihinde <http://ieeexplore.ieee.org/document/6857843/> adresinden erişildi.
- Jo, B. W., ve Khan, R. M. A. (2018). An Internet of Things System for Underground Mine Air Quality Pollutant Prediction Based on Azure Machine Learning. *Sensors*,

18(4), s. 930. 15 Ekim 2019 tarihinde <https://www.mdpi.com/1424-8220/18/4/930> adresinden erişildi.

Johnson, K. W., Torres Soto, J., Glicksberg, B. S., Shameer, K., Miotto, R., Ali, M., Ashley, E., vd. (2018). Artificial Intelligence in Cardiology. *Journal of the American College of Cardiology*, 71(23), ss. 2668–2679.

Johnson, M., ve Gu, Y. (2017). Estimating Airport Operations at General Aviation Airports Using the FAA NPIAS Airport Categories. *International Journal of Aviation, Aeronautics, and Aerospace*, 4(1), s. 4. 17 Aralık 2019 tarihinde <https://doi.org/10.15394/ijaaa.2017.1151> adresinden erişildi.

Kagermann, H., Helbig, J., Hellinger, A., ve Wahlster, W. (2013). Recommendations for implementing the strategic initiative INDUSTRIE 4.0: Securing the future of German manufacturing industry; final report of the Industrie. *Google Scholar*. 11 Ocak 2019 tarihinde [https://scholar.google.com/scholar\\_lookup?title=Recommendations for Implementing the Strategic Initiative INDUSTRIE 4.0%3A Securing the Future of German Manufacturing Industry%3B Final Report of the Industrie 4.0](https://scholar.google.com/scholar_lookup?title=Recommendations+for+Implementing+the+Strategic+Initiative+INDUSTRIE+4.0%3A+Securing+the+Future+of+German+Manufacturing+Industry%3B+Final+Report+of+the+Industrie+4.0) adresinden erişildi.

Kamble, S. S., Gunasekaran, A., ve Sharma, R. (2018). Analysis of the driving and dependence power of barriers to adopt industry 4.0 in Indian manufacturing industry. *Computers in Industry*, 101, ss. 107–119. 11 Ocak 2019 tarihinde <https://linkinghub.elsevier.com/retrieve/pii/S0166361518301453> adresinden erişildi.

Kesayak, B. (2018). Endüstri Tarihine Kısa Bir Yolculuk. *Türkiyenin Endüstri 4.0 Platformu*. 11 Ocak 2019 tarihinde <https://www.endustri40.com/endustri-tarihine-kisa-bir-yolculuk/> adresinden erişildi.

Koc, I., ve Arslan, E. (2018). Demand Forecasting for Domestic Air Transportation in Turkey using Artificial Neural Networks. *2018 6th International Conference on Control Engineering & Information Technology (CEIT)* içinde (ss. 1–6). IEEE. 21 Kasım 2019 tarihinde <https://ieeexplore.ieee.org/document/8751869/> adresinden

erişildi.

- Koçak, H. (2011). Efficiency Examination of Turkish Airport With Dea Approach. *International Business Research*, 4(2). Canadian Center of Science and Education. 18 Kasım 2019 tarihinde <http://www.ccsenet.org/journal/index.php/ibr/article/view/7908> adresinden erişildi.
- Köse, U. (2017). *Yapay Zekâ Tabanlı Optimizasyon Algoritmaları Geliştirilmesi (Development of Artificial Intelligence Based Optimization Algorithms)*. Selçuk Üniversitesi.
- Lasi, H., Fettke, P., Feld, T., ve Hoffmann, M. (2014). Industry 4.0. *Business & Information Systems Engineering*, 6(4). 11 Ocak 2019 tarihinde <https://aisel.aisnet.org/bise/vol6/iss4/5> adresinden erişildi.
- Lee, I., ve Lee, K. (2015). The Internet of Things (IoT): Applications, investments, and challenges for enterprises. *Business Horizons*, 58(4), ss. 431–440. Elsevier. 11 Ocak 2019 tarihinde <https://www.sciencedirect.com/science/article/pii/S0007681315000373?via%3Dihub> adresinden erişildi.
- Lee, J. H., Shin, J., ve Realf, M. J. (2018). Machine learning: Overview of the recent progresses and implications for the process systems engineering field. *Computers and Chemical Engineering*, 114, ss. 111–121. Elsevier Ltd.
- Malighetti, P., Paleari, S., ve Redondi, R. (2009). Airport classification and functionality within the European network. *Problems & Perspectives in Management*, 7(1), ss. 183–196. 17 Aralık 2019 tarihinde [https://www.researchgate.net/publication/41083514\\_Airport\\_classification\\_and\\_functionality\\_within\\_the\\_European\\_network](https://www.researchgate.net/publication/41083514_Airport_classification_and_functionality_within_the_European_network) adresinden erişildi.
- Marta, Z. (2017). Cluster Analysis of World's Airports on the Basis of Number of Passengers Handled (Case Study Examining the Impact of Significant Events). *Statistika-Statistics and Economy Journal*, 97(1), ss. 74–88. 17 Aralık 2019 tarihinde <https://www.czso.cz/csu/czso/statistika-statistics-and-economy-journal-no-12017> adresinden erişildi.



- Mayer-Schönberger, V., ve Cukier, K. (2013). *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. John Murray.
- Mayer, R. (2016). Airport classification based on cargo characteristics. *Journal of Transport Geography*, 54, ss. 53–65. Elsevier Ltd. 20 Nisan 2020 tarihinde <https://linkinghub.elsevier.com/retrieve/pii/S0966692316300096> adresinden erişildi.
- Microsoft. (2019). Azure Machine Learning Microsoft Docs. *Microsoft Azure Documentation*. 18 Şubat 2020 tarihinde <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/machine-learning-initialize-model-regression> adresinden erişildi.
- Miller, D. D., ve Brown, E. W. (2018). Artificial Intelligence in Medical Practice: The Question to the Answer? *American Journal of Medicine*, 131(2), ss. 129–133. Elsevier. 15 Ocak 2019 tarihinde <https://www.sciencedirect.com/science/article/pii/S0002934317311178> adresinden erişildi.
- Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267, ss. 1–38. Elsevier B.V.
- Mitchell, T. M. (2006). The Discipline of Machine Learning. 13 Ocak 2019 tarihinde <http://www.cs.cmu.edu/~tom/pubs/MachineLearning.pdf> adresinden erişildi.
- Nigam, R., ve Govinda, K. (2017). Cloud based flight delay prediction using logistic regression. *2017 International Conference on Intelligent Sustainable Systems (ICISS)* içinde (ss. 662–667). IEEE. 21 Kasım 2019 tarihinde <https://ieeexplore.ieee.org/document/8389254/> adresinden erişildi.
- OurAirports. (2019). OurAirports. <https://ourairports.com/>. 12 Şubat 2020 tarihinde <https://ourairports.com/> adresinden erişildi.
- Pereira, A. C., ve Romero, F. (2017). A review of the meanings and the implications of the Industry 4.0 concept. *Procedia Manufacturing*, 13, ss. 1206–1214. Elsevier. 11 Ocak 2019 tarihinde

<https://www.sciencedirect.com/science/article/pii/S2351978917306649?via%3Dihub> adresinden erişildi.

Pereira, T., Barreto, L., ve Amaral, A. (2017). Network and information security challenges within Industry 4.0 paradigm. *Procedia Manufacturing*, 13, ss. 1253–1260. Elsevier. 11 Ocak 2019 tarihinde <https://www.sciencedirect.com/science/article/pii/S2351978917306820?via%3Dihub> adresinden erişildi.

Roberto, I. (2018). Differences Between AI and Machine Learning and Why it Matters. 12 Ocak 2019 tarihinde <https://medium.com/datadriveninvestor/differences-between-ai-and-machine-learning-and-why-it-matters-1255b182fc6> adresinden erişildi.

Rodríguez-Déniz, H., ve Voltés-Dorta, A. (2014). A frontier-based hierarchical clustering for airport efficiency benchmarking. *Benchmarking: An International Journal*, 21(4), ss. 486–508. Emerald Group Publishing Ltd. 20 Nisan 2020 tarihinde <https://www.emerald.com/insight/content/doi/10.1108/BIJ-09-2012-0057/full/html> adresinden erişildi.

Russell, S., ve Norvig, P. (2010). *Artificial Intelligence A Modern Approach Third Edition*. (M. Hirsch, T. Dunkelberger, M. Haggerty, & A. Michael, Ed.) Pearson (Third.). Prentice Hall.

Sakthi, D. (2017). What are the main differences between artificial intelligence and machine learning? Is machine learning a part of artificial intelligence? - Quora. 12 Ocak 2019 tarihinde <https://www.quora.com/What-are-the-main-differences-between-artificial-intelligence-and-machine-learning-Is-machine-learning-a-part-of-artificial-intelligence> adresinden erişildi.

Sammut, C., ve Webb, G. I. (Ed.). (2010). *Encyclopedia of Machine Learning*. *Encyclopedia of Machine Learning*. Boston, MA: Springer US. 15 Nisan 2020 tarihinde <http://link.springer.com/10.1007/978-0-387-30164-8> adresinden erişildi.

Sarkis, J. (2000). An analysis of the operational efficiency of major airports in the United States. *Journal of Operations Management*, 18(3), ss. 335–351. Elsevier Science

Publishers B.V. 6 Kasım 2019 tarihinde [http://doi.wiley.com/10.1016/S0272-6963\(99\)00032-7](http://doi.wiley.com/10.1016/S0272-6963(99)00032-7) adresinden erişildi.

Sarkis, J., ve Talluri, S. (2004). Performance based clustering for benchmarking of US airports. *Transportation Research Part A: Policy and Practice*, 38(5), ss. 329–346. 17 Aralık 2019 tarihinde <https://linkinghub.elsevier.com/retrieve/pii/S0965856403000958> adresinden erişildi.

Schmidt, R., Möhring, M., Härting, R.-C., Reichstein, C., Neumaier, P., ve Jozinović, P. (2015). Industry 4.0 - Potentials for Creating Smart Products: Empirical Research Results. *International Conference on Business Information Systems* içinde (ss. 16–27). Springer, Cham. 11 Ocak 2019 tarihinde [http://link.springer.com/10.1007/978-3-319-19027-3\\_2](http://link.springer.com/10.1007/978-3-319-19027-3_2) adresinden erişildi.

Semih, Ş. (2018). Yapay Zeka, Makine Öğrenimi ve Derin Öğrenme Arasındaki Farklar. 12 Ocak 2019 tarihinde <https://www.endustri40.com/yapay-zeka-makine-ogrenimi-ve-derin-ogrenme-arasindaki-farklar/> adresinden erişildi.

Sennaroglu, B., ve Varlik Celebi, G. (2018). A military airport location selection by AHP integrated PROMETHEE and VIKOR methods. *Transportation Research Part D: Transport and Environment*, 59, ss. 160–173. Elsevier Ltd. 6 Kasım 2019 tarihinde <https://linkinghub.elsevier.com/retrieve/pii/S1361920917306181> adresinden erişildi.

Sobie, C., Freitas, C., ve Nicolai, M. (2018). Simulation-driven machine learning: Bearing fault classification. *Mechanical Systems and Signal Processing*, 99, ss. 403–419. Elsevier Ltd.

Staub, S., Karaman, E., Kaya, S., Karapınar, H., ve Güven, E. (2015). Artificial Neural Network and Agility. *Procedia - Social and Behavioral Sciences*, 195, ss. 1477–1485.

Stichhauerova, E., ve Pelloneova, N. (2019). An Efficiency Assessment of Selected German Airports Using the DEA Model. *Journal of Competitiveness*, 11(1), ss. 135–151. Tomas Bata University in Zlin. 18 Kasım 2019 tarihinde

<https://www.cjournal.cz/index.php?hid=clanek&bid=archiv&cid=320&cp=adresinden> erişildi.

Sunil, R. (2016). A Complete Tutorial which teaches Data Exploration in detail. *Analytics Vidhya*, 14 Ocak 2019 tarihinde <https://www.analyticsvidhya.com/blog/2016/01/guide-data-exploration/> adresinden erişildi.

Tack, C. (2019). Artificial intelligence and machine learning | applications in musculoskeletal physiotherapy. *Musculoskeletal Science and Practice*, 39(0), ss. 164–169. Elsevier.

Tjahjono, B., Esplugues, C., Ares, E., ve Pelaez, G. (2017). What does Industry 4.0 mean to Supply Chain? *Procedia Manufacturing*, 13, ss. 1175–1182. Elsevier. 11 Ocak 2019 tarihinde <https://www.sciencedirect.com/science/article/pii/S2351978917308302?via%3Dihub> adresinden erişildi.

Trenggonowati, D. L., Ulfah, M., Ekawati, R., ve Yusuf, V. A. (2019). Organization clustering airports using K-Means clustering algorithm. *IOP Conference Series: Materials Science and Engineering*, 673, s. 012081. 17 Aralık 2019 tarihinde <https://iopscience.iop.org/article/10.1088/1757-899X/673/1/012081> adresinden erişildi.

TÜİK. (2019). Türkiye İstatistik Kurumu. <http://www.tuik.gov.tr/Start.do>. 14 Şubat 2020 tarihinde <http://www.tuik.gov.tr/Start.do> adresinden erişildi.

Turing, A. M. (1950). Computing Machinery and Intelligence. *Source: Mind, New Series*, 59(236), ss. 433–460. Oxford University Press on behalf of the Mind Association.

Wall, L. D. (2018). Some financial regulatory implications of artificial intelligence. *Journal of Economics and Business*, 100(May), ss. 55–63. Elsevier.

Willmott, C., ve Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 30(1), ss. 79–82. Inter-Research. 13 Nisan 2020 tarihinde

<http://www.int-res.com/abstracts/cr/v30/n1/p79-82/> adresinden erişildi.

Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. J., vd. (2008). Top 10 algorithms in data mining. *Knowledge and Information Systems*, 14(1), ss. 1–37. 12 Ocak 2019 tarihinde <http://link.springer.com/10.1007/s10115-007-0114-2> adresinden erişildi.

Xia, Q., Zhang, Z., ve Zhang, B. (2019). Airport Role Orientation Based on Improved K-means Clustering Algorithm. *Advanced Hybrid Information Processing* (ss. 299–309). 17 Aralık 2019 tarihinde [http://link.springer.com/10.1007/978-3-030-36405-2\\_30](http://link.springer.com/10.1007/978-3-030-36405-2_30) adresinden erişildi.

Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., ve Gao, R. X. (2019). Deep learning and its applications to machine health monitoring. *Mechanical Systems and Signal Processing*, 115, ss. 213–237. Elsevier Ltd.

Zhou, K., Taigang Liu, ve Lifeng Zhou. (2015). Industry 4.0: Towards future industrial opportunities and challenges. *2015 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)* içinde (ss. 2147–2152). IEEE. 11 Ocak 2019 tarihinde <http://ieeexplore.ieee.org/document/7382284/> adresinden erişildi.

## ÖZGEÇMİŞ

### KİMLİK BİLGİLERİ

**Adı Soyadı** : Ahmet Arif ÇOLAKOĞLU  
**Doğum Yeri** : HATAY  
**Doğum Tarihi** : 15.04.1991  
**E-posta** : aarifcolakoglu@gmail.com

### EĞİTİM BİLGİLERİ

**Lise** : Halil İbrahim Evren Anadolu Lisesi  
**Lisans** : İstanbul Ticaret Üniversitesi / Endüstri Mühendisliği  
**Yüksek Lisans** : Pamukkale Üniversitesi / Sayısal Yöntemler Programı  
**Yabancı Dil ve Düzeyi** : İngilizce / B1

### İŞ DENEYİMİ:

- Mental Hr & Consultancy, Ankara, 2017 – 2019, Yönetim Danışmanlığı (Uzman Pozisyonunda Mevcut Durum Analizi, İş Analizi ve İş Geliştirme, Organizasyonel Yapılanma, Bireysel Güçlü Yönler Analizi ve Değerlendirme, Memnuniyet Anketi vb. çalışmalarda TAI, Türk Telekom, TRT, Enerji Bakanlığı gibi kurumlara danışmanlık hizmeti verilmiştir.)
- Tmekatronik Müh. San. ve Ltd. Şti, İstanbul, 2014 – 2015, Ar-ge (Üniversite bünyesinde KOSGEB ve Sanayi Bakanlığı hibe destekli projede TÜBİTAK’a Hibrit İnsansız Hava Aracı yapımında takım lideri olarak görevler üstlenilmiştir. Ayrıca CNC ve benzeri otomatik kontrollü makinelerin tasarlanması ve üretilmesi süreçlerinde bulunulmuştur.)

### ARAŞTIRMA ALANLARI:

- Dijital Optimizasyon ve Sürdürülebilir Operasyonlar, Makine Öğrenmesi, Derin Öğrenme, Katmanlı Üretim Teknolojisi, Robotik Süreç Otomasyonu, Sürü İnsansız Hava Araçları...

