

**T.C.**  
**BAHÇEŞEHİR ÜNİVERSİTESİ**

**TRAFİK KAZASI OLMA  
OLASILIĞININ HESAPLANMASI**

**Yüksek Lisans Tezi**

**MEHMET EREN BEKİN**

**İSTANBUL, 2018**

**T.C.  
BAHÇEŞEHİR ÜNİVERSİTESİ**

**FEN BİLİMLERİ ENSTİTÜSÜ  
BİLGİ TEKNOLOJİLERİ**

**TRAFİK KAZASI OLMA  
OLASILIĞININ HESAPLANMASI**

**Yüksek Lisans Tezi**

**MEHMET EREN BEKİN**

**Tez Danışmanı: Doç. Dr. M. ALPER TUNGA**

**İSTANBUL, 2018**

T.C.  
BAHÇEŞEHİR ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ  
BİLGİ TEKNOLOJİLERİ

Tezin Adı: Trafik Kazası Olma Olasılığının Hesaplanması

Öğrencinin Adı Soyadı: Mehmet Eren BEKİN

Tez Savunma Tarihi: 04.01.2018

Bu tezin Yüksek Lisans tezi olarak gerekli şartları yerine getirmiş olduğu Fen Bilimleri Enstitüsü tarafından onaylanmıştır.

Prof. Dr. Nafiz ARICA  
Enstitü Müdürü  
İmza

Bu tezin Yüksek Lisans tezi olarak gerekli şartları yerine getirmiş olduğunu onaylarım.

Doç. Dr. M. Alper TUNGA  
Program Koordinatörü  
İmza

Bu Tez tarafımızca okunmuş, nitelik ve içerik açısından bir Yüksek Lisans tezi olarak yeterli görülmüş ve kabul edilmiştir.

Jüri Üyeleri

Tez Danışmanı  
Doç. Dr. M. Alper TUNGA

Üye  
Yrd. Doç. Dr. Batu SALMAN

Üye  
Doç. Dr. Cihangir ÖZEMİR

İmzalar

  
-----  
  
-----  
  
-----

## TEŐEKKÜR

Bugüne kadar tüm çalışmalarımızda fikirlerimize deęer veren ve bu fikirlerin gerekleşmesine her zaman olanak saęlayan , tez konusu seçimimde de desteęini esirgemeyen Bilgi ve İletiřim Teknolojileri Çözüm Geliřtirme Müdürü Sayın Korhan KUYU ve Bilgi Teknolojilerinden Sorumlu Genel Müdür Yardımcımız Sayın Mehmet ABACI' ya , tez konunun güçlenmesinde önemli katkıları bulunan , güven ve inancını esirgemeyen deęerli Hasar Yönetimi Müdürlüęü Yöneticimiz Sayın Emin AYIK' a, yüksek lisans derslerimden tezimi sonuçlandırmaya kadar geçen sürede desteęini sürekli hissettięim Sayın Doç. Dr. M. Alper TUNGA' ya, bu alıřma sırasında kendilerine zaman ayıramadıęım sevgili eřime ve biricik oęluma gösterdikleri sabır ve anlayıřtan dolayı teőekkür ederim.

İstanbul, 2018

Mehmet Eren BEKİN

## ÖZET

### TRAFİK KAZASI OLMA OLASILIĞININ HESAPLANMASI

Mehmet Eren BEKİN

Bilgi Teknolojileri

Tez Danışmanı: Doç. Dr. M. Alper TUNGA

Mayıs 2017, 51 sayfa

Bu çalışmada dünyada milyonlarca insanın ölümüne sebep olan trafik kazalarının gerçek verileri incelenerek, konum ve sürücü verisi bazlı trafik kazalarının olma olasılığını ortaya çıkartabilecek başarılı bir veri modeli oluşturmak ve sürücülere risk öncesi bilgilendirme sağlayabilmektir.

Çalışmada kullanılan veriler sigortacılık sektörü firmalarından temin edilmiş olup, gerekli izinler doğrultusunda model kayıtları oluşturulmuştur. Bu kapsamda 460 kayıt bulunmaktadır. Mobil cihazlarla ölçülebilen ivme ve hız değerleri, fotoğraf oluşturma konum ve tarihleri, sürücüye ait profil verileri ve duygusal analiz verilerinin kazaya risk oluşturmada hesaplanması, hava durumu bilgisi ve konumlarda gerçekleşmiş kazaların kümelenmesi ile makine öğrenmesi trafik kazası olma olasılıklarını ortaya koymuştur. Veri modellerini oluşturmak için gerekli algoritmaların başarılı ölçme oranları incelenmiş ve en doğru yöntem seçilmiştir. Bu oluşan çalıştırılmış model ile de anlık olarak dijital uygulamalar ile entegrasyonları sağlanarak bir kişiye ait trafik kazası riski içeren haritalar çıkartılabilir, sürücüler kendileri için riskli bir bölgeye yaklaştığında sesli uyarılar ile daha dikkatli olması sağlanabilir.

**Anahtar Kelimeler:** Makine Öğrenmesi, Kaza, Mobil cihazlar, Dijital

## ABSTRACT

### CALCULATION OF THE PROBABILITY OF TRAFFIC ACCIDENT

Mehmet Eren BEKİN

Information Technologies

Thesis Supervisor : Doc. Dr. M. Alper TUNGA

December 2017, 51 pages

This study examines the actual data of traffic accidents causing the deaths of millions of people in the world and is able to provide a successful data model that can reveal the possibility of traffic accidents based on location and driver data and to provide pre-risk information to the drivers.

The data used in this study were obtained from the insurance sector companies and model registrations were established in line with the required permits. This includes 460 records. The acceleration and speed values that can be measured with the mobile devices, the location and date of photo creation, the calculation of profile data of the driver and the calculation of the emotional analysis data for risk to the accident, the clustering of the weather information and the accidents occurred in the locations and the machine learning revealed the possibility of traffic accidents. Successful measurement rates of the algorithms required to construct the data models have been examined and the most accurate method has been chosen. With this running model, it is possible to instantaneously provide integration with digital applications, so that maps containing risk of traffic accidents belonging to a person can be removed, and the drivers can be more careful with sound alerts when they approach a risky zone for themselves.

**Keywords:** Machine Learning, Mobile Phones, Accident, Digital

## İÇİNDEKİLER

<b>TABLolar</b> .....	<b>viii</b>
<b>ŞEKİLLER</b> .....	<b>ix</b>
<b>KISALTMALAR</b> .....	<b>x</b>
<b>1.GİRİŞ</b> .....	<b>1</b>
<b>2. LİTERATÜR TARAMASI</b> .....	<b>5</b>
<b>2.1 MAKİNE ÖĞRENMESİ</b> .....	<b>5</b>
<b>2.1.1 Makine Öğrenmesi Tanımı</b> .....	<b>5</b>
<b>2.1.2 Öğrenme Türü</b> .....	<b>6</b>
<b>2.1.2.1 Gözetimli-Gözetimsiz Öğrenme (Supervised – Unsupervised Learning)</b> .....	<b>6</b>
<b>2.1.2.1.1 Sınıflama ve Regresyon</b> .....	<b>6</b>
<b>2.1.2.1.2 Kümeleme</b> .....	<b>7</b>
<b>2.1.2.1.3 Birliktelik Kuralları</b> .....	<b>8</b>
<b>2.1.3 Özellik Çıkarımı (Feature Extraction)</b> .....	<b>8</b>
<b>2.1.4 Kullanım Alanları</b> .....	<b>8</b>
<b>2.1.5 Sigortacılıkta Makine Öğrenmesi</b> .....	<b>10</b>
<b>2.2 DİJİTAL KULLANIMLAR</b> .....	<b>11</b>
<b>2.2.1 Dijitalleşmenin Tanımı</b> .....	<b>11</b>
<b>2.2.2 Mobil Uygulamalar</b> .....	<b>11</b>
<b>2.2.3 Sigortacılıkta Mobil Uygulamalar</b> .....	<b>14</b>
<b>2.3 TRAFİK KAZASI ÇALIŞMALARI</b> .....	<b>16</b>
<b>2.3.1. Çalışma 1</b> .....	<b>16</b>
<b>2.3.2. Çalışma 2</b> .....	<b>16</b>
<b>2.3.3. Çalışma 3</b> .....	<b>17</b>
<b>2.3.4 Çalışma 4</b> .....	<b>17</b>
<b>2.3.5 Çalışma 5</b> .....	<b>18</b>
<b>2.3.6 Çalışma 6</b> .....	<b>18</b>
<b>2.3.7 Çalışma 7</b> .....	<b>19</b>
<b>3. VERİ VE YÖNTEMLER</b> .....	<b>20</b>

<b>3.1 VERİ KÜMESİ TANIMI</b> .....	20
<b>3.2 VERİ KÜMESİNİN HAZIRLANMASI</b> .....	20
<b>3.2.1 Veri Kümesinin İçeriği</b> .....	20
3.2.1.1 Sürücü Verileri .....	20
3.2.1.2 Hava Durumu Verileri.....	23
3.2.1.3 Lokasyon Verileri.....	23
3.2.1.4 Kaza Detay Bilgi Verileri.....	24
<b>3.2.2 Machine Learning Çalışmaları</b> .....	25
3.2.2.1 Parametrelerin Değerlendirilmesi .....	25
3.2.2.3 Referans Değerlerin Oluşturulması.....	27
<b>3.2.3 Azure Machine Learning Platformunun Kullanılması</b> .....	27
3.2.3.1 Veri Kümelerinin Oluşturulması.....	27
3.2.3.2 Meta Data Düzenleme İşlemi .....	29
3.2.3.3 Veri Kümesi Kolon Ayıklama İşlemi.....	29
3.2.3.4 Veriyi Bölme (Split Data) İşlemi.....	30
3.2.3.5 Veri Kümesi Birleştirme (Join Data) İşlemi .....	31
3.2.3.6 Algoritma Seçme İşlemi.....	31
3.2.3.7 Model Eğitime (Train Model) İşlemi.....	32
3.2.3.8 Skor Model (Score Model) İşlemi .....	33
3.2.3.9 Değerlendirme Modeli (Evaluate Model) İşlemi .....	34
<b>3.3 MAKİNA ÖĞRENMESİ YÖNTEMLERİ</b> .....	35
3.3.1 Gözetimli Öğrenme.....	35
3.3.2 Gözetimsiz Öğrenme .....	36
3.3.3 Önerme Yöntemi ile Öğrenme.....	36
<b>4. BULGULAR</b> .....	38
4.1 Matchbox Recommender Yöntemi .....	38
4.2 Sınıflama Yöntemleri.....	40
4.3 Regresyon Yöntemleri.....	43
<b>5.TARTIŞMA VE SONUÇ</b> .....	46
<b>KAYNAKÇA</b> .....	48
<b>ÖZGEÇMİŞ</b> .....	51



## TABLÖLAR

Tablo 3.1: Sürücü Veri Kümesi Açıklamaları .....	23
Tablo 3.2: Hava Durumu Veri Kümesi Açıklamaları.....	24
Tablo 3.3: Hasar Detay Lokasyon Veri Kümesi Açıklamaları.....	24
Tablo 3.4: Hasar Kaza Detay Veri Kümesi Açıklamaları.....	25
Tablo 3.5: Sürücü Veri Kümesi Parametre Açıklamaları.....	27
Tablo 3.6: Hasar Lokasyon Veri Kümesi Parametre Açıklamaları.....	27
Tablo 3.7: Hasar ve Sürücü İlişki Veri Kümesi Parametre Açıklamaları.....	28



## ŞEKİLLER

Şekil 1.1: Emniyet Genel Müdürlüğü Yıllara göre Ölü ve Yaralı Sayıları.....	3
Şekil 1.2: Makine Öğrenmesi Trafik Kazası Verileri İşleme Grafiği.....	4
Şekil 2.1: Mobil Kullanım Anketine Katılım Sağlayan Ülkeler.....	12
Şekil 2.2: Günlük telefona bakma rakamları.....	13
Şekil 2.3: Mobil Uygulama kullanıcı davranışları 1.....	13
Şekil 2.4: Mobil Uygulama kullanıcı davranışları 2.....	14
Şekil 3.1: Lokasyon bazlı uygulamalar ile örnek sürücü segmentleştirme.....	22
Şekil 3.2: Mobil uygulama ile sürüş kabiliyetlerinin ölçülmesi.....	23
Şekil 3.3: Sürücü Detay Bilgileri Azure Platform Görünümü.....	30
Şekil 3.4: Kaza Detay Bilgileri Azure Platform Görünümü.....	30
Şekil 3.5: Azure Metadata Düzenleme İşlemi.....	31
Şekil 3.6: Azure Veri Kümesinden Kolon Seçme.....	32
Şekil 3.7: Azure Data Ayıklama Yöntemi.....	32
Şekil 3.8: Azure Veri Kümesi Birleştirme Yöntemi.....	33
Şekil 3.9: Azure Model Eğitime İşlemi.....	34
Şekil 3.10: Azure Skor Model İşlemi.....	35
Şekil 3.11: Azure ML Algoritmaları.....	37
Şekil 4.1: Azure Matchbox Recommender Yöntemi ile Gerçekleştirilen Çalışma.....	40
Şekil 4.2: Azure Matchbox Recommender ile Gerçekleşen Sonuç Veri Kümesi.....	41
Şekil 4.3: Azure ile Gerçekleşen Çalışmadaki Sınıflandırma İşlem Akışı.....	42
Şekil 4.4: Azure ile Gerçekleşen Çalışmadaki Sınıflandırma İşlem Öğrenme Akışı....	43
Şekil 4.5: Azure ile Gerçekleşen Çalışmadaki Regresyon İşlemi.....	44
Şekil 4.6: Azure ile Gerçekleşen Çalışmadaki Regresyon İşlem Öğrenme Akışı.....	44

## KISALTMALAR

GPS	:	Global Positioning System
INC	:	Insurance Company (Sigorta řirketi)
KTT	:	Kaza Tespit Tutanađı
ML	:	Machine Learning (Makine Öğrenmesi)
NLP	:	Natural Language Processing
SBM	:	Sigorta Bilgi ve Gözetim Merkezi



## 1.GİRİŞ

Bilgi, dijital uygulamaların gelişmesi, akıllı cihazların özelliklerinin artması ile birlikte insanlara çok hızlı akışın sağlanabildiği ve kolaylıkla saklanarak geleceğimiz hakkında yol belirleyebileceği bir hale gelmiştir. Bugün, dünyada gerçekleşen haberleri anında, maç skorlarını saniyesinde ve olabilecek tüm olayların nasıl gerçekleşebileceğini öğrenebilmek için mobil cihazlardan, kişisel bilgilerden toplanan verilerin birlikte kullanılarak insanlara sunulması aşaması oldukça ilerlemiştir.

Geçmişten bugüne hızla veriler büyümekte ve sektörel ya da endüstriyel alanlarda başka verilerle artık anlamlı olan, yapay zeka ürünleri ile de bu verileri işleyerek geleceğe yönelik hareketlerde, kararlarda sonuç ortaya konulan bir hale gelmiştir.

Büyük veri kullanımı, tüm sektörlerde sadece oluşacak değil, oluşan ve saklanan kayıtlarının en doğru analizlerle doğru sonuçlara ulaşmasıdır. Konusunda hesaplanan olasılıklar doğru yol haritası ile kurumların vizyonlarına hangi aşamaların daha doğru olacağını kararını vermeye yardımcı olmaktadır.

Makine öğrenmesi, işlenecek olan tüm verilerin sayısallaştırması ve geçmişte öğretilmiş (çalışılmış) veri kütüphanesinde yakınlık tanınması ile gerçekleşen bir yöntemdir. Bu veriler içerisinde el yazısından kitap yazısına, sesten görüntüye, videodan GPS bilgilerine kadar detayları içerebilir. Bu veriler frekanslarına bakılarak sayısallaştırılmaktadır. Günümüzde birçok kurum ve kuruluşta makine öğrenmesini ihtiyaç duydukları süreçlerde olmazsa olmaz bir kısım haline gelmiştir.

Makine öğrenmesi ve veri madenciliğinin kullanım sektörleri incelendiğinde, güvenlik, bankacılık, sigortacılık, pazarlama, tıp, telekomünikasyon, e-ticaret gibi birçok alanda aktif olarak iş süreçlerinde entegrasyonların sağlandığı görülmektedir.

Sigortacılık sektörü üretim ve hasar olmak üzere müşterilerine hizmet vermektedir. Bu hizmetler şirketlerin müşterilere güvenlerini sağlamaya çalıştığı platformlardır. Hızlı

hizmet esas alınmaktadır. Sigorta zorunluluğun olması da tüm kaza ve üretim verilerinin detaylıca saklanması sağlamaktadır. Sigorta şirketlerinde ve sigorta havuz kurumlarında (Sigortalı bilgilerinin saklandığı ve haklarının korunduğu devlet kuruluşları) büyük çevre (network) incelemesi yapılması olasıdır. Bu sayede makine öğrenmesi ve veri madenciliği sigortacılık sektöründe tüm verilerin işlenerek, şirketlerin insan operasyonları ile yürüttükleri inceleme ve analiz süreçlerini çok kısa sürelerde bitirmelerine olanak sağlayacaktır.

Bunların yanı sıra sigortacılık sektöründe saklanmakta olan veriler trafikte bulunan insanların daha güvenli ulaşım gerçekleştirmesinde kullanılabilir, araç sürüşleri ve riskleri hakkında bilgi verebilir duruma gelmesi hedeflenmektedir. Gerçekleşen kazaların hangi lokasyonda, kazanın gerçekleşme tipi, kaza fotoğrafları, araç hız oranları, sürücülerin profilleri, hava durumu vb. bilgilerinin kullanılarak kişiye özel risk haritaları çıkartılabilir.

Türkiye’de Emniyet Genel Müdürlüğü’nün verilerine göre 2016 yılında 1.182.491 adet trafik kazası gerçekleşmiş olup 7.300 ölü, 303.812 yaralı mevcuttur. Bu değerlerin yüksek olması çözüm olarak yeni teknolojilerin gerekliliğini ortaya koymaktadır. Ör Şekil 1.1 de görüldüğü üzere kaza sayılarında yıllara göre düzenli bir artış bulunmaktadır. Ölüm sayısı da 2002 yılından günümüze yıllık ortalama da üç katına çıkmıştır.

**Şekil 1.1: Emniyet Genel Müdürlüğü Yıllara göre Ölü ve Yaralı Sayıları**

Yıl Year	Toplam kaza sayısı Total number of accidents	Maddi hasarı kaza sayısı Accidents involving material loss only	Ölümlü, yaralanmalı kaza sayısı Accidents involving death and personal injury	Ölü sayısı <sup>(1)</sup> Killed persons <sup>(1)</sup>			Yaralı sayısı Number of persons injured
				Toplam Total	Kaza yerinde At accident scene	Kaza sonrası Accident follow-up	
2002	439 777	374 029	65 748	4 093	4 093	-	116 412
2003	455 637	388 606	67 031	3 946	3 946	-	118 214
2004	537 352	460 344	77 008	4 427	4 427	-	136 437
2005	620 789	533 516	87 273	4 505	4 505	-	154 086
2006	728 755	632 627	96 128	4 633	4 633	-	169 080
2007	825 561	718 567	106 994	5 007	5 007	-	189 057
2008	950 120	845 908	104 212	4 236	4 236	-	184 468
2009	1 053 346	942 225	111 121	4 324	4 324	-	201 380
2010	1 106 201	989 397	116 804	4 045	4 045	-	211 496
2011	1 228 928	1 097 083	131 845	3 835	3 835	-	238 074
2012	1 296 634	1 143 082	153 552	3 750	3 750	-	268 079
2013	1 207 354	1 046 048	161 306	3 685	3 685	-	274 829
2014	1 199 010	1 030 498	168 512	3 524	3 524	-	285 059
2015	1 313 359	1 130 348	183 011	7 530	3 831	3 699	304 421

Kaynak: Emniyet Genel Müdürlüğü ve Jandarma Genel Komutanlığı  
Source: General Directorate of Public Security and General Command of Gendarmerie  
(1) Ölü sayıları 2015 yılına kadar sadece kaza yerinde tespit edilen ölümleri kapsarken, 2015 yılından itibaren trafik kazasında yaralanıp sağlık kuruluşuna sevk edilenlerden kazanın sebep ve tesiriyle 30 gün içinde ölenleri de kapsamaktadır.  
(1) Until year 2015 figures on persons killed include the deaths only at the accident scene however since year 2015 figures on persons killed also include the deaths within 30 days after the traffic accidents due to related accident and its impacts for people injured and sent to health facilities.  
- Bilgi yoktur.  
- Denotes magnitude null.

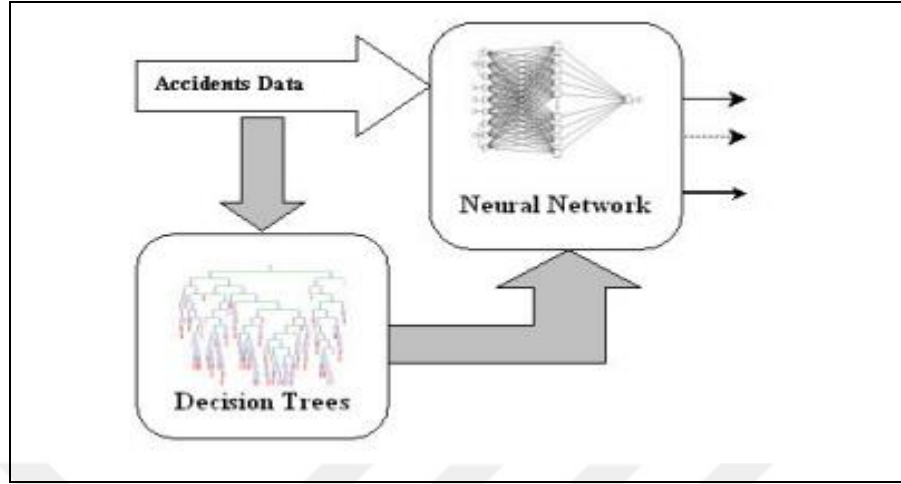
Kaynak: Emniyet Genel Müdürlüğü

Bilgi teknolojilerinin gelişmesi ve yol trafiğinde meydana gelen dönüşüm hem çok hassas trafik bilgisi talebini artırdı. Hassasiyeti sağlamak için ve bir karayolu ağında daha kapsamlı trafik koşulları, yaygın olarak kullanılan trafik sensörleri mevcut trafik durumunun ölçülmesi verimsizdir. Diğer veri kaynakları (kamaralar, gps, cep telefonları) mevcut ölçümlerin sağladığı bilgileri tamamlamak için aşamalı olarak kullanılır. (Nejdet DOGRU - Abdulhamit SUBASI 2013)

İnsanlık tarihinde göçebe yaşam ile ortaya çıkan ulaşım ihtiyacı katlanarak artmaktadır. Karayolu taşımacılığı ulaştırma hizmetlerinin büyük bir bölümünü oluştururken beraberinde çözümü güç problemler getirmektedir. Büyük maddi kayıplar ile birlikte toplum üzerinde olumsuz sosyal etkileri olan trafik kazaları karayolu ulaşımının çözülmesi zorunlu problemlerindedir. (Tarık Kaya 2015)

Bu çalışmada dünyada milyonlarca insanın ölümüne sebep olan trafik kazalarının gerçek verileri incelenerek, insanların canlarını korumak en öncelikli hedefdir. Konum ve sürücü verisi bazlı trafik kazalarının olma olasılığını ortaya çıkartabilecek başarılı bir veri modeli oluşturmak, mobil uygulamalarla sürücülere erken bilgi vermektir.

**Şekil 1.2 : Makine Öğrenmesi Trafik Kazası Verileri İşleme Grafiği**



Veri çalışması için Sigorta Bilgi Merkezi KTT ve Hasar verileri ile kazayı yapan kişilere ait model verileri ile henüz kaza yapmamış Anadolu Sigorta A.Ş. müşteri profillerinin verileri kullanılarak hazırlanmıştır.

## 2. LİTERATÜR TARAMASI

### 2.1 MAKİNE ÖĞRENMESİ

Günümüz teknolojileriyle birbiriyle ilişki kurulmayı bekleyen veriler oluşmaktadır. Makine öğrenmesi, geçmişte oluşan bu verileri kullanarak gelecek hakkında olabilecekleri hesaplama yöntemidir.

#### 2.1.1 Makine Öğrenmesi Tanımı

Bugün elektronik olarak üretilen veri için Web tabanlı veri analiz yöntemleri kullanılmaktadır. Makine öğrenimi, verileri otomatik olarak algılayabilen ve sonra ortaya çıkmış desenleri kullanarak gelecek verileri tahmin edebilen bu yöntemleri geliştirir. ( Kevin P. Murphy 2009)

Günümüzde, makine öğrenme algoritmaları, bilgisayarların insanlarla iletişim kurmasına, otomobilleri özerk olarak sürmesine, spor maçı raporları yazmasına ve yayınlamasına ve terörist zanlıları bulmasına olanak tanır. Makine öğreniminin çoğu endüstriyi ve onların içerisindeki işleri ciddi şekilde etkileyeceğine inanıyorum, bu yüzden her yöneticinin makine öğreniminin ne olduğunu ve nasıl geliştiğini en azından kavraması gerekir. (Bernard Marr 2016)

Arthur Samuel, 1950 yılında dama oyunu geliştirmiştir. Geliştirdiği dama oyunu programlamadan farklı olarak tüm verileri analiz ederek iyi ve kötü hamleleri öğrenmiştir. Çalışmanın sonunda bilgisayar Arthur Samuel' den daha iyi oynayan bir duruma gelmiştir. Arthur Samuel, makine öğrenmesini direk programlamadan bilgisayar kaynaklarına bir işi öğretme olarak bahsetmiştir.

Makine öğrenimi çevresel durum gözlemleri ve geçiş tabanlı kuralların eşdeğer olduğu öğrenme işlerinin ve öğrenmenin otomasyonudur. Bu geniş bir sahadır, yani sadece örneklerden öğrenme değil aynı zamanda takviye öğrenimi ve öğretmenle öğrenim bu



sahaya girmektedir. Bir öğrenme algoritması veri kaynakları ve onun beraberinde yer alan giriş bilgisi ve sonuçlardan oluşur. Makine öğrenimi önceki örnekleri ve sonuçları inceler ve bu işleri nasıl yeniden yapacağını öğrenir ve yeni durumlar hakkında genellemeler yapar. (Ömer Akgöbek, Fuat Çakır 2009)

## **2.1.2 Öğrenme Türü**

Makine öğrenmesi gözetimli(supervised) ve gözetimsiz(unsupervised) öğrenme olmak üzere iki kısımdan oluşmaktadır.

### **2.1.2.1 Gözetimli-Gözetimsiz Öğrenme (Supervised – Unsupervised Learning)**

Gözetimli öğrenme, eğitim verisi oluşturarak makine öğrenmesini takipli sağlamaya yarayan yöntemdir. Bu türdeki amaç işaretlenmiş veriler ile çıkan sonuçların eşleştirilmeye çalışılmasıdır.

Birden çok gözetimli öğrenme algoritmasının birden çok veri kümesi üzerinde karşılaştırılarak belirli bir başarı ölçütüne göre en iyisinin istatistiksel olarak anlamlı (şans eseri oluşamayacak kadar büyük) biçimde bulunması, ya da daha genel bir tanımla iyiden kötüye doğru sıralanması hem örüntü tanıma hem veri madenciliği açısından önemli bir konudur. Birden çok algoritmayı karşılaştırmak, birden çok veri kümesi üzerinde karşılaştırma yapmak, ya da hata yerine başka ölçütler kullanmak ancak son yıllarda yapay öğrenme yazınında yer almaya başlamıştır. (Ethem Alpaydın 2013)

Gözetimsiz veri madenciliğinde bir öğrenme süreci yoktur. Model oluşturken algoritmayı yönlendirebilecek önceden bilinen sonuçlar yoktur. Genelde tanımlayıcı modellerle sonuçlansa da tahminleyici de olabilirler (Oracle, 2014).

#### **2.1.2.1.1 Sınıflama ve Regresyon**

Sınıflama ve regresyon, önemli veri sınıflarını ortaya koyan veya gelecek veri eğilimlerini tahmin eden modelleri kurabilen iki veri analiz yöntemidir. Sınıflama kategorik değerleri tahmin ederken, regresyon süreklilik gösteren değerlerin tahmin edilmesinde kullanılır (Serhat ÖZEKES 2013).

Sınıflama ve regresyon modellerinde kullanılan başlıca teknikler şunlardır:

- a- Karar Ağaçları (Decision Trees)
- b- Yapay Sinir Ağları (Artificial Neural Networks)
- c- Genetik Algoritmalar (Genetic Algorithms)
- d- K-En Yakın Komşu (K-Nearest Neighbor)
- e- Bellek Temelli Nedenleme (Memory Based Reasoning)
- f- Naive-Bayes

Karar ağaçları, kurulması, yorumlanmalarının kolay olması, veri tabanı entegrasyonun rahatlıkla sağlanabilmesi ve güvenilirliklerinin iyi olması nedenleri ile sınıflama modelleri içerisinde en yaygın kullanıma sahip tekniktir. Karar ağacı, adından da anlaşılacağı gibi bir ağaç görünümünde, tahmin edici bir tekniktir.

#### **2.1.2.1.2 Kümeleme**

Kümeleme, veriyi sınıflara veya kümelere ayırma işlemidir. (Chameleon 1999) Kümeleme modelinde, sınıflama modelinde olan veri sınıfları yoktur. Bazı uygulamalarda kümeleme modeli, sınıflama modelinin bir önişlemi gibi görev alabilmektedir.

Literatürde pek çok kümeleme algoritması bulunmaktadır. Kullanılacak olan kümeleme algoritmasının seçimi, veri tipine ve amaca bağlıdır. Genel olarak başlıca kümeleme yöntemleri şu şekilde sınıflandırılabilir. (Serhat ÖZEKES)

- a- Bölme yöntemleri (Partitioning methods)
- b- Hiyerarşik yöntemler (Hierarchical methods)
- c- Yoğunluk tabanlı yöntemler (Density-based methods)
- d- Izgara tabanlı yöntemler (Grid-based methods)
- e- Model tabanlı yöntemler (Model-based methods)

### 2.1.2.1.3 Birliktelik Kuralları

Birliktelik kuralları, büyük veri kümeleri arasında birliktelik ilişkileri bulurlar. Toplanan ve depolanan verinin her geçen gün gittikçe büyümesi yüzünden, şirketler veri tabanlarındaki birliktelik kurallarını ortaya çıkarmak istemektedirler. Büyük miktardaki mesleki işlem kayıtlarından ilginç birliktelik ilişkilerini keşfetmek, şirketlerin karar alma işlemlerini daha verimli hale getirmektedir. (Serhat ÖZEKES)

### 2.1.3 Özellik Çıkarımı (Feature Extraction)

Bazı durumlarda veriye ait birçok özellik bilinse de verinin kümesini, sınıfını, değerini belirleyen özelliklerinin hangileri olduğu bilinmeyebilir. Bu durumlarda tüm özellik kümesinin bir alt kümesi seçilmesi (özellik seçimi) ya da bu özelliklerin birleşimlerinden yeni özelliklerin elde edilmesi işlemidir. (Merve Bozo, <http://e-bergi.com/y/makine-grenimi>)

Özellik çıkarımı için amaç karmaşık olan veriyi basit hale getirmeye çalışmaktır. Doğru olarak çıkarım sağlanmış özellikler sonucu ciddi oranda etkilemesi sebebiyle, özellik çıkarım işlemi iyi bir analiz ve işe hakim kişilerle beraber gerçekleştirilmelidir.

### 2.1.4 Kullanım Alanları

Makine öğrenmesi birçok sektörde kullanılmakta olup, en yaygın kullanım örnekleri aşağıdaki gibi sıralanmıştır.

- a- **El yazısı tanıma:** Yazılan el yazısının kime ait olduğu ve başka bir yazı ile eşleşmeye sahip olup olamayacağı bilgisinin ortaya konmasında önemli rol oynamaktadır.
- b- **Kredi ve Kredi Kartı Başvurusu Değerlendirme:** Kişilerin önceki finansal hareketlerinden kredi skorlarının hesaplanmasında, geleceğe yönelik ödeme durumunda finansal akışın sağlanıp sağlanamayacağına kadar gerekli olasılıkların hesaplanmasında büyük rol oynamaktadır.
- c- **Suistimal Tespiti:** Banka ve sigorta firmalarında suistimale yatkın kişilerin tespitinde kullanılmaktadır. Gerekli network analizleri ile hasarlardaki uzak ilişkilere kadar detay araştırması sağlanabilmektedir.

- d- **E-ticaret:** İnternet üzerinden ürün olarak farklı ürünlerin pazarlamasında da önemli rol oynamaktadır.
- e- **Tıp:** En önemli kullanım alanlarından birisidir. Hastalıklara ait laboratuvar sonuçları ile toplu değerlendirmeler ve çözüm odaklı incelemeler sonucunda geleceğe yönelik adım atmada önemli rol oynar. Gen mikro dizilimlerini incelemektedir.
- f- **Güvenlik:** Göz retina taraması, ses tanıma, parmak izi tanıma gibi güvenliği artırmaya yönelik adımların sağlanması esas alınmaktadır.
- g- **Ses tanıma:** Özellikle telekomünikasyon formlarında ses tanıma ile işlem yapmayı sağlayacak ve bu tanıma işlemi ile aktivasyonu sağlayarak müşterilere hızlı bilgi sağlama hizmeti sunmaktadırlar.
- h- **İmza ve Resim gerçeklik**
- i- **Kötü içerikli mail ayıklama:** Büyük kuruluşlarda her gün oluşan operasyonel mail okuma ve cevaplama işlemlerini makine öğrenmesi ile kategorileştirerek, otomatik cevapların verilmesi sağlanmaktadır.
- j- **Fotoğraf & Video İşleme:** Resim ve videolarda bulunan nesnelere tanınmasında önemli rol oynamaktadır.
- k- **Duygusal Durum Anlama:** Video ve resimlerde bulunan kişilerin duygularını ortaya koyabilmeye yarayan yöntemdir. Özellikle polis incelemesinde ve suistimal süreçlerinde kullanılmaktadır.
- l- **Ar-ge çalışmaları (sürücüsüz araç vb.):** En çok makine öğrenmesine bağlı olarak yapay zekaya ihtiyaç duyulan çalışmalardır. Sürücü olmadan bir aracın neler yapması gerektiği ve ani karar alma mekanizması için kullanılmaktadır.
- m- **Doğal Dil İşleme (NLP):** Artan teknoloji ile kullanılmaya başlanan fotoğraf ile yazı anlama ya da chat-bot uygulamaları ile müşterilerin yazdığı yazıları anlamaya yarayan ve gerekirse tercüme etmeye yarayan bu uygulama makine öğrenmesi ile farklı noktalara gelmiştir.
- n- **Eğitim:** Öğrencilerin başarı durumlarını inceleyerek, başarısızlıklardaki sebepleri bulmaya yarayan yöntemlerin başında gelmektedir.

### 2.1.5 Sigortacılıkta Makine Öğrenmesi

Sigortacılık alanında makine öğrenmesi birçok farklı konuda kullanılmaktadır.

- a- **Suistimal (Fraud) Tespiti:** Sigorta şirketleri ve Sigorta Bilgi Merkezi gibi havuz kuruluşları, hasar verilerinde uygunsuzlukları tespit etmeye çalışmakta ve potansiyel suistimal şüphelilerini ortaya koymak için makine öğrenmesi yöntemlerini kullanmaktadır. Bu kapsamda dolandırıcılığın önüne geçilmesi hedeflenmektedir.
- b- **Çapraz Satış:** Tıpkı e-ticaret sektöründe olduğu gibi bir ürünü satın alan müşteri profillerinin başka ürünlere de olan hassasiyeti ve yeni ürün alabilecek potansiyeldeki kişilerin analiz edilerek acentelere ve satış uzmanlarına yönlendirmelerde bulunmaktadır.
- c- **Risk Hesaplamaları:** Araştırma ve inceleme ile sigortalanması gereken yerlere ait risk oranlarının hesaplanması için makine öğrenmesi kullanılmaya başlanmıştır.
- d- **Kötü içerikli mailler:** Diğer sektörlerde de olduğu gibi Sigorta şirketlerine mail ile gelen yüksek sayıda mailin kümelenmesini sağlamaktadır.
- e- **Resim İşleme:** Hasar fotoğraflarından hasar oranını bulmaya ve buna göre de iş süreçlerinin ilerlemesini sağlamaktadır.
- f- **Tarife Hesaplama:** dönemlik hasar oranlarının artış ve azalışına göre öğretilmiş hesaplama verileri ile gelecekte nasıl tarifeler oluşturulacağı bilgisini sağlamaktadır.

## **2.2 DİJİTAL KULLANIMLAR**

### **2.2.1 Dijitalleşmenin Tanımı**

Dijitalleşme birçok açıklamada bahsedildiği üzere, birçok bilgi teknolojileri alt yapılarının entegrasyonunu kurmak ya da bu sistemlere sahip olmak yerine, dijital ile oluşturulan kaynakları, yeni gelir, büyüme ve şirkete daha fazla değer sağlayacak operasyonel sonuçlara dönüştürme sürecidir. Diğer bir söylem ile dijitalleşme, kuruma ait yeni iş modelleri oluşturmak, bilgi, şirket kaynakları ve dijital teknolojileri yeni yöntemlerle birlikte kullanarak mükemmeye yakın müşteri deneyimleri oluşturmak, en ideal ve en yeni hizmetleri, hizmet sonrası çalışmaları mümkün kılmak ve şirket kaynaklarını çok daha etkin kullanmak için teknolojiyi bu kaynaklara uygulamak anlamına gelir. Süreçler hızlıca ve en basit yöntemlerle dijitalleştirildiğinde sahip olunacak tek değer, şirketin koruduğu iş modellerini ve uygulamaların değerlerini değiştirmeden kullanılmakta olan sistemi kopyalayarak teknolojiyi uygulamak olur. Bu sayede verimlilik artar ancak büyüme beklentilerini sağlayacak bir kapasiteye sahip olunmamaktadır. (Accenture 2015)

Dijitalleşme, şirketlerin ve müşterilerinin deneyimlerinin farkındalığını ortaya koyarak, hizmetlerini onlar özelinde yeni teknoloji platformlarında gerçekleştirebilmektir.

### **2.2.2 Mobil Uygulamalar**

İnsanların ihtiyaçları ve taleplerine yönelimleri teknolojinin gelişmesiyle değişir. Bunların en başında akıllı cihazlar ve mobil uygulamalar gelmektedir. Mobil uygulamalar aracılığı ile telefonlar yalnızca sesli görüşme cihazından çıkarak rutin tüm işlemlerimizi anında sağlamaya yarayan cihazlar haline gelmiştir. Artı olarak da belirli özellikler sayesinde GPS ile yol bulmasını, görüntülü konuşma yapmasını ve günlük harcanan sportif aktivite verilerinin hesaplanmasını sağlamıştır. Artan mobil uygulama seçenekleri ile de insanlar için eskiden gerekli olan eforların kullanılmasına gerek kalmadan işlerin tamamlanması sağlanmış olup, boş vakit kavramının artmasıyla da mobil cihazlara yönelim gün geçtikçe yükselmektedir.

Aşağıdaki şekillerde belirtilen anket çalışması sonucunda Türkiye’de mobil uygulamaların ne kadar kullanıldığı bilgisi ve fazlalığını fark edebilirsiniz. Deloitte

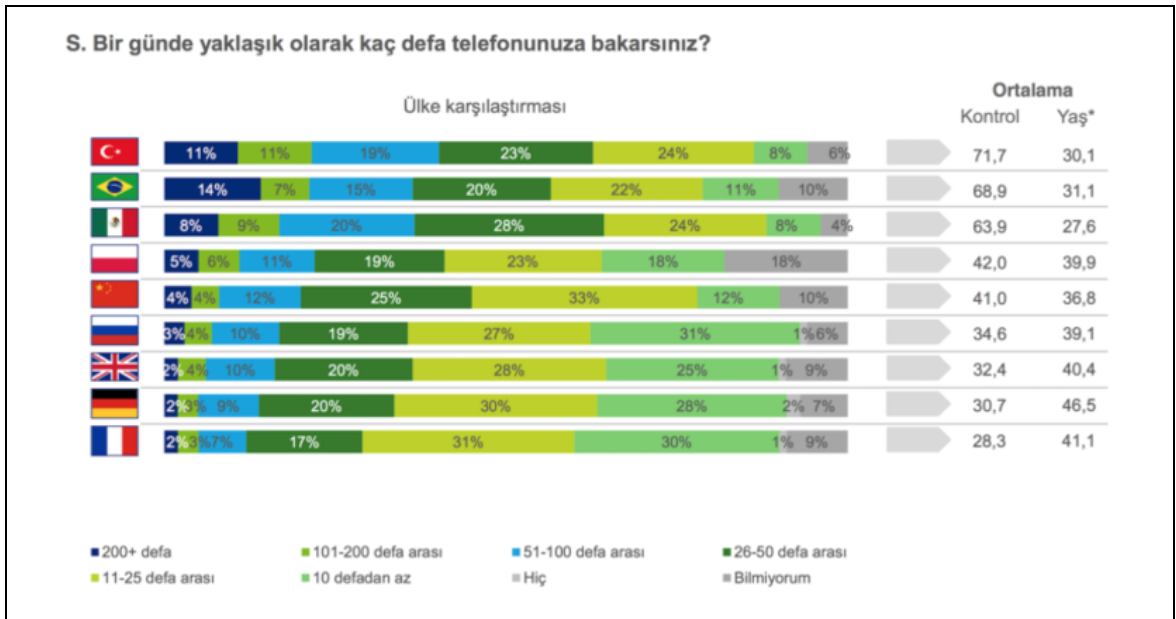
Global Mobil Kullanıcı anketine göre hazırlanan bu verilerde en çok telefonlarını kontrol eden oranının Türkiye’de olduğu gözükmektedir. Boş vakitlerde akıllı telefonlar için harcanan zamanda neler yapıldığı konusunda Şekil 2.3 ve Şekil 2.4 de oranlarıyla belirtilmiştir.

**Şekil 2.1 : Mobil Kullanım Ankete katılım sağlayan ülkeler**



Kaynak: Deloitte Global Mobil Kullanıcı Anketi

**Şekil 2.2 : Günlük telefona bakma rakamları**



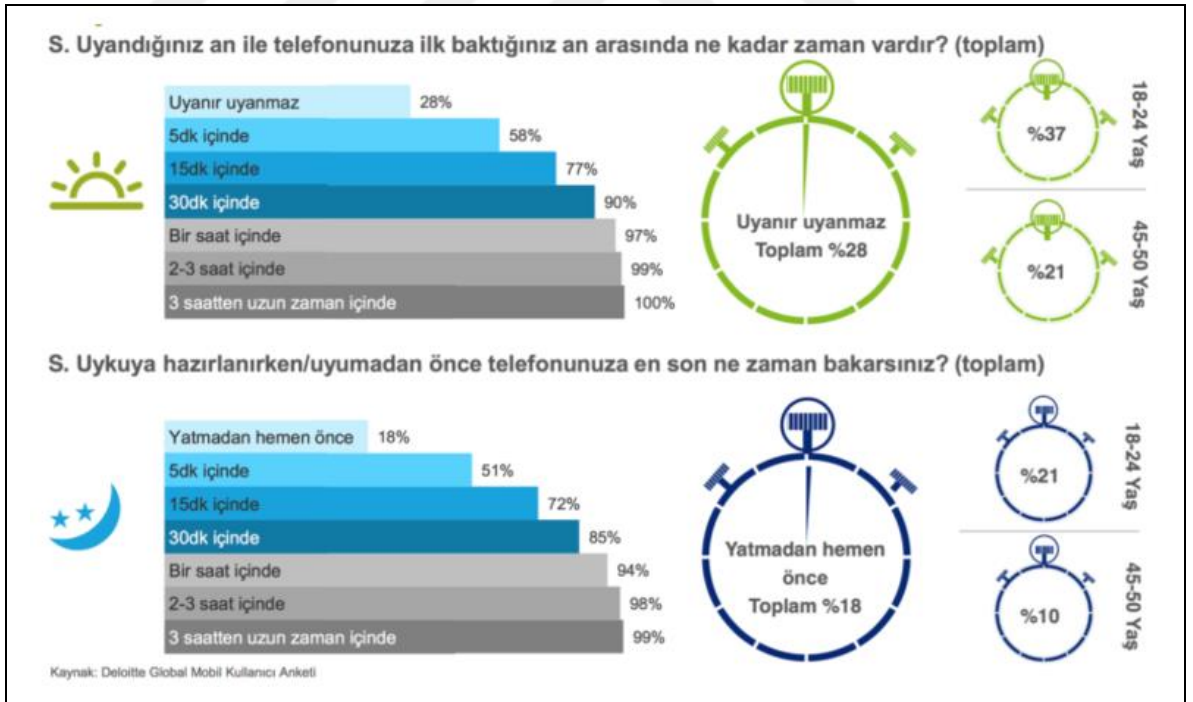
Kaynak: Deloitte Global Mobil Kullanıcı Anketi

Şekil 2.3 : Mobil Uygulama kullanıcı davranışları 1



Kaynak: Deloitte Global Mobil Kullanıcı Anketi

Şekil 2.4 : Mobil Uygulama kullanıcı davranışları 2



Kaynak: Deloitte Global Mobil Kullanıcı Anketi



Mobil uygulamaların gelişmesi, akıllı cihaz kullanımlarının artması ile beraber sosyal medyada var olmak insanlar için olmazsa olmaz bir duruma sebebiyet vermiştir. Bugün günlük aksiyonlar sonrasında kişiler tarafından fikirlerini ve ruhsal durumlarını ortaya koyan yazılar paylaşılmaktadır.

Tugay Özgirgin yaptığı tez çalışması sonucu olarak, yapılan çalışmanın en önemli adımı olan nitelik çıkarımı; kişi, kurum ve ürün hakkında sosyal medyadan elde edilen verilere duygu analizi yapıldıktan sonra yapılan nitelik çıkarımı, ilgili şey hakkında bir yargı elde edilebileceğini göstermiştir. Bu sonuçla ilgili kurum veya kişinin gelecekteki stratejisini kullanıcı yorumlarına bakarak yönlendirebilmesine olanak sağlayacağı düşünülmektedir. (Tugay Özgirgin 2016) Sosyal medyadan elde edilecek duygu analizi sonucu Trafik kazası önleme işlemimizin önemli bir parçası olacaktır.

### 2.2.3 Sigortacılıkta Mobil Uygulamalar

Teknolojinin ilerlemesi, mobil uygulamalarda kullanıcı tecrübelerinde hızla gelişmesi ile bu duruma ayak uydurmaya çalışan sigorta firmaları sektörde de farklı adımlar atarak müşterilerine hizmet sunmaya devam etmektedirler.

Sigorta firmaları üretim ve hasar hizmetleri için mobil uygulamalar geliştirmekte ve en kısa sürede işlemlerini sağlayabilmek için farklı yöntemler kullanmaktadırlar. Chat-bot teknolojisi gibi hızlı ve farklı sonuçlar üreten yöntemlerle alternatif dağıtım kanalları oluşturmaya çalışılmaktadır.

Sigortacılık sektöründeki örnek inovatif uygulamalar ve oluşturulan çözümler;

- a) **Lemonade**: Bu ürün 2016 yılında New York'ta üretime başlayan tüm hizmetlerini Chat-bot & Machine Learning üzerinden sağlayan, robot yazılımlarla müşterisine 6 saniye gibi bir rekor ile hasar ödemesi yapabilen bir uygulamadır. Ayrıca kazancının belirli bir kısmını müşteriler tarafından seçilen yöntemlerle bağış yapan ya da kazancı müşterilerine dağıtan bir firmadır. Kağıt kullanmadan tüm işlemleri mobil uygulama üzerinden tamamlayan, müşterinin kendini anlattığı video ile müşterisinin duygusal yönlerini ortaya çıkartarak

içerisinde suistimal olup olmadığı bilgisini araştıran, lokasyon ile verilen adres bilgisine ait evin risk durumunu ortaya çıkartan bir uygulamadır.

- b) **State Farm – Driver Feedback:** Amerika Ulusal Karayolu Trafik Güvenliği İdaresi (the National Highway Traffic Safety Administration) ‘ne göre sert ivmelenme, sert yavaşlama ve güçlü sapmalar en riskli trafik davranışlarıdır. Akıllı telefonlarda bulunan ivmeölçer, hız ölçer ve GPS ile müşterilerinin sürüşlerini skorlayarak bir sonraki sürüşlerinde iyileştirmeleri için bilgiler sağlamaktadır. Bu uygulamanın aynı alt yapısında Türkiye’de de birkaç firma mobil uygulamasında kullanılmaktadır.
- c) **SBM:** Türkiye’deki örneklerden en önemlisi olan Mobil KTT, uzun bir süredir araç sahiplerinin kaza yapması durumunda kağıtlara doldurarak karşılıklı anlaşmayı sağladığı tutanağı dijital ortamda hızlıca oluşturulmasını sağlamaktadır. İmza, kaza fotoğrafları, sürücü belgesi, ruhsat gibi tüm evrakların bilgileri bu uygulama ile sağlanmaktadır. Sektörde inovasyon yaklaşımının önemini çok hızlı ortaya koymuştur.
- d) **Zurih – What IF?:** Bu uygulama ile müşterilerinin sigortalamak istedikleri konutların risk analizini çıkartarak farklı sigorta planları önermektedir. Sigortalanacak kapsananın özelliklerini ortaya koymaktadır.
- e) **Diğer:** Sigorta firmaları operasyonlarını hızlıca sağlayabilmek için mobil uygulamalar oluşturmuştur. Türkiye’de de Anadolu Sigorta **Sigortam Cepte**, Allianz Sigorta **Allianz’ım** , Axa Sigorta **MyAxa** ve **AxaİyiSürücü** gibi uygulamalarla sektörde yer bulmuştur.

## **2.3 TRAFİK KAZASI ÇALIŞMALARI**

### **2.3.1. Çalışma 1**

Nejdet DOGRU-Abdulhamit SUBASI (2013), veri madenciliği teknikleriyle kullanan gerçek zamanlı trafik kazası tespit yöntemi sunmuştur. Çalışma yaptıkları yöntem, mikroskopik trafik simülasyon ortamında değerlendirilmiştir. Kazaların trafik verilerinde anormallik olarak görülebileceği ve bu anormallikleri tespit etmek için veri madenciliği tekniklerinin kullanılabilmesi gösterilmiştir. Bir trafik kazası tespit edilmez, olay hakkında diğer araçlara bilgi verilebilir ve bu, sürücülerin daha fazla olaydan kaçınmak veya rotalarını değiştirmeleri için bir karar verme şansı vermektedir. Kazaları saptamak, aynı zamanda yol görevlilerinin sorunu hemen ortadan kaldırmasına da yardımcı olacak. Buna ek olarak önerilen çalışma olası kaza ile ilgili tahmini konum bilgilerini sağlayabilmektedir.

Bu çalışmada verilen kararlar kaza sonrasındaki hayati süreçlerin düzgün devamlılığı için önemlidir. Ve arkasından gelecek olan riskleri de bertaraf etmek amaçlanmaktadır. Trafikte sürecin olağan gitmemesi durumunda değerlendirilmesi esas alınmaktadır.

### **2.3.2. Çalışma 2**

Tarik Kaya (2015), Güvenli trafik düzeninin sağlanabilmesi için sürücülerin bilinç düzeylerinin artırılması ve kaza riski konusunda bilgilendirilmesi gerekmektedir. Bu çalışma da KTT üzerinden alınan veriler ile kaza riski tahmini yapılmaktadır. 9 ölçüt, 5000 adet gerçekleşmiş trafik kaza verisi kullanılarak, meteoroloji veri tabanından alınan hava raporları verileri ile birleştirilmiş ve veri seti ortaya çıkarılmıştır. Trafik kazalarının sebepleri arasında cinsiyet, yaş aralığı gibi faktörler ile ilgili olduğu önemle vurgulanmıştır. Bu çalışma sonucunda trafik kazalarının önemli etmenlerinden olan aşırı hız, alkollü araç kullanımı, tehlikeli ve dikkatsiz araç kullanma gibi faktörler daha çok genç erkekler tarafından gerçekleştirildiği tespit edilmiştir. Çalışma kapsamında geliştirilen kaza riski tahmin sistemi ile sürücülerin ve polis merkezlerinin düzenli olarak bilgilendirilmesini sağlayacak bir uygulama geliştirmişlerdir. Kaza riskinin arttığı durumların bilinmesi ve uyarı sistemlerinin devreye girmesi kazaların azalmasını sağlayacağını belirtmiştir.

Geliştirmiş olduğu kaza riski hesaplama sisteminde beş adet algoritma kullanılmakta olup, sınıflandırma algoritmalarından CART(SimpleCart), C 4.5(J48), Bayesyen(NaiveBayes), Regresyon(OnerR), K en yakın komşu (IBk) uygulanarak, en iyi sonucu veren C 4.5 algoritması kullanılmıştır. Kaza riski doğru pozitif olarak tahmin edilen örnek sayısının, tüm pozitif tahminlere oranını veren kesinlik ve duyarlılık değerleri en yüksek olan C 4.5 algoritmasında 0.906 sonucu elde edilmiştir.

Çalışma sonunda ise veri setinin büyütülmesi gerekliliği ve bu sayede daha net sonuçlara ulaşılabileceği belirtilmiştir. Bu çalışmayı inceledikten sonra trafik kazası önceden anlama için veri modelimini oluştururken sürücü (evli mi ?, yaptığı iş, hız tutkunluğuna sahip mi?,duygusal durumu vs.), hava şartları, yol bilgileri (çift şerit , tek şerit , yokuş vb.) neler olduğunu tespit ederek daha doğru veri setleri oluşturmayı amaçlanmıştır.

### **2.3.3. Çalışma 3**

Beshah ve Hill (2010), çalışmalarında Etiyopya Trafik verileri üzerinden sağlamıştır. Yol faktörünün önemini vurgulayarak yaptığı çalışmada yola ait bazı karakteristik özelliklerin kazaya etkisini ortaya koymuştur. Çalışmada, karar ağacı, NaiveBayes ve K-en yakın komşu teknikleri seçilerek sonuçları incelenmiştir. Veriler incelendiğinde sürücüdenden ziyade dış etkenlere ait verilerin titizliği ile karşılaşılmaktadır. Çalışmasında makine öğrenmesini sağlamak için Weka yazılımı kullanılmıştır. Test edilen verilerin sonuçlarına bakıldığında bu üç model arasında anlamlı olmayan farklar görülmektedir. Sonuç olarak ise PART algoritması sonuçlarını en doğru sonuçlara ulaştığını belirtmiştir.

### **2.3.4 Çalışma 4**

Çinicioğlu ve arkadaşları (2013), Silivri Bölge Trafik Şube Müdürlüğü ve İlçe Jandarma Trafik Tim Komutanlığı'ndan sağladıkları maddi hasarlı trafik kaza tespit tutanaklarından veri seti oluşturmuşlardır. Bayes ağı ile öğretimi sağlayarak tahminleme çalışması yaptıklarında %71,45 doğru tahminleme yaptığı ve modele göre de %17 ilerleme kaydettiği belirtilmiştir. Ancak Türkiye'de KTT verilerinin doğru girilmemesi ve bu sebeplede oluşan kirliliği veri setleri çalışmaları zorlandırmıştır. Silivri Trafik

Müdürlüğünden alınan verilerin anlaşmasız Kaza Tespit Tutanaqları olması sebebiyle, anlaşmalı kaza bilgileri bu modele dahil edilememiştir. Çalışmasında gerçek kaza bilgileri veri setiyle test edilmesi, trafik kazaları birçok yönüyle inceleyen güncellenebilir model oluşturması açısından önem taşımaktadır.

Kaza tespit tutanağındaki veriler ve açıklama alanlarının makine öğrenmesi ile tek başına değerlendirilmesi açıklama alanlarına giriş ile sadece esas alınması gerekmektedir. Bu sebeple mobil uygulamalarda bu verileri pekiştirmek oranlarımızın daha sonuca yakın çıkmasını sağlayacaktır.

### **2.3.5 Çalışma 5**

Mehmet Akif ÇELİK (2010) yaptığı çalışmada, Akıllı Trafik Sistemlerini inceleyerek aşırı hızın etkileri üzerinde durmuştur. Akıllı Ulaşım Sistemleri trafik altyapısına ve araçlar için güvenliği artırma, ulaşımı kolaylaştırma çalışmaları olduğu açıklanmıştır. Trafik lambaları, uzaktan kontrollü elektronik bilgilendirme panoları, kritik ulaşım bilgilerinin toplanması ve değerlendirilmesi, otomatik yol denetim sistemleri Akıllı Ulaşım Sistemleri uygulamalarına örnek olarak verilebilmektedir. Çalışma da “Akıllı Trafik Sistemleri” ’nin tanımı, içeriği, kullanılan teknolojiler ile trafik sıklıklar ve trafik güvenliğine ne gibi etkilerde bulunduğu araştırılarak sonuçları sunulmuştur. Antalya ve Konya illeri çevresinde belirtilen akıllı uygulamanın kapsamında kullanılan ledli göstergelerin yaralanma ve ölüm sayılarındaki iyileştirici etkisi belirtilmiştir. Bu uygulamanın Türkiye’de nasıl kullanılabileceği konusunda belirli önerilerde bulunmuştur.

### **2.3.6 Çalışma 6**

Nuriye KABAKUŞ ve arkadaşları 2012 yılında yaptıkları çalışmada, Erzurum’da meydana gelen kazaların Coğrafi Bilgi Sistemleri ile değerlendirilmesini sağlamıştır. Çözümler üretebilmek için analiz aşamasında Erzurum ili ilçe bazında 19 tane alana ayrılmıştır. CBS yardımıyla elde edilen haritalar incelendiğinde kazaların en çok merkezde meydana geldiği görülmüştür. Aşağıda belirtilen parametreleri de çıkartarak en çok kazaya sebebiyet veren durumları oluşturduğundan bahsetmişlerdir.

- a) Nüfus
- b) Sosyo-ekonomik durum
- c) Gelişmişlik düzeyi
- d) Kayıtlı araç sayısı
- e) Endüstri kuruluşlarının sayısı
- f) Trafik yoğunluğu

Kazaların en çok merkez ve merkeze yakın ilçelerde (başta Pasinler) meydana geldiğini, kazalarında trafik hacmi, nüfus, gelişmişlik seviyesi, sosyo-ekonomik düzeyin etkisiyle araç sahipliliğinin fazla olması sebebiyle meydana geldiğini ispatlamak istemişlerdir.

### **2.3.7 Çalışma 7**

Miao Chong ve arkadaşları,1995-2000 yılları arasında otomobil kazalarına ait verileri inceleyerek, sinir ağı ve karar ağacı yöntemleri ile sürücülerin yaralanma şiddetini tahmin etmek için araştırmalarda bulunmuşlardır. Deneyle sonunda ölümcül ve ölümcül olmayan yaralanmalarına ait oluşturulan modellerin, diğer denedikleri türlerden daha performanslı olduğunun farkına vardılar. Ölüm seviyesinde sınıflandırmanın sosyal olarak da, ekonomik olarakda ülkelerde büyük maliyetlere yol açtığıın altını çizmektedirler. Ancak çalışma yaptıkları kaza bilgileri içerisinde yüzde 67 verinin hıza ait bilgisinin bulunmaması sebebiyle aşırı hızın etkisinin bu verinin eklenerek model oluşturmasında daha sağlıklı olacağı kanısına varmaktadırlar.

### **3. VERİ VE YÖNTEMLER**

#### **3.1 VERİ KÜMESİ TANIMI**

Bu çalışmada kullanılacak olan veriler, sigorta firmasına hasar yaptığını ihbar ederek araçlarını yaptıran sigortalılara ait verileri kullanılarak hazırlanmıştır.

Hasar yapılan kaza yeri ve sürücüler olmak üzere iki (2) farklı veri kümesi üzerinden çalışma yapılmıştır. 1500 adet lokasyon bazlı olayı HASAR ve 1000 adet kaza yapan SÜRÜCÜ profillerini oluşturan veriler olarak belirtilecektir.

#### **3.2 VERİ KÜMESİNİN HAZIRLANMASI**

Trafik Kazasını Olma Olasılığının hesaplanması üzerine yapılacak bu çalışmada farklı zamanlarda gerçekleşmiş hasar kayıtlarının toparlanması ve değerlendirilmesi sonucunda oluşturulmuştur. Her iki veri kümesinde de (HASAR ve SÜRÜCÜ) kadar parametre yer almaktadır. Veri kümesi içerisinde sürücüye ait veriler, riskin olduğu günün hava durumu, riske ait lokasyon verileri, fiziksel şartlar ve riskin gerçekleşmesi durumunda gerekli olan detay bilgileri mevcuttur.

##### **3.2.1 Veri Kümesinin İçeriği**

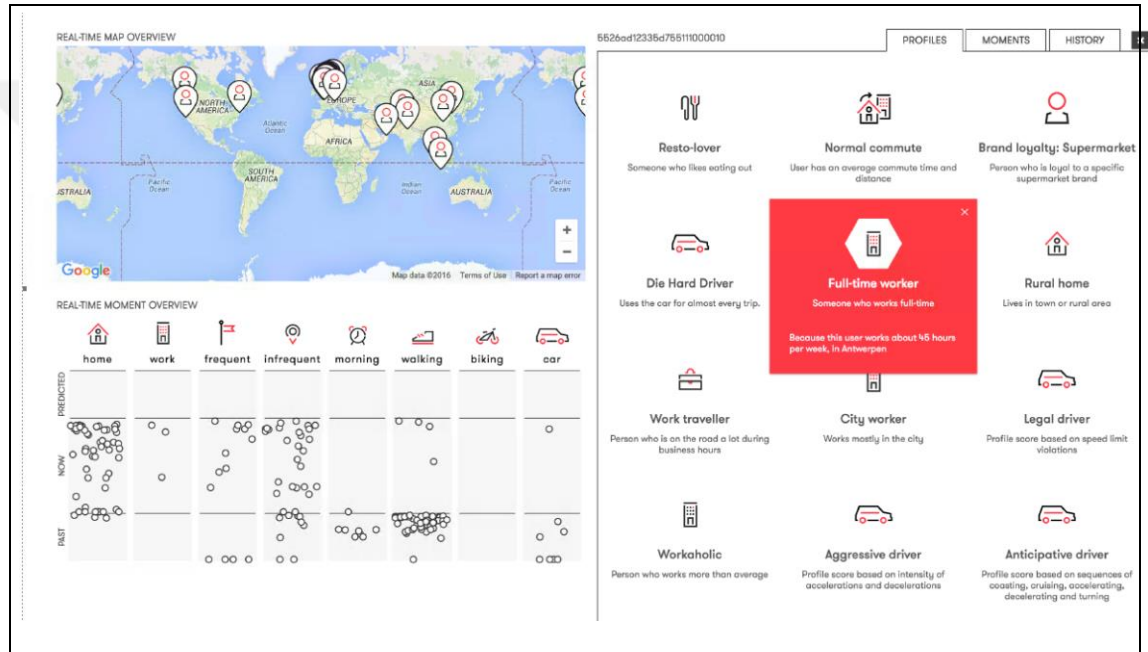
###### **3.2.1.1 Sürücü Verileri**

Riski oluşturan ya da kaza gerçekleştiren sürücülerin detay bilgilerini içermektedir. Riski arttıracak parametreler içerisinde, yaş aralığı, medeni durumu, sürüş kabiliyetleri, günlük araç kullanım oranı, sürüş esnasındaki duygusal durum gibi verileri içermektedir.

Trafikte ilerleme sırasında arka planda çalışan mobil uygulamalar ile sürücünün login olması durumunda kimlik numarasına ilişkili olarak doğum tarihini, medeni durumunu, cinsiyetini öğrenilebilmektedir. Meslek bilgisini poliçe yaptırma aşamasında beyan edilmektedir. Sigortacılıkta poliçe gerçekleştirme işlemi sırasında bu data oldukça az giriş sağlanmaktadır, bu sebeple sürücülerini segmentleştirmek adına yapılacak çalışmalar

için yine dijital çözümler getirilebilir ve lokasyon bazlı uygulamalar yardımıyla gün içerisinde kişilerin bulunduğu lokasyonların zaman aralıklarına göre çoğunlukları ile kategorileştirilmesi sağlanmaktadır. Segmentin önemi, özellikle ticari araç sigortalarında, mesleğin taksici, minibüs – otobüs şoförü olması oldukça önemli olup, risk faktörünü artırıcı sebeplerdendir. Şekil 3.1’de Insurtech lokasyon bazlı Satiencia adlı örnek bir uygulamadan sürücü profillerine göre kategorilendirme üzerine bir görsel içermektedir.

**Şekil 3.1 : Lokasyon bazlı uygulamalar ile örnek sürücü segmentleştirme**

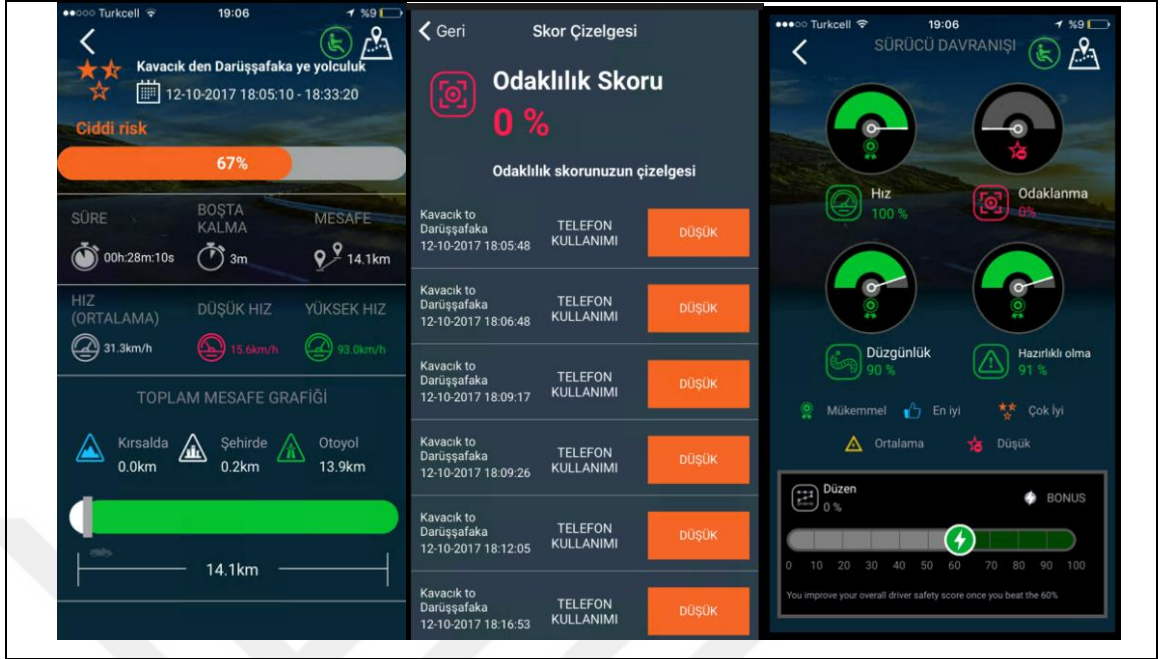


Kaynak: Satiencia App

Hız oranı ve günlük araç kullanımı, telefonda arka planda çalışacak olan ivme ölçer ve hız ölçer ile hesaplanabilmektedir. Telefon konuşma oranı ise, uygulama arka planda çalışırken, hız göstergesi de yürümeden oldukça yüksek hız olduğu durumlarda, telefon ile konuşma süresi ve sıklığı ölçülerek yapılan yol zamanına göre oranlanarak riski ne derecede sürüşte sahip olduğunu belirtmektedir. Sürücü davranışları dijital çözümler ile tespit edilebilir, kaza yapma riski modelinde önemli yer almaktadır. Şekil



**Şekil 3.2 : Mobil uygulama ile sürüş kabiliyetlerinin ölçülmesi**



Kaynak: Precision Telematic App

Duygusal durum eğer ki olması durumunda uygulamaya kaydını sosyal medya uygulamaları üzerinden sağlayarak o gün içerisinde en son paylaşılan yorumun neticesinde değerlendirilebilir.

**Tablo 3.1 : Sürücü Veri Kümesi Açıklamaları**

	Sürücü Alan Adı	Açıklama
1	surucu_id	Sürücünün sistemde tanımlı id'si
2	yas	Sürücünün yaşı
3	medeni_durum	Sürücünün medeni hali (evli, bekar)
4	telefon_orani	Sürücünün araç ile giriş sırasında ne kadar telefon ile görüştüğü yüzdesi
5	hiz_orani	Sürücünün araç ile geçtiği sokak ve caddelere göre hız ortalaması oranı
6	duygusal_durum	Sürücünün kaza anında ve risk durumunda duygusal durumu
7	gunluk_arac_kullanim	Sürücünün risk anına kadar araç kullanım zamanı
8	cinsiyet	Sürücünün cinsiyeti
9	meslek	meslek

### 3.2.1.2 Hava Durumu Verileri

Hava durumu verileri risk durumunda oldukça önemlidir. Özellikle eğimli, engebeli yollarda mevsimlere göre kaza gözlenme oranlarının artması, kar yağışlı ya da soğuk olan il ve ilçelerde yolların buzlanma riski de göz önünde bulundurularak kategorilendirilmiştir. Risk anında hava durumu bilgisi IOS ve Android telefonlarda Weather Api adı altında çalışan ve anlık bulunan lokasyonun hava durumunu verecek şekilde kullanılmaktadır.

Geçmişe ait verileri de Meteoroloji web sitesinde temin edilebilmektedir.

**Tablo 3.2 : Hava Durumu Veri Kümesi Açıklamaları**

	Hava Durumu Alan Adı	Açıklama
1	tarih	Hava durumunun tarihi
2	saat	Hava durumunun saati
3	hava_durum	Hava yağış durumu (yağmurlu, karlı, güneşli,ılık)

### 3.2.1.3 Lokasyon Verileri

Hasarın gerçekleşmesi ya da risk sırasında sürücünün bulunduğu adres bilgisi oldukça önemlidir. Sokak ya da caddenin hız sınırının ne kadar olduğu, riskin meydana gelebileceği saat içerisinde sokakta trafiğin ne oranda olduğu bilgileri riskin gerçekleşmesi bakımından oldukça önemlidir.

Yol tipi, bir riskin gerçekleşmesinde ve hasar nedeninin nasıl olabileceği konusunda ön bilgi sağlamaktadır. Sürüş alışkanlıklarının cadde ve sokakların yapılarına ne kadar uygun olup olmadığı önemli bir faktördür.

Enlem ve boylam bilgisi risk sırasında telefonlardan gönderilen konum bilgisi verileri alınmaktadır. Geçmişe ait hasar kayıtlarının enlem boylam verileri Sigorta Bilgi Merkezi'nde tutulabildiği kadarı ile mevcut olup, ayrıca 3rd parti olarak uygulamalara yerleştirilen SBM'nin sigorta firmalarına açtığı lokasyon bulma servisi, Google Place Api ya da Yandex SDK'ları ile full text adres bilgilerine ait enlem boylam bilgileri üretilebilmektedir. Sürücülere ait risk haritalarını çıkartmak için oldukça önemli ve zengin içerikli bir bilgidir.

**Tablo 3.3 : Hasar Detay Lokasyon Veri Kümesi Açıklamaları**

	<b>Hasar Detay</b>	<b>Açıklama</b>
1	hasar_id	Hasara ait değer
2	full_text_adres	Adres bilgisi
3	enlem	Adres Enlem Bilgisi
4	boylam	Adres Boylam Bilgisi
5	yol_tipi	Adresin bulunduğu lokasyonun fiziksel durumu (kavşak, tek şerit yol)
6	hiz_siniri	Sokak ya da caddenin sahip olduğu hız sınırı
7	yol_kullanim_orani	Adresin son 1 saat içerisinde kullanım değerleri

#### 3.2.1.4 Kaza Detay Bilgi Verileri

Hasara ait veri modeli, kazanın tüm detaylarını kapsayan parametrelerden oluşmaktadır. Bu değerler, kazanın gerçekleşmesine temel oluşturan tüm risk değerlerinin kapsamındadır.

Kazaya ait değerleri incelediğimizde,

**Tablo 3.4 : Hasar Kaza Detay Veri Kümesi Açıklamaları**

	<b>HASAR Kaza Detay Alan Adı</b>	<b>Açıklama</b>
1	risk_id	Riske ait id değeri
2	surucu_id	Sürücüye ait id değeri
3	risk_tarihi	Risk ve hasara ait tarih bilgisi
4	risk_saati	Risk ve hasara ait saat bilgisi
5	hasarli_parca	Aracın kaza sırasında aldığı hasarın hangi bölümde gerçekleştiği bilgisi(kapı, kaput, bagaj)
6	hiz	Sürücünün risk sırasındaki hızı
7	arac_tipi	Araç modeli riski değiştirir
8	hasar_sebebi	Hasar'ın gerçekleşme nedeni

### 3.2.2 Machine Learning Çalışmaları

Trafik Kazasını Önceden Tahmini üzerine oluşturulacak model için öncelikle gelen parametrelerin gerekli temizleme ve düzenleme işlemlerinden geçirilerek stabil hale getirilmesi gerekmektedir.

Eğitim veri kümeleri ile oluşturulacak kütüphanenin düzgün çalışması, verilerin incelenerek temizlenmesi, referans alarak belirli kuralların kapsanacağı kümelerin detay analizleri yapılacaktır. Referans verilerin gruplanması rolünde Anadolu Sigorta Hasar Yönetim Müdürlüğü tarafından yardım alınmıştır.

#### 3.2.2.1 Parametrelerin Değerlendirilmesi

Sürücü bilgilerinde yaş , telefon\_orani, hiz\_orani ,gunluk\_arac\_kullanım, alkol\_var\_mi parametreleri numeric değerlerdir. Sürücü verileri içerisinde duygusal\_durum parametresi sosyal medya içeriği olup olmaması ile ilişki olarak NULL değer gelebilir. Bu değer olmadan da ölçüm yapılabilmelidir.

**Tablo 3.5 : Sürücü Veri Kümesi Parametre Açıklamaları**

	Sürücü Alan Adı	Parametre Açıklaması
1	surucu_id	Sürücü Id's Guid bir değerdir
2	yas	0-100 arasında değerler alacaktır.
3	medeni_durum	E-Evli , B- Bekar
4	telefon_orani	0-100 arasında değer alacaktır.
5	hiz_orani	0-100 arasında değer alacaktır.
6	duygusal_durum	1-Mutlu , 2-Üzgün , 3-Stresli ,4- Endişeli ,5-Düşünceli, 6-Keyifli
7	gunluk_arac_kullanım	0-100 arasında değer alacaktır.
9	cinsiyet	E-Erkek , K-Kadın
10	meslek	Meslek grupları bilgileri text halinde gelecektir.

Adres veri setinde bulunan değerler içerisinde hiz\_siniri, kullanım\_orani, enlem ve boylam değerleri float değerler alacaktır. Adres verileri içerisinde full\_text\_adres verisi

NULL gelebilir yalnızca enlem, boylam verisi gelebilir. Bu sebeple bu veri olmadan da ölçüm yapılabilir.

**Tablo 3.6: Hasar Lokasyon Veri Kümesi Parametre Açıklamaları**

	Lokasyon Alan Adı	Parametre Açıklaması
1	risk_id	Guid Id
3	enlem	Float değer
4	boylam	Float değer
5	yol_tipi	Adresin bulunduğu lokasyonun fiziksel durumu (kavşak, tek şerit yol)
6	hiz_siniri	Int değer 0-200
7	kullanim_orani	Float değer 0-100

**Tablo 3.7: Hasar ve Sürücü İlişki Veri Kümesi Parametre Açıklamaları**

	HASAR Kaza Detay Alan Adı	Açıklama
1	risk_id	Guid Değer
2	surucu_id	Guid Değer
3	risk_tarihi	DateTime - Date
4	risk_saati	DateTime - Hour
5	hasarli_parca	(kapı, kaput, bagaj)
6	hiz	0-200
7	arac_segment	K-Kamyon O-Otobüs M-Minibus D-Dolmuş T-Taksi A-Otomobil
8	hasar_sebebi	Hasar'ın gerçekleşme nedeni

### **3.2.2.3 Referans Değerlerin Oluşturulması**

Parametrelere ait verileri değerlendirdiğimizde sayısal ve karakter içeren değerler olduğu görülmektedir. Kategori şeklinde hazırlanmış parametreler için hasar uzmanlarından destek alınarak referans değerler oluşturulmuştur. Referans değerler risk oranlarını etkilemesinden dolayı ihtiyaçlar doğrultusunda değişebilecektir. Özellikle hasar ve sürücü etkileşimini içeren derece gibi alanlara ait skorlamaların güncellenebilir olması, eğitilmiş dataya oluşturacağı etkisi oldukça önemlidir.

Elde edilen verinin referans değerler aracılığı ile dönüştürülmesi ile ilgili olarak gerekli tablolar yaratılmıştır. Referans tabloların oluşturulma süreçleri detaylandırılmıştır.

### **3.2.3 Azure Machine Learning Platformunun Kullanılması**

Microsoft Azure Machine Learning, tahmine dayalı modelleri analiz çözümleri olarak hızlı bir şekilde oluşturmayı ve dağıtmayı mümkün kılan bulut tabanlı ve tahmine dayalı analiz hizmetidir. Kullanıma hazır algoritma kitaplığıyla çalışabilir, bunları kullanarak internete bağlı bir bilgisayarda model oluşturabilir ve tahmine dayalı çözümünüzü hızlıca sağlayabilmektedir. (Microsoft, 12.07.2017)

Azure platformunda veri kümesini oluşturmak için csv ya da arff dosya formatında hazırlanması gerekmektedir. Proje geliştirmesi sırasında .csv dosya türü tercih edilmiş olup, üç farklı veri kümesi oluşturulmuştur.

- a) Hasar Detay Bilgileri
- b) Sürücü Detay Bilgileri
- c) Sürücü Hasar ilişkisi

#### **3.2.3.1 Veri Kümelerinin Oluşturulması**

Azure ML platformu, eğitim verilerinin kullanılabilmesi için oluşturulan csv dosyalarının sisteme yüklenmesi işlemi sonrasında içeriklerinin kontrol edildiği çalışma alanıdır. Sistem eklenen veri kümelerini hızlıca analiz ederek, boş değerleri bulunan sütunları tespit etmekte ve uyarılmaktadır. Bu sayede çalışmayı ilerletmeden önce hangi datalarda düzenleme yapılması gerekliliğini ortaya koyabilir ve veriler içerisindeki değişkenliklerin az ya da çok olduğunu grafiklerle tamamlamaktadır. Şekil 3.3 ve Şekil 3.4 'de olduğu gibi sisteme yüklenmiş olan Sürücü ve Kaza Detay Bilgileri alanlarında

hangi değerlerden kaç kayıt olduğu, boş kayıt sayısı, veri türü gibi tüm bilgileri erişilebilmektedir.

**Şekil 3.3 : Sürücü Detay Bilgileri Azure Platform Görünümü**

rows	columns								
339	9	SURUCUID	ENLEM	BOYLAM	YAS	CINSIYET	MEDENI_DURUM	SURUCU_SEGMENT	ARAC_TIPI
view as									
		KMQUCBNCPTG9E108	40.933279	29.318969	25-35	E	B	Calisan	Automobile
		ONHRSLPOAQBHTTC	40.888626	29.237694	45-65	E	E	Emekli	Automobile
		TM6D6R1FY6FBE54A	40.868723	29.265367	25-35	E	B	Calisan	Automobile
		13HXQ948WH5T5HPJ	40.941203	29.322781	35-45	E	E	Sofor	Truck
		GNMKS2NNUW362Q0P	40.923895	29.290197	25-35	E	B	Sofor	Minibus
		D8883YKQ3C7B8QZE	40.914504	29.322912	25-35	E	B	Calisan	Automobile
		U2W67SYVOL4T04XA	40.932362	29.341612	25-35	E	B	Calisan	Automobile
		PE28F99JZTV1SN1O	40.93337	29.309816	25-35	E	B	Calisan	Automobile
		ZZDAJYKED232I0S	40.898806	29.267123	25-35	E	B	Calisan	Automobile
		Q9KN09XTVS7ZGZJZ	40.855947	29.291293	25-35	E	B	Sofor	Truck
		LH4J3IOJR4CHS9CQ	40.86164	29.278184	25-35	E	B	Isci	Van
		3GOSWZHVSL9ZY3EF	40.936109	29.328899	25-35	E	B	Calisan	Automobile
		Z3CUWK510SY55NF4	40.929605	29.307915	25-35	E	B	Sofor	Bus

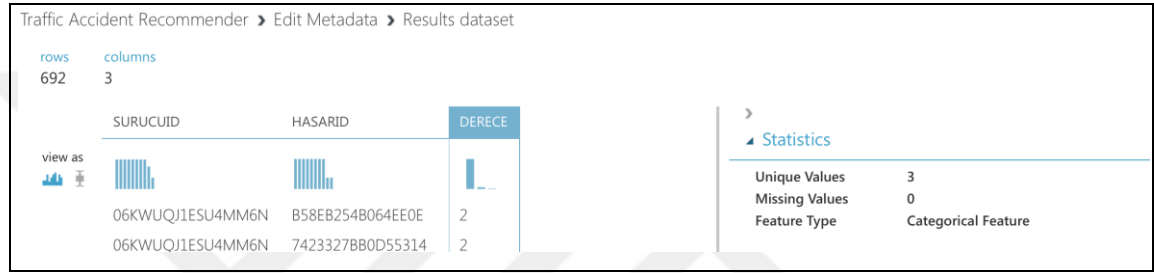
**Şekil 3.4 : Kaza Detay Bilgileri Azure Platform Görünümü**

rows	columns								
407	12	HASARID	ENLEM	BOYLAM	SAAT	HAVA_DURUMU	MEVSIM	ALKOL_DURUMU	YOL_TIPI
view as									
		EBC9D5906ECB6A70	40.922454	29.316629	2017-12-11T00:00:00	II k	Sonbahar	H	Cadde
		9F7F248134754039	40.896002	29.250288	2017-12-11T00:00:00	II k	Sonbahar	H	Cadde
		4587AEB243905436	40.943437	29.317616	2017-12-11T00:00:00	II k	Sonbahar	H	Kopr
		1A9BFB00EACD3878	40.891853	29.260496	2017-12-11T13:30:00	II k	Sonbahar	H	Cadde
		CCE3F95CE8EB67FC	40.891853	29.260496	2017-12-11T00:00:00	II k	Sonbahar	H	Sokak
		341D557D2C6E4005	40.891853	29.260496	2017-12-11T12:30:00	II k	Sonbahar	H	Cadde
		C41028831419BE3E	40.891853	29.260496	2017-12-11T00:00:00	II k	Sonbahar	H	Cadde
		C1DF24BE92A794AB	40.891853	29.260496	2017-12-11T00:00:00	II k	Sonbahar	H	Cadde
		93360F1747F7E59D	40.955396	29.350181	2017-12-11T01:10:00	II k	Sonbahar	H	Cadde
		7DA75DF0B51EB798	40.857235	29.288818	2017-12-11T00:00:00	II k	Sonbahar	H	Cadde
		8394049B7F067858	40.888382	29.238649	2017-12-11T00:00:00	Ya	Sonbahar	H	Cadde

### 3.2.3.2 Meta Data Dzenleme İşlemi

Veri kümelerinin içerikleri sistemin gerektirdiği özelliklere sahip olamayabilir. Bu sebeple Meta data düzenleme işlemi ile birlikte özellikle kategorileştirilmesi gereken önemli değerlerin düzenlenmesi, dataların formatlarında problemlerinin giderilmesi gibi birçok temeli hazırlayacak önemli işlemleri sağlar. Şekil 3.5 'de görüldüğü üzere belirlenen kolon bilgileri mevcut veri tipinden kategorik özellik tipine dönüştürülmesi sağlanmıştır.

Şekil 3.5 : Azure Metadata Dzenleme İşlemi



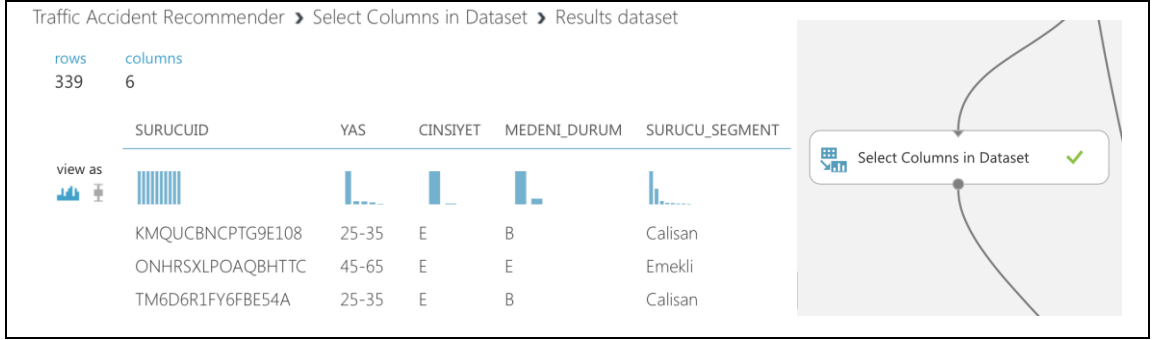
### 3.2.3.3 Veri Kümesi Kolon Ayıklama İşlemi

Sistemde projeye eklenen veri kümelerinin ve onlara ait tüm kolonlarının kullanılmayacağına kanaat verildiğinde, öğrenme işlemi sırasında işlem ağırlığını azaltması için "Select Columns In DataSet" başlığı altında işlem gerçekleştirilmektedir. Bu işlem ile hızlıca önemli kolonları seçerek iş akışına devam edecek veri kümesi oluşturulmaktadır. Şekil 3.6 'da görüldüğü üzere iş akışı olarak sisteme entegre edilen veri kümesi kolon ayıklama özelliği ile ihtiyaç duyulan kolonlara indirgenebilir ve işlemlere daha düzgün bir öğrenme datası oluşturmak adına önemli bir seçme işlemi sağlanmaktadır.

Azure ML platformunda yukarıdan aşağıya yerleştirilen iş akışları için kullanılacak sütunların sınırlandırılması gerekmesi durumunda veya gereksiz sütunları kaldırarak veri kümesinin boyutunu azaltmak için bu özellik yararlı olmaktadır.



### Şekil 3.6 : Azure Veri Kümesinden Kolon Seçme

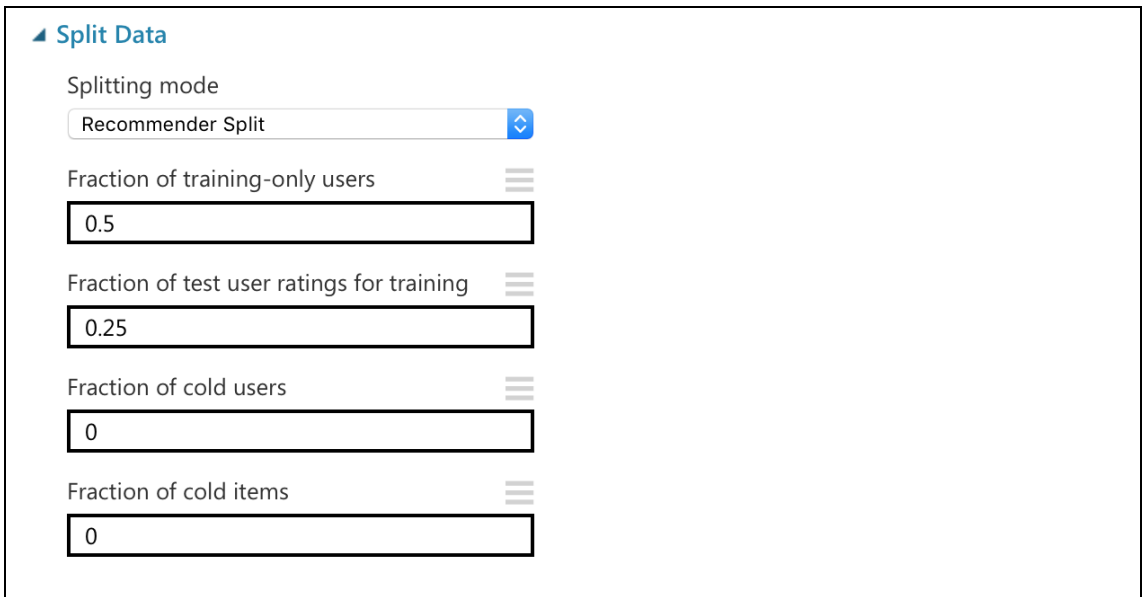


### 3.2.3.4 Veriyi Bölme (Split Data) İşlemi

Split data işlemi, modelin oluşturulması için eklenmiş olan data setinde, test datalarının belirtilen oranlarda ayrıştırılarak oluşturulan modelin kıyaslama yapılması amacıyla kullanılan yöntemdir.

Bu işlem, verileri eğitim ve test setlerine ayırmanızı gerektiğinde özellikle yararlıdır. Verilerin bölünme biçimi ihtiyaca göre özelleştirilebilmektedir. Bazı seçenekler verilerin randomize edilmesini desteklemektedir; diğerleri model türü için uyarlanmıştır. Şekil 3.7 'de Azure ML platformunda veri kümesini nasıl ayrıştırılması konusunda mevcut değerler ve parametreler gösterilmektedir.

### Şekil 3.7 : Azure Data Ayıklama Yöntemi



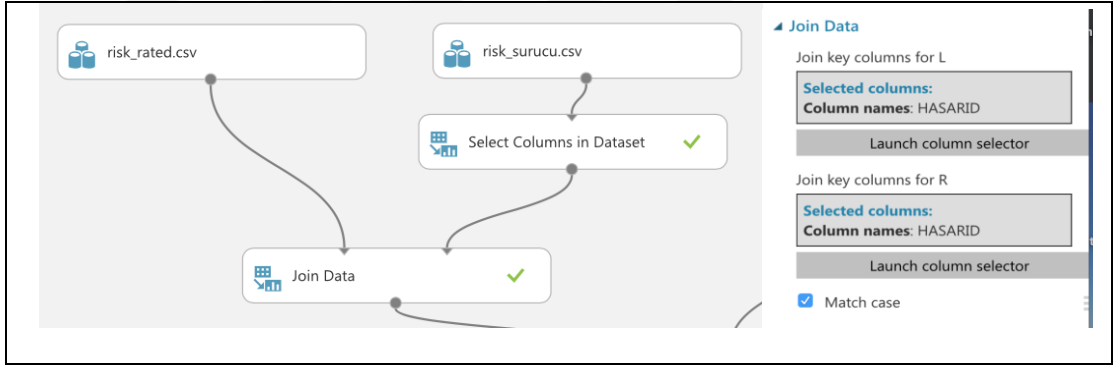
### 3.2.3.5 Veri Kümesi Birleştirme (Join Data) İşlemi

Join Data işlemi, çalışma alanına veri kümeleri olarak eklenen csv dosyalarındaki dataların sql işlemi gerçekleştirir gibi belirlenen bir ya da birden fazla değerler ile modelin tek bir yapıda oluşmasını sağlamaktadır.

Bu işlemde, veritabanı stili birleştirme işlemi kullanarak iki veri kümesini birleştirmek için Azure ML Studio' daki Veri Kümesi Birleştirme modülü kullanılmaktadır.

İki veri kümesindeki bir birleştirme gerçekleştirmek için, bunlar tek bir anahtar sütunla ilişkilendirilmelidir. Birden fazla birleştirme işlemi tek bir işlemde desteklenmemektedir. Şekil 3.8 ile birden fazla veri kümesinin sistem tarafından nasıl birleştirildiği, ayrıca seçilen kolon verisinin de “key” değeri olarak nasıl seçildiğini göstermektedir.

**Şekil 3.8 : Azure Veri Kümesi Birleştirme Yöntemi**



### 3.2.3.6 Algoritma Seçme İşlemi

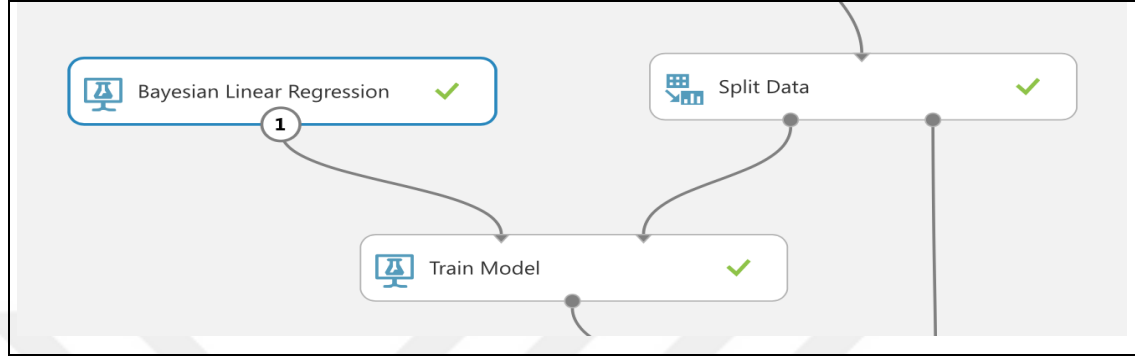
Azure Machine Learning Studio platformunda ML kütüphanesi oluşturabilmek için gerekli olan tüm algoritmalar sistemde bulunmaktadır. Eğitim çalışmasını gerçekleştirebilmek için projeye eklenen veri kümeleri ile düşünülen algoritmalar beraberinde eğitilerek, ileride gelecek data'lara yol gösterici olacaktır.

Azure Machine Learning Platformunda Gözetimli ve Gözetimsiz öğrenme üzerine iş akışları mevcuttur.

### 3.2.3.7 Model Eğitime (Train Model) İşlemi

Bu çalışmada seçilen algoritma işlemi ile split işlemi gerçekleştirilmiş veri kümesini belirtilen iş akışına entegrasyonunu sağlayarak, modelin eğitilmesini sağlamaktadır.

**Şekil 3.9 : Azure Model Eğitime İşlemi**



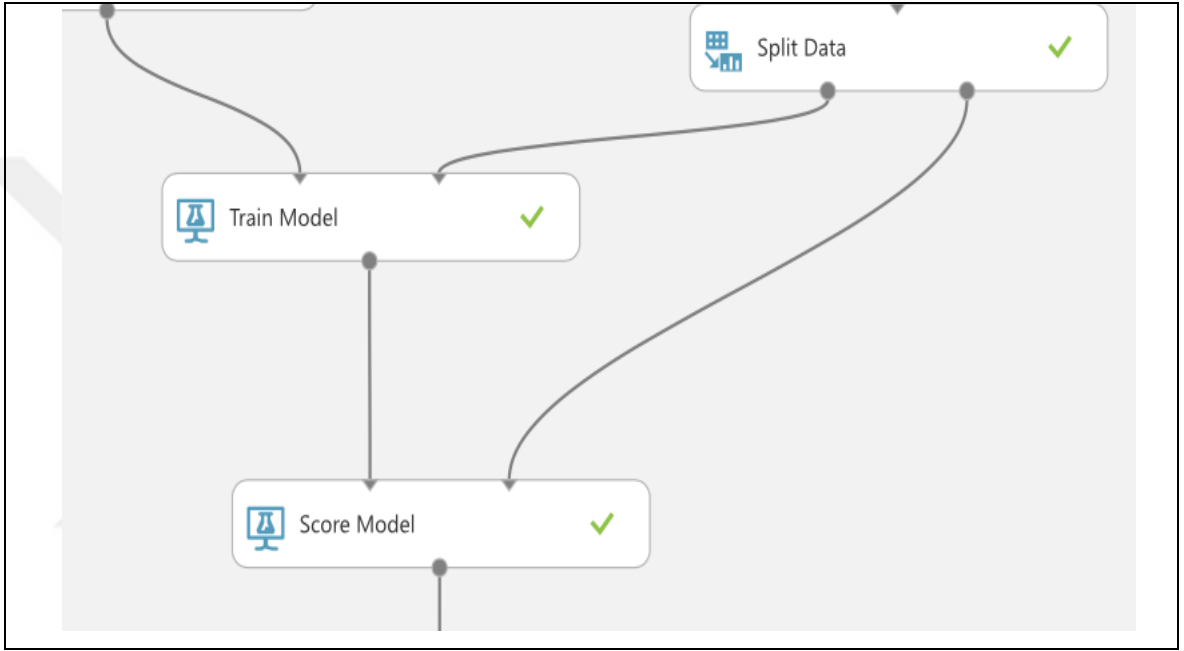
Bir sınıflandırma veya regresyon modeli eğitimi, denetlenen bir çeşit makine öğrenmedir. Bu çalışma ile oluşan sonuçlar, gelen yeni kalıpları öğrenmek için geçmiş verileri içeren bir veri kümesi sağlanması gerektiği anlamına gelmektedir. Veriler, hem öngörmeye çalışılan sonucu hem de ilgili değişkenleri içermektedir. Makine öğrenme modeli, istatistiksel tasarımları çıkarmak ve bir model oluşturmak için verileri kullanır. Detaylıca incelendiğinde, önceden yapılandırılmış bir modelin (regresyon algoritması, karar ağacı modeli veya diğer makine öğrenme modülü gibi) bağlanması gerekmektedir. Şekil 3.9 da olduğu gibi sistem üzerinden önceden oluşturulmuş ve kullanıma hazır algoritmanın mevcut ayrıştırılmış kayıtlar ile eğitilebilmesi için Train Model iş akışı sisteme eklenmiştir.

Çıktı olan model, tüm verileri işleyerek eğitildikten sonra çıktı, değerlendirip değerlendirilecek veya tahminler oluşturmak için kullanılacak eğitilmiş bir modeldir.

### 3.2.3.8 Skor Model (Score Model) İşlemi

Bu çalışmada, “Train Model” akışı sonrasında dışa verilen ve sistem tarafından hazırlanmış, anlamlandırılmış, veri kümesi ile Split işlemi sonrası eğitim dataları ile karşılaştırılmada kullanılmak üzere ayrılan veri kümesinin birlikte çalıştırılmasını sağlamaktadır. (Şekil 3.10)

Şekil 3.10 : Azure Skor Model İşlemi



Eğitilmiş bir sınıflandırma veya regresyon modeli kullanarak tahminler üretmek için Skor Modeli kullanabilirsiniz. Tahmini değer, modele ve eklenen parametre verilerine bağlı olarak birçok farklı biçimde gerçekleşebilir.

Çalıştırılan eğitim kütüphanesinden çıkan değerlendirmeleri skorlamak için bir sınıflandırma modeli kullanıyorsanız, Skor Modeli kümeleme ile hazırlanmış ML projesi için öngörülen bir değeri ve tahmin edilen değer olasılığını çıkarır.

Regresyon modelleri için Skor Modeli sadece tahmin edilen sayısal değeri üretir.

Skor Modeli'ni kullanarak sonuçlara ilişkin skor seti oluşturduktan sonra, modelin doğruluğunu (performansı) değerlendirmek için kullanılan bir metrik kümesi oluşturmak için puanlanan veri kümesini mevcut modele bağlayarak analiz edilebilmektedir.

Bu çalışmalar ile oluşan sonuçları bir veri kümesine kaydedebilir veya çıktıyı tahmini web servis bir parçası olarak kullanılabilir.

### **3.2.3.9 Değerlendirme Modeli (Evaluate Model) İşlemi**

Değerlendirme Model işlemi ile çalıştırılan modelin sonuç istatistiklerinin analizi yapılmaktadır. Bu kapsamda hata olarak tahmin edilen değer ile gerçek değer arasındaki farkı temsil eder. Genellikle mutlak değerlerini ya da bu fark kare genellikle hesaplanan boyunca tüm örneklerde, hata toplam büyüklüğünü yakalamak için tahmin edilen ve gerçek değer arasındaki farkı bazı durumlarda negatif olması gerektirir. Bu açıklamalara istinaden tahmine dayalı bir regresyon modeli kendi tahminlerin gerçek değerlerle gelen ortalama sapmasını ölçümleyerek bilgi vermektedir. Alt hata değerlerini tahminleme yaparken daha doğru bir model analiz etme anlamına gelmektedir. Genel bir hata ölçümü modelin verilere mükemmel uyduğunu 0 anlamına gelir (Microsoft 2017).

Değerlendirme Modeli işlemi sonrasında model tarafından açıklanan rakamlar, değişim oranı olarak yorumlanabilir. Yüksek bir oranda bu durumda, daha iyi 1 burada bir mükemmel bir çalışma olduğunu gösterir.

Eğitimli bir modelin doğruluğunu ölçmek için Değerlendirme Modeli kullanılmaktadır. Bir modelden oluşan skorları içeren bir veri kümesi proje tarafından sağlanması ile birlikte, Değerlendirme Modeli modülü bir dizi endüstri standardı değerlendirme metriğini hesaplamaktadır. Değerlendirme Modeli tarafından döndürülen metrikler, değerlendirmekte olduğunuz model türüne bağlıdır.

### 3.3 MAKİNA ÖĞRENMESİ YÖNTEMLERİ

Bu çalışmada hasar detay ve sürücü verileri ile belirlenen noktalarda sürücülerin kaza yapma risklerini hesaplamayı ve aynı zamanda da sistem tarafından bir sürücüye hangi hasar yerlerinde kaza yapabileceğini öneren bir yapı oluşturuldu.

Bu çalışmayı gerçekleştirirken hem sınıflandırma hem de regresyon yöntemleri kullanılmıştır. Bir sürücünün kaza yapabileceği diğer noktaları önerebilmesi için de Microsoft Azure platformunda kullanılmakta olan önerme yöntemlerinden MatchBox Recommender eğitim akışı denenmiştir.

#### 3.3.1 Gözetimli Öğrenme

Etiketlenmiş gözlemlerden öğrenme sürecidir. Etiketler, algoritmaya gözlemleri nasıl etiketlemesi gerektiğini öğretir. Örneğin, içerisinde sigorta yaptır ya da sigorta poliçesi kes gibi cümleler var ise poliçe işlemi olduğunu anlamasına olanak sağlayacaktır.

Sınıflandırma (Classification): Her bir inceleme işlemi bir kategori atması yapar. Örneğin; bir eklenen evrakın fatura mı yoksa başka bir doküman tipi mi olup olmadığını sınıflandırma ile anlamlandırılabilir. Azure platformunda bu türde ML çalışması için Şekil 3.11’de olduğu gibi fonksiyonlar bulunmaktadır.

Regresyon (Regression): Her inceleme için eğitilen modele bakarak reel bir değer tahmini yapar. Azure platformunda Şekil 3.11 ‘de olduğu gibi Bayesian Linear, Boosted Desicion Tree, Desicion Forest gibi regresyonu sağlayan öğrenmeye dayalı tüm algoritmalar mevcuttur.

### Şekil 3.11 : Azure ML Algoritmaları

Classification	
Multiclass Decision Forest	
Multiclass Decision Jungle	
Multiclass Logistic Regression	
Multiclass Neural Network	
One-vs-All Multiclass	
Two-Class Averaged Perceptron	
Two-Class Bayes Point Machine	
Two-Class Boosted Decision Tree	
Two-Class Decision Forest	
Two-Class Decision Jungle	
Two-Class Locally-Deep Support Vector Machine	
Two-Class Logistic Regression	
Two-Class Neural Network	
Two-Class Support Vector Machine	

Regression	
Bayesian Linear Regression	
Boosted Decision Tree Regression	
Decision Forest Regression	
Fast Forest Quantile Regression	
Linear Regression	
Neural Network Regression	
Ordinal Regression	
Poisson Regression	

### 3.3.2 Gözetimsiz Öğrenme

Etiketsiz gözlemlerden öğrenme sürecidir. Algoritmanın kendi kendine keşifler yapması, gözükmeyen örüntüleri keşfetmesi beklenir.

Kümeleme (Clustering): Gözlemleri homojen bölgelere ayırır. Örneğin: bir okuldaki öğrenci gruplarını tespit etmek.

Boyut Azaltımı (Dimensionality Reduction): Gözlemlerin mevcut özellik sayısını az ve öz hale indirir, bize en iyi öğrenme imkanı sunar.

### 3.3.3 Önerme Yöntemi ile Öğrenme

Machine Learning önerme sisteminin ana amacı kullanıcılarına bir veya daha fazla öğe önermektir. Bir öğeye örnek olarak bir film, restoran, kitap veya şarkı olabilir. Bir kullanıcı, bir öğe tercihi olan bir kişi, kişi grubu veya başka bir varlık olabilir.

Önerme sisteminde iki temel yaklaşım vardır. Birincisi hem kullanıcıları hem de öğeleri kullanan içeriğe dayalı yaklaşımdır. Kullanıcılar yaş ve cinsiyet gibi özelliklerle tanımlanabilir ve yazar ve üretici gibi özellikler tarafından tanımlanabilir. İçerik tabanlı

tavsiyelerin tipik örnekleri çöpçatanlık siteleri üzerindeki sistemlerdir. İkinci yaklaşım, yalnızca kullanıcıların ve öğelerin tanımlayıcılarını kullanan ve kullanıcıların öğelere verdiği derecelerin (seyrek) bir matristen bu varlıklar hakkında örtülü bilgileri elde eden işbirlikçi filtreleme yöntemidir. Derecelendirdiğimiz öğelerin bir listesine sahibiz.

Önerme sistemi, ortak çalışan filtrelemeyi içerik tabanlı bir yaklaşımla birleştirir. Sorun, iyi bilinen "soğuk başlangıç" sorununa işaret ediyor. Bununla birlikte, belirli bir kullanıcıdan yeterli sayıda oy topladıktan sonra, tek başına özelliklerinden ziyade belirli derecelendirmelerine dayanarak kendileri için tamamen kişiselleştirilmiş tahminler yapılmasını sağlamaktadır. Böylece, içerik temelli önerilerden işbirlikçi filtrelemeye kadar sorunsuz bir geçiş olmaktadır. Kullanıcı veya öğe özellikleri bulunmasa bile, Azure platformunda bulunan Matchbox Recommender eğitim akışı ile yine de ortak filtreleme modunda çalışmaktadır.



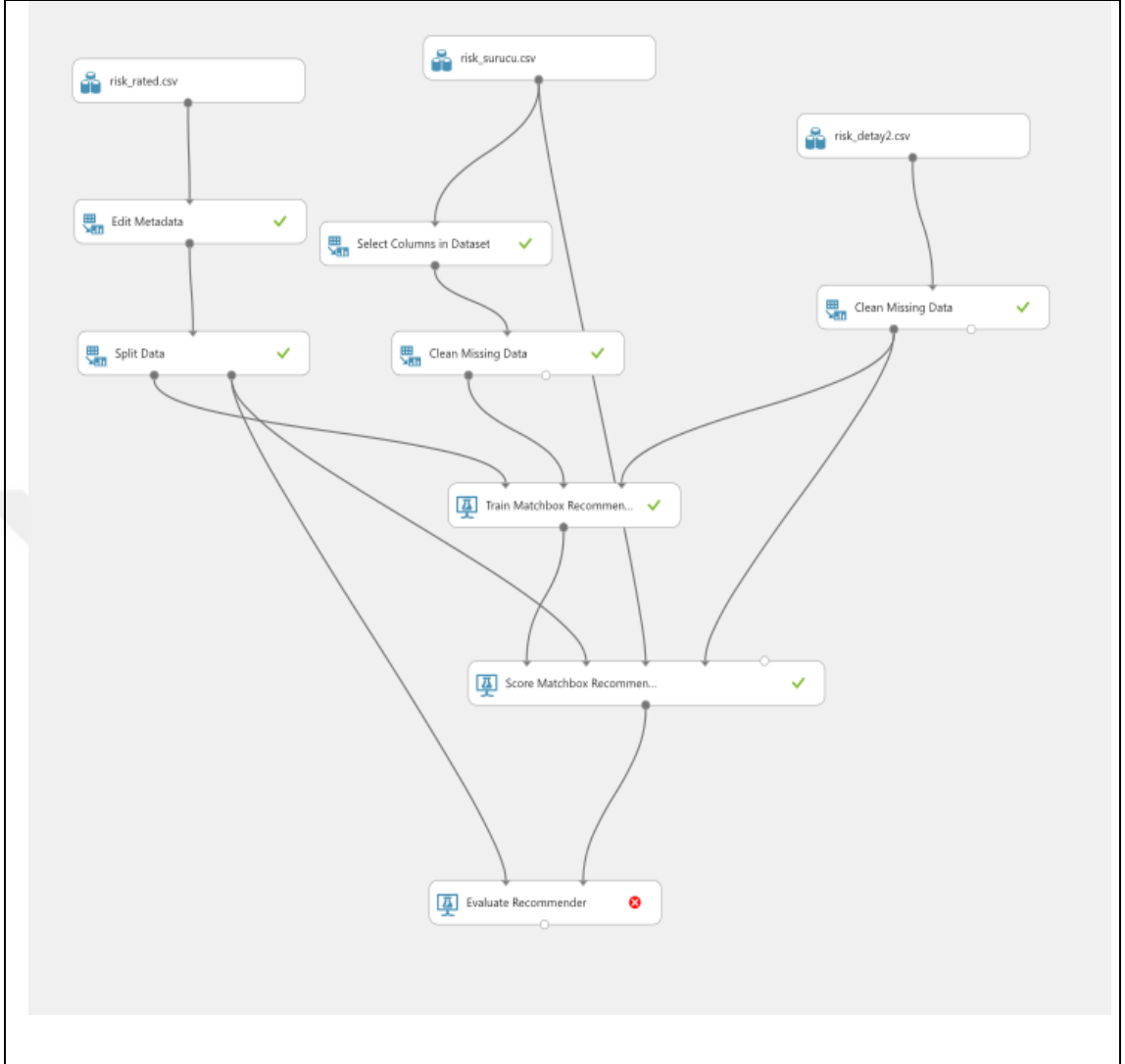
## 4. BULGULAR

Bu çalışmada, bir hasarın oluşmasına sebebiyet veren veri kümesi, sürücü ve lokasyon detayları olmak üzere iki farklı veri kümesine dönüştürülmüştür. Veri kümesi içerisindeki niteliklerin oluşturulacak olan modele etkisi incelenerek, en doğru parametre grubu elde edilmeye çalışılmıştır. Bu parametre grubu dışında kalan nitelikler veri kümelerinden çıkarılmıştır. Bu çalışma ile birlikte ML yöntemleri uygulanmıştır. Bu çalışmalar ile birlikte elde edilen bulguların detayları paylaşılmaktadır.

### 4.1 MATCHBOX RECOMMENDER YÖNTEMİ

Trafik Kazasının olma olasılığını ve sürücülere ait bir risk haritası oluşturabilmek için iki farklı ML yöntemi izlenmiştir. Azure ML Matchbox Recommender yöntemi ile sürücülerin profillerine uyan kaza lokasyon örneklerini önermektedir. Bu sayede geçmişte yaptıkları kaza detayları ya da web servis ile gelecek anlık veri modellerindeki dataların kullanılarak, bu profillere uyan benzer kaza noktalarında sürücülerin dikkatli olması önerilmektedir. Şekil 4.1’de görülen yöntem ile kaza ve sürücü ilişkisi derecesi, sürücü detayları ve kaza detayları ayrı ayrı üç veri kümesinde projeye eklenmiştir.

**Şekil 4.1 : Azure Matchbox Recommender Yöntemi ile Gerçekleştirilen Çalışma**



Bu çalışma ile elde edilen sonuç kümesinde 68 sürücü için 5 ayrı hasar lokasyonunda profil kesişimleri görülmüş olup, sürücülerin kaza yaptığı lokasyonlardan farklı olarak dikkat edilmesi gereken noktaları belirlemiştir. Şekil 4.2 'de bu çalışmaya ait çıkan sonuçların listesi görülmektedir.

## Şekil 4.2 : Azure Matchbox Recommender ile Gerçekleşen Sonuç Veri Kümesi

Traffic Accident Recommender > Score Matchbox Recommender > Scored dataset

rows: 68, columns: 6

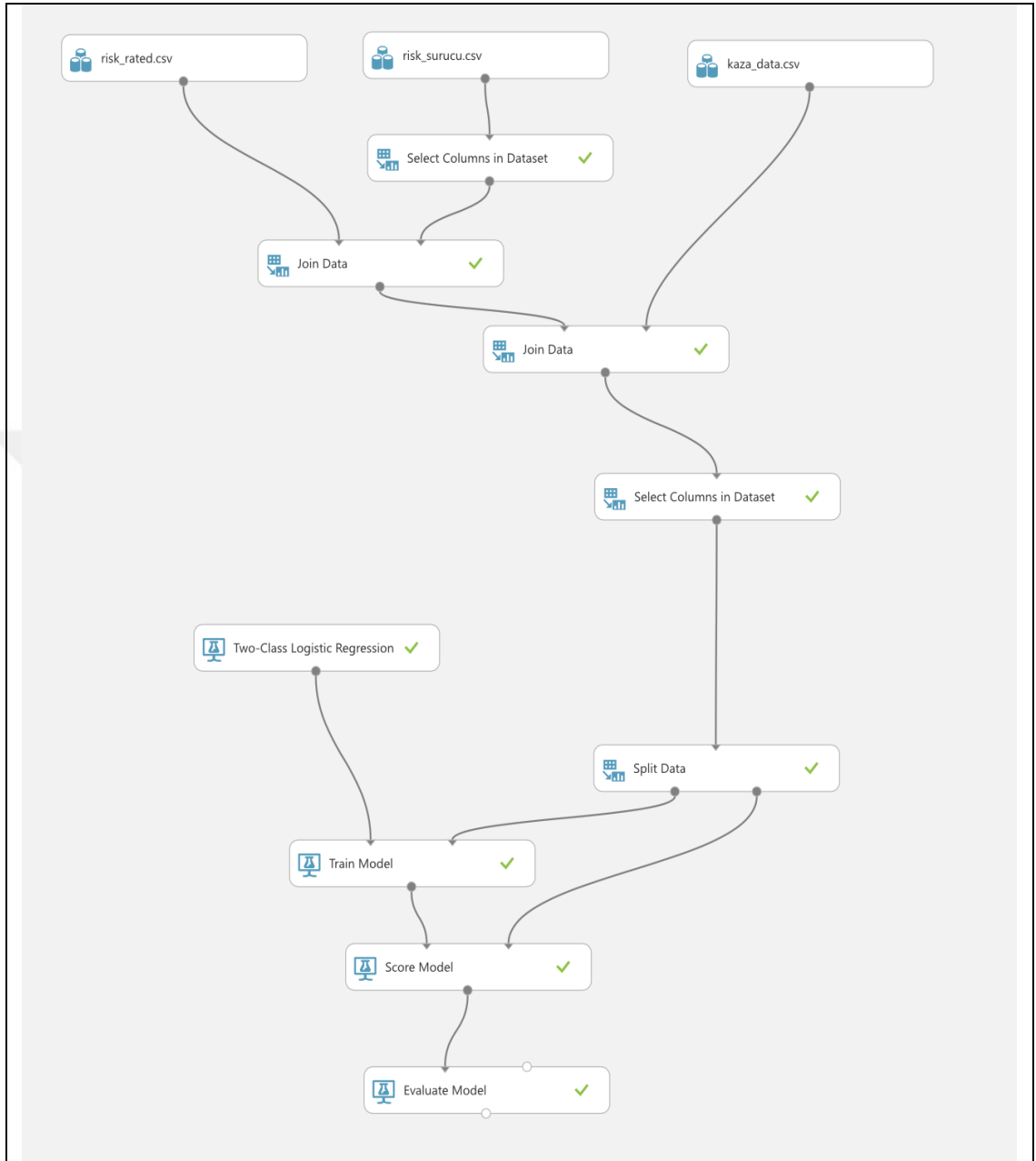
view as: [Bar Chart Icon]

User	Item 1	Item 2	Item 3	Item 4	Item 5
HNPJU3YTIIRYNAM	C41028831419BE3E	9638CBF882A4E860	F67333B1AA6DFF89	C1180DFF4DA45180	ECFCEB395AD43F3A
O2ZDLL4ABQNCIQ7W	C41028831419BE3E	9638CBF882A4E860	F67333B1AA6DFF89	C1180DFF4DA45180	ECFCEB395AD43F3A
Z33RNVJ6I92VNVQ8K	C41028831419BE3E	9638CBF882A4E860	F67333B1AA6DFF89	C1180DFF4DA45180	ECFCEB395AD43F3A
VT5XU2D5OQX38S3T	C41028831419BE3E	9638CBF882A4E860	F67333B1AA6DFF89	C1180DFF4DA45180	ECFCEB395AD43F3A
SP6MSQS962V42W86	F67333B1AA6DFF89	9638CBF882A4E860	C41028831419BE3E	C1180DFF4DA45180	ECFCEB395AD43F3A
BSZIS7AZZWL7YEI7	C41028831419BE3E	9638CBF882A4E860	F67333B1AA6DFF89	C1180DFF4DA45180	ECFCEB395AD43F3A
M9BFP9C50P7YB3TQ	C41028831419BE3E	9638CBF882A4E860	F67333B1AA6DFF89	C1180DFF4DA45180	ECFCEB395AD43F3A
98AXRLHW1UHOM73N	C41028831419BE3E	9638CBF882A4E860	F67333B1AA6DFF89	C1180DFF4DA45180	ECFCEB395AD43F3A
Q9KN09XTVS7ZGZJZ	F67333B1AA6DFF89	C41028831419BE3E	9638CBF882A4E860	C1180DFF4DA45180	ECFCEB395AD43F3A
H40PYSEV80P8A8P2	C41028831419BE3E	9638CBF882A4E860	F67333B1AA6DFF89	C1180DFF4DA45180	ECFCEB395AD43F3A
QMU7CY3JZ26CORLQ	F67333B1AA6DFF89	9638CBF882A4E860	C41028831419BE3E	C1180DFF4DA45180	ECFCEB395AD43F3A
B87DORHWNF0Y765H	F67333B1AA6DFF89	C41028831419BE3E	9638CBF882A4E860	C1180DFF4DA45180	ECFCEB395AD43F3A
NZDSRXKDG40ZZKU	C41028831419BE3E	9638CBF882A4E860	F67333B1AA6DFF89	C1180DFF4DA45180	ECFCEB395AD43F3A
D3K29ZPA2DJ9MTRS	C41028831419BE3E	9638CBF882A4E860	F67333B1AA6DFF89	C1180DFF4DA45180	ECFCEB395AD43F3A
7RV5TPPGFQWQBBS	C41028831419BE3E	9638CBF882A4E860	F67333B1AA6DFF89	C1180DFF4DA45180	ECFCEB395AD43F3A
6CB5OTMCHV5MZL4M	F67333B1AA6DFF89	C41028831419BE3E	9638CBF882A4E860	C1180DFF4DA45180	ECFCEB395AD43F3A
D02Y689B4DUSO2RW	9638CBF882A4E860	C41028831419BE3E	F67333B1AA6DFF89	C1180DFF4DA45180	ECFCEB395AD43F3A
1QMP7L70W4CJCOL	C41028831419BE3E	9638CBF882A4E860	F67333B1AA6DFF89	C1180DFF4DA45180	ECFCEB395AD43F3A
DGQR8BQBLMB53CB9	9638CBF882A4E860	C41028831419BE3E	F67333B1AA6DFF89	C1180DFF4DA45180	ECFCEB395AD43F3A
TU155HX5VCLFRJNL	C41028831419BE3E	9638CBF882A4E860	F67333B1AA6DFF89	C1180DFF4DA45180	ECFCEB395AD43F3A
...	...	...	...	...	...

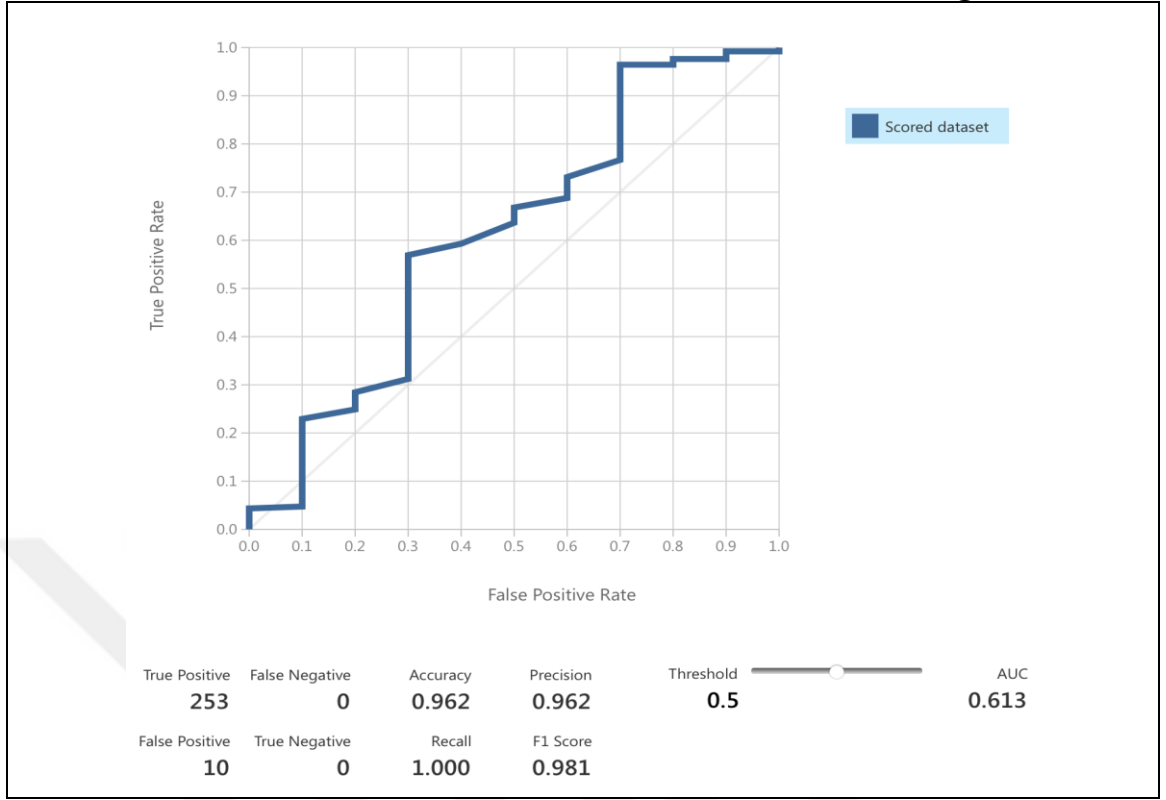
## 4.2 SINIFLAMA YÖNTEMLERİ

Bu çalışmada kullanılan veri kümelerinin gerekli birleştirme işlemlerinden sonra hasar\_sebebi değerine göre sınıflandırma çalışması yapılmıştır. Bu sayede kaza lokasyonlarındaki hasarın gerçekleşme nedenini ortaya koyacak ve bu tarz sürücülerin o lokasyonlarda hasar gerçekleştirme oranlarını sunacaktır. Fakat hasar kaydının az olması sebebiyle yalnızca Aracın Çarpması hasar sebebini doğru tahminleyebilmektedir. Two Class Logistic Regression işlemi ile Aracın Çarpması hasar sebebi en iyi şekilde tahminlenebilmektedir. Şekil 4.3’de belirtildiği üzere üç (3) ayrı modeldeki gerekli kayıtların birleştirilerek Two Class Logistic Regression ile eğitilmesi üzerine işlemler belirtilmiştir. Bu ML iş akışının sonunda ise Şekil 4.4 ‘de de belirtildiği üzere modelin doğruluk ve hassasiyet değeri 0,962 olarak sonuçlanmıştır. Sınıflandırma tipindeki işlemlerde belirli lokasyonları hasar sebeplerine göre kategorize ederek, oluşacak risk durumunda kişinin profiline göre sınıflandırılan lokasyonlar uyarı şeklinde gösterilecektir.

**Şekil 4.3 : Azure ile Gerçekleşen Çalışmadaki Sınıflandırma İşlem Akışı**



**Şekil 4.4 : Azure ile Gerçekleşen Çalışmadaki Sınıflandırma İşlem Öğrenme Akışı**

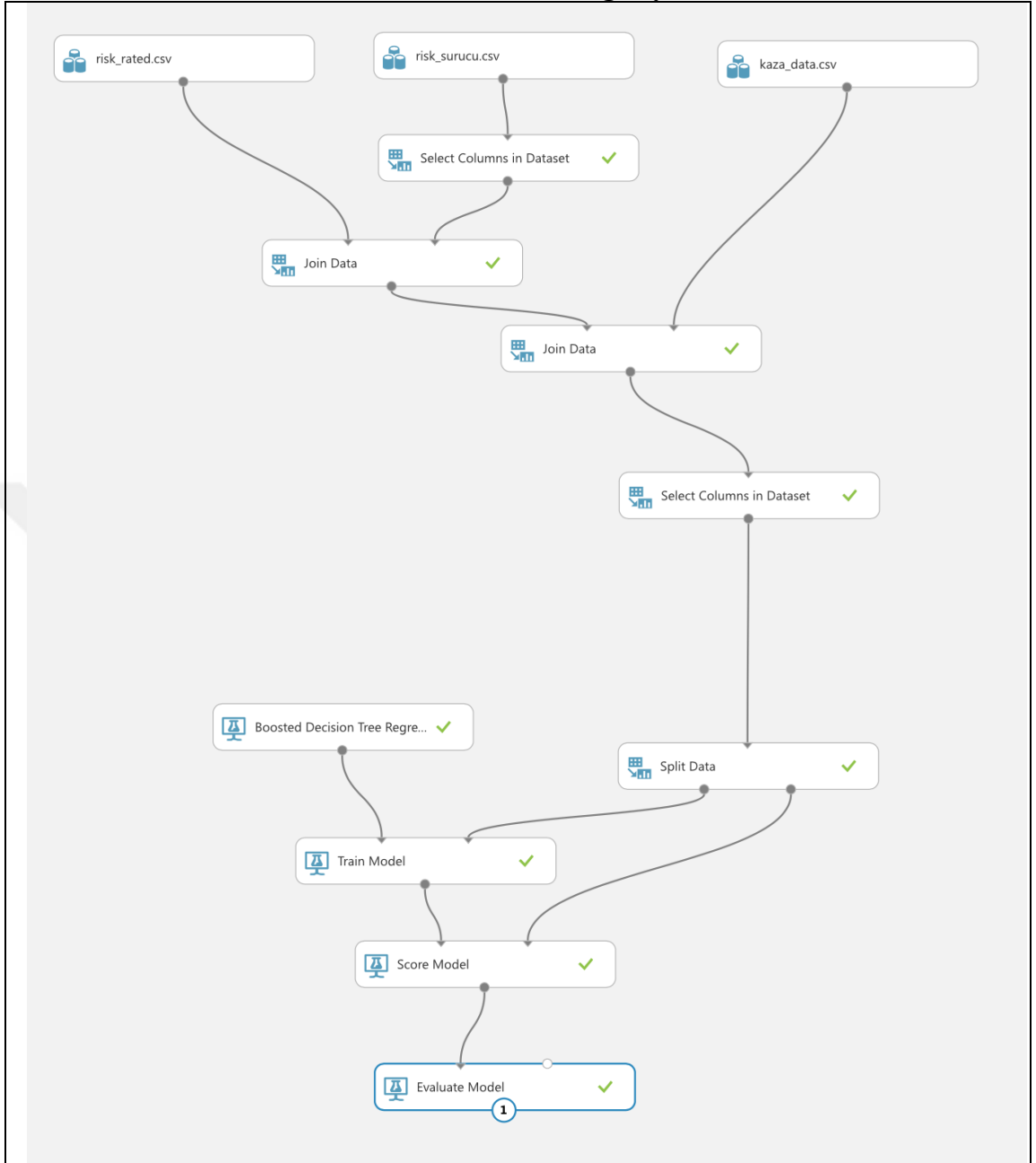


### 4.3 REGRESYON YÖNTEMLERİ

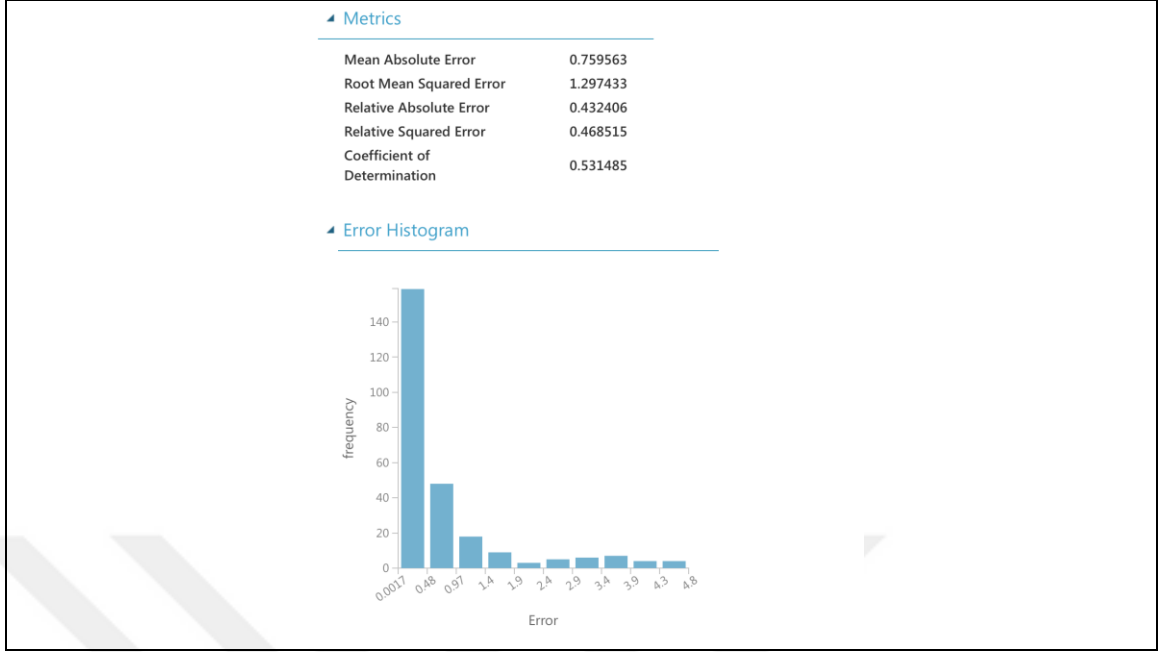
Sürücülerin risk durumları her an gerçekleşebilir ve farklı farklı senaryolar hayata geçebilir. Sigortacılık sektöründeki ihbarlarla gelen değerlerin kaza senaryolarını ne gibi farklılıklara yol açtığını görmekteyiz. Haliyle kaza yapma şekilleri de takvim ilerledikçe değişiklik gösterebilmektedir. Sürücü ve lokasyon bazındaki verilerin değişkenliklerini sistemin düzenli öğrenebilmesi içinde regresyon algoritmalarında çalışmalar yapılmış olup, en iyi öğrenme derecesi Boosted Decision Tree algoritması ile sağlanmıştır. Ancak Pendik olarak ele alınan datanın azlığı sebebiyle oranın düşük olması, verinin kalitesi ve eğitim datasının daha fazla olması durumunda doğru sonuçlara erişebileceğinin göstergesidir.



Şekil 4.5 : Azure ile Gerçekleşen Çalışmadaki Regresyon İşlemi



Şekil 4.6 : Azure ile Gerçekleşen Çalışmadaki Regresyon İşlem Öğrenme Akışı





## 5.TARTIŞMA VE SONUÇ

Bu çalışmada, sigortacılık(insurtech) çözümleriyle uygulamalar ile birlikte kişilerin kaza durumlarını değerlendirerek, yaklaştıkları lokasyonlarda, geçmişde olmuş kaza verilerine göre, sürücülere kaza yapma risklerini belirtecek model oluşturulması hedeflenmiştir.

Çalışmada Pendik bölgesinde Eylül 2017 yılından itibaren gerçekleşmiş kaza bilgileri veri olarak kullanılmıştır. Bu model kapsamında 25 adet parametre bulunmaktadır. Bu parametrelerin analizleri sonrasında 20 adet veri değerli olarak seçilmiş olup modelde kullanılmıştır. Her bir parametre için referans değer aralıkları belirlenmiştir.

Model için gerekli parametrelerin seçilmesi ve referans dataların oluşturulması konusunda sırasında kullanılacak parametrelerin belirlenmesi aşamasında Hasar Yönetiminde bulunan hasar uzmanı tarafından değerlendirilerek hazırlanmıştır.

Parametrelerin belirlenmesi sonrasında veri kümelerine, her bir sınıflandırma kategorisinden en az bir yöntem olacak şekilde Two Class Logistic Regression, Two Class Average Perceptron denenmiştir. Uygulanan algoritmaların doğruluk oranları incelendiğinde tüm veri kümeleri üzerinde Two Class Logistic Regression algoritmasının en başarılı algoritma olduğu görülmektedir. Regresyon çalışmalarında da Linear, Bayes ve Boosted Decision Tree algoritmaları denenmiş olup en iyi sonuç ise Boosted Decision Tree algoritması ile sağlanmıştır. Kaza kayıtlarında ve split edilerek modelde kullanılmayan ancak kaza verilerinin gerçekleştiğini bildiklerimizdeki oranların yüksek olması sonuçlarının başarılı olduğu bilgisini sağlamaktadır.

Bu çalışma ile birlikte değerlerin anlamlı olması için kullanılan hasar verilerini Öneri sistemine entegre ettiğimizde sürücü ve davranışları eğitildiğinde 68 adet lokasyonda dikkat etmesi gereken sürücüler belirtilmiştir. Bu sayede sürücülere ait bir risk haritası hazırlanacaktır. Veriyi tanımanın vermiş olduğu bilgi birikim ile gelen sonuçların doğruluğu rahatlıkla kontrol edilebilmektedir.

İleride bu çalışma ile birlikte, bir kullanıcı telefonu arka planda çalışırken sürücü davranışlarını ölçebilecek ve kendisine riskli gelen noktalara yaklaştıkça telefonda uyarı bildirimleri gelecektir.

Bu çalışmada görünen riskin sürücüye beyan edilmesi ile belki de bir can da olsa bir kişiyi kurtarabilmek olacaktır.



## KAYNAKÇA

### *Kitaplar*

Murphy, Kevin P. 2012. Machine Learning A Probablistic Perspective.

Olson, D. L. & Delen, D., 2008. Advanced data mining techniques. Berlin:Springer.



## ***Diğer Yayınlar***

- Accenture. ,2015, Accenture Dijitalleşme Endeksi Türkiye Sonuçları  
[http://www.tbv.org.tr/core/uploads/page/document/1100\\_18031611540.pdf](http://www.tbv.org.tr/core/uploads/page/document/1100_18031611540.pdf)  
[erişim tarihi 28 Nisan 2017]
- Akgöbek,Ö. , Çakır, F. ,2009, Veri Madenciliğinde Bir Uzman Sistem Tasarımı.  
*Yüksek Lisans Tezi.Harran Üniversitesi*  
[http://ab.org.tr/ab09/kitap/akgobek\\_cakir\\_AB09.pdf](http://ab.org.tr/ab09/kitap/akgobek_cakir_AB09.pdf) [erişim tarihi 10 Nisan 2017]
- Aladağ E., 2015,Makine Öğrenmesi Nedir?, <http://www.emrealadag.com/makine-ogrenmesi-nedir.html>. [erişim tarihi 09 Aralık 2017]
- Beshah,T., Hill,S., Mining Road Traffic Accident Data to Improve Safety: Role of Road- elated Factors on Accident Severity in Ethiopia. *Yüksek Lisans Tezi*  
Department of Information Science<sup>1</sup> , Operations and Information Management  
Department<sup>2</sup> Addis Ababa University, Ethiopia<sup>1</sup> , University of Pennsylvania,  
Philadelphia, PA
- Bozo, M.,2014, Makine Öğrenimi. <http://e-bergi.com/y/makine-grenimi> [Erişim tarihi  
01 Mayıs 2017]
- Chameleon: Hierarchical Clustering Using Dynamic Modeling, Karypis G. - Han E.-  
Kumar V, IEEE Computer, 1999 : 68-75
- Çinicioğlu,N. , Atalay,M. , Yorulmaz,H. ,2013,Trafik Kazaları Analizi için Bayes  
Ağları Modeli Esması. *Yüksek Lisans Tezi İstanbul, Türkiye 2013*
- Deloitte., 2015, Deloitte Global Mobil Kullanıcı Anketi.  
<https://www2.deloitte.com/content/dam/Deloitte/tr/Documents/technology-media-telecommunications/deloitte-global-mobil-kullanici-anketi-2015-f.pdf> [erişim  
tarihi 27 Nisan 2017]
- Dogru,N., Subası, A. , 2013 , Comparison of clustering techniques for traffic accident  
detection.*Yüksek Lisans Tezi*. Department of Information Technologies, Faculty of  
Engineering and Information Technologies, International Burch University,  
Sarajevo, Bosnia and Herzegovina
- Kabakuş,N. , Tortum, A., 2012, Erzurum'un ilçelerinde meydana gelen trafik  
kazalarının coğrafi bilgi sistemleri ile değerlendirilmesi.*Yüksek Lisans Tezi*.  
Erzurum,,Atatürk Üniversitesi, Mühendislik Fakültesi, İnşaat Mühendisliği  
Bölümü.

- Kaya,T., 2015, Makine öğrenme yöntemleri ile Trafik Kazaları için risk tahmini yapabilen web tabanlı bir yazılım. *Yüksek Lisans Tezi*. Gazi Üniversitesi
- Marr,B.,2016, A Short History of Machine Learning. <https://www.forbes.com/sites/bernardmarr/2016/02/19/a-short-history-of-machine-learning-every-manager-should-read/#575a4d1415e7> [erişim tarihi 01 Nisan 2017]
- Microsoft, Azure Platform, 2017, <https://msdn.microsoft.com/en-us/library/azure/> [erişim tarihi 12 Aralık 2017]
- Nizam,H. , Akın,S. , Sosyal Medyada Makine Öğrenmesi ile Duygu Analizinde Dengeli ve Dengesiz Veri Setlerinin Performanslarının Karşılaştırılması. *Yüksek Lisans Tezi*
- Özekes, S., Veri Madenciliği Modelleri Ve Uygulama Alanları İstanbul Ticaret Üniversitesi Dergisi
- Özgirgin, T., 2016, Sosyal Medyada Duygu Analizi Ve Nitelik Çıkarımı. *Yüksek Lisans Tezi*. Yıldız Teknik Üniversitesi
- Prof. Dr. Alpaydın,E. (2013) Gözetimli Öğrenme Algoritmalarının İstatistiksel Yöntemlerle Karşılaştırılması.*Proje*

## ÖZGEÇMİŞ

**Adı Soyadı:** Mehmet Eren BEKİN  
**Doğum Yeri ve Yılı:** Altındağ 1988  
**Yabancı Dili:** İngilizce  
**İlk Öğretim:** Türkan Yamantürk İlköğretim Okulu (Ankara)  
**Orta Öğretim:** Ayrancı Yabancı Dil Ağırlıklı Lisesi (Ankara)  
**Lisans:** Atılım Üniversitesi Bilgisayar Mühendisliği (Ankara)  
**Yüksek Lisans:** Bahçeşehir Üniversitesi  
**Enstitü Adı:** Fen Bilimleri Enstitüsü  
**Program Adı:** Bilgi Teknolojileri  
**Çalışma Hayatı:**  
Kariyer.Net A.Ş. Yazılım Uzmanı 01.2013 – 04.2014  
Anadolu Sigorta A.Ş. Hasar Süreçleri Yazılım Uzmanı 04.2014 – 06.2017  
Anadolu Sigorta A.Ş. İnovasyon Yazılım Uzmanı 06.2017 – Halen