

T.C.
ERZİNCAN BİNALİ YILDIRIM ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

TRAFİK KAZALARININ MAKİNE ÖĞRENMESİ VE FARKLI
İSTATİSTİKSEL YÖNTEMLER İLE MODELLENMESİ:
ANKARA İLİ ÖRNEĞİ

Nurşah ALTINTAŞ

Danışman: Doç. Dr. Halim Ferit BAYATA

İNŞAAT MÜHENDİSLİĞİ
ANABİLİM DALI

ERZİNCAN

2022

Her Hakkı Saklıdır

ÖZET

Yüksek Lisans

TRAFİK KAZALARININ MAKİNE ÖĞRENMESİ VE FARKLI İSTATİSTİKSEL YÖNTEMLER İLE MODELLENMESİ: ANKARA İLİ ÖRNEĞİ

Nurşah ALTINTAŞ

Erzincan Binali Yıldırım Üniversitesi
Fen Bilimleri Enstitüsü
İnşaat Mühendisliği Anabilim Dalı

Danışman: Doç. Dr. Halim Ferit BAYATA

Dünyada gelişen teknolojiyle birlikte araç sayılarında meydana gelen artışlardan dolayı trafik problemleri oluşmuş, dolayısı ile ölümlü yaralanmalı ve hasar kayıtlı kazalardaki artışların çözümleri zaruri hale gelmiştir. Bu bakımdan trafik kazaları üzerinde yapılan tahmin modellemeleri gün geçtikçe önem kazanmış ve makine öğrenmesi ile oluşturulan tahmin modellemelerinde sıklıkla kullanılmaya başlanmıştır.

Bu çalışmada Trafik Daire Başkanlığı'nın ölümlü yaralanmalı kaza tutanaklarından elde edilen 2013-2020 yılları arasında Ankara iline ait 86.694 veri incelenmiş ve analiz edilmiştir. Makine öğrenmesi modeli haricinde, veri seti içerisinde açıklayıcı nitelik taşıyan istatistiksel kapsamda Ankara ili içerisinde kazaların genel karakterini sergilemek amacı ile birçok sonuç elde edilmiştir. Kaza tahmin modellemeleri oluşturulurken; Lojistik regresyon (Logistic Regression), KNN (K En Yakın Komşu Sınıflandırması), Destek vektör makinesi (Support Vector Machine- SVM), Karar ağacı (Decision Tree- DT), Rastgele Orman (Random Forest-RF), Ekstrem Gradyan Artırma (XGBoost), Hafif Gradyan Artırma Makinesi (Lightgbm) algoritmaları kullanılmıştır. Veri seti ölümlü kazalar açısından Sınıflandırma Problemi olarak değerlendirilmiştir. Bu yaklaşımdan dolayı veri setinde ölü sayısı itibari ile dengesizlik (unbalance) durumu söz konusudur. Bu açıdan modelleme aşamasında veri setinin orijinal hali üzerinde modelleme kurulmuş, daha sonra dengesizlik durumundan kaynaklı olarak SMOTE işlemi uygulanmış ve modellemeler tekrar oluşturulmuştur. Söz konusu işlemlere ilişkin sonuçlar Doğruluk (Accuracy) ve Duyarlılık (Recall) metriğine göre değerlendirilmiştir. Elde edilen sonuçların literatür çalışmaları ile uyumlu olduğu görülmüştür.

2022, 68 Sayfa

Anahtar Kelimeler: Makine öğrenmesi, Sınıflandırma problemi, Smote, Trafik kazaları, Yapay sinir ağları

ABSTRACT

MSc Thesis

MODELİNG OF TRAFFIC ACCİDENTS WITH MACHINE LEARNING AND DIFFERENT STATİSTICAL METHODS: THE CASE OF ANKARA PROVİNCE

Nurşah ALTINTAŞ

Erzincan Binali Yıldırım University
Graduate School of Natural and Applied Sciences
Department of Civil Engineering

Supervisor: Assoc. Prof. Dr. Halim Ferit BAYATA

Due to the developing technology in the world, traffic problems have appeared subsequent to the increase in the number of vehicles, and therefore solutions for the increase in accidents with fatal injuries and damage have become essential. In this respect, prediction models on traffic accidents have gradually gained importance and have been started to be used frequently in predictive models created with machine learning.

In this study, 86.694 data belonging to Ankara province between the years of 2013-2020 obtained from the Decedents of accidents with fatal injuries of the General Directorate of Police and the Traffic Department were examined and analyzed. Apart from the machine learning model, many results were obtained in order to reveal the general characteristics of the accidents in the province of Ankara within the statistical scope with an explanatory nature in the data set. While creating accident prediction models, Logistic Regression, KNN (K Nearest Neighbor Classification), Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), Extreme Gradient Boost (XGBoost), and Light Gradient Increasing Machine (Lightgbm) algorithms were used. The data set was evaluated as the Classification Problem in terms of fatal accidents. Due to this approach, there was unbalance in the data set in terms of the number of deaths. In this respect, modeling was established on the original form of the data set during the modeling phase, and then SMOTE process was administered due to the unbalance and the models were reestablished. The results related to the available processes were evaluated according to the Accuracy and Recall metrics. The results obtained were noticed to be compatible with the studies in the literature.

2022, 68 Pages

Keywords: Artificial neural network, Classification problem, Machine learning, Smote, Traffic accidents

TEŞEKKÜR

Tez sürecim esnasında bilgi ve birikimleri ile yolumu aydınlatan saygıdeğer danışmanım Doç. Dr. Halim Ferit BAYATA' ya ve deneyimlerinden faydalandığım Saygıdeğer hocam Dr. Fatih HATTATOĞLU' na, aynı zamanda tez yazım sürecinde tecrübeleri desteklerini esirgemeyen Sayın Dr. Öğretim Üyesi Muhammed Ali ÇOLAK' a ve Araştırma Görevlisi Ahmet Oğuz DEMİRİZ' e, ayrıca bu süreçte varlığı ile her an yanımda hissettiğim sevgili aileme teşekkürlerimi borç bilirim.

Nurşah ALTINTAŞ

Ağustos, 2022

İÇİNDEKİLER

	Sayfa
ÖZET.....	i
ABSTRACT.....	ii
TEŞEKKÜR.....	iii
ŞEKİLLER LİSTESİ	v
TABLolar LİSTESİ.....	viii
SİMGELER ve KISALTMALAR	ix
1. GİRİŞ	1
2. KAYNAK ÖZETLERİ	3
3. KURAMSAL TEMELLER.....	8
3.1 Ulaşım	8
3.1.1 Ülkemizde karayolu ulaşımının genel durumu.....	9
3.1.2 Ülkemizde ve dünyada trafik kazaları	10
3.1.3 Trafik kazalarına etki eden parametreler	11
3.2. Makine Öğrenmesi	13
3.2.1 Yapay sinir ağları.....	18
3.2.1.1 Yapay sinir ağı bileşenleri	21
3.2.1.2. Yapay sinir kullanım alanları.....	22
3.2.1.3. Yapay sinir ağları modelleri.....	22
3.2.1.4. Yapay sinir ağları eğitimi.....	25
4. MATERYAL METOT	27
5. ARAŞTIRMA BULGULARI.....	33
5.1. Makine Öğrenmesi Model Değerlendirme Sonuçları	33
5.2. Kaza Lokasyonlarının Haritada Gösterilmesi	59
6. SONUÇLAR	62

ŞEKİLLER LİSTESİ

Sayfa

Şekil 3.1. Makine Öğrenmesi ve Alt Grupları (Meltek ve Kayaalp, 2020).....	15
Şekil 3.2. Yapay Sinir Ağları Genel Yapısı (Kubat, 2015).....	18
Şekil 3.3. Tek katmanlı algılayıcı modeli (Öztemel, 2006).....	23
Şekil 3.4. İleri Besleme Ağ Yapısı.....	23
Şekil 3.5. Çok Katmanlı Algılayıcı Modellemesi (Kabalcı, 2014).....	24
Şekil 3.6. Geri beslemeli yapay sinir ağı.....	24
Şekil 5.1. Yıllık Ölümlü Yaralanmalı Kaza Sayısı.....	33
Şekil 5. 2. Aylık Ölümlü Yaralanmalı Kaza Sayısı.....	34
Şekil 5.3. Saatlik Ölümlü Yaralanmalı Kaza Sayısı.....	34
Şekil 5.4. İlçe Bazında Gerçekleşen Ölümlü Yaralanmalı Kaza Sayısı.....	35
Şekil 5.5. Kazanın Gerçekleştiği Bölgenin Çevresel Özellikleri.....	35
Şekil 5.6. Kazanın Gerçekleştiği Yolun Tipi.....	36
Şekil 5.7. Kazanın Gerçekleştiği Yolun Sınıfı.....	36
Şekil 5.8. Kazanın Gerçekleştiği Yolun Yatay Geometri Özellikleri.....	37
Şekil 5.9. Kazanın Gerçekleştiği Yolun Düşey Geometri Özellikleri.....	37
Şekil 5.10. Kazanın Gerçekleştiği Yolun Kavşak Geometrisi.....	38
Şekil 5.11. Kazanın Gerçekleştiği Geçit Geometrisi.....	38
Şekil 5.12. Gün Durumuna Göre Kaza Gerçekleşme Sayıları.....	39
Şekil 5.13. Hava Durumuna Bağlı Olarak Kaza Sayıları.....	39
Şekil 5.14. Kazanın Gerçekleştiği Yol Yüzey Durumu.....	40
Şekil 5.15. Kazanın Gerçekleşme Tipi.....	40
Şekil 5.16. Araç Sayılarına Kazanın Gerçekleşme.....	41

Şekil 5.17. Kazaların Gerçekleştiği Çarpışma Yeri.....	41
Şekil 5.18. Kazaya Etkiyen Yol Kusuru.....	42
Şekil 5.19. Kazanın Gerçekleştiği Noktada Otokorkuluk Grafiği.....	42
Şekil 5.20. Kazanın Gerçekleştiği Noktada Yaya Yolu Grafiği.....	43
Şekil 5. 21. Kazanın Gerçekleştiği Noktada Yol Şerit Çizgisi.....	43
Şekil 5.22. Kazanın Gerçekleştiği Noktada Trafik İşareti Tablosu.....	44
Şekil 5. 23. Kazanın Gerçekleştiği Noktada Yol Aydınlatma Grafiği.....	44
Şekil 5.24. Kazanın Gerçekleştiği Noktada Hasar Gören Unsur.....	45
Şekil 5.25. Kazanın Gerçekleştiği Noktada Yol Çalışması Durumu.....	45
Şekil 5.26. Gerçekleşen Kaza Sonrası Sürücü Durumu.....	46
Şekil 5. 27.Gerçekleşen Kaza Sonrası Yaya Ölüm Sayısı.....	46
Şekil 5.28. Gerçekleşen Kaza Sonrası Yaralanan Sürücü Sayısı.....	47
Şekil 5.29. Gerçekleşen Kaza Sonrası Yaralanan Yolcu Grafiği.....	47
Şekil 5.30. Gerçekleşen Kaza Sonrası Yaya Yaralı Grafiği.....	48
Şekil 5.31. Gerçekleşen Kaza Sayılarına göre ölüm sayıları.....	48
Şekil 5.32. Gerçekleşen Kaza Sonrası Toplam Yaralı Sayısı.....	49
Şekil 5.33. Ölüm İle Sonuçlanan Kazalar.....	49
Şekil 5.34. Yaralanma İle Sonuçlanan Kazalar.....	50
Şekil 5.35. Kazaya Karışan Araç Cinsi.....	50
Şekil 5.36. Araç Kullanım Amacı.....	51
Şekil 5.37. Kazaya Karışan Sürücü Durumu.....	51
Şekil 5. 38. Kazaya Karışan Sürücü Yaşı.....	52
Şekil 5.39. Kazaya Karışan Sürücü Cinsiyeti.....	52
Şekil 5. 40. Kazaya Karışan Sürücü Ehliyet Durumu.....	53

Şekil 5.41. Kazaya Karışan Sürücü Eğitim Durumu.....	53
Şekil 5.42. Gerçekleşen Kazada Sürücü Hatası.....	54
Şekil 5.43. Gerçekleşen Kaza Noktasında Yolun Kaplaması.....	54
Şekil 5.44. Smote İşlemi Öncesi Algoritmaların Doğruluk (Accuracy) Sonuçları.....	55
Şekil 5.45. Smote İşlemi Öncesi Algoritmaların Duyarlılık (Recall) Değerleri.....	55
Şekil 5.46. Smote İşlemi Öncesi Algoritmaların F1_ Skor Değerleri.....	56
Şekil 5.47. Smote İşlemi Sonrası Algoritmaların Doğruluk (Accuracy) Sonuçları.....	57
Şekil 5.48. Smote İşlemi Sonrası Algoritmaların Duyarlılık (Recall) Değerleri.....	58
Şekil 5.49. Smote İşlemi Sonrası Algoritmaların F1_ Skor Değerleri.....	58
Şekil 5.51. Ankara Ve Çevre İlçeleri Dâhil Kaza Noktaları Genel Görünüm (2013-2020 Yılları).....	60
Şekil 5.52. Ankara İli Kaza Noktaları İçin Örnek Bir Kavşak (2013-2020 Yılları).....	61
Şekil 5.53. Ankara İli Kaza Noktaları Örnek Bir Şehir İçi Yol (2013-2020 Yılları).....	63

TABLÖLER LİSTESİ

	Sayfa
Tablo 3.1. Geleneksel Yöntemler ile Ysa Arasındaki Farklar (Crone, 2004).....	19
Tablo 3.3. Geleneksel Yöntemler ile Ysa Arasındaki Farklar (Mcneill ve Anderson, 1992).....	20
Tablo 4.1. Ölümlü Yaralanmalı Trafik Kaza Tutanaklarından Elde Edilen Veri Seti ¹ ..	27
Tablo 4.2. Ölümlü Yaralanmalı Trafik Kaza Tutanaklarından Elde Edilen Veri Seti ² ..	28
Tablo 4.3. Veri Analizi Kategorileri (Kakde ve Agrawal, 2018).....	28
Tablo 4. 4. Karmaşıklık Matrisi.....	31
Tablo 5.1. Smote İşlemi Öncesi Algoritmaların Değerleri.....	56
Tablo 5.2.Smote İşlemi Sonrası Algoritmaların Değerleri.....	59

SİMGELER ve KISALTMALAR

Kısaltmalar

ANN	Artificial Neural Network (Yapay Sinir Ağları)
DN	Doğru Negatif
DP	Doğru Pozitif
DSÖ	Dünya Sağlık Örgütü
EDA	Exploratory Data Analysis (Keşifsel Veri Analizi)
EGM	Emniyet Genel Müdürlüğü
KNN	K-Nearest Neighbors
LightGBM	Light Gradient Boosting Machine (Hafif Gradyan Artırma Makinesi)
P	Precision
R	Recall
SVM	Support Vector Machines (Destek Vektör Makinesi)
SMOTE	Synthetic Minority Oversampling Method (Sentetik Azınlık Aşırı Örnekleme Yöntemi)
XGBoost	eXtreme Gradient Boosting (Ekstrem Gradyan Artırma)
YN	Yanlış Negatif
YP	Yanlış Pozitif
YSA	Yapay Sinir Ağları
YOGT	Yıllık Ortalama Günlük Trafik

1. GİRİŞ

Trafik kazaları günlük ortalama 3.242 kişi, yıllık olarak da 1,2 milyon kişinin ölümüne, 20 ila 50 milyon kişinin yaralanmasına veya sakat kalmasına sebep olmaktadır. Dünya üzerinde görülen tüm ölümler içerisinde trafik kazaları %2,1 pay almakta ve ölüm nedenleri listesi içerisinde 11. Sırada gelmektedir. Bu oranlar orta ve düşük gelir seviyeli ülkelerde daha da artmaktadır (Who, 2004).

Türkiye de 2019 yılına ait trafiğe kayıtlı motorlu araç sayısı 23.156.975 iken 2020 yılında bu sayı 24.195.880'e kadar yükselmiştir. Trafikte giderek artan araç sayısı beraberinde birçok trafik probleminin de artmasına sebep olmuştur. Trafik kazaları günümüzde hala devam eden büyük problemdir. DSÖ'nün raporuna göre karayollarındaki trafik kazaları iyi bir trafik planlaması çalışması ile azaltılabilir. Planlı bir trafik ağı, doğru trafik yönetimi ve karayolu tasarımı ile efektif çözümler üretilerek araçlar, yolcular ve yayalar can kaybından korunabildiği gibi maddi hasardan aynı doğrultuda kaçınabilirler.

Bu çalışmada trafik kazalarını analiz etmek ve çözüm bulabilmek amacı ile Trafik kazalarına etkiyen parametreler ve bu parametrelerin birbiri arasındaki bağlantılarının analizi yapılarak tahminleme modelleri geliştirilmiştir.

EGM verilerine göre ülkemiz de 2020 yılın da 983.808 kazanın meydana geldiği ve bunların 150.275 adetinin ölümlü yaralanmalı, 833.533'ü maddi hasarlı trafik kazası olduğu görülmüştür. TÜİK verilerinde 2020 yılında Türkiye'de gerçekleşen kazaların bir önceki yıla oranla %11,1 oranında azaldığı görülmektedir (Hemdil., 2021).

Bu çalışmada Ülkemizde ve Dünya da büyük problem olan trafik kazalarına karşı etkin bir çözüm üretebilmek anlamında birtakım analizler yapılarak, Ankara ilinde gerçekleşen trafik kazalarına sebep olan etkenleri tespit etmek ve bu etkenler arasındaki etkileşimini incelemektir.

Yapılan çalışmada Ankara ili sınırları içerisinde gerçekleşen ölümlü yaralanmalı trafik kaza tutanakları Trafik Daire Başkanlığından alınarak modellemeler yapılmıştır. YOGT verileri KGM' nün yayınlamış olduğu trafik hacim haritalarından, Ankara ili nüfus

bilgileri, meydana gelen kaza sayıları, kayıtlı motorlu araç sayıları TÜİK'ten elde edilmiştir.

Trafik kazalarında can ve mal kayıplarını ve toplumda meydana getirdiği sosyal ekonomik sorunları engellemek için her bir parametrenin detaylı incelenmesi gerekmektedir. Bu sebeple Yapılan analizlerde problemin temelinde yatan parametreler belirlenmiş ve istatistiksel çözüm önerileri sunulmuştur.



2. KAYNAK ÖZETLERİ

Geçmişte trafik kazaları ile alakalı çalışmalar aşağıda literatür araştırması sonucu elde edilmiş ve sunulmuştur.

Kartal vd., (2011) yaptıkları bu çalışma da Kayseri, Sivas ve Yozgat illerindeki trafik kazalarını incelemiş, lojistik regresyona tanımlanacak veriler sıklık ve tanımlamalı istatistikleri çıkartılmış daha sonrasında ise belirtilen iller göz ardı edilerek, tüm kaza tutanaklarında ulaşılan veriler analizde kullanılmıştır. Modelleme sonucu için iki değerlendirme kriteri oluşturulmuş olup bunlar, kaza sonucunun yaralanma ile sonuçlanması diğeri ise ölümle sonuçlanmasıdır. Trafik kazalarında hata arttıkça sonuçların ölümle sonuçlanma oranının arttığı gözlemlenmiştir. Bölünmemiş yollarda gerçekleşen kaza sayısının bölünmüş yollara göre daha fazla olması buna örnektir. Yapılan çalışmada Kayseri ilinde meydana gelen kazaların diğer illere oranla daha fazla olduğu modellenen sonucu elde edilmiştir.

Budak ve Erpolat , (2012) yaptıkları çalışmada Kredi Riski Tahmininde Yapay Sinir Ağları ve Lojistik Regresyon Analizi Karşılaştırılması çalışmasında doğru sınıflandırma oranları karşılaştırmasında sinir ağları yaklaşımı ile lojistik regresyon modelinin başarı oranını %95,17 bulmuşlardır.

Jiawei vd., (2011) 1950'lerin başında K En Yakın Komşu (KNN) algoritması öne sürülmüştür fakat ilerleyen süreçlerde modellemeler de kapsamlı veriler kullanılmaya başlanması ile KNN 'nin yetersiz kaldığı, sonuçları elde etmekte zaman kaybı yaşattığı görülmüştür. Günümüzde Bilgisayarların yüksek performans sergilemesi ve artan bilgi işlem gücü ile popülerliğini geri kazanmıştır.

KNN algoritması tahminleme modelinde tüm veri setinde kendine en yakın komşu veriyi K değerini temel alarak çalışır. Buradaki K değeri anlamına bakılacak veri sayısıdır. Modellemeye bir veri seti daha eklenmeye çalışıldığında algoritma en yakın K kadar eleman içerisinde en yakın olanları belirleyerek modellemeye dahil eder. Mesafe ölçütü içinse Öklid uzaklığı kullanılır. Modellemelerde seçilen K sayının değerinin sınıflandırma başarısında önemli rol oynamaktadır. Modellemede kullanılan K değerinin seçimi değerlendirmeye tabi tutulan veri setine bağlı olarak değişkenlik

göstermekte olup, yapılan çalışmalarda K değerinin en büyük değer seçilmesi sonucunda sınıflandırma doğruluğunun büyük oranda doğruluğu arttırdığı gözlenmiştir.

Meltek ve Kayaalp (2020), Yaptığı çalışmada sınıflandırma problemlerinde Destek Vektör Makineleri (SVM) son zamanlarda ön plana çıktığını, KNN ile kıyaslandığında Yüksek ve etkin makine öğrenmesi algoritmalarından biri olarak göze çarptığını belirtmişlerdir. Bunun En önemli avantajlarından biri öğrenme aşamasına işlem sayısını azaltarak hızlı sonuçlar elde etmesi olduğunu belirtmişlerdir. Büyük veri setlerinde büyük avantajlar sağlayan bu algoritma optimizasyon temellidir. Destek Vektör Makinesinin literatürde birçok alanda kullanıldığı yapılan araştırmalarda elde edilmiştir. Ancak destek vektör makinelerinin sınıflandırma işlemlerinde diğer yöntemlere göre daha çok tercih edilmektedir. Denetimli öğrenme yöntemine dayanan destek vektör makineleri doğrusal ve doğrusal olamayan sınıflandırma yöntemleri yapabildiği belirtilmiştir.

Tan P vd., (2016), yaptıkları çalışmada Makine öğrenmesinde Ele alınan herhangi bir veriyi gruplandıran ve bu verilerin her birinin birbiri ile bağlantılarını belirleyerek anlamlı veri kümeleri oluşturan Karar Ağacı yöntemi kullanarak yapmaktadır. Anlamlı veri kümelerini ağaç diyagramı şekline dönüştüren, oluşturduğu anlamlı veri kümeleri ile çıktı veri kümeleri oluşturan bu yöntem denetimli öğrenme yöntemlerinden biridir. Girdi veri kümeleri ile çalışan ve elde edilen çıktı veri kümeleri arasında bağlantıyı keşfettiği görülmüştür.

Kuşkapan ve Çodur (2021), Hata ölçütlerinin WEKA analiz programı ile yapıldığı çalışmada, ülkemizde gerçekleştiren trafik kaza verileri kullanılarak yıllara göre ölüm ve yaralanma sınıflandırılması yapılması için kullanılan çok katmanlı algılayıcı, regresyon ve K En Yakın Komşu yöntemlerini kullanılmışlardır. Elde edilen sonuçlarda birden fazla kriter için K en yakın komşu algoritmasının diğerlerine kıyasla daha iyi performans verdiği görülmüştür.

Özden ve Acı (2016), yaptıkları çalışma da 2005 ve 2014 yılları arasında Adana ilinde gerçekleşmiş olan trafik kazaları ve meteorolojik verileri ile yapılan tahmin modelinde kaza sayısı ve yaralı sayısı tahmin edilmek için modellemeler yapmışlardır. İleri Beslemeli Çok Katmanlı Yapay Sinir Ağı, Fonksiyon Uydurma Yapay Sinir Ağı,

Genelleştirilmiş Regresyon Yapay Sinir Ağı, Regresyon Ağacı, Destek Vektör Makinesi ve Çoklu Doğrusal Regresyon Analizi yöntemlerini kullanılmışlardır. Kaza sayısı ve yaralı sayısı tahmininde Destek Vektör Makinesinin başarılı sonuçlar verdiği görülmüştür.

Galal ve Awadalla (2012), yaptıkları çalışmada Trafik kazası kayıplarını karşılaştırmalı olarak tahmin etmek için Yapay Sinir Ağı (YSA) ve regresyon tekniklerini uygulamışlardır. Her iki yaklaşım da 1991'den 2009'a kadar nüfus, kayıtlı araba sayısı ve diğer ilgili faktörler hakkındaki tarihsel verileri kullanarak kaza kayıplarını modellemişlerdir. Yapay sinir ağı tahminleri, tahminler ve gözlemlenen veriler arasında maksimum %1,84'lük bir farkla veriler için en iyi uyumu sağlamıştır. Çalışma, YSA'larının kaza kayıplarının analizi ve tahmini için güçlü bir araç sağladığını göstermiştir. Kazaların başlıca nedenleri, sürücü davranışları, araç filosu ve koşulları, yol ağı kusurları, hız sınırı ihlali, emniyet kemeri kullanım ihmali ve trafik kurallarının uygulanmamasına bağlanmıştır.

Chong vd., (2004), yaptıkları çalışmada Üç makine öğrenimi yaklaşımı kullanarak trafik kazaları sırasında meydana gelen yaralanmaların ciddiyetini tahmin etmek için bazı modeller sunmuşlardır. Hibrit öğrenme yaklaşımları, karar ağaçları ve karar ağaçları ile sinir ağlarını içeren eşzamanlı bir hibrit model kullanılarak eğitilmiş sinir ağları düşünülmüştür. Deney sonuçları, hibrit karar ağacı-sinir ağı yaklaşımı olarak kabul edilen makine öğrenimi paradigmaları arasında bireysel yaklaşımlardan daha iyi performans gösterdiğini ortaya koyduğunu tespit etmişlerdir.

Mussone vd., (1999), yaptıkları çalışmada Milano'daki araç kazalarının analizi ile ilgili bir model geliştirmek için yapay sinir ağlarının (YSA) kullanımına dayalı alternatif bir yöntemi açıklamışlardır. Farklı senaryolar kullanan kentsel kavşakların tehlike derecesi, YSA modeli ile ölçülmüştür. Metodoloji, kentsel araç kazalarının modellemesini YSA'nın yenilikçi kullanımıyla ele almamızı sağlayan ilk sonuçtur. Diğer sonuçlar model çıktılarıyla ilgilidir: kavşak karmaşıklığı, kavşak düzenlemesine bağlı olarak daha yüksek bir kaza indeksi belirleyebilir. Yayaaların üzerinden geçme ile ilgili en yüksek endeks, gece saatlerinde sinyalizasyon yapılmayan kavşaklarda meydana geldiği görülmüştür.

Altın Yavuz vd., (2021), yaptıkları çalışmada Trafik kazalarının sebebinin belirlenmesi, trafik kazalarının önlenmesi açısından oldukça önemli olduğunu vurgulayarak geçmişte meydana gelen kazalara sebep olan etkenleri belirlemek trafik kazaları azaltmak için oldukça önemli olduğunu belirtmişlerdir. Doğru bir modelleme ile kaza analizi yapılarak bu nedenler belirlenmelidir. Yapılan çalışmada 2012 ve 2016 yılları içerisinde Antalya ili içerisinde gerçekleşen 3181 trafik kazası ele verileri ele alınarak makine öğrenmesi yöntemleri ile modelleme kurulmuştur. Ele alınan makine öğrenmesi yöntemleri içerisinde En yüksek doğruluk oranını Naive olduğunu belirlemişlerdir

Bayata ve Hattatoğlu (2011), 1974 -2007 yılları arasında Erzincan ili içerisinde ceza alan sürücülerin sayısı ile kaza sayıları ele alınarak yapılan modellemede ceza sayısının kaza sayılarını etkileyip etkilemediğini ölçmek istemişlerdir. Çok değişkenli istatistiksel analiz ve yapay sinir ağları (YSA) yöntemleri ile yapılan modelleme sonucunda ceza sayılarının artması kaza sayılarının azalması üzerinde bir etki göstermemiştir. Yapay sinir ağlarının R^2 değerinin çok değişkenli istatistiksel analize göre daha yüksek, ortalama karesel hatasının ise minimum olduğu elde edilmiştir. Bulunan sonuçlar ışında YSA ‘nen daha başarılı bir model olduğu görülmüştür.

Bayata ve Hattatoğlu (2011), yaptıkları çalışmada Erzincan ilinde 1985-2010 yılları arasında meydana gelen yıllık trafik kaza sayıları kullanılarak farklı istatistiksel yöntemlerle modelleme yapılmışlardır. Modellerde trafik kaza sayıları bağımlı değişken olarak kabul edilmiştir. Aynı yılları kapsayan nüfus, araç sayısı, ölü sayısı ve yaralı sayıları ise bağımsız değişken olarak kabul edilmiştir. Bu veriler Yapay Sinir Ağları ve çok değişkenli zaman serileri yöntemleri ile modellenmiştir. YSA yönteminde trafik kazaları değerleri [0 ile 1] arasında bir ölçeklemeye tabii tutulmuştur. Zaman serisi analizi ile seri durağan hale getirilmiştir. Etki tepki analizi ve varyans ayrıştırması yapılarak nüfus ve araç sayısındaki bir birimlik şokların kaza sayısı üzerindeki etkileri araştırılmıştır. Kaza sayılarını etkileyen değişkenler arasında, araç sayısı nüfusa göre istatistiksel olarak daha anlamlı çıkmıştır. Yöntemler kıyaslanırken ortalama karesel hata (OKH), belirlilik katsayısı (R^2) ve Akaike bilgi kriterine (AIC) göre en iyi model belirlenmiştir. Yöntemler arasında YSA yönteminin R^2 değeri diğer yöntemlere göre daha yüksek, ortalama karesel hatası (OKH) ise daha düşük çıkmıştır. Bu sonuçlara göre YSA yönteminin istatistiksel olarak daha başarılı olduğu görülmüştür.

Mohammed vd., (2018), yaptıkları çalışmada Kavşak ve buna benzer noktalarda kaza tahmini oluşturulmasında kaza tahmin modelleri oldukça önem taşımaktadır. Kaza tahmini oluşturulurken ele alınan verilerin geçmiş verileri içermesi ve kapsamlı olması gerekmektedir. Söz konusu çalışma trafik kaza tahmin modellerinin sınıflandırılması yapılmış olup Çoklu Doğrusal Regresyon, Poisson Regresyon ve Conway-Maxwell Poisson Regresyon Modeli ve Yapay Sinir ağları kullanılmıştır.

Kulkula B. vd (2019), İngiltere de 2016 yılı içerisinde Calderdale gerçekleşen trafik kaza verileri ile Trafik kazasına sebep olan önemli faktörleri elde etmek ve kazaların boyutlarını incelemek amacı ile sınıflandırma performansı yüksek olan Bulanık Mantık, Rastgele Orman, Radyal Temel Fonksiyon Ağı, Hiyerarşik Vektör Nicelermeyi Öğrenme Çok Katmanlı Perceptron ve Naive Bayes algoritmaları incelenmiştir. Bulanık mantık yönteminin başarılı sonuçlar elde ettiği görülmüştür.

Dong S. vd., (2022), yaptıkları çalışmada Pakistan içerisinde 2015-2019 yılları arasında Peşaver'den Rahim Yar Han'a kentsel ve kırsal kesimlerinde meydana gelen trafik kaza verileri ele alınarak Cat Boost (Kedi Artırma), LightGBM (Hafif Gradyan Artırma Makinesi), Adaptive Boosting (Uyarlamalı Güçlendirme) algoritmaları ile modellemeler oluşturulmuştur. Model değerlendirmelerinde Doğruluk (Accuracy) metriği baz alınmış olup LightGBM (Hafif Gradyan Artırma Makinesi) yönteminin diğer algoritmalara kıyasla daha doğru sonuçlar verdiği görülmüştür.

Yassin vd., (2020), trafik kaza tahmin modellemesi Etiyopya için yapılan çalışmada K-ortalama kümeleme ve Rastgele orman algoritmaları ile modellemeler kurulmuş her iki model değerlendirmesi için Doğruluk (Accuracy) metriği baz alınmıştır.

Aydın (2022), Dengesiz sınıflandırma bir model için ele alınan veri seti içerisinde azınlıkta bulunan verilerin çok az örneğinin olması durumu olarak tanımlamış ve Müşteri Kaybı Tahmini yapabilmek adına Makine Öğrenmesi yöntemlerinden faydalanılan Müşteri Kaybı Tahmininde Sınıf Dengesizliği Problemi Çalışmasında dengesiz veri probleminin çözümünde Aşırı Örnekleme (Over Sampling) veya Sentetik Azınlık Aşırı Örnekleme Yöntemi (SMOTE) yöntemlerinin öncelikle tercih edilebilecek yöntemler olduğu kanısına varılmıştır.

3. KURAMSAL TEMELLER

3.1 Ulaşım

Ulaşım bir nesnenin veya kişinin bulunduğu yerden farklı bir yere aktarılmasıdır. İnsanlık tarihinin ilk dönemlerinden itibaren var olan bu faaliyet türü zaman içerisinde çeşitli değişimler ile günümüzdeki noktasına evrilmiştir. M.Ö 7000’li yıllarda Mezopotamya kültüründe insanlığın gelişmesi için ulaşım ihtiyaçları artış göstermiştir. Roma kültürü için de ulaşım çok büyük önem arz ettiği için yollara önem verilmiştir. Ulaşımında yeniçağda yeni bir soluk kazanmıştır. Hayvanlar, arabalar ulaşım için bir araç olarak kullanılmaya başlanmıştır. 2 tekerlekli araçların kullanımı M.Ö. 2000’li yıllara. 3 tekerlekli araçların kullanımı ise M.Ö. 200’lü yıllara dayanmaktadır. Uygarlıkların değişmiş ve artan teknoloji ile kullanılan araçlara gelişim göstermiştir. Tekerleğin icadı ulaşımına yeni bir yaklaşım kazandırmıştır. İçten patlamalı motorlar günümüzde hala varlığını sürdürmekte olup Almanların çalışmaları ile piyasaya sunulmuştur (Bulak H. , 2010;Sungur vd, 2014).

Ulaşım ağları günümüzde Karayolu ulaşımı, denizyolu ulaşımı, havayolu ulaşımı ve demiryolu ulaşımından oluşmaktadır. Yıllar boyunca toplumlar ve insanlar arasındaki bağlantıyı. İletişimi ve ticareti sağlayan kervan yolları, O bölgenin coğrafi koşullarının uygun olduğu bölgelerden geçmiştir (Ökse, 2005).

Ulaşımında çeşitli enstrümanların kullanılması farklı ulaşım şekillerine yol açarak, farklı ulaşım sistemlerinin oluşmasına, alternatiflerin ortaya çıkmasına sebep olmuştur. Değişik ulaşım araçlarının kullanımının ise birbirinden farklı sebepleri mevcuttur. Jeopolitik, coğrafi, ekonomik ölçütler ulaşım araçlarının kullanım dağılımını büyük ölçüde etkilemiş ve bu noktalardaki ulaşım ağının planlaması bu faktörlere göre gelişmiş ve günümüzdeki halini almıştır. Genel olarak topoğrafik yapı ulaşım ağı planlamasını büyük ölçüde etkilemektedir. Bu bağlamda bu etkileşimden en çok etkilenen karayolu ulaşımı olmakta ve karayolu ağları planlamaları topoğrafik özelliklerden büyük ölçüde etkilenecek günümüzdeki şeklini almıştır.

Ulaşım ağının geçtiği bölgenin yükselti ve eğimi oldukça önemlidir. Güzergâhın belirlenmesinde önemli olan bu etmenler aynı zamanda karayolu ve demiryolu ulaşımını

olumsuz yönde de etkileyebilmektedir. Türkiye de dağlar doğu batı yönünde uzanmaktadır buna bağlı olarak iç kesimlere ulaşımı zorlaştırmaktadır (Aydın ve Oral , 2018).

3.1.1 Ülkemizde karayolu ulaşımının genel durumu

Son dönemlerde Osmanlı imparatorluğunda demiryolları iç kesimler ve kıyı kesimler arasındaki bağlantıyı sağlamak amacı ile kullanılan önemli bir enstrüman olmuştur. 19. Yüzyılın ortasından itibaren demiryolları hatlarının kurulabilmesi için birçok girişimci, gerekli adımları atmış dönemin padişahından ayrıcalık istemiştir (Avcı S, 2005).

Cumhuriyetin ilanından II. Dünya Savaşı'nın sonuna kadar olan dönemde ulaşım politikası içerisinde demiryolu çalışmaları birinci sırada yer almıştır. Hazırlan bütçelerde demiryolu yapımı önemle ele alınmış bir politika haline getirilmiştir. Demiryolu politikası Cumhuriyetin ilk yıllarında altın çağını yaşamıştır. Ülke kalkınmasını ve savunmasını sağlamak adına atılmış bir adım olan demiryolu politikası o dönemde ülke çıkarlarına büyük hizmet vermiştir (Özdemir, 2006).

1950'li yıllara gelindiğinde ulaşım kültürü değişmeye başlamış demiryolu ağırlıklı yolcu taşımacılığı yerini kapıdan kapıya taşıma kolaylığı taşıyan karayolu taşımacılığına bırakmıştır. Bu değişimle birlikte demiryolu taşımacılığı yük taşımacılığındaki payı giderek azalmıştır. Söz konusu oran 1955 dönemi öncesi %50 iken kara yolu taşımacılığının payı her geçen yıl artmıştır. 1995 Yılında karayolu taşımacılığının yük taşıma oranı %95 olmuştur. Yolcular ise yıllar itibari ile karayolu taşımacılığına yönelerek 1970 de yolcuların %91'i Karayolunu tercih etmiştir (Parıldar , 2014).

Türkiye geçmişten günümüze gelişen yapısıyla her ne kadar diğer ulaşım alanlarında gelişme gösterse de günümüz itibari ile karayolu ulaşımı Türkiye'de önemli bir yer kaplamaktadır. Karayolu taşımacılığı ile Türkiye'nin her noktasına ulaşım sağlanması bu taşımacılık türünü diğerlerinden daha önemli kılmakta ve bu alandaki yatırımların ilerlemesine sebebiyet vermektedir. Ancak karayolu trafiği ile birlikte trafikteki araç sayısında aynı oranda artış gözlemlenmekte ve bu beraberinde olumlu ve olumsuz sonuçları doğurmaktadır.

3.1.2 Ülkemizde ve Dünyada trafik kazaları

Paris'te 1771 yılında kedi kendine hareket eden anlamını taşıyan otomobil Cugnot tarafından buharlı bir araç olarak üretilmiştir. 1821 yılın da ise buharlı posta arabası İngiliz Griffilt tarafından üretilmiştir. Yine İngiltere sınırları içerisinde 824'te Hill'in ve 1831'de Hannock'un yaptıkları buharlı posta araçları karayolu araçlarının ilk örnekleridir. Karayolu taşımacılığında gerçekleşen ve sonucu ölümle biten ilk kaza ise İngiltere de görülmüştür. İlerleyen yıllarda araç talebindeki artışa bağlı olarak artan araç sayısı trafik kazalarının artışına sebep olmuştur. Otomotiv sektöründeki gelişim ile birlikte araçların km başına düşen hızları giderek artmıştır. Oluşan kazalara karşı olarak ülkeler yeni bir trafik güvenliği politikası belirlemek mecburiyetinde kalmış buna uygun yasalar düzenlemiştir. 20. yy.'da ABD devletleri bu gereksinimlerin farkında olarak trafik kurlarını yasalaştıran ilk devlet olmuştur (Bulak H, 2010).

Türkiye Ülke bazında Trafik kazası nedeni ile oluşan ölümlerin yaklaşık yarısını oluşturan on ülke arasında yer almaktadır. Toplum sağlığı açısından trafik kazaları gündemde olması gereken bir konudur. Trafik kazalarına sebep olan birçok etmen olsa bile en temelde sürücü, yolcu, yaya trafik kazası içerisinde en büyük yapa sahip olan 3 faktördür. Bu etmenler arasında sürücü trafik kazalarında %95 ile en büyük paya sahiptir. Hız, alkol, madde bağımlılığı, uykusuzluk, emniyet kemeri kullanılmaması, kask kullanılmaması trafik kazalarına neden olan etmenlerdir. Trafik kazalarını engellemek için oluşturulan düzenlemeler yerinde ve düzgün uygulanmaktadır (Sungur vd, 2014).

Türkiye'de her geçen yıl trafikteki araç sayısı artmaktadır. “ Türkiye'de 2020 yılında bir önceki yıla göre trafiğe kaydı yapılan taşıt sayısı %54,8 artarak 1 milyon 38 bin 905 adet olurken, trafikten kaydı silinen taşıt sayısı %87,3 azalarak 48 bin 328 adet oldu. Böylece 2020 yılında trafikteki toplam taşıt sayısında 990 bin 577 adet artış gerçekleşti” (Demircigil, 2020).

“Ülkemiz karayolu ağında 2020 yılında toplam 983 bin 808 adet trafik kazası meydana geldi. Bu kazaların 833 bin 533 adedi maddi hasarlı, 150 bin 275 adedi ise ölümlü yaralanmalı trafik kazasıdır. Yıl içerisinde meydana gelen ölümlü yaralanmalı trafik kazalarının %75,8'i yerleşim yeri içinde %24,2'si ise yerleşim yeri dışında meydana

geldi. Türkiye'de 2020 yılında bir önceki yıla göre trafikteki motorlu kara taşıtı sayısı %4,3 artarken, toplam kaza sayısı %15,8, ölümlü yaralanmalı kaza sayısı %14,1, maddi hasarlı kaza sayısı %16,1, toplam ölü sayısı %11,1, kaza yerindeki ölü sayısı %13,0, kaza sonrası ölü sayısı %9,5 ve yaralı sayısı %20,1 azaldı'' (Hemdil, 2021).

İnsan ölümlerinin sebepleri olarak sıralandığında ülkemizde trafik kazaları 3. sırada yer almaktadır. Bu kadar acı sonuçlar karşısında maalesef ülkemizde trafik kazaları ile ilgili bir üst kurul veya iletişim sağlayacak bir yapı bulunmamaktadır (Bayata , 2010).

Artan kent nüfusları ile ve yerleşim yeri içerisinde artan araç sayısı trafik problemlerinin görülmesine yol açmıştır. Trafik eğitimin yetersizliği kazaların artışında rol aldığı gibi birçok faktör kazaların sebebini oluşturmaktadır. Covid-19 salgını ile gelen sokağa çıkma yasakları kent içi hareketliliğin azalması beraberinde getirdi. Şehir içi trafiğin azalması ve trafikte olan araç sayısındaki düşüş trafik kazalarının azalmasına neden olmuştur. Fakat sokağa çıkma yasaklarının bitmesi ve bireylerin salgından korunmak amacı toplu taşıma tercih etmemesi ilerleyen süreçlerde, düşen trafik kaza sayılarının yükselmesine sebep olacak gibi duruyor.

3.1.3 Trafik kazalarına etki eden parametreler

Dünya üzerindeki nüfus arttıkça ulaşım ihtiyacı da aynı oranda artmaktadır. Ulaşımında en çok tercih edilen karayolu taşımacılığının karşılanabilmesi için araç sayısındaki artış beraberinde trafik kazaları gibi ciddi sonuçlara sebebiyet vermektedir. Günümüzdeki tablo hala istenilen düzeyde trafik bilincinin oluşmadığını göstermektedir. Bu açıdan ülkeler kendi içerisinde trafik planlamaları yapmakta ve önemli adımlar atmaktadır.

Karayolu ağının genişlemesi ve daha konforlu hale gelmesi ile artan talep karayolu üzerindeki taşıt sayısının artışına neden olmuştur. Bu gelişme ile karayolu ağı üzerinde trafik yoğunluğunun yıllar itibari ile arttığı gözlenmiştir. Trafik mühendisliği bu noktada elzem bir ihtiyaç haline gelmiş ve ulaşım ağlarının daha sistemli bir şekilde düzenlenmesi ihtiyacı doğmuştur. Karayollarının bakımı, işletmesi ve mevcut yollarının iyileştirilmesi için gerekli çalışmalar yapılmakta ve lokasyonlara uygun kavşak tasarımları trafik yoğunluğu olan bölgeleri bu açıdan rahatlatacaktır. Fakat bazen bu çalışmalar dahi yetersiz kalmaktadır. Artan talebe cevap vermekte zorluk çeken ulaşım

ağında dış etmenler devreye girdiğinde trafik probleminin ciddi boyutlara taşındığını görebiliriz. Meteorolojik koşullar ve trafiğin temel etmenleri olan sürücü, yaya, yolcu trafik kaza oranını büyük oranda etkilemektedir.

Yıllık ortalama günlük trafik karayolları üzerinde yıllık aktif araç sayısını işaret etmektedir. Bu değişken birçok faktörün arasında çok önemli olmasından kaynaklı olarak modellemelerin birçoğunda kullanılmaktadır. Trafik yoğunluğundaki artışlar dolaylı yoldan o karayolu üzerindeki kazaların artmasına sebep olmaktadır.

Kazaların nedenlerini ulaşım ağını kullanan insanlar, kullanılan yolun durumu ve çevresel faktörleri ele alarak değerlendirebiliriz. Sürücülerin durumunu incelemek ve değerlendirme faktörleri arasında ele almak kazaların sebebini belirlemek açısından uygun bir görüş olacaktır. Sürücülerin sosyo-kültürel durumu, bulundukları yerin coğrafi özellikleri, psikolojik ve fiziksel durumları, sürüş esnasında sergiledikleri beceri kazalar üzerinde etkilidir. Bahsedilen özelliklerin tümü kazanın gerçekleşmesi için bir neden olabilmektedir. Bu durumların hepsi tek başına ele alınarak araştırma konusuna bahis olabilmektedir. Her bir durum için farklı sonuçlar elde edilebilir (Selimoğlu, 2014).

Trafik kazaları çeşitli faktörlerden kaynaklanmaktadır. Trafik kazalarına neden olan etkenler arasında o andaki hava koşulları da yer almaktadır. Hava şartlarına bağlı olarak trafik kazalarına göre ölüm oranlarında farklılık var. Trafik kazalarına bağlı ölümlerin sayısını azaltmak için hava şartlarına bağlı olarak meydana gelen trafik kazalarının görülme sıklığını tahmin etmek gerekmektedir. Bu bildiride, hava koşullarından kaynaklanan trafik kazalarının tahmin etmek için bir model önerilmiştir. Markov süreçleri teorisine tahmine dayalı modelleme uygulandı. Önerilen model için gerçek veriler uygulanarak, trafik kazalarının görülme sıklığını tahmin etmek için uygulamadaki olay sayısı ile karşılaştırılmıştır (Chung vd., 2013).

Kazaların neden olduğu faktörlerden olan insan faktörü analizi diğer parametrelerin aksine daha karmaşık ve literatürde spesifik çalışmaların olduğu bir alandır. İnsanların anlık duygu durum tahminlerini ve değişen çevresel, psikolojik durumunu incelemek şu anki şartlarda pek mümkün olmadığı için bu parametrenin analizi diğer parametrelerden daha zordur.

Her ne kadar sürücünün durumu kazaların oluşumunda etkili rol oynasa da araç teknik donanımı kaza değerlendirmelerinde dikkate alınması gereken bir husustur. Kullanım amacına göre sınıflandırılmış araçların, düzenlenen kurallara uyması kazaların önlenmesi açısından oldukça önemlidir. Kazalarda insan etkileri kadar yolun kalitesi de etken faktörlerden olup. Yol ağlarının bakım ve onarımının yapılması gerekmektedir. Bölgenin coğrafi özellikleri ile hava durumu ve çevre yapısı kazalar üzerinde etkilidir (Selimoğlu, 2014).

Makine öğrenmesi günümüz teknolojisinin gelişmesi ile birçok alanda kullanılmaya başlamıştır. Kayıt altında tutulan verilerin artması ve akabinde bu verilerin analiz edilme süreçleri eski yöntemlerle oldukça zor bir hal alırken makine öğrenmesi ile artık bu yöntemler rafa kaldırılmıştır. Daha etkin analizler veriler her ne kadar büyük olursa olsun makine öğrenmesi ile daha şekillendirilebilir ve düzenli bir analize yol açmıştır.

3.2. Makine Öğrenmesi

“Kâinatın başlangıcı ve yaşamın başlangıcının ardından evren üzerindeki en büyük olay yapay zekânın ortaya çıkışıdır.” BBC ile röportajında dijital fiziğin öncülerinden ve MIT Bilgisayar Bilimleri laboratuvar yöneticilerinde olan Edward Fredkin yapay zekâ için bu ifadeleri kullanmıştır (Pirim, 2006).

Yapay zekâ aslında insan beyninin algoritmasının taklitlidir. Beynin algoritmalarının bilgisayar sistemlerine tanıtılması ile bu yeteneği kazanan modellemeler var olan problemlere çözüm üretmek, tahminleme yapabilme gibi yetenekleri kazanır (Atınç, 2017).

Yapay zekâ teknolojisi her geçen gün artarak gelişmektedir. Yapay zekâ teknolojileri günümüzde birçok alanda kendine yer bulmuş ve birçok problemin çözümünde önemli rol oynamaktadır. Bu teknolojinin gelişim sürecinde doğadan esinlenilmiş insanın genetik yapısı ve sinir sistemi taklit edilmiştir. Uzman sistemler, genetik algoritmalar, bulanık mantık, yapay sinir ağları, makine öğrenmesi gibi teknikler yapay zekâ teknolojisi olarak adlandırılmaktadır.

Gün geçtikçe büyüyen problemlerin çözümü aşamasında makine öğrenmesi bilimi tercih edilen bir alternatif olmuştur. Büyük Miktardaki Verilen El ile İşlenmesi Oldukça Zordur. Makine Öğrenmesi Bu Büyük Verilerin İşlenmesi ve Geleceğe Dair Tahmin Modellemeleri Oluşturulabilmektedir. Verinin İncelenip İçerisinde İşe Yarayan Bilginin Çıkarılmasına Veri Madenciliği Denilmektedir.

Makine öğrenmesi de bilgisayarın eğitilmesi ile oluşturulan algoritmalar sonucu işler. Buradaki öğrenmeyi şu şekilde tanımlayabiliriz “geçmiş tecrübeler sayesinde gelecekte meydana gelebilecek olaylar hakkında tahmin oluşturmak ve ortaya çıkabilecek problemlere çözüm üretmektir.” bilgisayar bu tecrübeyi ele alınan ve makineye işlenen verilerden almaktadır.

Makine öğrenmesi temelinde şu işlemleri gerçekleştirmektedir,

- Sınıflandırma
- Kümeleme
- Regresyon
- Tahmin

Makine öğrenmesini kullanabilmek için elimizdeki verileri doğru ve anlamlı bir hale getirmek gerekmektedir. Bu verileri anlamlı hale getirdikten sonra kurulan algoritma ile makine öğrenmesi tahminleme modeli oluşturacaktır. Çeşitli analizlerle makine öğrenmesinin doğruluğu ölçülür eğer istenilen sonuç elde edilmezse tekrar bir algoritma oluşturmak gerekmektedir. Doğruluğu en yüksek tahminlemeye ulaşıncaya kadar modelleme tekrar edilir.

Makine Öğrenmesi denetimli öğrenme, denetimsiz öğrenme ve pekiştirmeli öğrenme olmak üzere üç sınıfa ayrılmaktadır. Rastgele orman modeli sınıflandırma analizleri içinde kullanılmakta olup, sınıflandırma Şekil 3.1’de görüldüğü gibi kategorize edilebilir.



Şekil 3.1. Makine Öğrenmesi ve Alt Grupları (Meltek ve Kayaalp, 2020)

Sınıflandırma problemlerinde denetimli öğrenme kullanılarak, eğitim sistemi içerisindeki amaç test edilen verilerle tahmin ve tanıma yapabilmektir. (Afrın ve Nahar) ‘‘Destek Vektör Makinesi, Yapay Sinir Ağları, Lojistik Regresyon, Basit Bayes, Multinom Basit Bayes, k-En Yakın Komşu, Rastgele Orman ve Karar Ağaçları algoritmaları denetimli öğrenme modeli oluşturulurken yaygın olarak kullanılan yöntemlerdendir’’ (Caruanac ve Niculescu-Mizil, 2006).

Denetimsiz Makine Öğrenmesi (UML: Unsupervised Machine Learning) modellemeler de kullanılan veriler ham halde yani yapay zekâya göre anlamlandırılmamış şekilde tanımlanmamaktadır. Denetimsiz öğrenmenin amacı sınıflandırma veya tanıma değildir. Kümeleme, tahminleme, olguların ölçülebilir niteliği içerisindeki ilişkilerin bulunması ve bu olguların ölçülebilir niteliği fazlası ile kapsamlıysa harcanan zamandan tasarruf etmek amacı ile Boyut azaltma gibi amaçlar ile kullanılmaktadır. Denetimsiz makine öğrenmesi ile elde edilen sonuçlar denetimli öğrenme modelleri için kullanılabilir (Chao, 2011).

Pekiştirmeli öğrenme, robot algısı ve hareketi, otomatik satranç oyuncusu ve otomatik araç sürüşü gibi karar verme problemlerini (genellikle bir dizi karar) çözmek için kullanılır (Chao, 2011).

KNN sınıflandırıcısı, bir örneğin sınıflandırmasının, vektör uzayında yakındaki diğer örneklerin sınıflandırılmasına en çok benzer olduğu varsayımına dayanır. Bayesian sınıflandırıcı gibi diğer metin sınıflandırma yöntemleriyle karşılaştırıldığında, kNN önceki olasılıklara dayanmaz ve hesaplama açısından verimlidir. Ana hesaplama, test belgesi için en yakın k komşuyu bulmak için eğitim belgelerinin sıralanmasıdır (Liao.ve Vemuri, 2002).

Karar ağacı (Decision Tree) denetimli makine öğrenmesinin alt dalıdır. Karar ağacı veri tabalarını kolayca birleştirip ve güvenilir sonuçlar elde edilmesini sağlar bu açıdan tüm parametrelere uygulanabilir olması ile yorumlana bilirliliği kolaydır. Karar ağaçları adından da anlaşılabilirliği gibi dallara ayrılarak toplamda 3 segmentten oluşur. Karar düğümü, dallar ve yapraklardan oluşur (Özekes ve Karakoç, 2002).

Rastgele Orman (Random Forest) bir sınıflandırma algoritması olup, temelinde karar ağacı yöntemini barındırmaktadır. Denetimli öğrenme yöntemlerinden biri olan bu yöntem, sınıflandırma ve regresyon için kullanılmaktadır. Elde edilen tüm bilgilerin ışığında en iyisini seçmeye yönelik olan bir modeldir (Breiman, 2001). Karar ağaçlarından oluşturulan bir orman içerisinde en iyi tahmin değerini seçer. Bu modellemede esas olan ormanın büyüklüğüdür. Veri seti ne kadar büyük olursa orman o kadar geniş ve elde edilen sonuçlar o kadar sağlam olacaktır. Bir karar ağacına kıyasla ormandaki verilerin büyüklüğü arttıkça elde edilen verinin doğruluğu o kadar artacaktır. Rastgele orman bu büyük veri seti içerisinde bir karar ağacına göre daha iyi sonuçlar verecek olacağı bu modelin ana mantığıdır.

Rastgele orman algoritmasında kullanıcıya her düğümde kullanılacak ağaç sayısı ve değişken sayısı sorulur. Her ağaç, eğitim verilerinden rastgele seçilen değişkenler tarafından oluşturulur ve algoritmanın “rastgele” olarak adlandırılması da buradan gelmektedir. Rastgele orman algoritması ile N adet ağacın en altında bulunan yapraklardan alınan değerlerden sonuçlar elde edilir ve verilerin değişkenlerinin ağırlıkları alınarak sonuç hesaplanır. Rastgele orman regresyon makine öğrenme türünde her ağaçtan ortalama sonuçlar alınarak sonuç değeri hesaplanır. Rastgele orman modeli daha sonra verilerin geri kalanı için de sonuçları tekrar hesaplamak için kullanılır (Meltek ve Kayaalp, 2020).

Lojistik regresyon analizinin kullanım amaç istatistikte kullanılan diğer model yapılandırma teknikleri ile aynıdır. En az değişkeni kullanarak en iyi uyuma sahip olacak şekilde sonuç değişkeni (bağımlı ya da cevap değişkeni) ile bağımsız değişkenler kümesi (açıklayıcı değişkenler, durumun başındaki değişkenler) arasındaki ilişkiyi tanımlayabilen ve genel olarak kabul edilebilir modeli kurmaktır. Yukarıda bahsedilen bağımsız değişkenleri genellikle birlikte değişkenler olarak adlandırılmaktadır (Coşkun vd., 2004).

XGBoost, Karar ağacı tabanlı bir makine öğrenmesi algoritmasıdır. Diğer yapay sinir ağı algoritmalarından daha iyi performans gösterme eğilimindedir. XGBoost, Extreme Gradyan Artırma teriminin kısaltılmış hali olup, temeli Gradyan Artırma ve karar ağacı algoritmalarına dayanan bir makine öğrenme tekniğidir. XGBoost algoritmasının orijinal hali Friedman tarafından 2002 yılında geliştirilmiştir (Dikker, 2017).

Sonrasında Washington Üniversitesi'nde iki araştırmacı olan Tianqi Chen ve Carlos Guestrin tarafından SIGKDD (Bilgi İşlem Makinalarının Bilgi Keşfi ve Veri Madenciliği Özel İlgi Grubu Derneği) 2016 konferansında makale olarak sunulmuştur ve makine öğrenme dünyasında çok popüler olmuştur (Chen ve Guestrin, 2016).

Bir XGBoost modeli eğitilirken iyi sonuçlar elde edebilmek için parametre ayarının yapıldığı bir algoritmadır. XGBoost'ta ne kadar iyi performans elde edildiğini gözlemlemek için bir modelleme yapıldıktan sonra diğer bir model oluşturulur. Oluşturulan ikinci model önceki model ile karşılaştırılır. İkinci model parametre ayarlaması yapılarak modellenir. Parametre ayarlamasındaki yaklaşım Gradyan Artırma yöntemine benzemektedir (Patrous, 2018). XGBoost içi dolu olmayan verileri model dışı bırakarak yüksek tahmin performansı gösterir. Model olarak ağaç toplulukları kullanması bakımından rastgele orman modellemesine benzemektedir. Aynı zamanda Rastgele orman modeli gibi bir denetimli öğrenme modelidir.

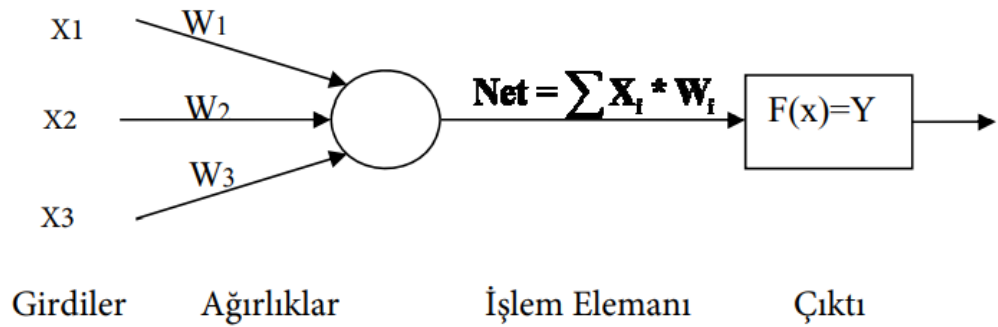
Karar ağaçları tek olgudan genel olguya varma yöntemi ile çalışmaktadır, LightGBM (Hafif Gradyan Artırma Makinesi) algoritması ise bu temele dayanmaktadır. Veri setlerinin her birinin başlangıçtaki ağırlığı aynıdır. Artırma özelliği ise Ağaç sayısının ve ormanın büyümesi ile birlikte ortaya çıkarak ağırlıkların değişkenlik göstermesidir. Yanlış sınıflandırılan verilerin ağırlığı artarak öğrenme zorluğu çeken noktalar

birleştirilir. Amaç daha güçlü bir yapı kurarak daha iyi sonuçlar elde etmektir (Arslan , 2020).

LightGBM en az 10.000 satır veri ile yani büyük veriler ile çalışmasına rağmen RAM'i minimum seviyede tüketmektedir. Hatta LGBM'nin küçük verilerde takılmalar yaşadığı bilinmektedir. Doğru sonucu elde etmeyi hedefleyen bu sistem 100'den fazla parametre kullanmaktadır. Diğer karar ağaçlarına göre yüksek performans sergilemesinin ve tercih edilen bir algoritma olmasının sebebi ise hem büyük verilerle çalışılabilir olması hem de parametre sayısına bağlı olarak elastiki bir algoritma olmasıdır (Ke G vd., 2007).

3.2.1 Yapay sinir ağları

Mcculloch ve pitts ile birlikte yapay zekâya adım atılmıştır. Bir fonksiyonun yapay sinir ağları ile birlikte hesaplanabileceğini öne sürmüşlerdir. Yapay sinir ağlarının birbiri ile bağlantılarını, ağırlıklarını değiştirmenin mümkün olacağı ilerleyen süreçlerde Hebb tarafından literatüre girmiş böylece yapay sinir ağlarının eğitilebileceği kanaatine varılmıştır. Bu eğitimden sonra teknik olarak yapay sinir ağları girdi verileri ile çıktı datalarının oluşmasını sağlamaktadır. Her bir sinir ağı birbiri ile ilişkili yapıda olup her bir bağlantı bir diğer nöronu etkilemektedir. Şekil 3.2'de bir yapay sinir ağının içerisinde ağırlıkların işlem süreci esnasında etkisi görülmektedir.



Şekil 3.2. Yapay Sinir Ağları Genel Yapısı (Kubat, 2015)

Yapay zekâ kavramı ortaya atıldığı yıllarda yapay sinir ağları ile ilişkilendirilmemiştir. 1950'lili yıllarda yapay zekâ kavramı yapa sinir ağlarından daha fazla araştırılan ve

üzerinde durulan bir alan olmuştur.1980'lere kadar yapay sinir ağları bir duraklama dönemi yaşamış daha sonrası ise bilgisayar teknolojisinin gelişmesiyle birlikte yapay sinir ağlarındaki ilerleme hız kazanmıştır. Bilgisayar kullanımına yapay zekâ yeni bir soluk getirerek, birden çok kaynağı kullanarak aynı anda işlemcide kullanılmasını, veri büyüklüğüne göre tercih edilen birden çok bilgisayar ile sistem kurulması ve adaptif yöntemlerin geliştirilmesine sebep olmuştur (Öztemel, 2006). Günümüzde hala birçok problemin çözümünde yapay sinir ağları kullanılmaktadır.

Bir sinir ağı milyonlarca nöronun bir araya gelmesinden meydana gelebilmektedir. İnsan beyni ve bilgisayar arasındaki farklar Tablo 3.1' de verilmiştir. İnsan beyninde 10^{10} adet nöron bulunmaktadır ve bu hücrelerin her biri $6 \cdot 10^{23}$ bağlantı içermektedir.

Tablo 3.1. İnsan Beyni İle Bilgisayarın Sayısal Olarak Karşılaştırılması (Crone, 2004)

	İNSAN BEYİNİ	BİLGİSAYAR
İŞLEMCİ HIZI	10^{-3} ms (0.25 MHz)	10^{-9} ms (0.25 MHz)
NÖRON / TRANSİSTÖR	10^{11} & 10^{33} Bağlantı	10^9 (Yonga)
AĞIRLIK	1500 g	Kilogramlarca
ENERJİ TÜKETİMİ	10^{-16} Joule	10^{-6} Joule
HESAPLAMA	100 Adım	Milyarlarca Adım

İnsan beyninden etkilenen bilim insanları yapay sinir ağlarını oluşturmuşlardır. Bu görüş McCulloch ve W.Pitts yapay sinir ağlarını üzerinde ilk çalışmalara başlamış bilim insanlarıdır. Nöronların yapısına benzetilerek oluşturulan bu sistem beynin özelliklerini taklit ederek onun gibi çalışan bir sistemdir. Öğrenme odaklı olan yapay sinir ağları doğrusal olmayan sistemlerde ya da sistem içerisindeki bilgilerin tam olmaması yahut eksik olması halinde yine de sistemi hedefe ulaştırmayı amaçlamaktadır. Yapay sinir ağlarının kullanım alanları oldukça fazladır görüntü ve ses tanıma, tahmin yapabilme, tıp, iletişim, trafik mühendisliği, endüstriyel alanlarda kullanılabilmektedir. (Kubat, 2015).

Yapay sinir ağları geleneksel yöntemlerden farklı ve daha gelişmiş yöntemler ile sonuçlar sunmaktadır. Geleneksel sistemlere kıyasla yapay sinir ağlarının çalışma sistematığına Tablo 3.2'de yer verilmiştir.

Tablo 3.2. Geleneksel Yöntemler ile Ysa Arasındaki Farklar (Mcneill ve Anderson, 1992)

KARAKTERİSTİK	GELENEKSEL HESAPLAMA UZMAN SİSTEMLER	YAPAY SİNİR AĞLARI
	Sıralı	Paralel
İşlem Stili		
Fonksiyonlar	Kurallar Kavramlar Ve Hesaplama Yoluyla, Mantıksal (Sol Beyin)	Resimler, Görüntüler, Kontroller Yoluyla Geşalt (Sağ Beyin)
Öğrenme Metodu	Kurallarla(Didaktik)	Örneklerle (Sokratik)
Uygulamalar	Muhasebe, Kelime İşlem, Matematik, Stok, Dijital İletişim	Sensör İşleme, Ses Tanıma, Desen Ve Desen Tanıma, Karakter Tanıma, Sınıflandırma

Öğrenme yetisi kazanan YSA'lar kendilerine tanıtılan verileri ezberleyerek, bilgiler arasında ilişkiler kurarak ve bu bilgileri tekrar hatırlayabilme gibi yetkinlikleri kazanır. Nasıl ki insan beyni içerisinde birbiri ile bağlantı kuran nöronlar var ise yapay zekâ insan beyninin bu özelliğini taklit edebilmektedir. Nöronlar kendilerine gelen sinyalleri bir sonraki nörona geçirmek için gerekli işlemlere tabii tutar. Her bir Nöron kendine gelen iletiye bazı tepkiler verir. Yapay sinir ağları ise bu nöron sistemine benzeyen bir sistemdir. Yapay nöron sisteminde ağırlıklar, toplama fonksiyonu, transfer fonksiyonu ve çıkış fonksiyonu bulunmaktadır (Şen, 2004).

Yapay sinir ağlarının temel işlevleri şu şekilde belirtilebilir:

- *Öngörü (Prediction) veya tahminleme:* İleriki satışlar, hava tahminleri, at yarışları, çevresel risk, ...
- *Sınıflandırma (Classification) ve Kümeleme (Clustering):* Müşteri profilleri, tıbbi teşhis, ses ve şekil tanıma, hücre tipleri ...
- *Kontrol (Control):* Erken uyarı için uçaklarda ses ve titreşim düzeyleri, (Uğur Ve Kınacı, 2006)

Yapay sinir ağları birden fazla basit elemanların bir araya geldiği paralel bağlantılı sistemlerdir. Ulaştırma mühendisliğinde sonra zamanlarda yapay sinir ağları

sistemlerinin kullanıldığı görülmektedir. Üst yapı tasarımı, ulaşım ağı planlaması, bakım ve onarımı, ulaşım ağı sistemlerinin işletilmesi gibi alanlarda kullanılmaktadır. Ulaşım mühendisliği ile alakalı birçok probleme çözüm üretebilme kabiliyetine sahip yapay sinir ağları artan trafik talebine karşılık verilmesi için yapılan çalışmalarda etkin trafik ağı ve yol tasarımı için başvurulan bir yöntemdir. (Kıyıldı, 2017).

Trafik kazasının oluşmasına neden olan birçok parametre söz konusudur. Bunların en temel şekilde karayolu, insan, taşıt ve çevresel koşullar olarak sınıflandırılabilir. Fakat kaza modellemesi kurulurken tüm bu parametrelerin sisteme tanıtılması ve uyarlanması mümkün olamamakla birlikte pek doğru sonuçlar elde edilmeyeceği açısından uygun değildir. Uygulamalarda elde edilen sonuçların güvenilir olması oldukça önemlidir bu açıdan modellemeler daha basit bir şekilde kullanılmalıdır. Yapılan modellerde nüfus ve araç sayıları akabinde bunların beraberinde getirdiği kaza sayısına bağlı olarak gelişen yaralı sayısı veri olarak kullanılabilir. (Kıyıldı, 2017).

3.2.1.1 Yapay sinir ağı bileşenleri

Yapay sinir ağları temel olarak iki bileşenden oluşur. Bunlar öğrenme algoritması diğeri ise bir nöronun aktif olup olmayacağına karar vermek için kullanılan aktivasyon fonksiyonudur. Bu bileşenler hakkında temel bilgiler aşağıda verilmiştir (Öztürk Ve Şahin, 2018).

Öğrenme Algoritması:

Verinin eğitime alınmasından, her bir verinin ağırlığının çıktı üzerindeki etkisinin belirlenmesi ve son olarak ele alınan verilerin hücrenin net girdisinin hesaplanmasına kadar geçen sürece öğrenme algoritması denilmektedir. Sonucun doğru elde edilebilmesi için verilerin doğrulukları kadar ağırlıklarının da doğru hesaplanması gerekmektedir. Yapay sinir ağı ne kadar doğru eğitilirse ağırlıklar aynı oranda doğruluk kazanır.

Aktivasyon Fonksiyonu:

Aktivasyon fonksiyonu sonuçların elde edildiği süreçtir. Belirlenen fonksiyonun doğru kurulması çıktıların doğruluk oranı açısından çok önemlidir. Çünkü aktivasyon süreci aşamasında girilen veriler işlenerek bu veriye karşılık gelen çıktı verilerinin elde

edildiği aşamadır. İyi bir aktivasyon fonksiyonun doğrusal, türevlenebilir ve alt sınırlı olmasıdır.

3.2.1.2. Yapay sinir kullanım alanları

Yapay sinir ağlarının uygulama alanlarına bakıldığında binlerce uygulamanın yapıldığı ve başarılı sonuçlar elde edildiği gözlenmektedir. Günlük yaşantımızın neredeyse her noktasında yapay sinir ağlarının kullanıldığını görmek mümkündür. Pek çok problemin çözümü noktasında aktif rol oynayan bu disiplinin genel hatları ile kullanıldığı alanları şu şekilde sıralayabiliriz.

Endüstriyel Uygulamalar:

Yapay Sinir Ağları'nın bir üretim sürecinin başından sonuna kadar tüm sürecin modellenmesi konusunda birçok uygulaması vardır. En geniş kullanılan alan olarak endüstriyel alan gösterilebilir. İmalattan güç istasyonuna kadar yapay sinir ağlarının varlığını görebiliriz.

Finansal Uygulamalar:

Banka kredilerinin değerlendirilmesi, makro ve mikro ekonomik tahminler, risk analizleri, Döviz kuru tahminleri gibi birçok finansal alanda yapay sinir ağları modelleri kullanılmaktadır.

Askeri Uygulamalar:

Yapay sinir ağları hedef tanıma, askeri uçakların uçuş yörüngesinin belirlenmesi, nesneleri ve görüntüleri ayırma gibi çeşitli konularda uygulanmıştır.

Sağlık Uygulamaları:

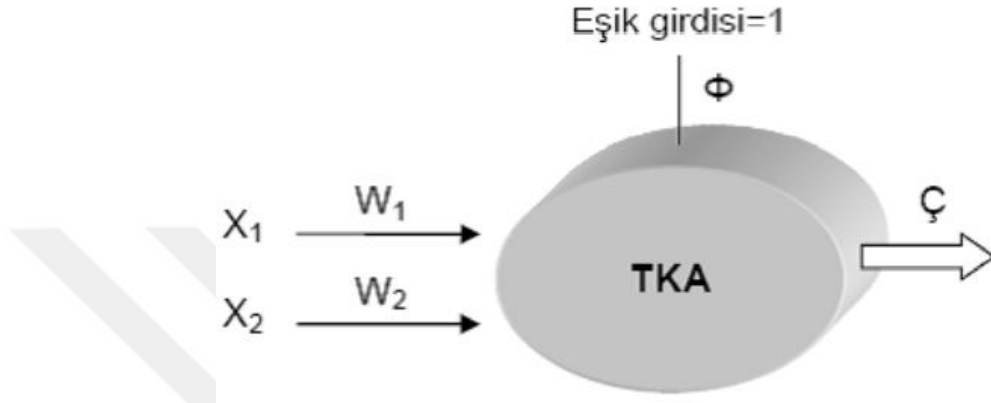
Hastalıkların teşhisi, kanser hücrelerinin belirlenmesi, tıbbi resim işleme gibi birçok sağlık alanında yine yapay sinir ağları kullanılmaktadır.

3.2.1.3. Yapay sinir ağları modelleri

Yapay sinir ağı modelleri Tek katmanlı algılayıcılar, İleri Beslemeli Yapay Sinir Ağları, Çok Katmanlı Algılayıcılar, Geri Beslemeli Yapay Sinir Ağları olmak üzere 4 ana başlıkta incelenmektedir.

Tek katmanlı algılayıcılar:

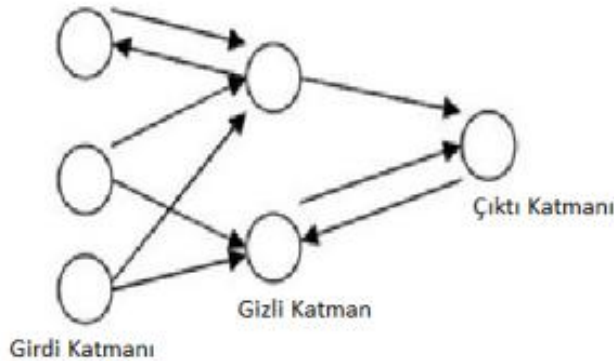
Tek katmanlı ağlar girdi ve çıktı katmanı olmak üzere 2 katmandan oluşur. Girdi katmanındaki tüm sinyaller çıkış katmanındaki tüm nöronlara bağlanmaktadır. Şekil 3.3’ te görüldüğü gibi doğrusal bir fonksiyon ile çalışan sistem 1 ya da -1 değerini almaktadır. Çıktı değerlerine göre sınıflandırma yapılmaktadır (Öztemel, 2006).



Şekil 3.3. Tek katmanlı algılayıcı modeli (Öztemel, 2006)

İleri Beslemeli Yapay Sinir Ağları:

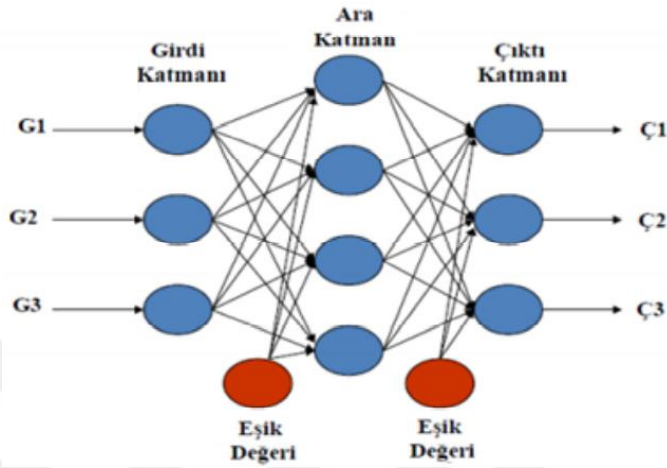
İleri beslemeli ağlar katmanlardan oluşmaktadır. Tek yönlü bir işleyiş gösteren bu yapılar, bir sonraki katmana aktarılırken ağırlıkları üzerinden giriş verilir. Giriş katmanındaki bilgi süreç boyunca bir değişikliğe uğramadan çıktı katmanına ulaşır. Şekil 3.4’ te görüldüğü gibi Bu ağ içerisinde gizli düğümler olsa bile veri bir sonraki katmana aktarılır. İleri beslemeli özelliğinden dolayı bilgi döngüye girmez ve süreç sonunda ağ çıkışı belirlenir (Rojas, 1996).



Şekil 3.4. İleri Besleme Ağ Yapısı

Cok Katmanlı Algılayıcılar:

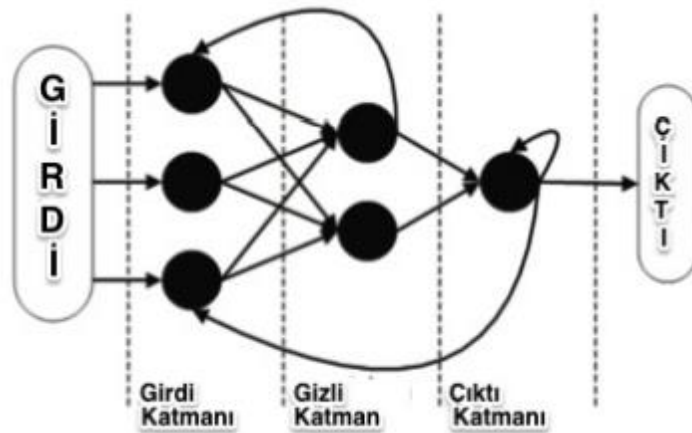
Bu model ileri beslemeli Yapay sinir ağları arasında en çok bilinen ve kullanılan yine de bir o kadar basit bir modeldir. Şekil 3.5’ te görüldüğü gibi çok katmanlı algılayıcılar girdi, ara ve çıktı katmanından oluşmaktadır. Tek katmanlı algılayıcılara kıyasla doğrusal olmayan olaylı tahmin edebilme yeteneğine sahiptir (Haykin, 2009).



Şekil 3.5. Çok Katmanlı Algılayıcı Modellemesi (Kabalıcı, 2014)

Geri Beslemeli Yapay Sinir Ağları:

Yapay sinir ağlarının ezberleme yeteneğinden doğan dinamik hafızaları bu sinir ağlarının temelini oluşturmaktadır. Şekil 3.6’da görüldüğü gibi Giriş, ara ve çıkış katmanlarından oluşan bu sistemde her bir katman bir önceki katmanın değerlerinden etkilenmektedir. Özellikle tahmin modellemelerinde kullanılan bu yöntem oldukça başarılıdır. (MacKay, 1992)



Şekil 3.6. Geri beslemeli yapay sinir ağı

3.2.1.4. Yapay sinir ağı eğitimi

Yapay sinir ağı modellemesi süresinde bilgiler girdi katmanından algoritmaya işlenir. Bilgi bu algoritma sürecinden çıktı katmanı olarak çıkar. Girdi ve çıktı katmanı arasında ki süreçte ara katmanlarda işlenir. Ara katmanlar da yapılan işlemler çıktı katmanının doğruluğunu belirler. Bu katmanda her bir girdinin ağırlığının doğru belirlenmesi gerekmektedir. Doğru ağırlığın belirlenmesi veri setinin doğru eğitilmesine bağlıdır. Yapay sinir ağlarında farklı algoritmalar kullanılarak Her bir girdi için farklı ağırlık belirlenerek en doğru sonucu ulaşıncaya kadar işlemler devam eder.

Öğrenme yeteneği, zekânın temel bir özelliğidir. Öğrenmenin kesin tanımını formüle etmek zor olsa da, YSA bağlamındaki bir öğrenme süreci, bir ağı belirli bir görevi verimli bir şekilde yerine getirebilmesi için ağ mimarisini ve bağlantı ağırlıklarını güncelleme sorunu olarak görülebilir. Ağ genellikle mevcut eğitim modellerinden bağlantı ağırlıklarını öğrenmelidir. Ağdaki ağırlıkları yinelemeli olarak güncelleyerek performans zaman içinde iyileştirilir. YSA'ların örneklerden otomatik olarak öğrenme yeteneği onları çekici ve heyecan verici kılmaktadır. YSA'lar, insan uzmanlar tarafından belirlenen bir dizi kuralı takip etmek yerine, verilen temsili örnekler koleksiyonundan temel kuralları (girdi-çıktı ilişkileri gibi) öğreniyor görünmektedir. Bu en önemlilerinden biri Yapay sinir ağlarının geleneksel uzman sistemlere göre avantajları. Bir öğrenme sürecini anlamak veya tasarlamak için önce bir sinir ağının çalıştığı ortamın bir modeline sahip olmalısınız, yani ağ için hangi bilgilerin mevcut olduğunu bilmelisiniz (Jain vd., 1996).

İstatistiksel metodolojilerin sınırlamaları ile Yapay Sinir ağları, Destek vektör Makineleri ve Derin öğrenme modelleri trafik güvenliği sorunları üzerinde uygulanmış moldeller olarak, büyük verilerin analizinde dahi modellemede esneklik tanınması ve doğru tahmin yeteneği le başa çıkmıştır (Dong vd. , 2018).

Ağın ağırlıkları belirlendikten sonra her bir ağırlığın ne anlama geldiği bilinmemektedir. O nedenle yapay sinir ağlarına “kara kutu” yakıştırması yapılmaktadır. Ağırlıkların tek tek ne anlama geldikleri bilinmemekle birlikte ağın girdiler hakkındaki kararını bu ağırlıkları kullanarak vermesi, ağın zekâsının bu ağırlıklarda saklandığı söylenebilir. Ağın bir olayı öğrenmesi o olay için en doğru yapay sinir ağı modelini seçmekle

mümkündür. Şu ana kadar birçok yapay sinir ağı modeli geliştirilmiştir. Bir yapay sinir ağının modelini şu bilgiler karakterize etmektedir (Öztemel, 2006).

Genelde yapay sinir ağı modellemelerinde verilerin %70'i eğitilirken %30'u verilerin doğru eğitilip eğitilmediği konusunda test verisi olarak saklanır.



4. MATERYAL METOT

Bu çalışmada Ankara ili içerisinde 2013-2020 yılları arasında gerçekleşen ve ölümlü yaralanmalı trafik kaza tutanaklarından elde edilen yaklaşık 86.694 kaza verileri incelenmiş ve analiz edilmiştir. Veri Seti Emniyet Genel Müdürlüğü'nden temin edilmiştir. EGM' den alınan kaza verilerine Tablo 4.1' de yer verilmiş olup kaza araç sürücü bilgisine Tablo 4.2'de yer verilmiştir.

Bu verilere ek olarak TÜİK ve KGM verilerinden yararlanılmıştır. Yıllık ortalama günlük trafik kayıtları KGM verilerinden elde edilirken, Ankara iline kayıtlı motorlu araç sayısının güncel verilerine TÜİK'ten ulaşılmıştır.

Tablo 4.1. Ölümlü Yaralanmalı Trafik Kaza Tutanaklarından Elde Edilen Veri Seti ¹

KAZA BİLGİSİ			
1	Kaza yılı	18	Yolun Yüzeyi
2	Kaza Saati	19	İlk Yardım
3	Kaza Koordinatları	20	Kaza Oluş Türü
4	Kaza Tarihi	21	Kaza Yol Kusuru
5	Kaza İlçesi	22	Çarpışma Yeri
6	Kaza Yerleşim Yeri/ Dışı	23	Araç Sayısına Göre Kaza
7	Yolun Tipi	24	Otokorkuluk
8	Yolun Kaplaması	25	Yaya Yolu
9	Yolun Sınıfı	26	Emniyet Şeriti Banket
10	Yolun Adı	27	Yol Şerit Çizgisi
11	Yatay Geometri	28	Trafik İşaret Levhası
12	Düşey Geometri	29	Trafik Levhası
13	Kavşak Geometrisi	30	Aydınlatma
14	Geçit Geometrisi	31	Trafik Görevlisi
15	Diğer Geometri	32	Görüşe Engel Cisim
16	Gün Durumu	33	Hasar Gören Unsur
17	Hava Durumu	34	Yolda Çalışma

Tablo 4.2. Ölümlü Yaralanmalı Trafik Kaza Tutanaklarından Elde Edilen Veri Seti ²

KAZA ARAÇ/ SÜRÜCÜ BİLGİSİ	
35	Araç Cinsi
36	Sürücü Kusuru
38	Sürücü Cinsiyeti
39	Kaza Sonucu Sürücü Durumu
40	Sürücü Yaşı
41	Sürücü Belgesi Var/ Yok
43	Sürücü Belgesi Sınıfı
44	Araç Kullanım Amacı
45	Sürücü Emniyet Kemer

Kullanılan parametreler makine öğrenmesi yöntemi olan yapay sinir ağları kullanılarak modellenmiştir. Python kütüphanesi kullanılarak, elde edilen veriler çeşitli yöntemler ile modellenmiş ve veri setinin %70 ile eğitim verisi, %30'u ise test verisi olarak kullanılmıştır. Kurulan makine öğrenmesi modellerinde yazılan kodlarda büyük bir kısmına Keşifsel Veri Analizi (Exploratory Data Analysis-EDA) yapılmıştır. Veri analizi 4 kategoriden oluşmakta olup bu kategorilere Tablo 4.3' de yer verilmiştir. Keşifsel Veri Analizi (EDA) veri seti içerisinde verileri temizlemeye yardımcı olan ve verileri anlamlandırmaya yönelik kritik bir süreçtir. Bu analiz kapsamında veri seti makine öğrenmesi modelinin kurulabileceği hale getirilmiştir.

Tablo 4.3. Veri Analizi Kategorileri (Kakde ve Agrawal, 2018)

1	Veri Temizleme
2	Keşifsel Veri Analizi
3	Doğrulayıcı Veri Analizi
4	Tahminleyici Veri Analizi

Makine öğrenmesi yöntemi olan ve algoritmaların eğitimi sağlayan Yapay sinir ağları ile modelleme oluşturulurken değişkenlerden bağımlı değişken olarak kaza sayısı, ölü sayısı ve yaralı sayısı, bağımsız değişken olarak ise nüfus, araç sayısı, yıllık ortalama

günlük trafik sayısı kullanılmıştır. YSA modellerinde bağımsız değişkenler girdi, bağımlı değişkenler ise çıktı olarak isimlendirilir.

Yapılan literatür araştırmalarında yapay sinir ağları modellemelerinin trafik kaza tahmin modellemesi oluşturmak için oldukça tercih edilen bir yöntem olduğu görülmektedir. Bu yöntemde çeşitli teknikler kullanılmaktadır. Öngörme teknikleri, nitel öngörme yöntemleri ve nicel öngörme yöntemleri olmak üzere iki şekilde sınıflandırılabilir. Her iki yöntemin çıkış noktası da ilgili değişkene ait gözlem değerleridir. Geçmiş ve şimdiki dönem gözlem değerlerinden, gelecek dönem gözlem değerleri belirli kurallar çerçevesinde öngörüler (Ataseven, 2013).

Bu çalışmada ile birlikte Ankara ilinde Trafik kazalarının bilançosunu azaltmak için kazalara sebep olan parametreler belirlenerek efektif çözüm yolları oluşturmak amacı ile sürücü kusuru, yol kusuru, yaya kusuru, gün durumu, çevresel faktörlerin etkisi ve araç kusurlarına göre yaralı sayısı ve kaza sayısı tahmin edilmiştir. Trafik kaza analizini kapsayan bu çalışma sınıflandırma problemi (classification problem) olarak değerlendirilmiştir. Makine Öğrenmesi (Machine Learning- ML) metotlarından; Lojistik regresyon (Logistic Regression), KNN (K En Yakın Komşu Sınıflandırması), Destek vektör makinesi (Support Vector Machine- SVM), Karar ağacı (Decision Tree- DT), Rastgele Orman (Random Forest-RF), Ekstrem Gradyan Artırma (XGBoost), Hafif Gradyan Artırma Makinesi (Lightgbm) algoritmaları kullanılarak tahmin modelleri kurulmuştur.

Modellemeler oluşturulurken ele alınan veriler de dengesizlik (unbalanced) durumu söz konusudur. Dengesizlik durumunda verilere Smote işlemi uygulanabilmektedir. Smote işlemi sırasında iki uygulama yapılmış olup bunlar; Oversampling (Aşırı Örnekleme) ve Under Sampling (Eksik Örnekleme) işlemleridir. Bu amaçla modelleme oluşturulurken veriler smote işleminden önce ve sonra olmak üzere çeşitli algoritmalar ile modellenerek karşılaştırılmıştır.

Yapılan çalışmada EGM Kayıtlarında tutulan Trafik kazası tespit tutanakları ile elde edilen 2013-2020 kaza veri setleri ile model oluşturulmuştur. Bu çalışma sınıflandırma problemi olarak değerlendirildi. Ele alınan veriler 2013 ve 2020 yılları arasında gerçekleşen ölümlü yaralanmalı trafik kaza tutanakları içerisinde alınan verilerin

%30'unun test çalışması (test veri), %70'inin ise modelleme çalışması test verisi (train data) için kullanılması uygun görülmüştür. PYHTON kütüphanelerinden kullanılan algoritmalar ise,

- Logistic Regression
- KNN
- SVM
- Decision Tree
- Random Forest
- XGBoost
- LightGBM

PYHTON kütüphaneleri ile oluşturulan algoritmalar da incelenen veriler açısından ölümlü kaza ile diğerlerinin oranları büyük bir fark olması sebebi ile dengesiz (unblance) için bu çalışmada smote işlemleri uygulanmıştır. Smote işlemleri kapsamında Smote Oversampling, Smote Overundersampling gibi metotlar uygulanmış ve tekrar bu metotlar uygulandıktan sonra yukarıda belirtilen algoritmalar bir kez daha test edilmiştir.

Makine öğrenmesi son yıllarda tahmin modellemeleri için oldukça sık kullanılmaktadır. Makine öğrenmesinin tercih edilen bir uygulama olmasının sebebi ise büyük verilerle çalışma konforunu sağlamış olmasıdır. Fakat büyük verilerinde kendi içinde dengesiz sınıflandırılması tahminleme modellerini daha karmaşık bir hale getirmekle birlikte sonuçların doğruluğunu etkilemektedir. Bu bağlamda verilere yeni bir yaklaşım kazandırılma ve doğru sonuçlar elde etmek için verileri kendi içerisinde sağlıklı sınıflandırmak gerekmektedir. Bu çalışmada ele alınan verilerin çok büyük olması hasebi ile verileri serilerini dengelemek ve ilk modellemelerden daha doğru sonuçlar elde etmek amacı ile oluşturulan ikinci modellemelerde kullanılan veri setine Sentetik Azınlık Örneklem Tekniği (SMOTE) yöntemler uygulanarak algoritmalar eğitilmiştir.

Smote Over Sampling sentetik veri üretilmesini sağlayan bir aşırı örneklem sürecidir, Smote overunder sampling ise temel olarak baskın olan verilerin düşük olana yakın veya eşit bir seviyeye düşürülmesidir. Smote algoritmasında azınlık sınıfına ait her örnek, Öklid mesafesini kullanarak K yakın komşu bulur ve bulduğu örneklerden

rastgele seçim yapar. Örnek ile K-En Yakın Komşu arasındaki fark alınır, 0-1 arasında rastgele bir sayı (α) seçilir ve bulunan fark ile çarpılır. (Dal vd.,2021)

Modellemede bahsi geçen algoritmaların doğruluk oranlarını elde etmek için kullanılan metrikler hakkında yapılan literatür araştırmalarında aşağıdaki bilgilere ulaşılmıştır;

Her bir hata metriği için kullandığımız algoritmaların ortak bir platformda performansını karşılaştırdığımız grafikler aşağıda verilmiştir. Olasılık modelini kullanan tüm sistemler, sisteme ait örnekleri dört ayrı başlık altında toplamaktadır. Bunlar Tablo 4.4’ te görüldüğü gibi Doğru Pozitif (DP), Yanlış Pozitif (YP), Doğru Negatif (DN) ve Yanlış Negatif (YN) değerleridir. Bu değerler, yapay zekâya tanıtılan veriler ile edilen tahmin sonucunun uyumlu olup olmadığını ölçen metriklerdir. Modellemeye tanıtılan verinin pozitif olduğu kabul edilirse ve elde edilen tahmin sonucunda değer pozitif ise DP, negatif ise YN değeri oluşmaktadır. Eğer modellemeye tanıtılan verinin negatif olduğu kabul edilirse, tahmin sonucu eğer pozitif ise YP, negatif ise YN değeri oluşmaktadır (Goute ve Gaussier , 2005).

Tablo 4. 4. Karmaşıklık Matrisi

Tahminlenen	
Gerçekleşen	True Positive (Doğru pozitif)
	False Positive (Yanlış Pozitif)
False Negative (Yanlış Negatif)	True Negative (Doğru Negatif)

Kesinlik (Precision): Doğru pozitif sonuçların, model üzerinde tahminleme yapıldığında elde edilen tüm pozitif sonuçlara oranı ile elde edilen Eş. 4.1’de görülmektedir.

$$P = \frac{DP}{DP+YP} \quad (4.1)$$

Duyarlılık (Recall): Doğru pozitif sonuçların, tahminleme modelinde elde edilen doğru pozitif ve yanlış negatif sonuçlarına oranı ile elde edilir. Gerçek durumu pozitif olarak kabul edilen verilerin ne kadarının pozitif olduğunu anlamak için kullanılan bir metriktir. Formül Eş. 4.2’ de görülmektedir.

$$R = \frac{DP}{DP+YN} \quad (4.2)$$

F1-Skoru: Yukarıda ele alınan hata metriklerinin harmonik ortalaması F-1 skorunu verir. Formül Eş. 4.3’te görülmektedir.

$$F1 \text{ Skoru} = 2 * \frac{P*R}{P+R} \quad (4.3)$$

Doğruluk (Accuracy): Doğru tahminlerin tüm tahminlere oranı olarak hesaplanan değer genellikle 0 ile 1 arasındaki değer 1’e yaklaşması doğruluk oranını arttırmaktadır. Formül Eş 4.4’ te verilmiştir.

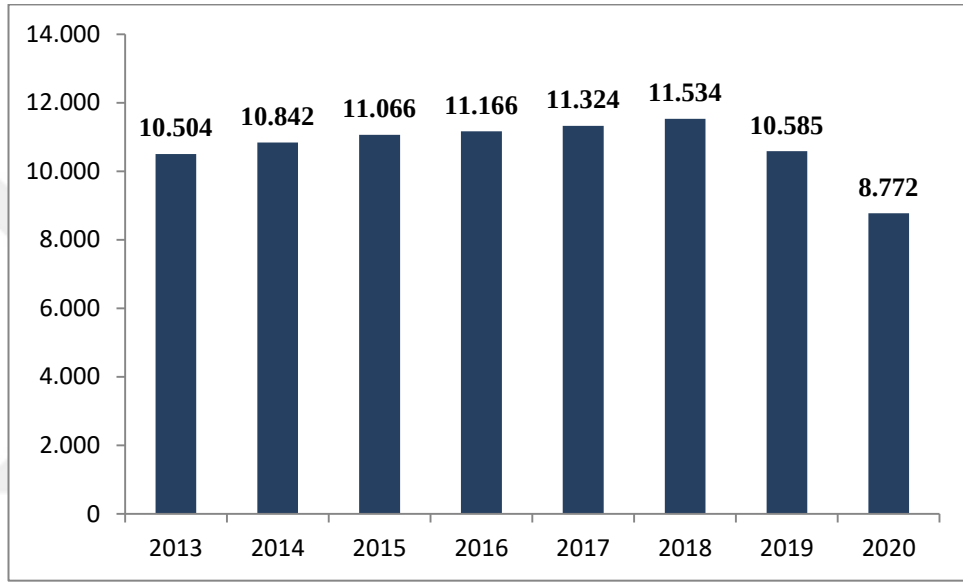
$$\text{Doğruluk oranı} = \frac{DP + DN}{DP + DN + YP + YN} \quad (4.4)$$

Elimizde birkaç modelin kesinlik ve doğruluk değeri mevcutsa en iyi modeli seçmek için duyarlılık metriği değerini esas alarak oluşturulan modelleri değerlendirmeye tabi tutarız.

5. ARAŞTIRMA BULGULARI

5.1. Makine Öğrenmesi Model Değerlendirme Sonuçları

Şekil 5.1' den Şekil 5.44' e kadar verilen şekil listeleri makine öğrenmesi modellemesi haricinde mevcuttaki verilerdir. Söz konusu verilerin grafik haline getirilmiş ve bu grafikler üzerinden değerlendirmeler yapılmış olup söz konusu şekiller açıklayıcı niteliktedir.

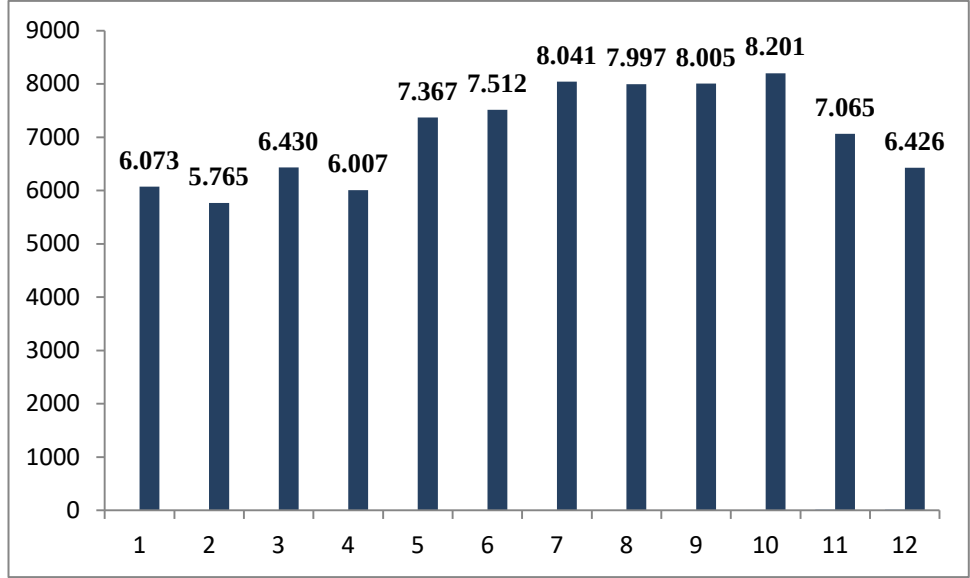


Şekil 5.1. Yıllık Ölümlü Yaralanmalı Kaza Sayısı

Şekil 5.1' e göre Yıllık ölümlü yaralanmalı kaza sayıları incelendiğinde 2020 COVID Salgını ile uzaktan eğitim, evden çalışma gibi esneklikler trafikteki araç sayısını azaltıcı yönde etki göstermiştir. Trafikteki aktif araç sayısının azalması ile kaza sayılarında gözle görülür bir düşüş görülmektedir.

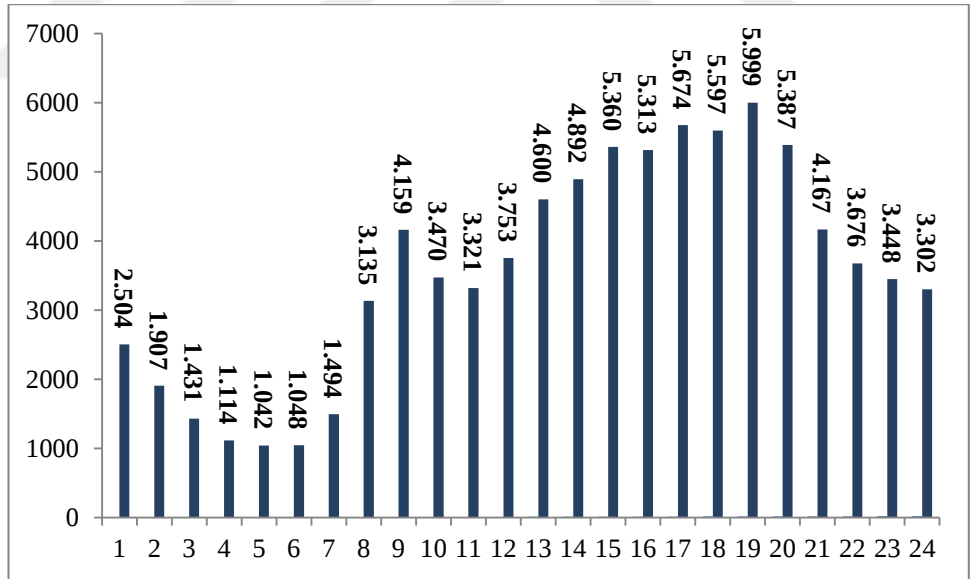
Şekil 5.2' ye göre Ankara ili içerisinde yaz aylarında artan tatil seyahatleri sebebi ile kaza sayılarının arttığı görülmüştür.

Şekil 5.3'e göre 2013-2020 yılları arasında saat bazında yapılan kaza sayıları analizinde mesai gidiş- dönüş saatleri arasında kaza sayılarının fazla olduğu görülmektedir.



Şekil 5. 2. Aylık Ölümlü Yaralanmalı Kaza Sayısı

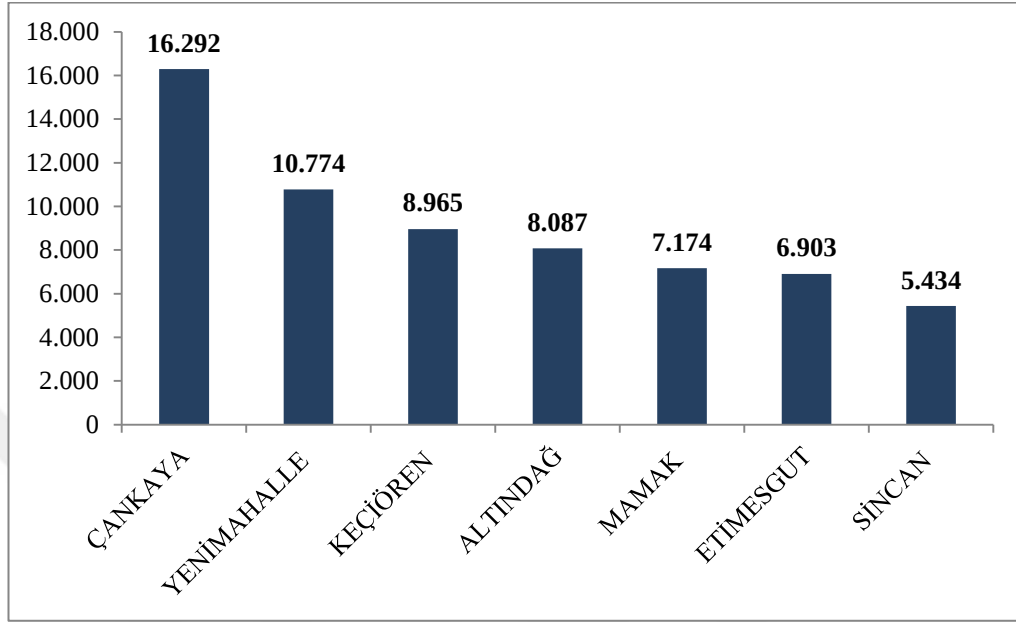
Şekil 5.3'e göre 2013-2020 yılları arasında saat bazında yapılan kaza sayıları analizinde mesai gidiş- dönüş saatleri arasında kaza sayılarının fazla olduğu görülmektedir.



Şekil 5.3. Saatlik Ölümlü Yaralanmalı Kaza Sayısı

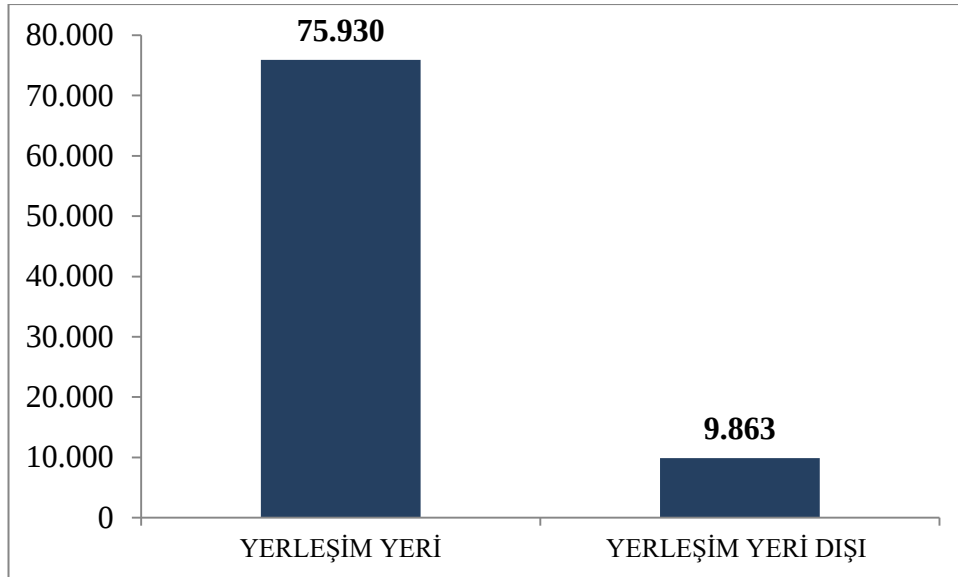
Şekil 5.4'e göre Ankara ili üzerinde yapılan analizde kaza sayılarının en çok Yenimahalle ve Çankaya ilçelerinde gerçekleştiği görülmektedir. Kurum ve kuruluşların bu ilçelerde yoğunluk göstermesi ve iki ilçenin yaşayan bireyler için cazibe merkezi

olması, nüfus artışına dolayısıyla trafik yoğunluğundaki artışa sebep olmaktadır. Bu ilçelerdeki kaza sayının diğer ilçelere kıyasla fazla olmasını bu yukarıda belirtilen sebeplere bağlamamız mümkündür.



Şekil 5.4. İlçe Bazında Gerçekleşen Ölümlü Yaralanmalı Kaza Sayısı

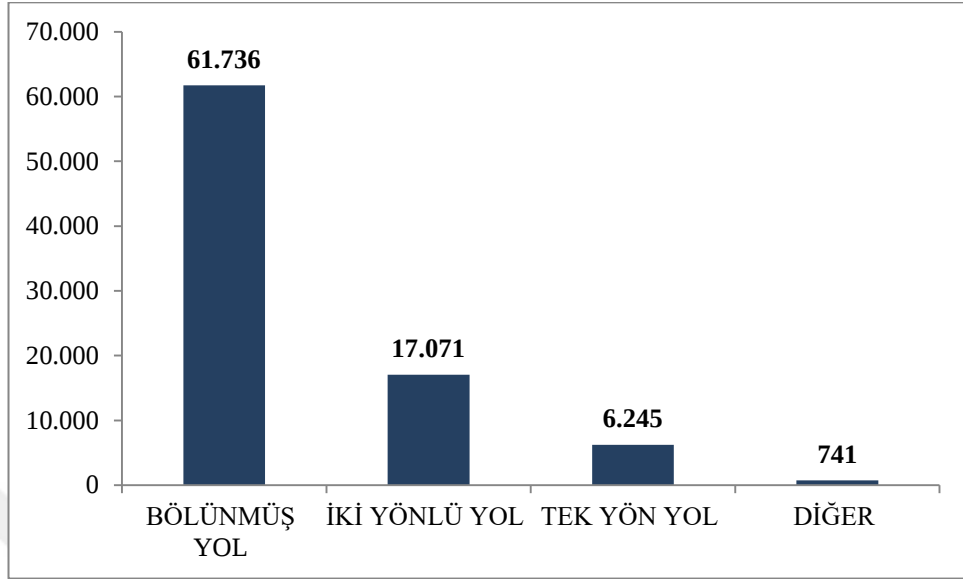
Şekil 5.5’ te görüldüğü gibi yerleşim yeri içinde trafik hacminin fazla olması sebebi ile yerleşim yeri içinde kaza sayılarının daha fazla olduğu görülmektedir.



Şekil 5.5. Kazanın Gerçekleştiği Bölgenin Çevresel Özellikleri

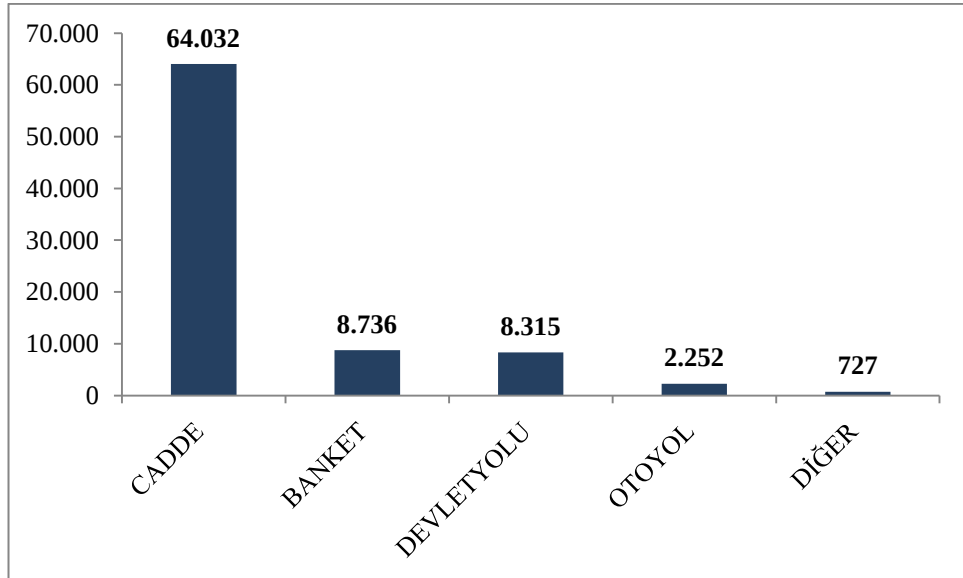
Ankara ili trafik ağı içerisinde ağırlıklı olarak Bölünmüş Yol şeklinde tasarlanmış olup

söz konusu özelliği taşıyan noktalarda trafik hacmi yüksek olduğu görülmektedir. Şekil 5.6’da görüldüğü gibi kaza sayılarının bu yollarda daha fazla olduğu gözlenmiştir.



Şekil 5.6. Kazanın Gerçekleştiği Yolun Tipi

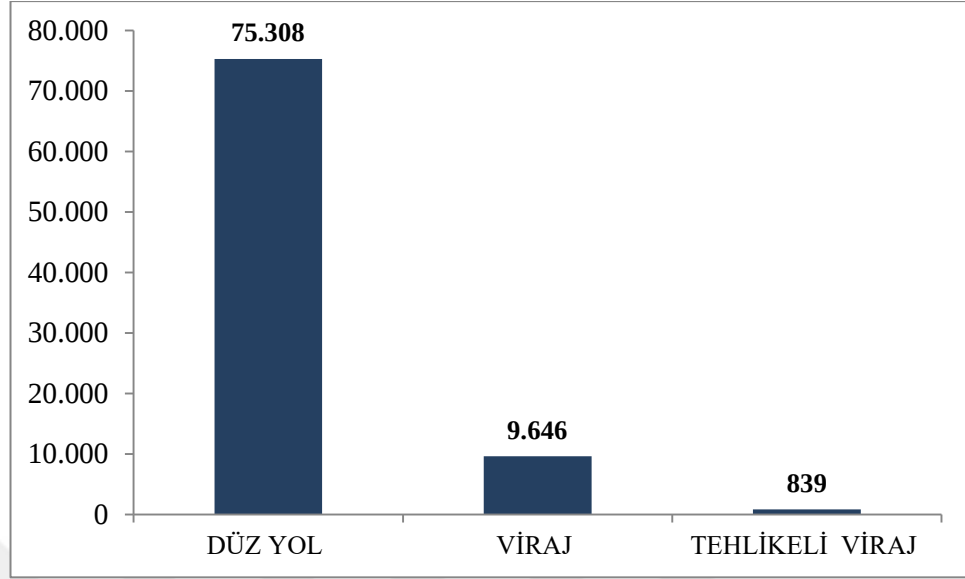
Yolun sınıfı değerlendirmesinde bazı sınıflar çok küçük olduğu için çıkartılmıştır. Şekil 5.7’ de görüldüğü gibi kazaların yüksek oranda caddelerde gerçekleşmiştir.



Şekil 5.7. Kazanın Gerçekleştiği Yolun Sınıfı

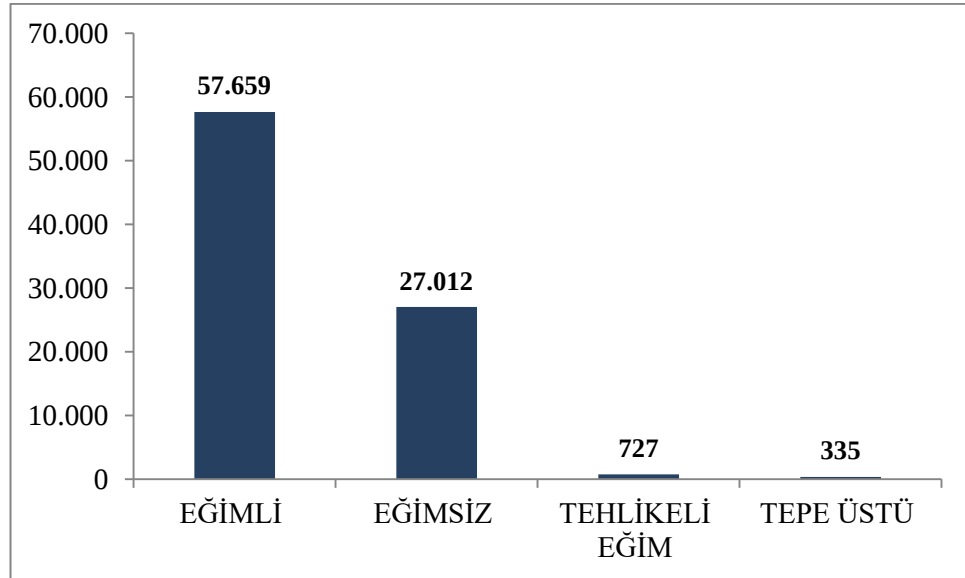
Ayrıca Ankara ili içerisinde düz yolda yoğun trafik etkisi ile kazaların ağırlıklı olarak

düz yolda gerçekleştiği Şekil 5.8’ de görülmektedir.



Şekil 5.8. Kazanın Gerçekleştiği Yolun Yatay Geometri Özellikleri

Şekil 5.9’ da kazanın gerçekleştiği düşey geometri özellikleri incelendiğinde Ankara da eğimli yolların fazla olması sebebi ile kaza sayılarının eğimli yollarda diğer düşey geometri özelliklerine kıyasla daha fazla olduğu gözlemlenmiştir.

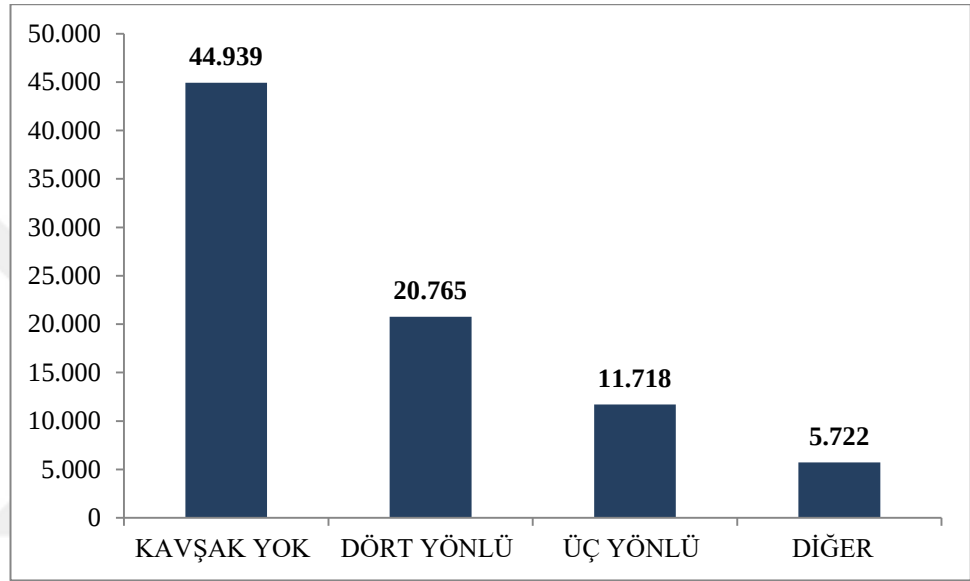


Şekil 5.9. Kazanın Gerçekleştiği Yolun Düşey Geometri Özellikleri

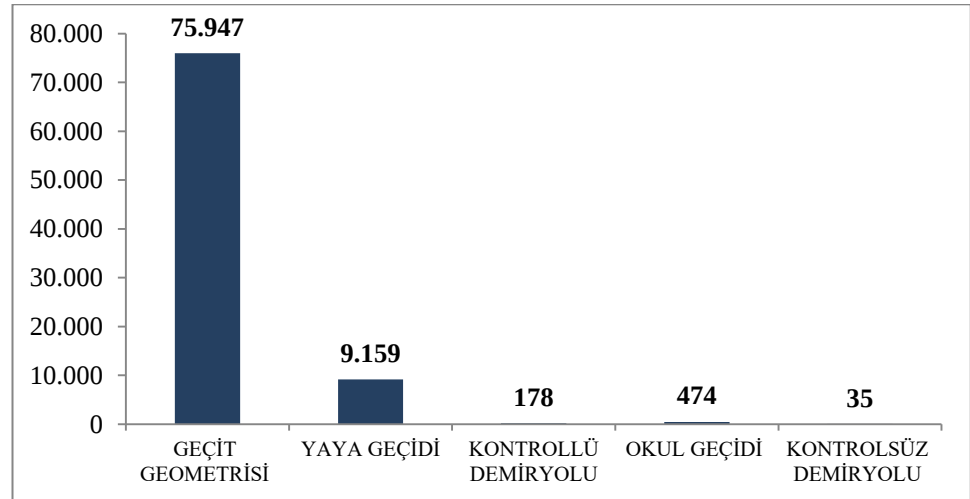
Kavşaklardaki kaza durumunu gösteren Şekil 5.10’da Ankara da ki kavşak arası

mesafelerin uzun olmasından dolayı kavşak noktalara kıyasla kavşak olmayan noktalarda kaza sayısının fazla olduğu gözlemlenmiştir. Kavşak tiplerinde ise en çok kaza 4 yönlü kavşak türünde gözlemlenmiştir.

Şehirlerarası yolların geçiş noktaları Ankara'nın tam merkezine denk gelmekte olup bu yollarda geçitler bulunmaktadır. Şekil 5.11'de görüldüğü gibi kazaların ağırlıklı olarak geçit geometrisinde gerçekleştiği gözlenmektedir.

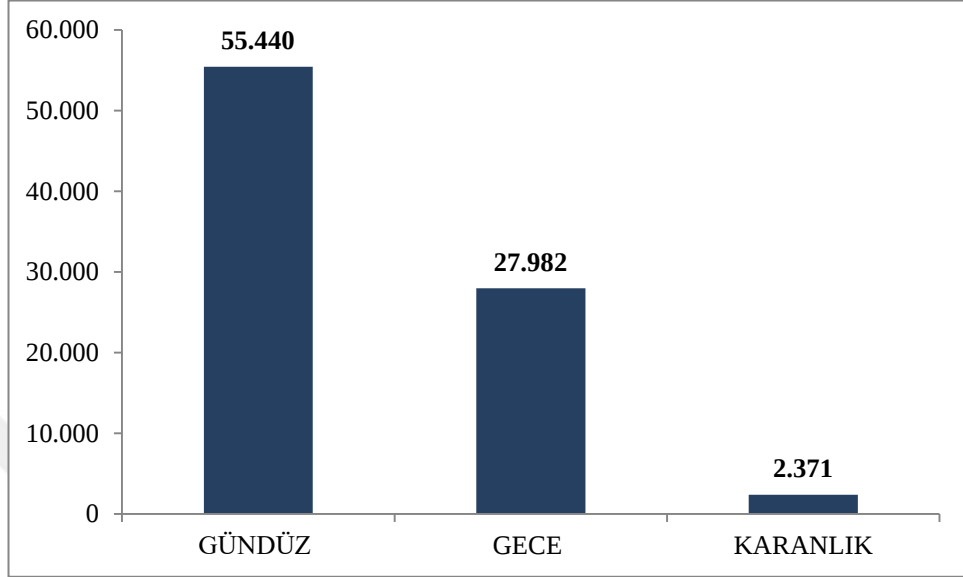


Şekil 5.10. Kazanın Gerçekleştiği Yolun Kavşak Geometrisi



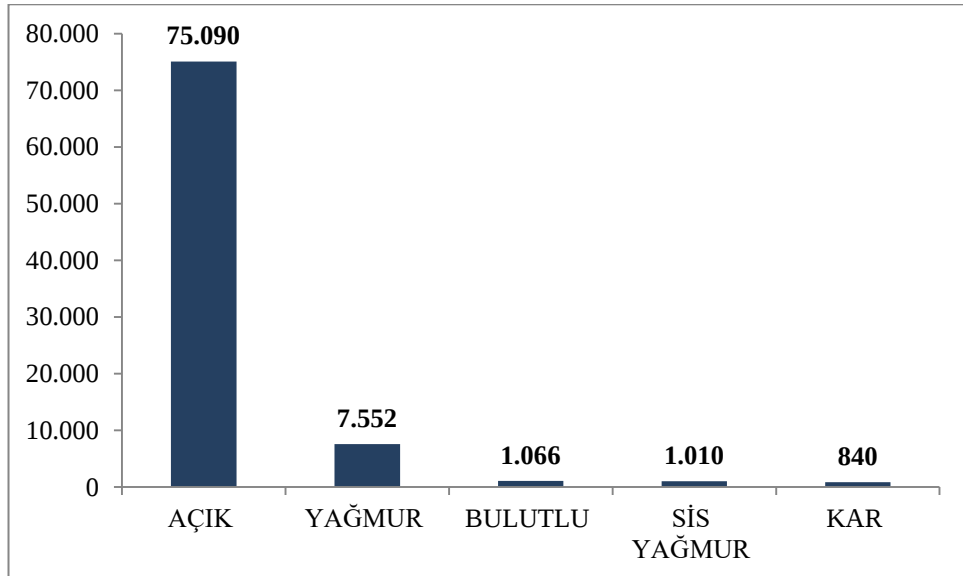
Şekil 5.11. Kazanın Gerçekleştiği Geçit Geometrisi

Trafik kazalarının ağırlıklı olarak şehir içerisinde gerçekleşmektedir. Gerçekleşen trafik kazalarının trafik yoğunluğunun en yoğun olduğu gündüz saatlerinde olduğu gözlenmiş olup Şekil 5.12’de görülmektedir.



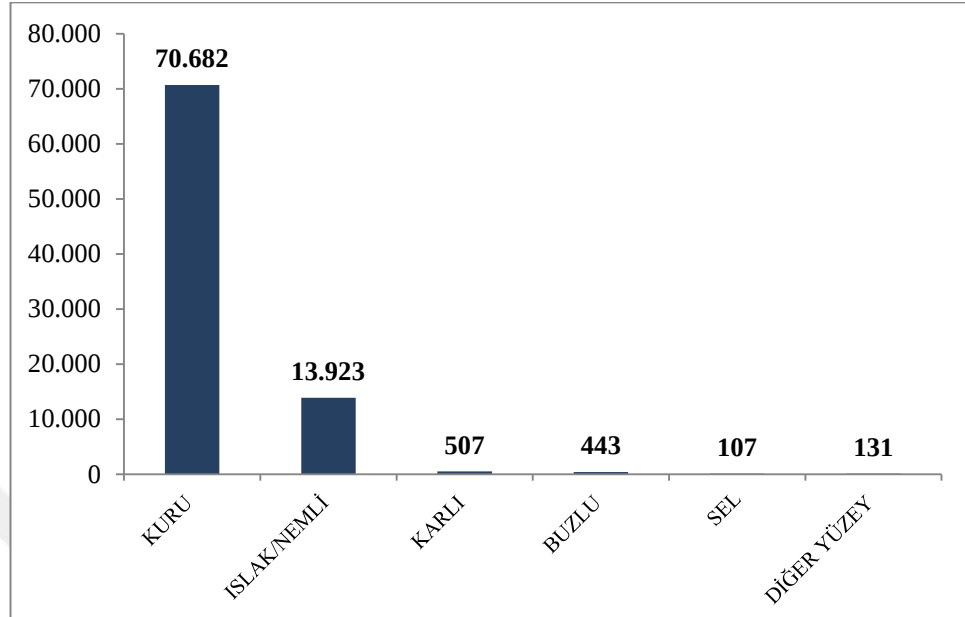
Şekil 5.12. Gün Durumuna Göre Kaza Gerçekleşme Sayıları

Ankara ili genel hava durumu ele alınırsa trafik yoğunluğunun en çok açık havada olduğu bu açıdan meydana gelen kazaların Şekil 5.13 ‘de hava açıkken olduğu görülmektedir. Yolum yüzeyinde mevsimsel geçişlerin etkisi gözlemleniyor.



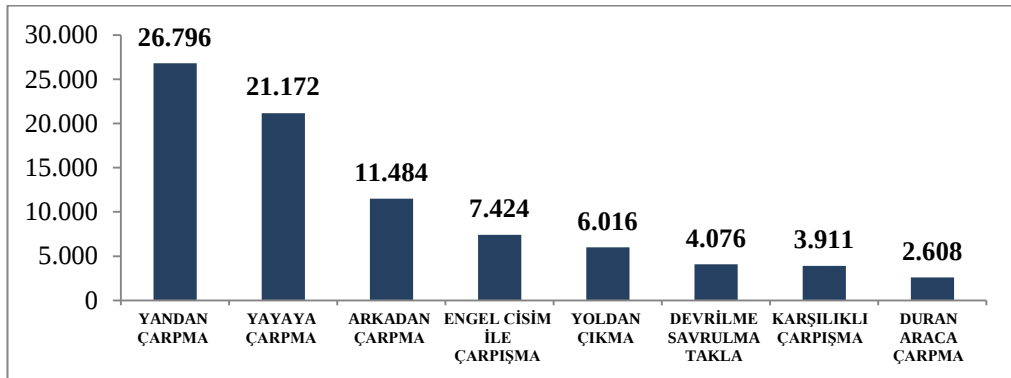
Şekil 5.13. Hava Durumuna Bağlı Olarak Kaza Sayıları

Ankaradaki yağış oranının azlığı yolun yüzeyine yansımış olup gerçekleşen kazaların ağırlıklı olarak kuru yüzeyde gerçekleştiği Şekil 5.14'te görülmektedir.



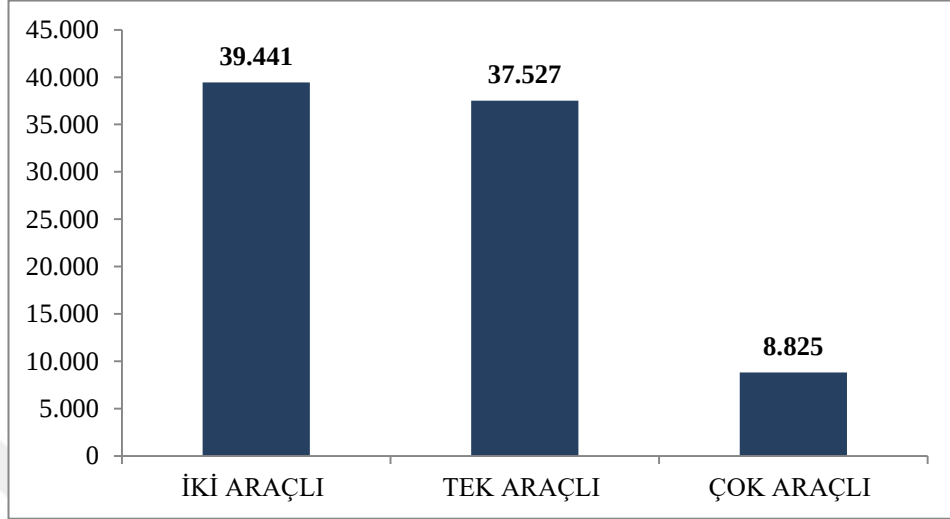
Şekil 5.14. Kazanın Gerçekleştiği Yol Yüzey Durumu

Şekil 5.15'te görüldüğü gibi kazalar çoğunlukla yandan çarpma şeklinde gerçekleşmiştir. Yandan çarpma şeklindeki kazalar Sollama hataları yüzünden meydana gelmektedir. Bu kaza çeşidini yayaya çarpma kaza tipi takip etmiştir. Yayaya çarpma kaza türünde birçok faktör ile değerlendirmede bulunabiliriz. Sürücü ve yayanın dikkatsizliği, yolun durumu gibi faktörler ile değerlendirme yapılabilir.

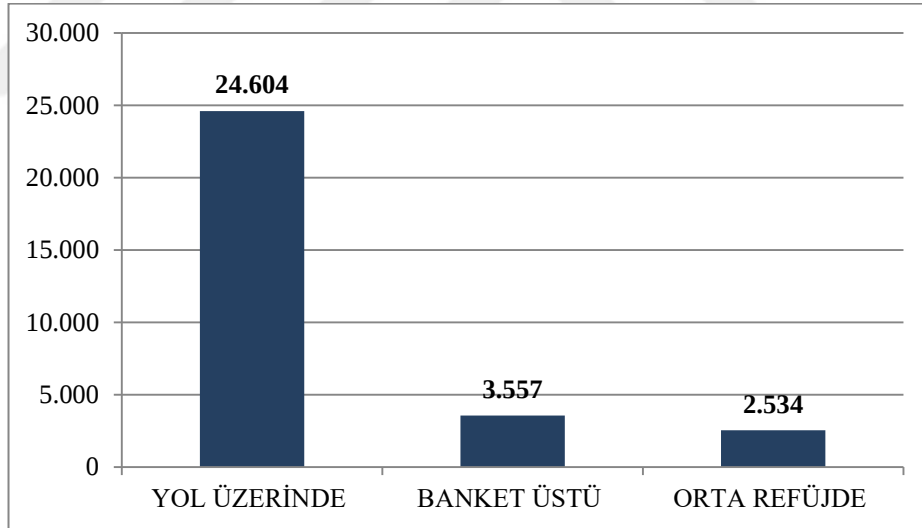


Şekil 5.15. Kazanın Gerçekleşme Tipi

Kazaların gerçekleşme sayılarına bakıldığında Şekil 5.16'ya bakıldığında İki Araçlı Trafik Kazalarının Çoğunlukta Olduğu Görülmektedir. Şekil 5.17'te ise kazaların en çok yol üzerinde gerçekleştiği görülmektedir.

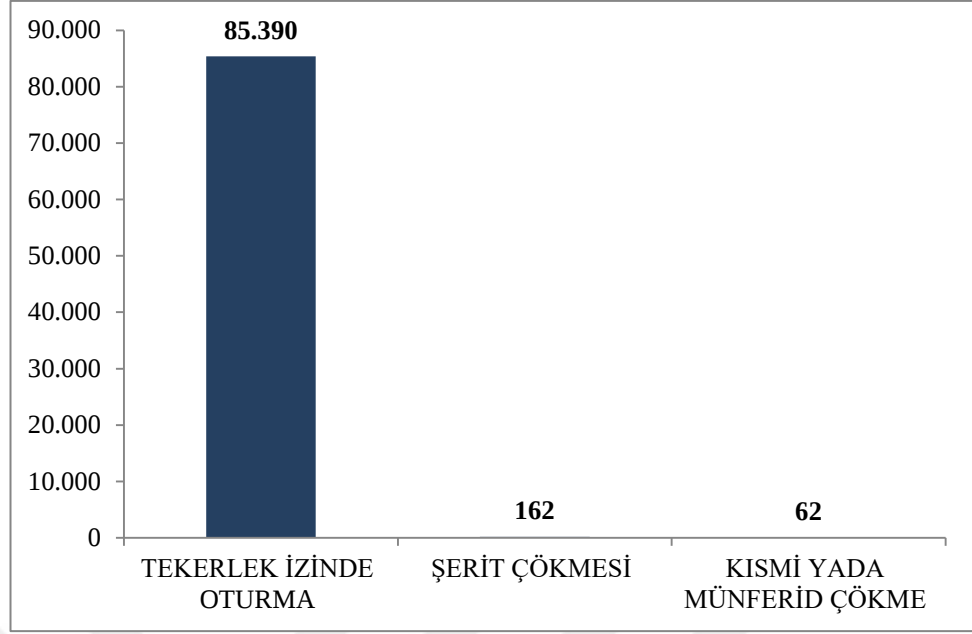


Şekil 5.16. Araç Sayılarına Kazanın Gerçekleşme

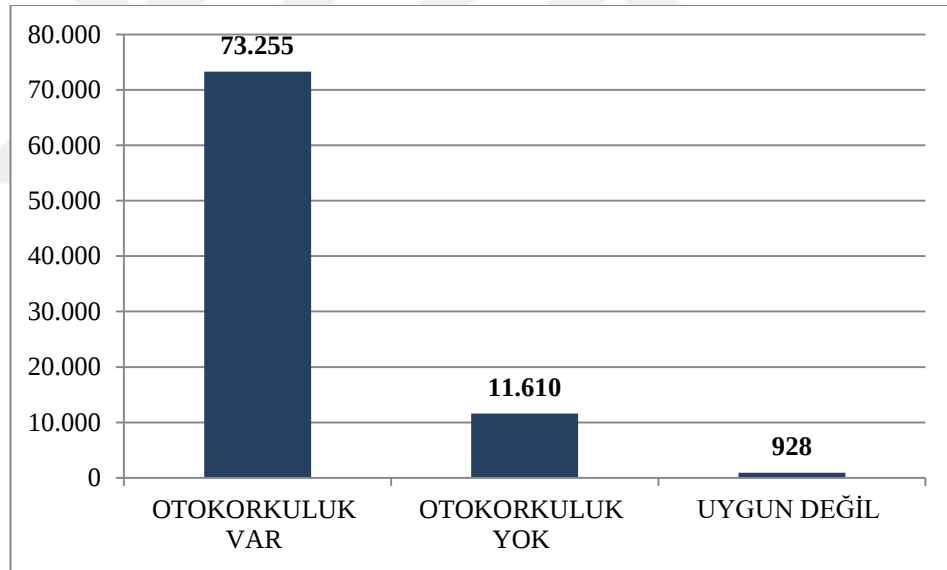


Şekil 5.17. Kazaların Gerçekleştiği Çarpışma Yeri

Analizin hemen hemen tamamında kazaya etkiyen yol kusurunun Tekerlek izinde oturma olduğu Şekil 5.18'de görülmektedir. Şehir içi yollarda korkuluk bulunmamaktadır ve kazalar çoğunlukla yerleşim yeri içerisinde görülmektedir. Bundan dolayı Şekil 5.19' da görüldüğü gibi Oto korkuluğun olmadığı yerlerde genel olarak kazalar daha fazla görülmektedir.



Şekil 5.18. Kazaya Etkiyen Yol Kusuru

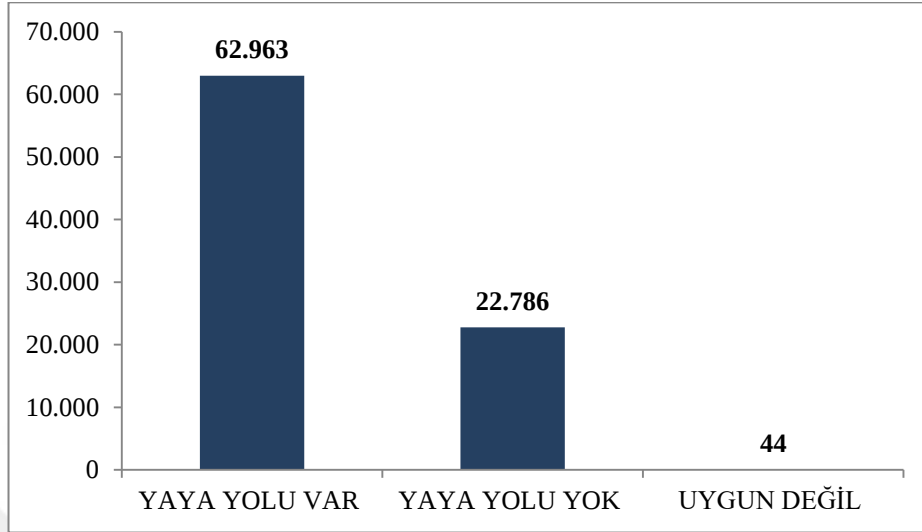


Şekil 5.19. Kazanın Gerçekleştiği Noktada Otokorkuluk Grafiği

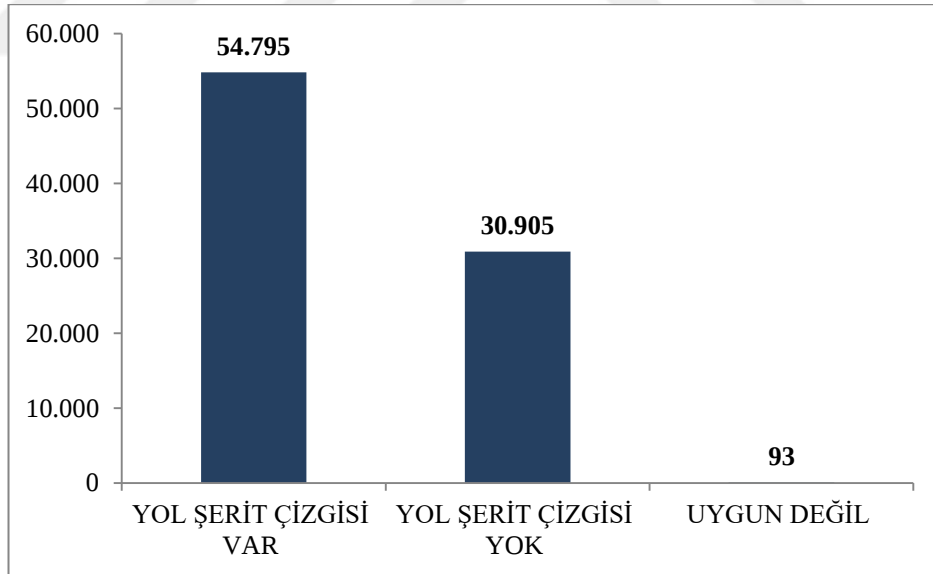
Yaya yolunun olduğu yerlerin kaza noktaları için önemli olduğu, kaza sayılarına bakıldığında gerçekleşen kazaların çoğunluğunun yaya yolu olan yerlerde meydana geldiği Şekil 5.20’de görülmektedir.

Yol şerit çizgisini olduğu yerlerde trafik yoğunluğunun fazla olmasından kaynaklı olarak kaza sayısının yüksek olduğu Şekil 5.21’de görülmektedir. Fakat şerit çizgisinin

olmadığı noktalarda da kazaların fazla olduğu görülmektedir. Dolayısı ile bu şekildeki yolların yeniden ele alınması gerekmektedir.

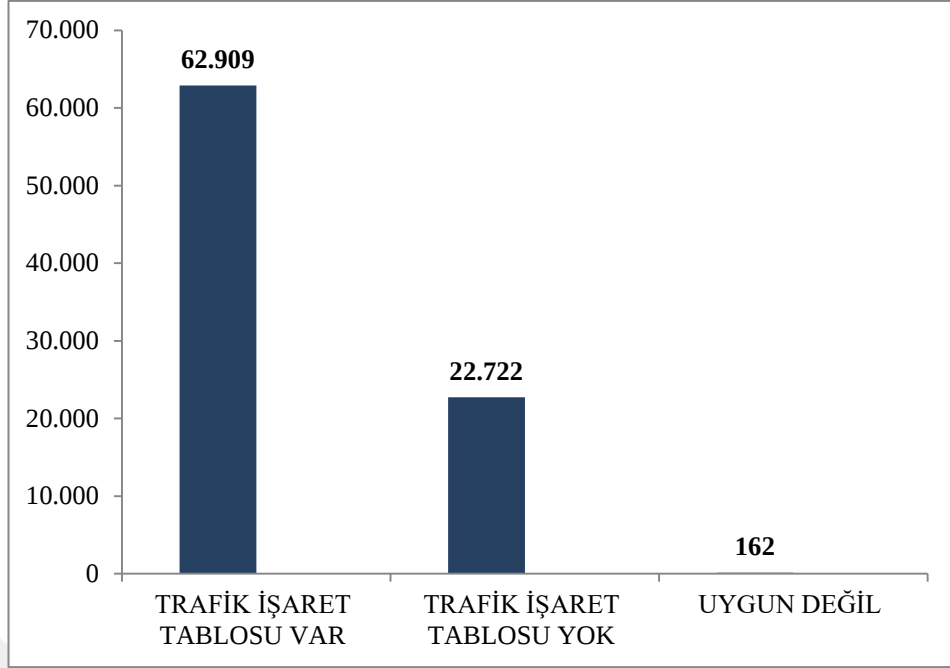


Şekil 5.20. Kazanın Gerçekleştiği Noktada Yaya Yolu Grafiği



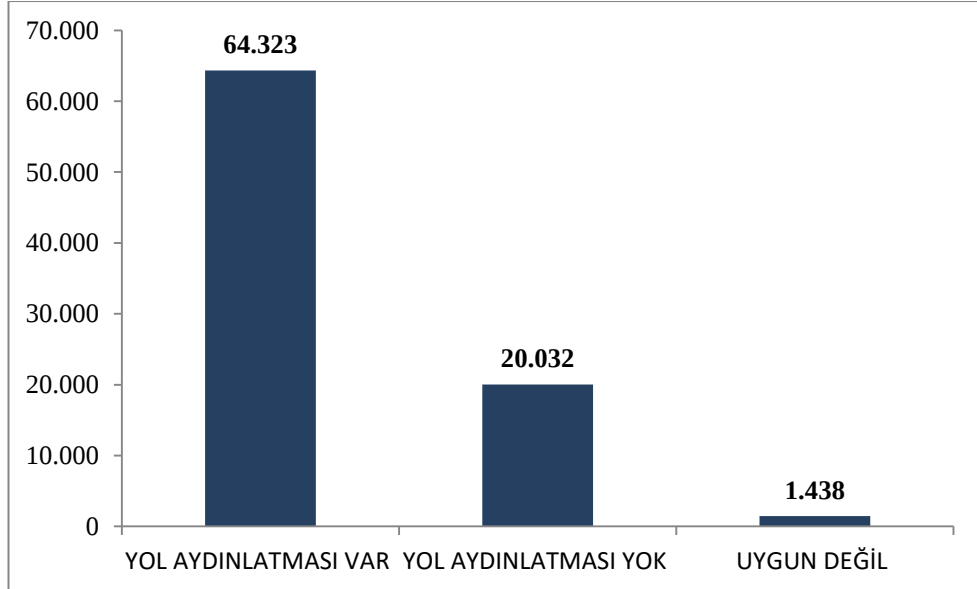
Şekil 5. 21. Kazanın Gerçekleştiği Noktada Yol Şerit Çizgisi

Şekil 5.22 'ye göre Trafik işaretlerinin olmadığı noktalar da kaza sayılarının fazla olması sebebi ile bu noktalara müdahale edilmesi gerektiği görülmektedir.



Şekil 5.22. Kazanın Gerçekleştiği Noktada Trafik İşareti Tablosu

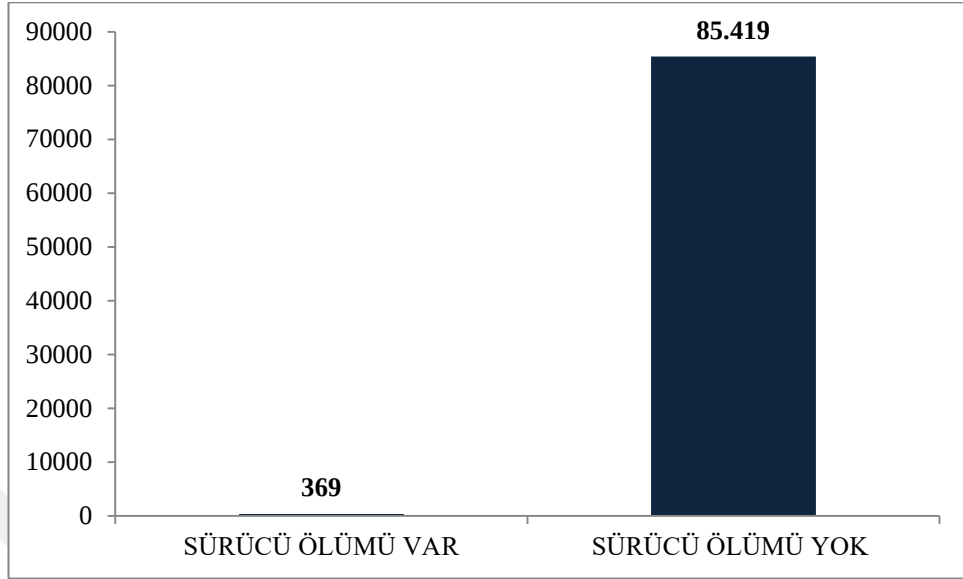
Şehir merkezindeki birçok noktada aydınlatma mevcuttur bu nedenle Şekil 5.23’ de görüldüğü gibi kazaların şehir merkezinde fazla olması ile kazaların bu noktalarda gerçekleşmesi ile ilişkilendirilmektedir.



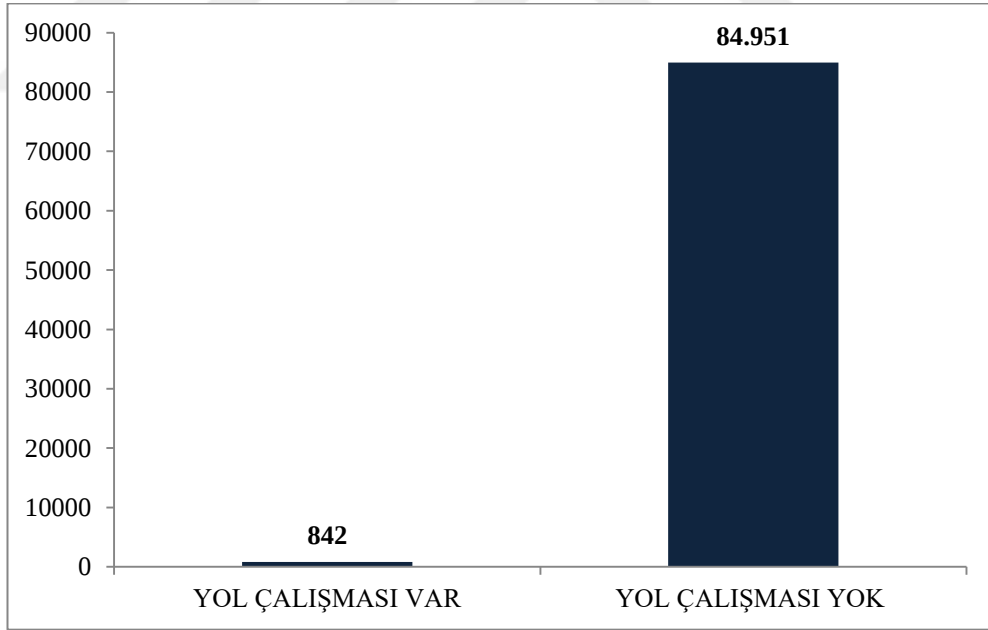
Şekil 5. 23. Kazanın Gerçekleştiği Noktada Yol Aydınlatma Grafiği

Şehir merkezindeki birçok noktada aydınlatma mevcuttur bu nedenle Şekil 5.23’ de

görüldüğü gibi kazaların şehir merkezinde fazla olması ile kazaların bu noktalarda gerçekleşmesi ile ilişkilendirilmektedir

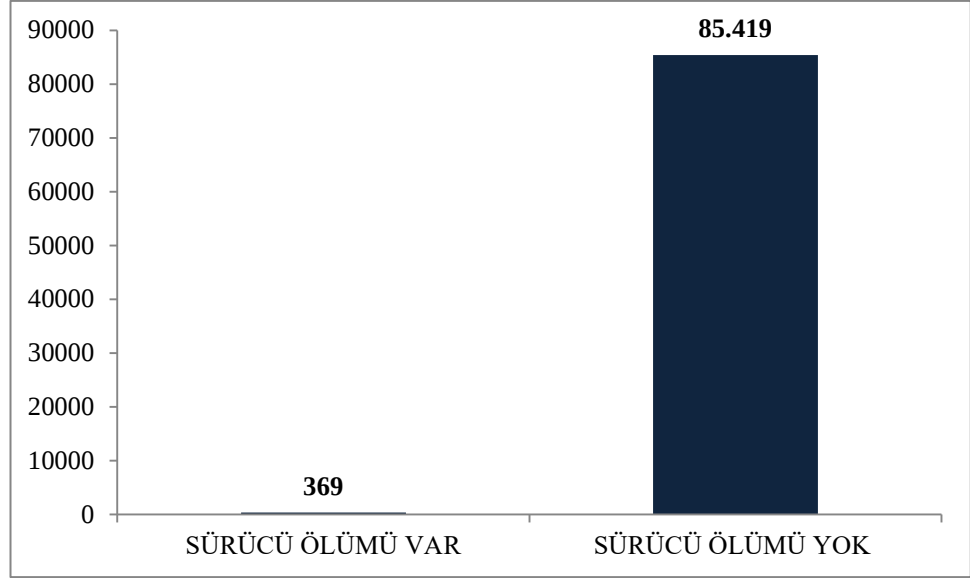


Şekil 5.24. Kazanın Gerçekleştiği Noktada Hasar Gören Unsur



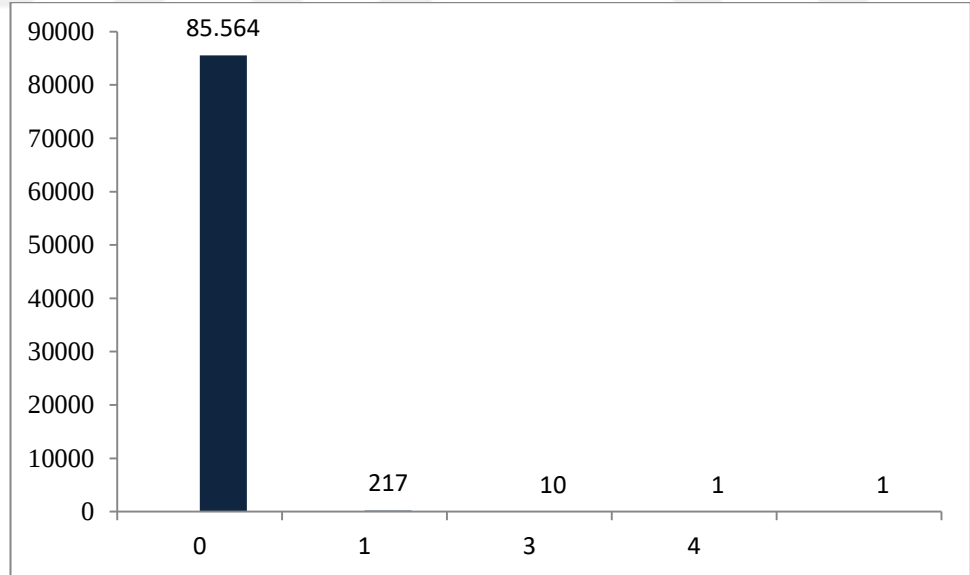
Şekil 5.25. Kazanın Gerçekleştiği Noktada Yol Çalışması Durumu

İncelediğimiz veri setinde 369 kazada sürücü hayatını kaybettiği Şekil 5.26'da görülmektedir.



Şekil 5.26. Gerçekleşen Kaza Sonrası Sürücü Durumu

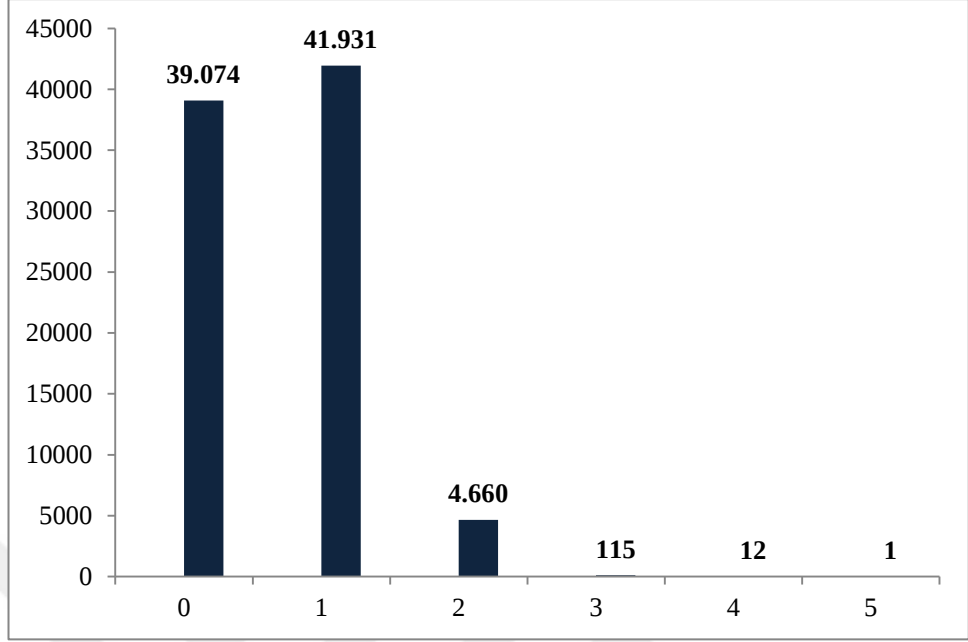
Şekil 5.27'ye göre 85.564 Kazada 1 yayanın, 217 kazada 1 yayanın, 10 kazada ise 3 yaya ölümü ile sonuçlanmıştır. Gerçekleşen kazalar incelendiğinde ağırlıklı olarak yaya ölümlerinin gerçekleşmediği görülmektedir.



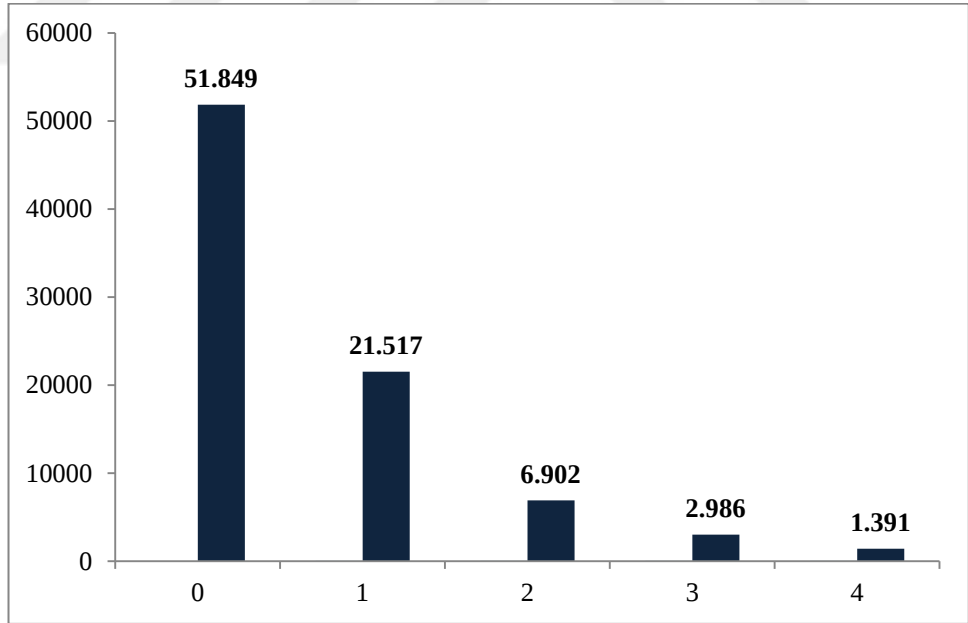
Şekil 5. 27.Gerçekleşen Kaza Sonrası Yaya Ölüm Sayısı

Şekil 5.28' e göre 39.074 Kazada 0 yaralı, 41.931 Kazada ise 1 yaralı olduğu görülmektedir. Şekil 5.29' a göre 51.849 Kazada 0, 21517 kazada 1, 6.902 kazada ise 2

kişinin yaralandığı görülmüştür.



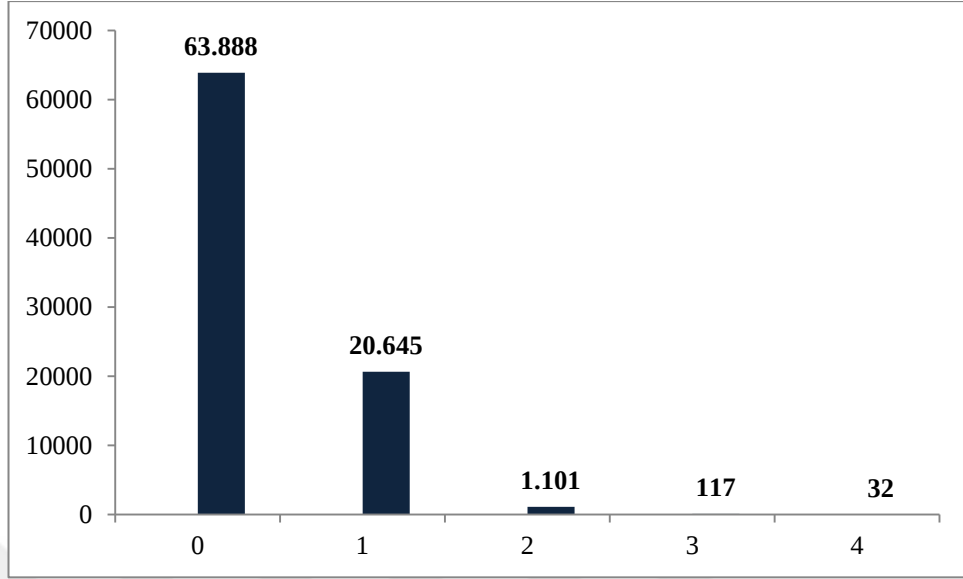
Şekil 5.28. Gerçekleşen Kaza Sonrası Yaralanan Sürücü Sayısı



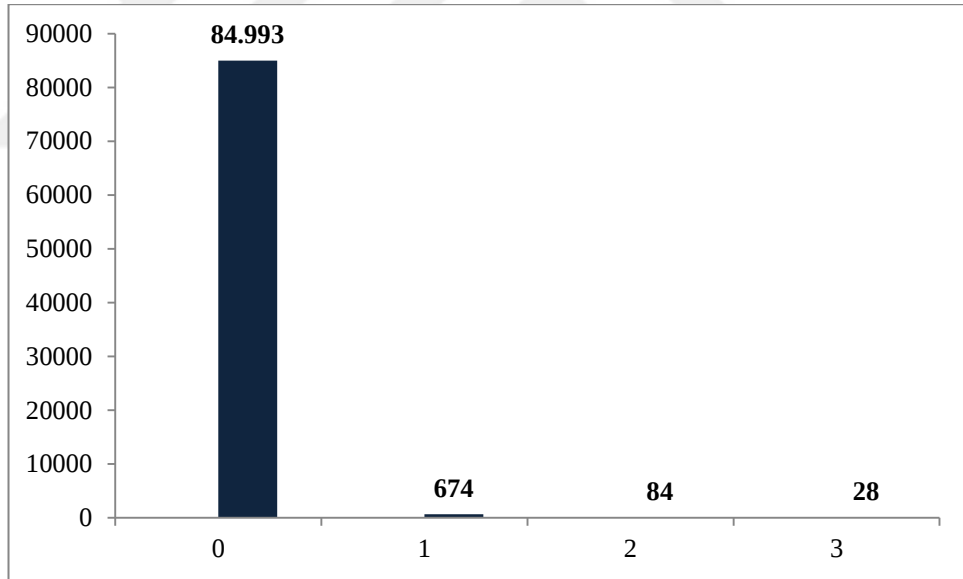
Şekil 5.29. Gerçekleşen Kaza Sonrası Yaralanan Yolcu Grafiği

Şekil 5.30'a göre 63.888 Kazada yaralı yaya olmadığı, 20.645 kayıta ise 2 kişinin yaralandığı elde edilmiştir. Şekil 5.31'e göre 84.993 kazada ölüm gerçekleşmezken, 674 Kazada 1 kişi ölmüştür. Şehir merkezinde gerçekleşen kazaların çoğu şiddetli

olmadığı için ölü sayısının düşük olduğu görülmektedir.

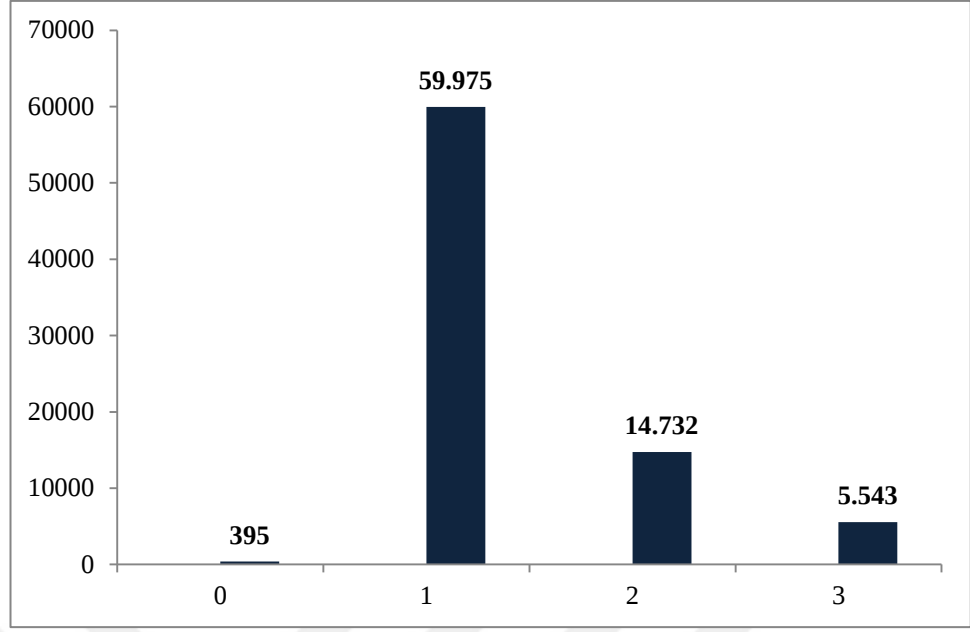


Şekil 5.30. Gerçekleşen Kaza Sonrası Yaya Yaralı Grafiği

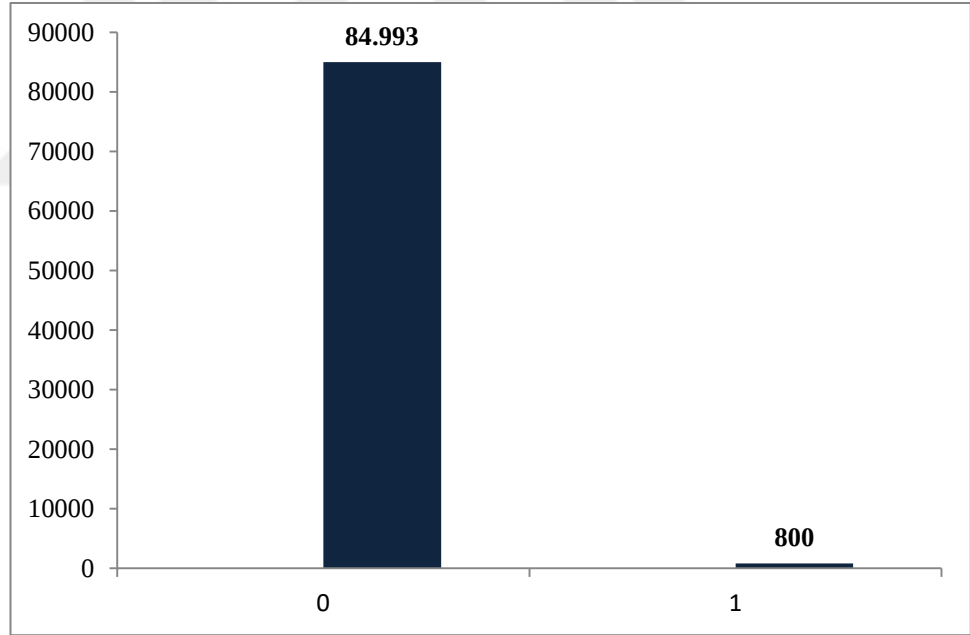


Şekil 5.31. Gerçekleşen Kaza Sayılarına göre ölüm sayıları

Şekil 5.32’de görüldüğü gibi 59.975 kazada en az 1 yaralı, 14.732 kaza da 2 yaralı bulunmaktadır. Şekil 5.33’e göre 84.993 Kişiden 0 kişi ölmüştür, 800 kişiden ise 1 kişi ölmüştür. Söz konusu tablo veri dengesizliğini göstermektedir.



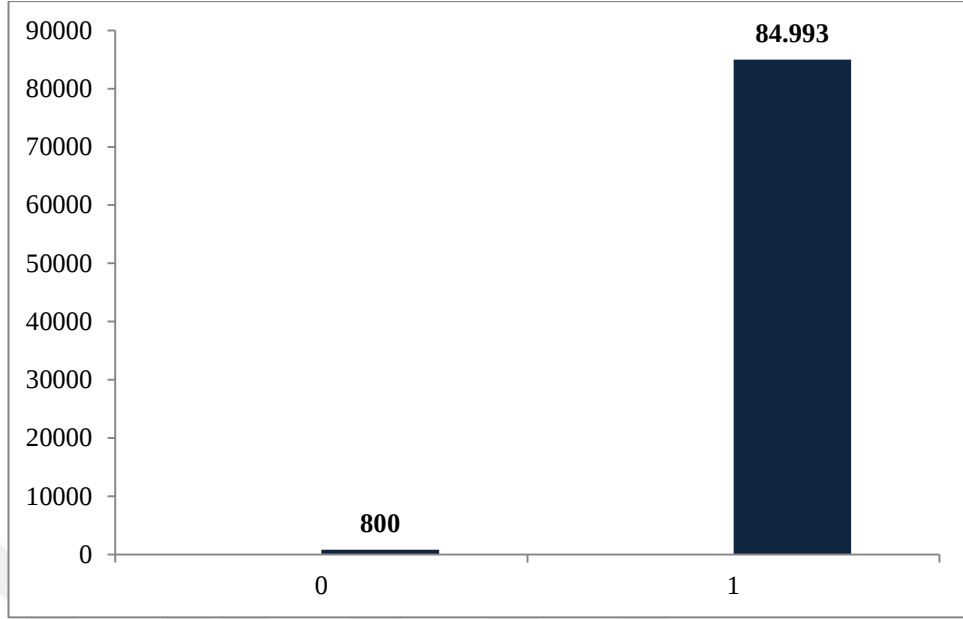
Şekil 5.32. Gerçekleşen Kaza Sonrası Toplam Yaralı Sayısı



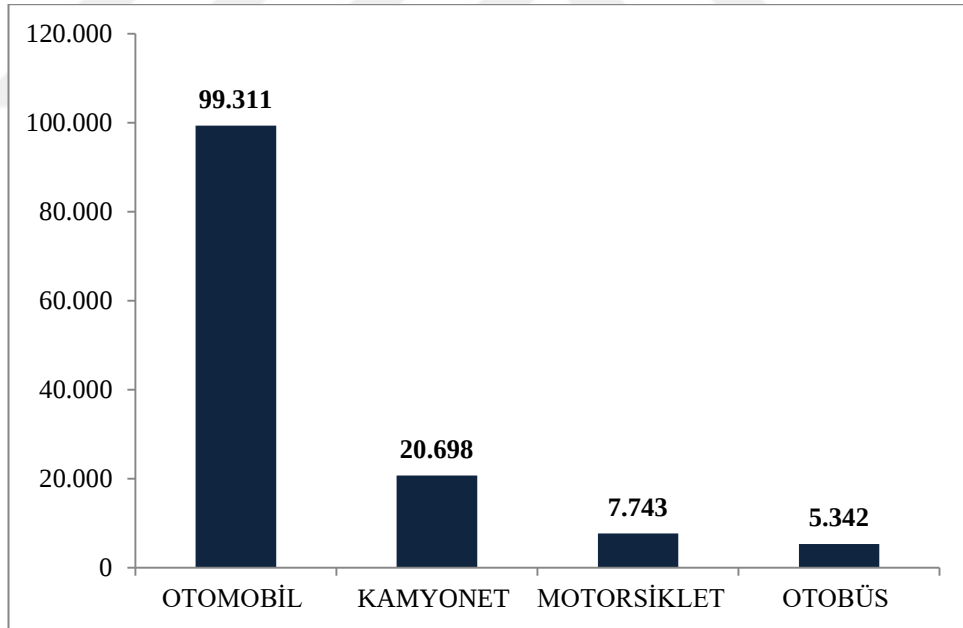
Şekil 5.33. Ölüm İle Sonuçlanan Kazalar

Ankara ili içerisinde 2013- 2020 yılları arasında gerçekleşen kazalarda Şekil 5.34'e göre 84.993 Kişiden 1 kişi yaralanmıştır, 800 kişiden ise yaralanan olmamıştır. Gerçekleşen kazalar da otomobillerin daha fazla kazaya karıştığı Şekil 5.35'te görülmektedir. Şekil 5.36'a göre Ankara ile içerisinde kazaya karışan araçların 123.812 adetinin kullanım amaçlarının özel olduğu, 19.418 adetinin ticari ve 731 adetinin emniyet amacı ile

kullanıldığı görülmektedir



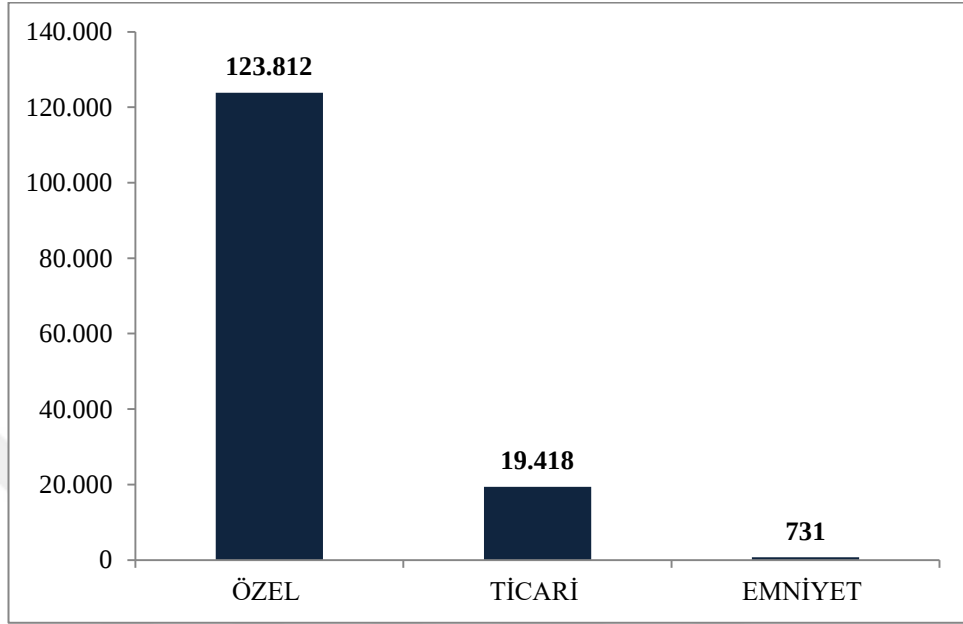
Şekil 5.34. Yaralanma İle Sonuçlanan Kazalar



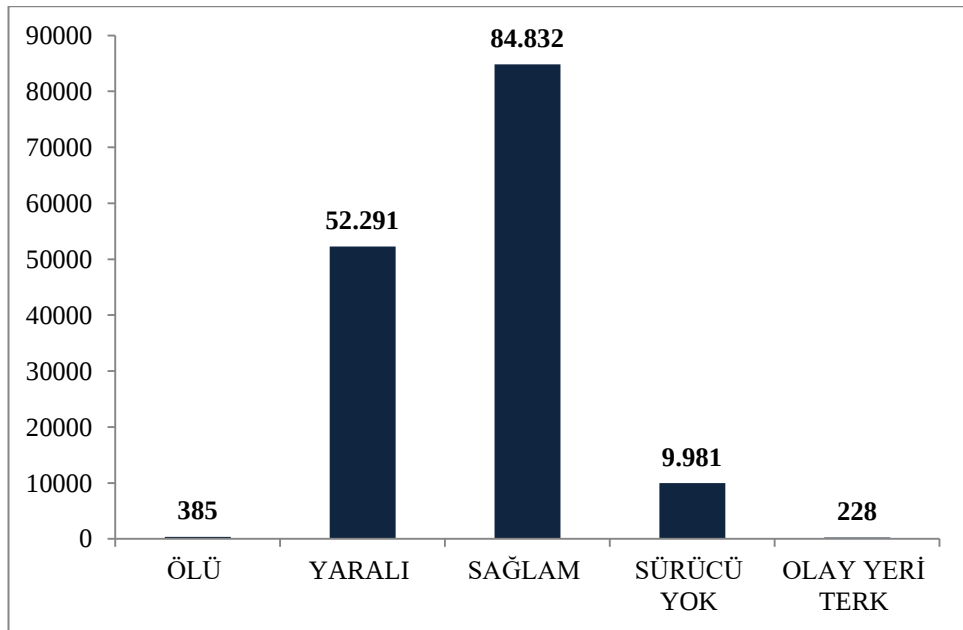
Şekil 5.35. Kazaya Karışan Araç Cinsi

Gerçekleşen 84.832 kazada sürücülerin sağlam, 52.291 kazada yaralı olarak kazaları atlattığı Şekil 5.37’de görülmektedir. Kaza sonrası ölüm gerçekleşen kaza sayılarının toplam kazalar içerisinde Cüz’i seviye de olduğu görülmektedir. Şekil 5.38’e göre 22-30

yaş arası en çok kazaya karışan kişiler olduğu gözlenmektedir. Kaza sayıları, söz konusu yaş grubu içerisinde hızlı araç kullanmaya meyilli ve tecrübe eksikliği ile ilişkilendirilmiştir.



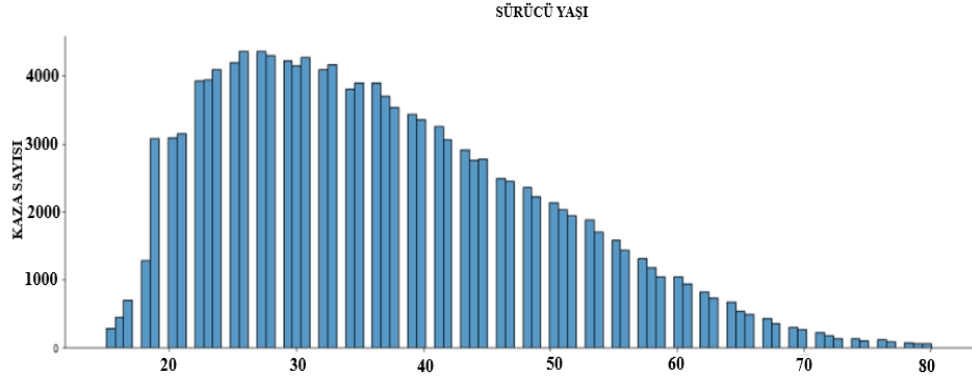
Şekil 5.36. Araç Kullanım Amacı



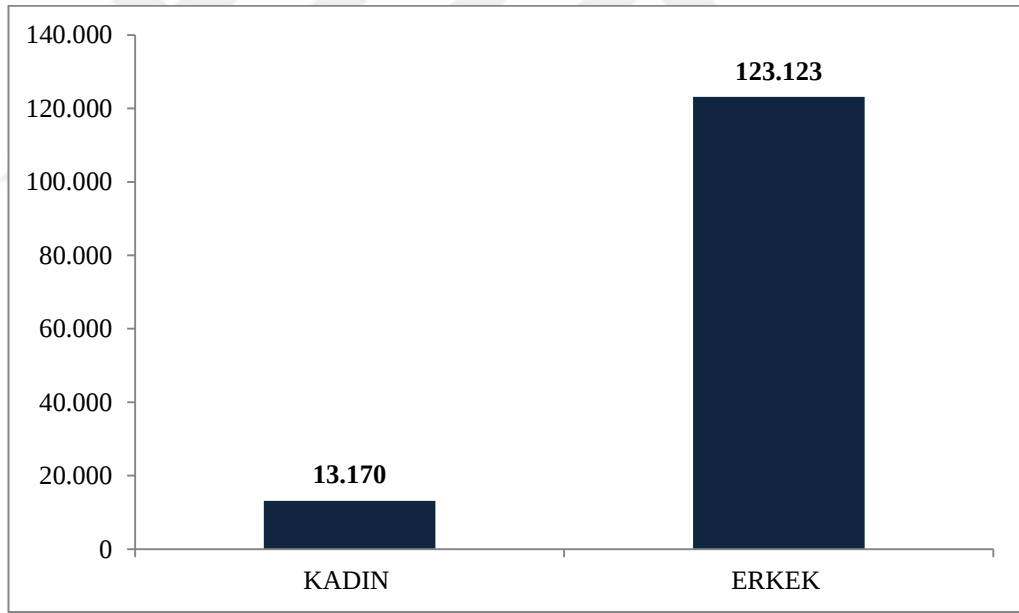
Şekil 5.37. Kazaya Karışan Sürücü Durumu

Trafikte aktif araç kullanan sürücülerin çoğunlukla erkek olmasından kaynaklı olarak

erkeklerin daha fazla kazaya karıştığı ve 2013-2020 yılları arasında 123.123 erkeğin kazaya karıştığı Şekil 5.39’da görülmektedir.



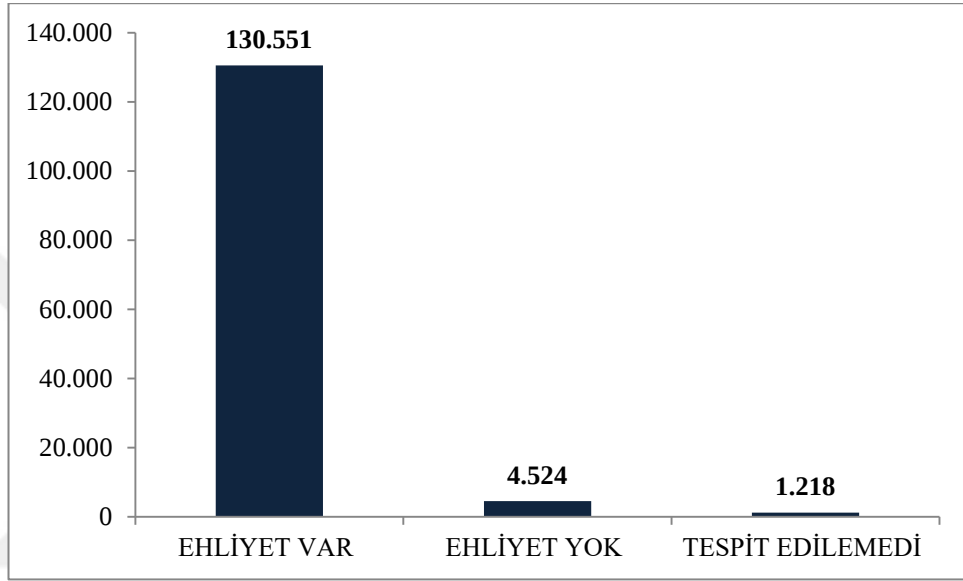
Şekil 5. 38. Kazaya Karışan Sürücü Yaşı



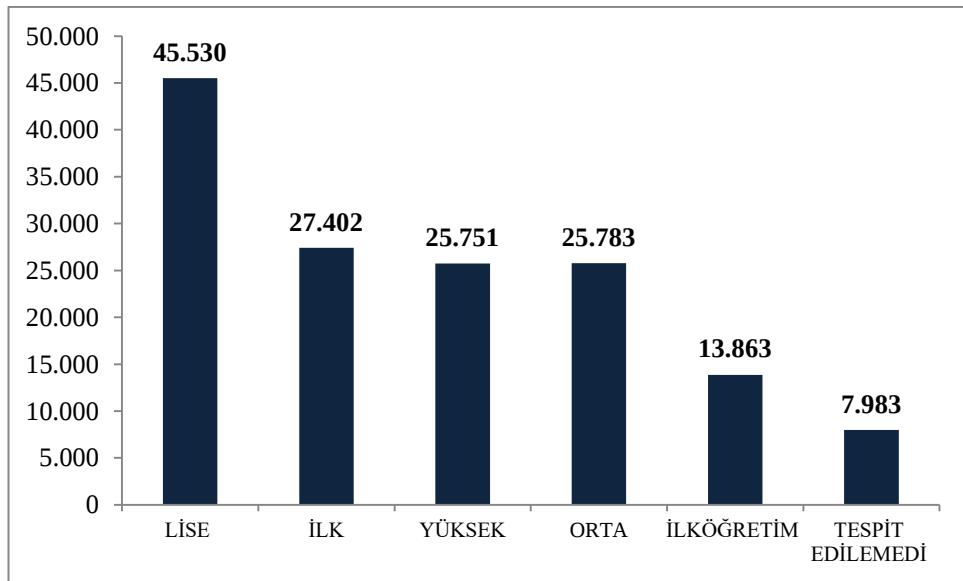
Şekil 5.39. Kazaya Karışan Sürücü Cinsiyeti

Kazaya karışan 130.551 sürücünün ehliyeti olduğu Şekil 5.40’ta görülmektedir. Kazaya karışan sürücü eğitim durumları Şekil 5.41 de verilmiş olup 45.530 kazada sürücülerin lise eğitimi, 72.240 kazada ilköğretim eğitimi, 25.751 kazada yüksek eğitim almış bireylerden oluştuğu görülmektedir. Şekil 5.42’ye göre 48.845 Kazada sürücülerin kusursuz olduğu elde edilirken, 39.338 kazada araç hızını yol, hava ve trafik şartlarına

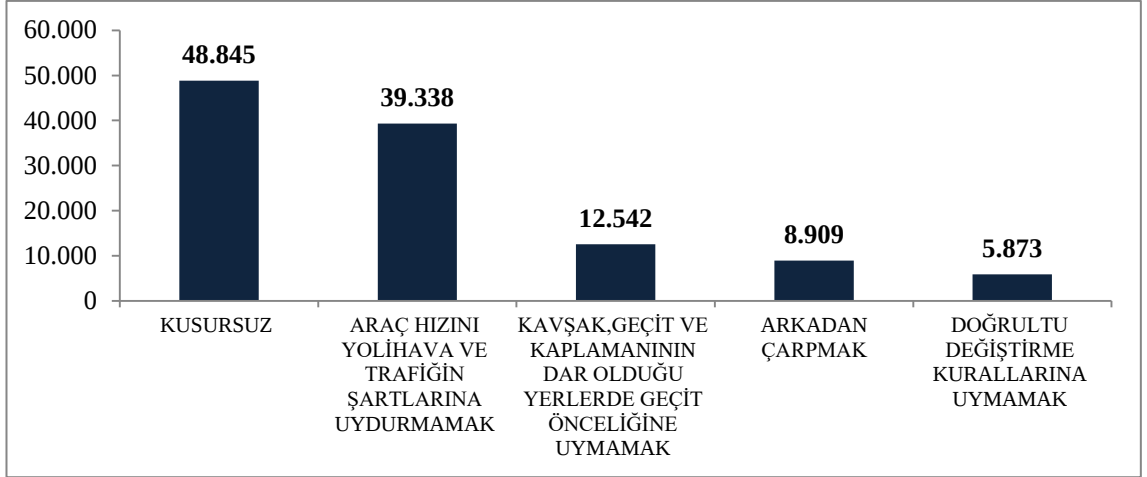
uyulmadığından, 12.542 kazada geçit önceliğine uyulmamasından, 8.909 kazanın arkadan çarpma şeklinde olduğu, 5.873 kazada ise dönüş kurallarına uyulmamasından dolayı kazaların gerçekleştiği gözlemlenmiştir. Ankara ili içerisinde yol kaplaması ağırlıklı olarak asfalt olması sebebi ile 83.972 kaza asfalt zeminde gerçekleşirken, 1.821 kazanın yol kaplaması asfalt zemin olmayan noktalarda gerçekleştiği Şekil 5.43’de görülmektedir.



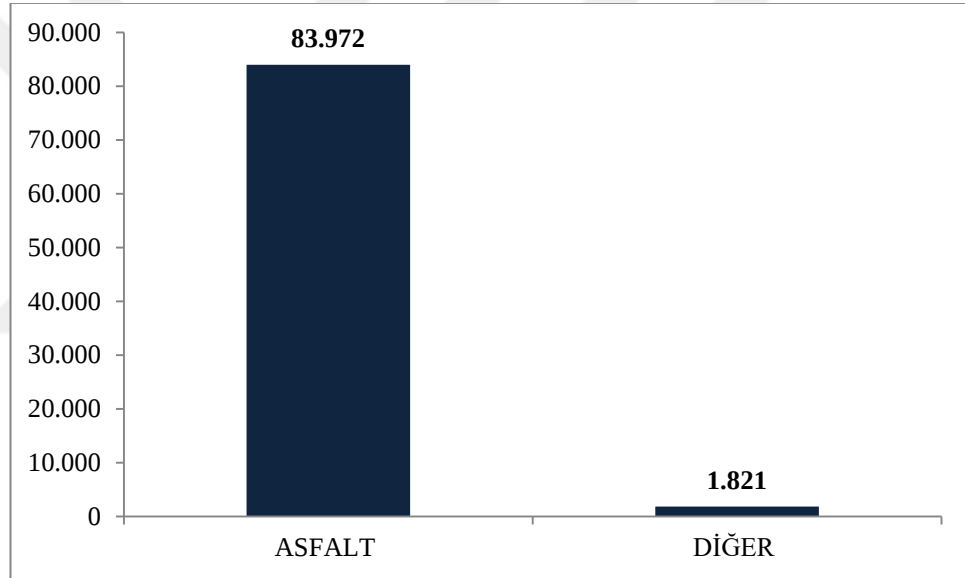
Şekil 5. 40. Kazaya Karışan Sürücü Ehliyet Durumu



Şekil 5.41. Kazaya Karışan Sürücü Eğitim Durumu



Şekil 5.42. Gerçekleşen Kazada Sürücü Hatası

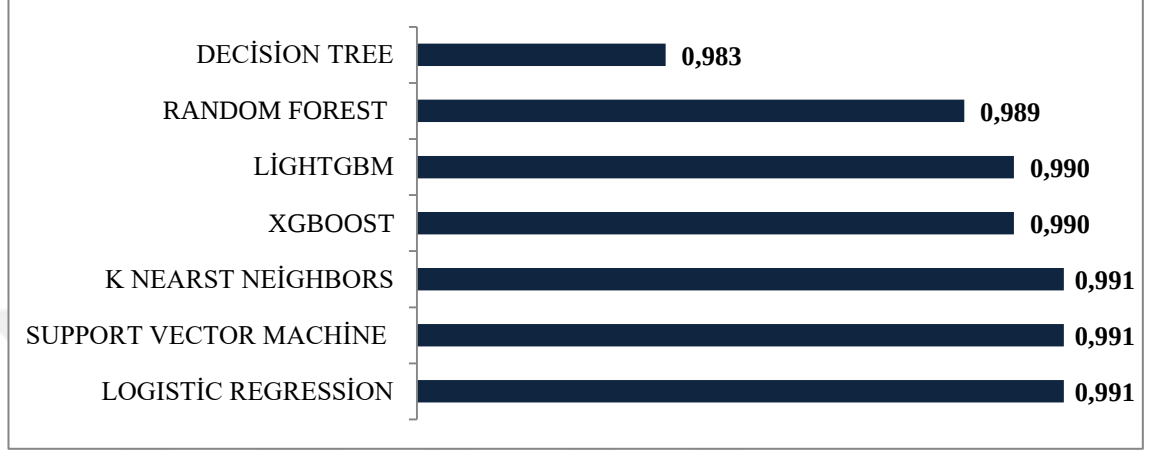


Şekil 5.43. Gerçekleşen Kaza Noktasında Yolun Kaplaması

Makine öğrenmesi algoritmaları ile yapılan model sonuçları aşağıda verilmiştir. Burada uygulanan modeller Logistic Regression (Lojistik regresyon), KNN (K En Yakın Komşu), SVM (Destek Vektör Makinesi), Decision Tree (Karar Ağacı), Random Forest (Rastgele Orman), XGBoost (Ekstrem Gradyan Artırma), LightGBM (Hafif Gradyan Artırma)'dır. Bu yapılan modellerle ilgili olarak aşağıda bazı hata metriklerine göre değerlendirmeler yapılmıştır.

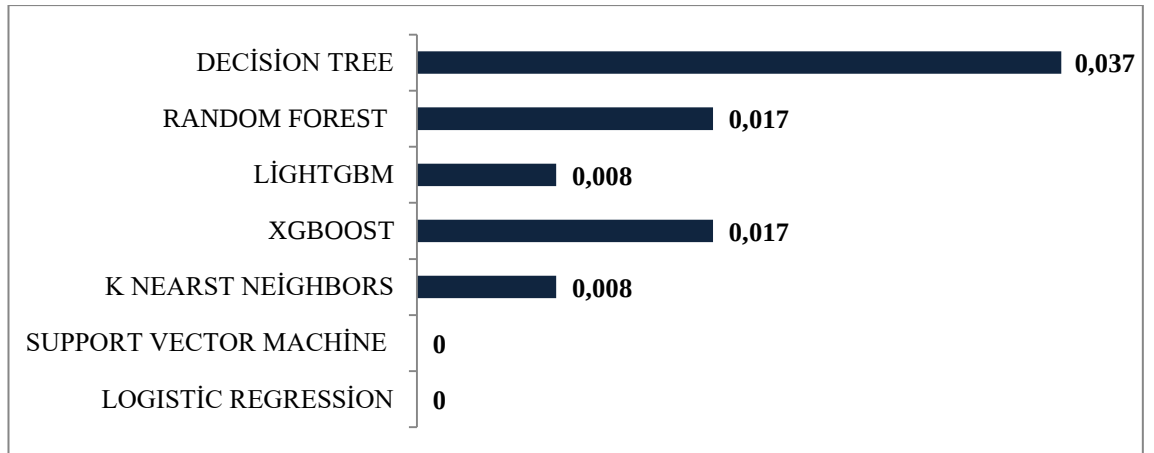
Smote işlemi uygulanmadan önce yapılan modelleme sonuçları:

Aşağıdaki sonuçlar veri dengesizliğinden dolayı yapılan Smote işleminden önce elde edilen sonuçlardır.



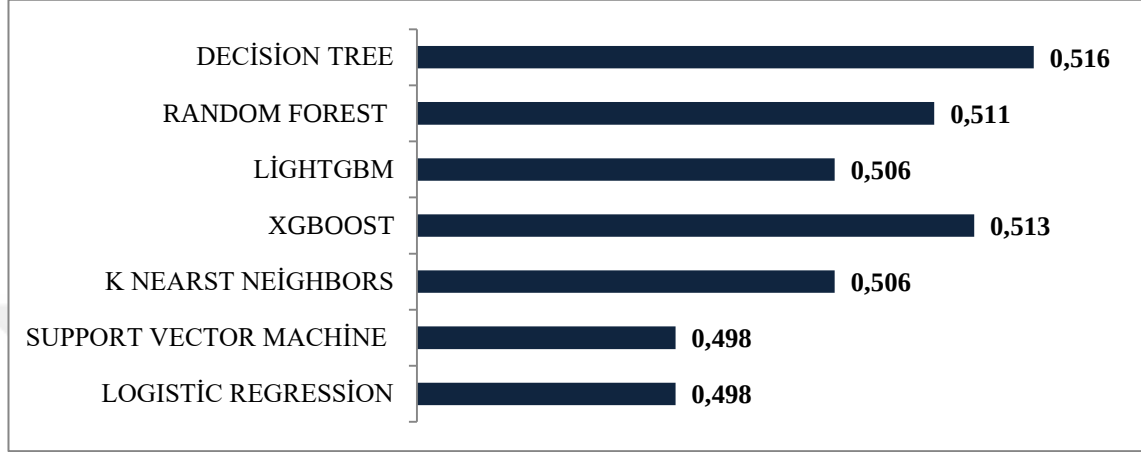
Şekil 5.44. Smote İşlemi Öncesi Algoritmaların Doğruluk (Accuracy) Sonuçları

Doğruluk (Accuracy) değeri modelleme sonucunda doğru tahmin edilen alanların toplam veri kümesine oranını göstermekte olup. Smote öncesi ele alınan verilerin Doğruluk (Accuracy) metrik sonuçları Şekil 5.44'te verilmiştir. Elde edilen sonuçlarda Lojistik regresyon (Logistic regression), Destek Vektör Makinesi (SVM) ve K En Yakın Komşu (KNN) yöntemlerinin başarılı sonuçlar verdiği görülmektedir.



Şekil 5.45. Smote İşlemi Öncesi Algoritmaların Duyarlılık (Recall) Değerleri

Duyarlılık (Recall) pozitif tahmin etmemiz gereken işlemlerin ne kadarını pozitif tahmin ettiğimizi gösteren bir metriktir. Bu metriğe göre en iyi sonucu veren Şekil 5.45'e göre Karar Ağacı (Decision Tree), Rastgele Orman (Random Forest) ve XGBoost (Ekstrem Gradyan Artırma) yöntemidir.



Şekil 5.46. Smote İşlemi Öncesi Algoritmaların F1_Skor Değerleri

F1_Skor dengesiz dağılıma sahip veri setleri için kullanılan kesinlik ve duyarlılığın harmonik ortalamasının alındığı modelleme sonuçlarıdır. F1_Skor modellemesinde ağırlığı düşük olan veri setleri ön plana çıkarken ağırlıkları yüksek olan değerler manipüle edilerek modelleme sonuçlarını alır. Bu metrikte Şekil 5.46'a göre en iyi sonucu Karar Ağacı (Decision Tree) algoritması vermiştir.

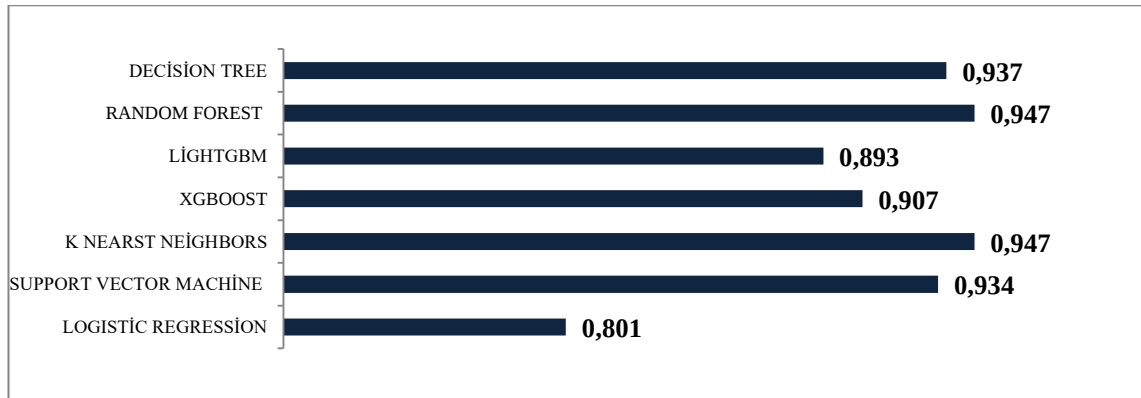
Tablo 5.1. Smote İşlemi Öncesi Algoritmaların Değerleri

	Accuracy	F1_Skor	Recall	Recall Macro
Decison Tree	0,983	0,516	0,037	0,515
Random Forest	0,989	0,511	0,017	0,508
Lightgbm	0,99	0,506	0,008	0,504
Xgboost	0,99	0,513	0,017	0,506
Knn	0,991	0,506	0,008	0,504
Svm	0,991	0,498	0	0,5
Logistic Regression	0,991	0,498	0	0,5

Tablo 5.1’de smote işleminden önce kurulan modellerin tüm hata metriklerinin sonuçları verilmiştir. Smote işleminden önce Doğruluk (Accuracy) metriğine göre her bir algoritmanın doğru sonuçlar verdiği görülmüştür. Literatürde yapılan birçok trafik kaza tahmininde Doğruluk (Accuracy) metriğine göre değerlendirme yapılmaktadır. Fakat bu çalışma Sınıflandırma problemi (Classification problem) olduğu için Duyarlılık (Recall) metriği üzerinden değerlendirme yapılması kanaatine varılmıştır. Bu değerlendirmeye göre Smote İşlemi Öncesi elde edilen veriler ile oluşturulan Tablo 5.1.’ de Duyarlılık (Recall) metriğine göre sırasıyla Karar Ağacı (Decision Tree), Rastgele Orman (Random Forest) ve XGBoost (Ekstrem Gradyan Artırma) yöntemlerinin smote işleminden önce başarılı sonuçlar verdiği görülmüştür.

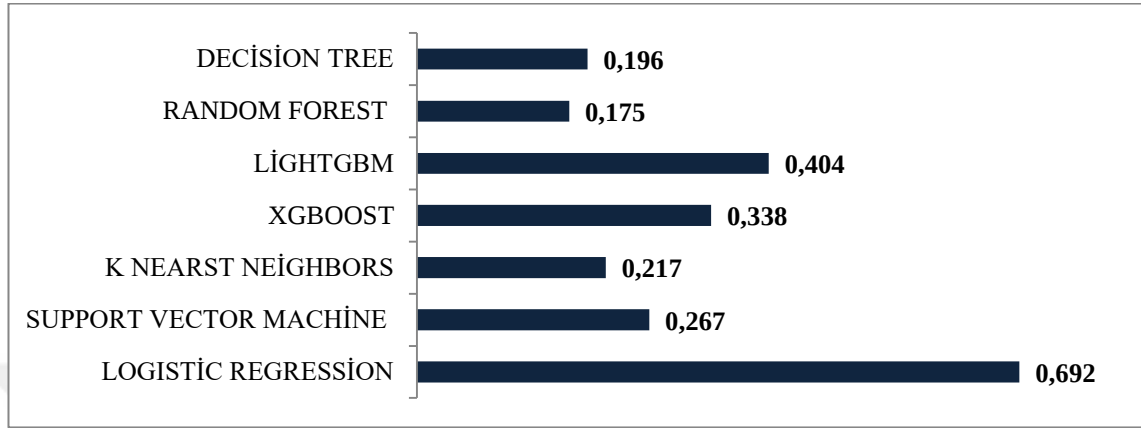
Smote işlemi uygulandıktan sonra yapılan modelleme sonuçları:

Elimizdeki veri setinde dengesiz (unbalance) durumu söz konusudur. Şekil 5.33’te görüldüğü gibi 84.993 kazada ölüm görülmezken, 674 kazanın 1 kişinin ölümü ile sonuçlandığı görülmektedir. Bu durum söz konusu veri dengesizliğine neden olmaktadır. Bu açıdan veri setine smote işlemi uygulanmıştır. Smote işlemi uygulandıktan sonra Logistic Regression (Lojistik regresyon), KNN (K En Yakın Komşu), SVM (Destek Vektör Makinesi), Decision Tree (Karar Ağacı), Random Forest (Rastgele Orman), XGBoost (Ekstrem Gradyan Artırma), LightGBM (Hafif Gradyan Artırma) modelleri tekrar kurulmuştur. Smote işleminden sonra oluşturulan modeller aşağıda verilmiştir.



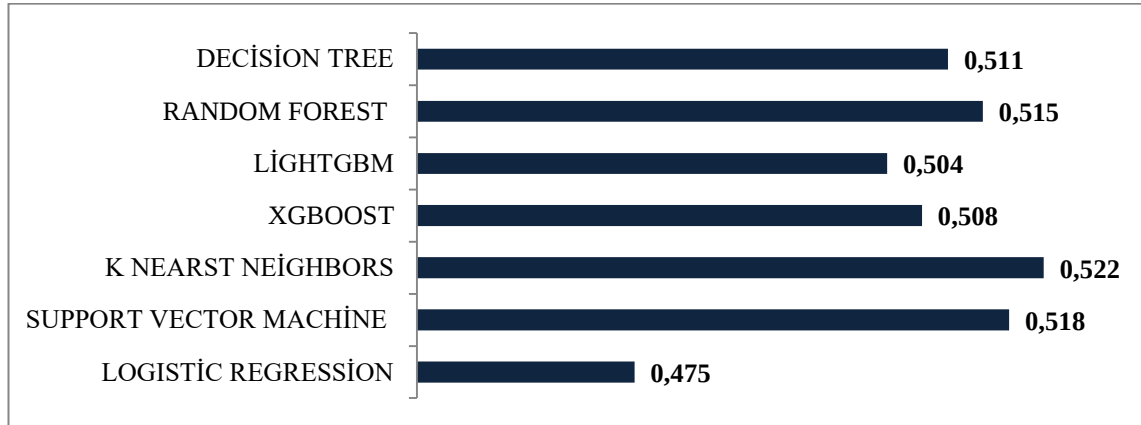
Şekil 5.47. Smote İşlemi Sonrası Algoritmaların Doğruluk (Accuracy) Sonuçları

Smote sonrası ele alınan veriler ile Doğruluk (Accuracy) sonuçları Şekil 5.47'ye göre Rastgele Orman (Random Forest) ve K En Yakın Komşu (K Nearest Neighbors) olduğu görülmektedir.



Şekil 5.48. Smote İşlemi Sonrası Algoritmaların Duyalılık (Recall) Değerleri

Şekil 5.48'de görüldüğü gibi smote işlemi sonrası, Duyalılık (Recall) metriğine göre sırasıyla en iyi sonucu veren Lojistik Regresyon (Logistic Regression) ve Hafif Gradyan Artıma Makinesi (Light Gradient Boosting Machine) yöntemidir.



Şekil 5.49. Smote İşlemi Sonrası Algoritmaların F1_ Skor Değerleri

Şekil 5.49'da görüldüğü gibi, F1 metriğe göre en iyi sonucu K En Yakın Komşu (KNN) algoritması vermiştir.

Tablo 5.2.Smote İşlemi Sonrası Algoritmaların Değerleri

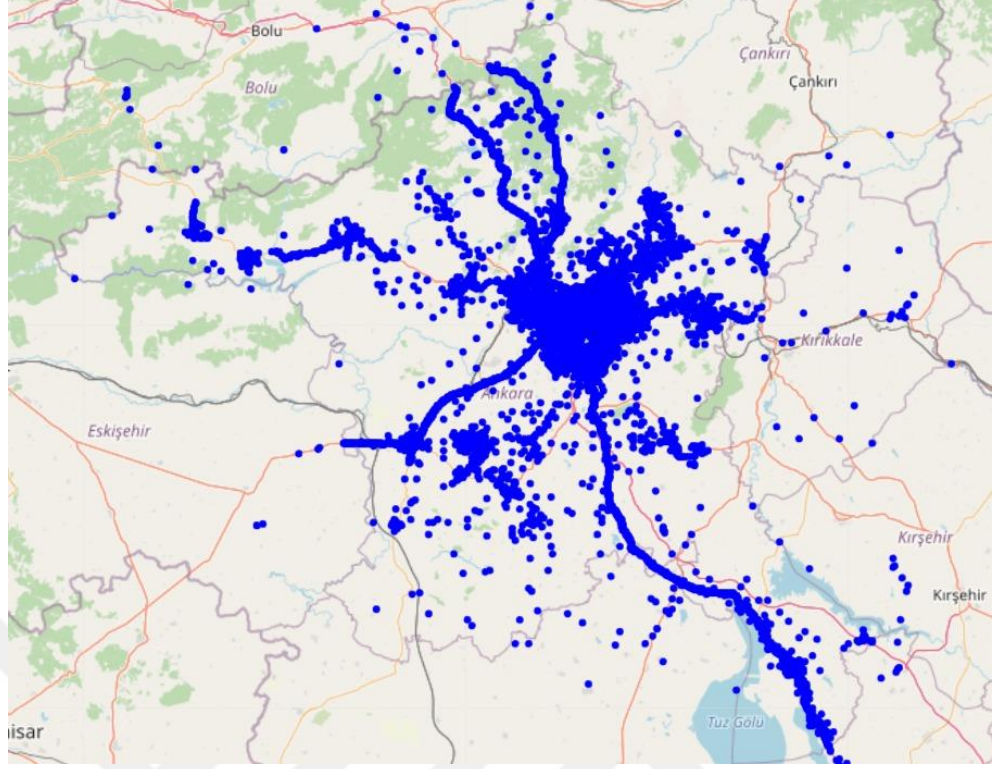
	Accuracy	F1_Skor	Recall	Recall Macro
Random Forest	0,947	0,515	0,175	0,565
Knn	0,947	0,522	0,217	0,585
Decison Tree	0,937	0,511	0,196	0,57
Svm	0,934	0,518	0,267	0,604
Xgboost	0,907	0,508	0,338	0,625
Lightgbm	0,893	0,504	0,404	0,651
Logistic Regression	0,801	0,475	0,692	0,747

Veri setine smote işlemi uygulandıktan sonra verilen algoritmalar ile eğitilmesi sonu elde edilen sonuçlar Tablo 5,2’de verilmiştir. Tablo5.2’de görüldüğü üzere Duyarlılık (Recall) metriğine göre sırasıyla Lojistik Regresyon (Logistic Regression) ve Hafif Gradyan Artıma Makinesi (Light Gradient Boosting Machine) yönteminin doğru sonuçlar verdiği görülmüştür. Smote işlemi sonrasında Doğruluk (Accuracy) metriği açısından modellerin her birinin tahmin gücünün çok iyi seviyelerde seyrettiği kanaatine varılmıştır.

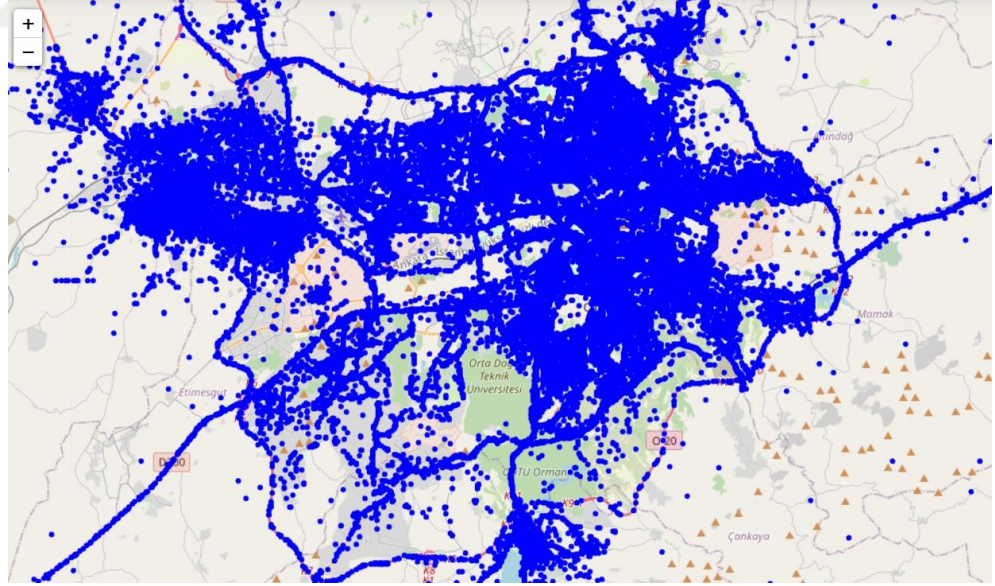
5.2. Kaza Lokasyonlarının Haritada Gösterilmesi

Ankara ilinde gerçekleşen kazaların genel durumu, kaza noktaları detaylandırılarak harita üzerinde lokasyonlar olarak işlenmiş, Şekil 5.50, Şekil 5.51, Şekil 5.52 ve Şekil 5.53’te verilmiştir.

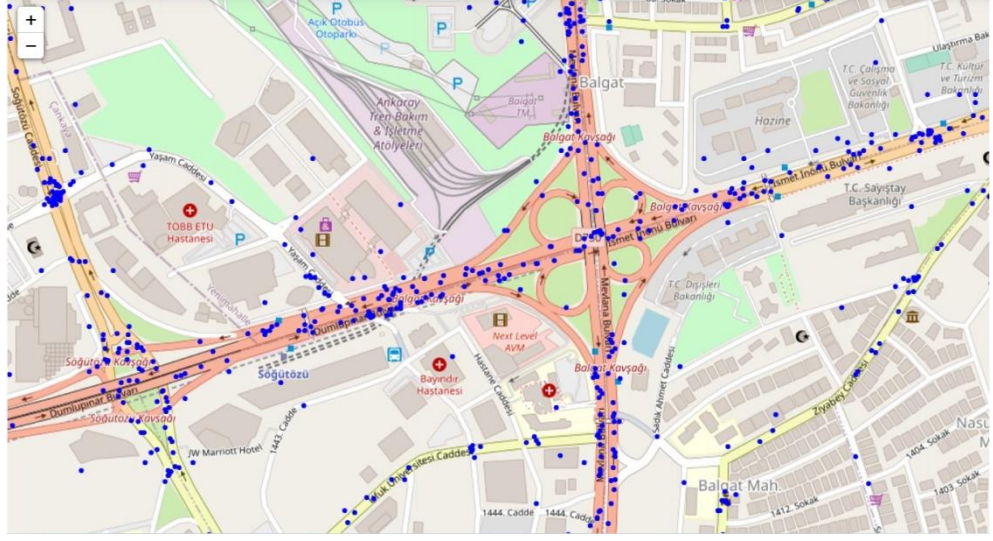
Bütün kazalar Python Kütüphanesi kullanılarak haritaya işlenmiş ve haritada çok başarılı sonuçlar elde edildiği görülmüştür. Şekil 5.53’te kaza lokasyonları detaylı olarak görülmektedir.



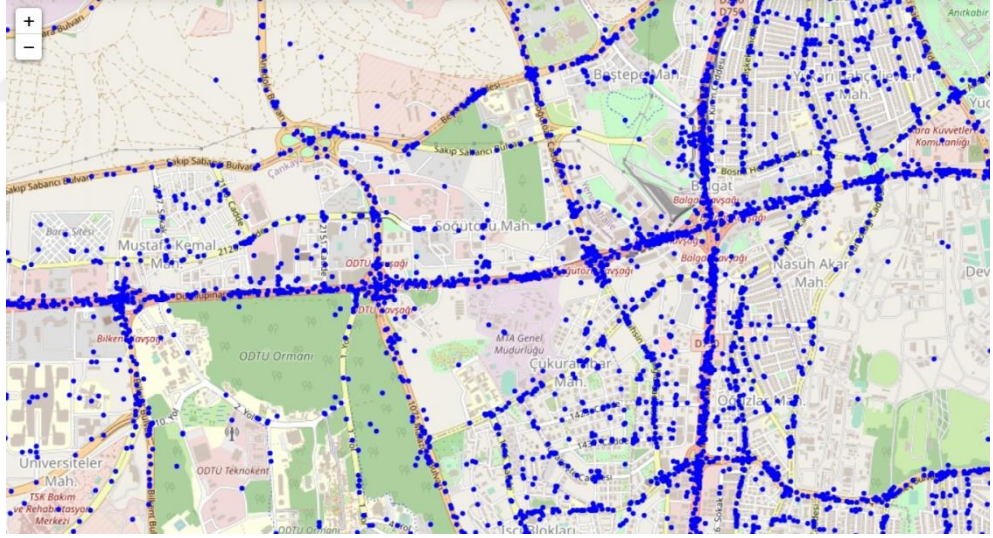
Şekil 5.50. Ankara Ve Çevre İlçeleri Dâhil Genel Görünüm



Şekil 5.51. Ankara Ve Çevre İlçeleri Dâhil Kaza Noktaları Genel Görünüm (2013-2020 Yılları)



Şekil 5.52. Ankara İli Kaza Noktaları İçin Örnek Bir Kavşak (2013-2020 Yılları)



Şekil 5.53. Ankara İli Kaza Noktaları Örnek Bir Şehir İçi Yol (2013-2020 Yılları)

6. SONUÇLAR

Makine öğrenmesi hayatın birçok noktasında, birçok problemin çözümünde kullanılmakta olan bir metottur. Trafik kazalarının modellenmesinde de başarılı uygulanabileceği görülmüştür. Bu çalışmada Makine Öğrenmesi ile trafik kazalarının modellenebileceği görülmüştür. Literatürde de buna benzer sonuçlar elde edilmiştir. Trafik kazaları ile alakalı tüm parametreler elde alındığında bu kazaların tahmin modellerinin kurulmasında Makine Öğrenmesi algoritmaları ile başarılı sonuçlar elde edilmiştir.

Smote işlemi öncesi ele alınan verilerin Doğruluk (Accuracy) metrik sonuçları incelendiğinde Lojistik regresyon (Logistic Regression), Destek Vektör Makinesi (SVM) ve K En Yakın Komşu (KNN) yöntemlerinin başarılı sonuçlar verdiği görülmektedir.

Smote işlemi sonrası ise Doğruluk (Accuracy) metrik sonuçları incelendiğinde Rastgele Orman (Random Forest) ve K En Yakın Komşu (K Nearest Neighbors) olduğu görülmektedir.

Smote işlemi öncesi iyi sonucu veren Karar Ağacı (Decision Tree), Rastgele Orman (Random Forest) ve XGBoost (Ekstrem Gradyan Artırma) yöntemi olduğu görülmektedir.

Duyallılık (Recall) metriğine göre sırasıyla en iyi sonucu veren Lojistik Regresyon (Logistic Regression) ve Hafif Gradyan Artırma Makinesi (Light Gradient Boosting Machine) yöntemi olduğu görülmektedir.

Ele alınan veriler dengesiz (unbalance) olduğu için modellemedeki sonuçları bir miktar etkilemiştir. Trafik kazası çalışmalarında Doğruluk (Accuracy) metriği öne çıkarılarak yapılan modellerin çok başarılı olduğu ifade edilmiştir. Bu çalışmada Doğruluk (Accuracy) metriği hem smote işlemi öncesi hem de smote işlemi sonrası yüksek seviyede elde edilmiştir. Fakat sınıflandırma problemlerinde (Classification Problem) Duyarlılık (Recall) metriğine bakılması doğru olacaktır. Literatür çalışmalarında

Duyarlılık (Recall) metriğine bakılmaması yanlış bir değerlendirme olduđu gör÷lmektedir.

Bu alıřmada hem Doğruluk (Accuracy) hem de Duyarlılık (Recall) metriđi verilmiř olup Duyarlılık metriđinde iyileřtirme yapılmıřtır.



KAYNAKLAR

- Afrin , F. and Nahar, I. (2015) "Incremental learning based intelligent job search system", Lisans Tezi, **Department of Computer Science & Engineering BRAC University**, Bangladeş
- Altın Yavuz , A., Ergül , B. ve Gündoğan Aşık , E. (2021) "Trafik kazalarının makine öğrenmesi yöntemleri kullanılarak değerlendirilmesi", **Uluslararası Mühendislik Araştırma Ve Geliştirme Dergisi**, 66-73.
- Arslan, Y. (2020) "Lightgbm algortiması ile yeni bir satış tahmin modelinin oluşturulması ve preakande sektöründe uygulanması", Yüksek Lisans Tezi **İstanbul Aydın Üniversitesi Fen Bilimleri Enstitüsü**, İstanbul, Türkiye.
- Ataseven , B. (2013) "Yapay sinir ağları ile öngörü modellemesi", **İstanbul Kültür Üniversitesi, İktisadi ve İdari Bilimler Fakültesi**, 101-115.
- Avcı , S. (2005) "Ulaşım coğrafyası açısından türkiye'nin ulaşım politikaları ve coğrafi sonuçları", **Ulusal Coğrafya Kongresi** , İstanbul, 88.
- Aydın , A. (2022) "Müşteri kaybı tahmininde sınıf dengesizliği problemi", **Politeknik Dergisi**, 351-360.
- Aydın, F. and Muhammed, O. (2018) "Türkiye'de karayolu ulaşımının tarihsel gelişimi", **Journal Of Awareness**, 257-266.
- Bayata H.F. and Hattatoğlu, F. (2010) "Yapay sinir ağları ve çok değişkenli istatistik yöntemlerle trafik kaza modellemesi", **Erzincan Üniversitesi Fen Bilimleri Enstitüsü Dergisi**, 207-219.
- Bayata, H., & Hattatoğlu , F. (2011) "Erzincan ili için farklı yöntemlerle trafik kaza tahmin modellemesi", **Erzincan Binali Yıldırım Üniversitesi Fen Bilimleri Enstitüsü Dergisi**, 31-46.
- Breiman, L. (2001) " Random Forest" **Kluwer Academic Publisher**, 5–32.
- Budak , H., & Erpolat , S. (2012) " Kredi riski tahmininde yapay sinir ağları ve lojistik regresyon analizi karşılaştırılması", **Online Academic Journal of Information Technology**, 23-30.
- Bulak, H. (2010) "Mersin ili merkezinde 2009-2006 yılları arasında meydana gelen trafik kazalarının incelenmesi", Yüksek Lisans Tezi , **Gazi Üniversitesi Sağlık Bilimleri Enstitü Kazaların Demografisi ve Epidemiyolojisi AnaBilim Dalı**

- Caruana, R. and Niculescu-Mizil, A. (2006) "An empirical comparison of supervised learning algorithms". *in Proceedings of the 23rd international*. 161-168.
- Chao, W.-L. (2011) "Machine learning tutorial" **Graduate Institute of Communication Engineering National Taiwan University**, Taiwan
- Chen , T., and Gusestrin , C. (2016) "Xgboost: a scalable tree boosting system" **KDD '16 Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**, New York, 785-794
- Chung, Y.-S., Park, K.-R., Kim, J.-M. and Kim, D.-H. (2013) "Study on predictive modeling of traffic accident". **Life Science journal**, 1-6
- Chong , M., Paprzycki, A. and Paprazcki, M. (2004) "Traffic accident data mining using machine learning paradigms" **Computer Science Department Oklahoma State University**
- Coşkun, S., Coşkun, A., Kartal, M., ve Bircan , H. (2004) "Lojistik regresyon analizinin incelenmesi ve dişhekimliğinde bir uygulaması", **Cumhuriyet Üniversitesi Diş Hekimliği Fakültesi Dergisi**.
- Crone, S. (2004) " " **Uluslararası Yapay Zeka Konferansı Bildiriler Kitabı**, Nevada/Las Vegas, 21-24
- Dal , A., Gümüş , İ., Güldal, S. and Yavaş, M. (2021) "A new resampling approach based on weighted geometric mean for unbalanced data" **Adıyaman Üniversitesi Mühendislik Bilimleri Dergisi**, 343-352.
- Demircigil, C. (2020, Ocak) "Tuik Haber Bülteni", **TÜİK**: <https://data.tuik.gov.tr/Bulten/Index?p=Motorlu-Kara-Tasitlari-Aralik-2020-37410>, 26.01.2021
- Dikker , J. (2017) " Boosted tree learning for balanced item recommendation in online retail", Master Thesis, **Netherlands: Endhoven University of Technology**
- Dong , S., Khattak , A., Ullah, I. and Hussian , A. (2022) "Predicting and analyzing road traffic injury severity using boosting-based ensemble learning models with shapley additive explanations", **International Journal of Environmental Research and Public Health**, 1-23
- Dong, C., Chunfu , S., Lİ , J., and Xiong, Z. (2018) " An improved deep learning model for traffic crash prediction ", **Journal of Advanced Transportation**, 1-13

- Galal, A., and Awadalla , T. (2012) "Characteristics and prediction of traffic accident casualties in sudan using statistical modeling and artificial neural networks" ***International Journal of Transportation Science and Technology***, 305-317.
- Goutte , C., And Gaussier , E. (2005) "A probabilistic interpretation of precision, recall and f-score, with implication for evaluation", ***A Probabilistic Interpretation of Precision, recall and F-Score, with Implication for Evaluation***, 345-359.
- Hall, P., Park, B., and Samworth , R. (2008) "Choice of neighbor order in nearest-neighbor classification" ***Annals of Statistics***, 2135–2152.
- Han, J., Kamber, M., & Pei, J. (2011) Data mining concepts and techniques third edition, ***Morgan Kaufmann***, Massachusetts
- Haykin, S. (2009) Neural Networks and Learning Machines, ***Pearson Education***, Londra .
- Hemdil, Ö. (2021) "Karayolu Trafik Kaza İstatistikleri" , 2020. ***TÜİK***: <https://data.tuik.gov.tr/Bulten/Index?p=Road-Traffic-Accident-Statistics-2020-37436,01.06.2021>
- Jain, A., Mao, J., and Mohiuddin, K. (1996) "Artificial neural network; a tutorial" ***IBM Almaden Research Center***, 34-35.
- Kabalıcı , E. (2014) "Yapay Sinir Ağı Ders Notları", ***Nevşehir Üniversitesi Mühendislik Mimarlık Fakültesi***
- Kakde , Y., and Agrawal, S. (2018) "Predicting survival on titanic by applying exploratorydata analytics and machine learning techniques ", ***International Journal of Computer Applications***, 32-38
- Kartal M., Kutlar A. ve Kartal M. (2011) "Logistik regresyon tekniği ile trafik kazalarını etkileyen risk faktörlerinin incelenmesi: sivas, kayseri, yozgat örneği", ***Ekonomik ve Sosyal Araştırmalar Dergisi***, 46-68.
- Ke G. , Meng, Q., Finley T , Wang T , Chen W , Ma W , and Liu T . (2017) " Lightgbm: a highly efficient gradient boosting decision tree", ***Microsoft Redmond, The University of Waikato***.
- Kıyıldı, K. (2017) "Türkiye için yapay sinir ağları yöntemi ile trafik kazası tahmini araştırması" ***International Symposium on Innovative Technologies in Engineering and Science***, 1643-1651.
- Kubat, C. (2015) Matlab Yapay Zeka Ve Mühendislik Uygulamaları, ***Abaküs***, İstanbul.

- Kumede , B., Zhang, F., Zhou, F., Hüseyin , S., Almasrı, A. and Assefa, M. (2019) "Classification of road traffic accident data using machine learning algorithms" **IEEE 11th International Conference on Communication Software and Networks (ICCSN)**, 682-687.
- Kuşkapan , E. ve Çodur , M. (2021) "Trafik kazalarının sınıflandırılmasında çok katmanlı algılayıcı ,regresyon ve en yakın komşuluk algoritmalarınınperformans analizi", **Politeknik Dergisi**, 1-373.
- Liao, Y. and Vemuri, V. (2002) " Using text categorization techniques for intrusion detection " **Computers & Security**.
- MacKay, D. (1992) " Bayesian interpolation", **California Institute of Technology** , 74-139.
- Mcneill, G., and Anderson, D. (1992) "Artificial neural networks technology" **Kaman Sciences Corporation**, Utica/New York.
- Meltek, S., and KAYAALP, K. (2020) Makine öğrenmesinde,teoriden örnek matlab uygulamalarına kadar destek vektör makineleri, **İksad**, Ankara.
- Mohammed , A., Ambak, K., Mosa A., and Syamsunur D. (2018) "Classification of traffic accident prediction models: a review paper", **International Journal of Advances in Science Engineering and Technology**, 35-38.
- Mussone , L., Ferrari , A. and Oneta, M. (1999) "An analysis of urban collisions using an artificial intelligence model" **Accident Analysis & Prevention**, 705-718
- Özdağ, M. (2019) "Derin öğrenme teknikleri kullanılarak anayol trafik analizi", Yüksek Lisans Tezi, **Karabük Üniversitesi Fen Bilimleri Enstitüsü**, Karabük, Türkiye
- Özdemir , N. (2006) "Cumhuriyet dönemi karayolu politikası (1923-1960)", Doktora Tezi, **Ankara Üniversitesi Sosyal Bilimler Enstitüsü**, Ankara
- Özden , C. ve Acı , Ç. (2018) "Makine öğrenmesi yöntemleri ile yaralanmalı trafik kazalarının analizi: adana örneği" **Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi**, 266-275.
- Özekes, S., ve Karakoç , E. N. (2002) "Makine öğrenmesi yöntemleriyle anormal ağ trafiğinin tespiti " **Marmara Üniversitesi Fen Bilimleri Dergisi**, 566-576.
- Öztemel, E. (2006) Yapay Sinir Ağları , **Papatya Yayıncılık Eğitim**

- Öztürk, K., & Şahin , M. (2018) Yapay Sinir Ağları Ve Yapay Zekâ'ya Genel Bir Bakış, **Takvim-İ Vekayi**, 32
- Parıldar , O. (2014) "Trafik kazalarının sınıflandırılmasında karar ağacı kullanımı bodrum ilçesi örneği", Yüksek Lisans Tezi , **Gazi Üniversitesi Fen Bilimleri Enstitüsü**, Ankara
- Patrous, Z. (2018) "Evaluating xgboost for user classification by using behavioral features extracted from smartphone sensors", Yüksek Lisans Tezi, **Kth Royal Institute Of Technology, Stockholm, Sweden**.
- Pirim, H. (2006) "Yapay zeka", **Journal Of Yaşar University**, 81-93.
- Rojas , R. (1996) Neural Networks-A Systematic, **Springer**, Berlin.
- Selimoğlu, E. (2014) "Trafik kazalarının nedenleri, sonuçları ve kazaların önlenmesine ilişkin öneriler", **Ankara Üniversitesi Ziraat Mühendisliği**, 51-54.
- Sungur, İ., Akdur R. ve Piyal, B. (2014) "Türkiye'deki Trafik kazalarının analizi", **Ankara Medical Journal**, 114-125.
- Şen , Z. (2004) Yapay Sinir Ağları, **Su Vakfı** .
- Tan, P., Steinbach, M., ve Karpatne, A. (2016) Introduction to Data Mining, **Pearson Addison Wesley**, Boston.
- Uğur, A. Ve Kınacı, A. (2006) "Yapay zeka teknikleri ve yapay sinir ağları kullanılarak web sayfalarının sınıflandırılması", **TOBB Ekonomi Ve Teknoloji Üniversitesi. Türkiye'de İnternet Konferansı Bildirileri**, 345-349, Ankara
- Who. (2004) World report on road traffic injury prevention, World Helat Organization :who.int/publication/i/item/worl-report-on-road-traffic-injury-prevention
- Yakut, E., Elmas, B. and Yavuz, S. (2014) "Yapay sinir ağları ve destek vektör makineleri yöntemi ile borsa endeksi tahmini", **Süleyman Demirel Üniversitesi İktisadi Ve İdari Bilimler Fakültesi Dergisi**, 139-157.
- Yassin, S. (2020) Road accident prediction and model interpretation using a hybrid K-means and random forest algorithm approach, **Springer**.
- Yılmaz , A. (2022) Yapay Zeka, **Kodlab**, İstanbul