

T.C.
ERZİNCAN BİNALİ YILDIRIM ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
İNŞAAT MÜHENDİSLİĞİ ANABİLİM DALI

MALATYA İLİ İÇİN TRAFİKTE KAZA ANALİZİ

Berivan AVCI

Danışman: Doç. Dr. HALİM FERİT BAYATA

TEZ JÜRİ ÜYELERİ

Doç. Dr. Halim Ferit BAYATA

Doç. Dr. Osman Ünsal BAYRAK

Dr. Öğr. Üyesi Fatih İrfan BAŞ

YÜKSEK LİSANS TEZİ

ERZİNCAN, 2026

© 2026 [Berivan AVCI]. Tüm hakları saklıdır.

Kabul ve Onay Sayfası

Doç. Dr. Halim Ferit BAYATA danışmanlığında, Berivan AVCI tarafından hazırlanan bu çalışma 16.01.2026 tarihinde aşağıdaki jüri tarafından İnşaat Mühendisliği Anabilim Dalı'nda Yüksek Lisans Tezi olarak kabul oybirliği (4/4) ile kabul edilmiştir.

Başkan : Doç. Dr. Osman Ünsal BAYRAK İmza:

Üye : Doç. Dr. Halim Ferit BAYATA İmza:

Üye : Dr. Öğr. Üyesi Fatih İrfan BAŞ İmza:

Bu tez Enstitü Yönetim Kurulunun / / 20.... tarih ve/..... sayılı kararı ile onaylanmıştır.

Prof. Dr. Kemal Volkan ÖZDOKUR
Enstitü Müdür V.

Not: Bu tezde kullanılan özgün ve başka kaynaklardan yapılan bildirişlerin, şekil ve tabloların kaynak olarak kullanımı, 5846 sayılı Fikir ve Sanat Eserleri Kanunundaki hükümlere tabidir.

Bilimsel Etięe Uygunluk Sayfası

Malatya İli İin Trafikte Kaza Analizi isimli —Yüksek Lisans tezim tarafımca intihal tespit programı ile incelenmiştir. Buna göre tezimde bilimsel etik ihlali ve intihal olarak nitelendirilebilecek herhangi bir durum olmadığını taahhüt ederim.

Bu alışmadaki tüm bilgilerin, akademik ve etik kurallara uygun bir biçimde elde edildiğini; aynı zamanda bu kural ve davranışların gerektirdiğı gibi, bu alışmanın özünde olmayan tüm materyal ve sonuçları tam olarak aktardığımı ve referans gösterdiğimi beyan ederim. 16/01/2026

(İmza)

Berivan AVCI

ÖZET

MALATYA İLİ İÇİN TRAFİKTE KAZA ANALİZİ

Berivan AVCI

Yüksek Lisans Tezi

Erzincan Binali Yıldırım Üniversitesi, Fen Bilimleri Enstitüsü,

İnşaat Mühendisliği Anabilim Dalı

Danışman: Doç. Dr. Halim Ferit BAYATA

2026, 95

Bu çalışma, Malatya ili kent içi karayolu ulaşım sisteminde meydana gelen trafik kazalarının şiddet düzeyini belirlemeye yönelik veri odaklı bir analiz gerçekleştirmektedir. Amaç, kazaların neden-sonuç ilişkilerini ortaya koymak ve makine öğrenmesi yöntemleri aracılığıyla kazaların olası sonuçlarını önceden tahmin edebilmektir. Bu kapsamda geniş ölçekli kaza verileri kullanılmıştır. Veriler üzerinde öncelikle eksik gözlem temizleme, ölçekleme ve kategorik değişkenlerin dönüştürülmesi gibi veri işleme adımları uygulanmıştır. Ardından, veri dengesizliğini gidermek amacıyla SMOTE (Synthetic Minority Oversampling Technique) yöntemi kullanılmış ve veri seti eğitim (%80) ve test (%20) olarak ikiye ayrılmıştır. Modelleme sürecinde lojistik regresyon, karar ağacı, rastgele orman, SVC, adaboost ve XGBoost algoritmaları kullanılarak kaza şiddeti tahmin edilmiştir. Modeller, doğruluk (accuracy), kesinlik (precision), duyarlılık (recall), F1 skoru ve Cohen's Kappa katsayısı gibi performans ölçütleriyle değerlendirilmiştir. Elde edilen sonuçlar, lojistik regresyon modelinin daha dengeli ve genellenebilir bir performans sergilediğini, XGBoost'un ise karmaşık örüntüleri yakalamada başarılı olduğunu göstermektedir. Karar ağacı modeli yüksek uyum gücüne rağmen aşırı öğrenme eğilimi göstermiştir. Analiz sonuçları, sürücü yaşı ve davranışı, yolun yapısı, yaya davranışı ve kazanın meydana geldiği yıl gibi değişkenlerin kaza şiddeti üzerinde belirleyici etkiler yarattığını ortaya koymaktadır. Çalışma, kent içi trafik güvenliği politikalarının veri temelli olarak şekillendirilmesi açısından önemli bir örnek oluşturmaktadır.

Anahtar Kelimeler: Trafik kazası, kaza şiddeti, makine öğrenmesi, veri analizi, lojistik regresyon, karar ağacı, XGBoost, smote

ABSTRACT

TRAFFIC ACCIDENT ANALYSIS FOR MALATYA PROVINCE

Berivan AVCI

Master's Thesis

Erzincan Binali Yıldırım University, Institute of Science and Technology,

Department of Civil Engineering

Advisor: Assoc. Prof. Halim Ferit BAYATA

2026, 95

This study conducts a data-driven analysis to determine the severity levels of traffic accidents occurring within the urban road transportation system of Malatya province. The main objective is to reveal the cause-effect relationships of traffic accidents and to predict their possible outcomes using machine learning methods. Within this scope, large-scale accident data obtained from the General Directorate of Security (EGM) were utilized, and data preprocessing steps including missing observation removal, feature scaling, and transformation of categorical variables were applied. Subsequently, the Synthetic Minority Oversampling Technique (SMOTE) was employed to eliminate data imbalance, and the dataset was divided into training (80%) and testing (20%) sets. During the modeling stage, traffic accident severity was predicted using Logistic Regression, Decision Tree, Random Forest, Support Vector Classification (SVC), AdaBoost, and XGBoost algorithms. Model performance was evaluated using accuracy, precision, recall, F1-score, and Cohen's Kappa coefficient. The results indicate that the Logistic Regression model demonstrated a more balanced and generalizable performance, while XGBoost was more successful in capturing complex patterns in the dataset, and the Decision Tree model exhibited a tendency toward overfitting despite its high fitting capability. The analysis results reveal that variables such as driver age and behavior, road characteristics, pedestrian behavior, and the year in which the accident occurred have significant effects on accident severity.

Keywords: Traffic accident, accident severity, machine learning, data analysis, logistic regression, decision tree, XGBoost, smote

TEŞEKKÜR

Bu çalışmanın hazırlanmasında her aşamada bilgi ve deneyimleriyle yol gösteren, akademik gelişimime önemli katkılar sağlayan danışmanım Doç. Dr. Halim Ferit Bayata'ya en içten teşekkürlerimi sunarım. Akademik süreçte değerli görüşleriyle katkı sağlayan Dr. Öğr. Üyesi Fatih İrfan Baş ve değerli meslektaşım Fatih Ahmet Deniz'e teşekkür ederim. Çalışmaya sağladıkları veri ve katkılar nedeniyle Karayolları Genel Müdürlüğü ve Malatya'daki ilgili birimlere şükranlarımı sunarım. Son olarak, bu süreçte her zaman yanımda olan aileme ve arkadaşlarıma destek ve anlayışları için teşekkür ederim.

Berivan AVCI

Ocak, 2026

İÇİNDEKİLER

ÖZET	i
ABSTRACT.....	ii
TEŞEKKÜR.....	iii
İÇİNDEKİLER	iv
TABLolar DİZİNİ.....	vi
ŞEKİLLER DİZİNİ	vii
SİMGELER VE KISALTMALAR DİZİNİ	x
1. GİRİŞ.....	1
1.1. Araştırmanın Amacı.....	3
1.2. Araştırmanın Önemi	4
1.3. Makine Öğrenmesi Yöntemleri.....	5
2. KAVRAMSAL ÇERÇEVE VE İLGİLİ ÇALIŞMALAR	7
2.1 Veri Temelli Değerlendirme Girişi.....	8
2.1.1 Kazazedelerin Demografik Özellikleri	8
2.2. İlgili Çalışmalar (Literatür Özeti)	11
3. YÖNTEM	16
3.1 Dummies (Kukla Değişken) Yöntemi ile Sayısallaştırma	17
3.2 Sınıf Dengesizliğinin Giderilmesi ve Veri Bölme (SMOTE).....	18
3.3 Verilerin Ölçeklendirilmesi (StandardScaler)	21
3.4 Kullanılan Makine Öğrenmesi Modelleri	22
4. BULGULAR.....	25
4.1. Betimsel (Tanımlayıcı) İstatistikler	26
4.2. Model Performans Değerlendirmeleri	57
4.2.1 Logistik regresyon modelinin performans değerlendirmesi.....	57
4.2.2 XGBoost Modelinin Performans Değerlendirmesi	62
4.2.3 Decision Tree (Karar Ağacı) Modelinin Performans Değerlendirmesi	66
4.2.4 Random Forest (Rastgele Orman) Modeli.....	70
4.2.5 SVC Modeli Performans Değerlendirmesi.....	74
4.2.6 AdaBoost Modeli Performans Değerlendirmesi	79
4.2.7 Gradient Boosting Modeli Performans Değerlendirmesi.....	83
5. SONUÇ ve ÖNERİLER	88
6. KAYNAKÇA.....	92

TABLÖLAR DİZİNİ

Tablo 1. Trafik Veri Seti Özellikleri.....	6
Tablo 2. Yaya Kusuru Türlerine Göre Trafik Kazası Dağılımı	9
Tablo 3. Yaya Kusuruna Göre Dağılım	10
Tablo 4. Temizlenmiş Veri Setine Ait Tanımlayıcı İstatistikler	17
Tablo 5. Değişkenlere Ait Özet İstatistikler.....	26
Tablo 6. Sayısal Değişkenlere İlişkin Temel İstatistikler	27
Tablo 7. Malatya İli Trafik Kazası Verisinde Kategorik Değişkenlerin Özellikleri.....	54
Tablo 8. Malatya İli Trafik Kazası Verisinde Sayısal Değişkenlerin Tanımlayıcı İstatistikleri.....	56
Tablo 9. Lojistik Regresyon Modelinde Öne Çıkan Özellikler ve Katsayı Yorumları	58
Tablo 10. Modellerin Eğitim ve Test Seti Performanslarının Karşılaştırması	90

ŞEKİLLER DİZİNİ

Şekil 1. Kaza-Yaya Kusuruna Göre Dağılım.....	8
Şekil 2. Kaza-Yaya Kusuruna Göre Dağılım.....	9
Şekil 3. Kaza-Yaya Kusuruna Göre Dağılım.....	10
Şekil 4. Seçilmiş Değişkenlere Ait Violin Plot Görselleştirmeleri	20
Şekil 5. Kazaların Yıllara Göre Dağılımı	29
Şekil 6. Kazaların Günün Saat Dilimine Göre Dağılımı.....	30
Şekil 7. Kazaların İlçelere Göre Dağılımı.....	32
Şekil 8. Kazaların Toplam Ölü Sürücü Sayısına Göre Dağılımı	33
Şekil 9. Kazaların Toplam Ölü Yolcu Sayısına Göre Dağılımı.....	34
Şekil 10. Kazaların Toplam Ölü Yaya Sayısına Göre Dağılımı	35
Şekil 11. Kazaların Toplam Yaralı Sürücü Sayısına Göre Dağılımı	37
Şekil 12. Kazaların Toplam Yaralı Yolcu Sayısına Göre Dağılımı	38
Şekil 13. Kazaların Toplam Yaralı Yaya Sayısına Göre Dağılımı	39
Şekil 14. Kazaların Yıllara Göre Dağılımı	40
Şekil 15. Kazaların Günün Saat Dilimine Göre Dağılımı.....	41
Şekil 16. Kazaların İlçelere Göre Dağılımı.....	42
Şekil 17. Kazaların Yol Şerit Tipine Göre Dağılımı.....	43
Şekil 18. Kazaların Yol Kaplama Türlerine Göre Dağılımı	44
Şekil 19. Kazaların Yol Sınıfına Göre Dağılımı	45
Şekil 20. Kaza Yapılan Yolların Yasal Hız Limitine Göre Dağılımı	46
Şekil 21. Kazaların GeoYatay Özelliklerine Göre Dağılımı.....	47
Şekil 22. Kaza Yapılan Yolların Düşey Eğimine Göre Dağılımı	48
Şekil 23. Kaza Kavşak Tipi Dağılımı	49
Şekil 24. Kaza Geçit Dağılımı	50
Şekil 25. Kaza Gün durumu Dağılımı	51
Şekil 26. Kaza Hava Durumu Dağılımı	52
Şekil 27. Kaza Yol Yüzeyi Dağılımı	53
Şekil 28. Kazaların Kaza Ay Değişkenine Göre Dağılımı.....	54
Şekil 29. Kaza Şiddeti Değişkeninin Dağılımı	57
Şekil 30. Logistik Regresyon Modeli Sınıflandırma Raporu (Classification Report Heatmap)	59
Şekil 31. Logistik Regresyon Modeli Karmaşıklık Matrisi (Confusion Matrix).....	60
Şekil 32. Logistik Regresyon Modeline Ait ROC Eğrileri	61
Şekil 33. XGBoost Modeli Karmaşıklık Matrisi	63

Şekil 34. XGBoost Modeli Sınıflandırma Raporu (Classification Report Heatmap).....	64
Şekil 35. XGBoost Modeline Ait ROC Eğrileri	65
Şekil 36. XGBoost Öğrenme Eğrisi (Learning Curve).....	66
Şekil 37. Decision Tree Sınıflandırma Raporu (Karar Ağacı)	68
Şekil 38. Decision Tree Karmaşıklık (Confusion) Matrisi	69
Şekil 39. Decision Tree ROC Eğrisi	70
Şekil 40. Random Forest Öğrenme Eğrisi	71
Şekil 41. Random Forest Sınıflandırma Raporu	72
Şekil 42. Random Forest Karmaşıklık Matrisi (Confusion Matrix)	73
Şekil 43. Random Forest ROC Eğrisi (ROC Curve)	74
Şekil 44. SVC Modeli Öğrenme Eğrisi	75
Şekil 45. SVC Modeli Sınıflandırma Raporu	76
Şekil 46. SVC Modeli Sınıflandırma Raporu ve Karmaşıklık Matrisi	77
Şekil 47. SVC Modeli ROC Eğrisi	78
Şekil 48. AdaBoost Modeli Öğrenme Eğrisi	79
Şekil 49. AdaBoost Modeli Sınıflandırma Raporu	80
Şekil 50. AdaBoost Modeli Karmaşıklık Matrisi	81
Şekil 51. AdaBoost Modeli ROC Eğrisi	82
Şekil 52. Gradient Boosting Modeli Öğrenme Eğrisi	84
Şekil 53. Gradient Boosting Sınıflandırma Raporu (Classification Report).....	85
Şekil 54. Gradient Boosting Karmaşıklık Matrisi (Confusion Matrix)	86
Şekil 55. Gradient Boosting ROC Eğrileri (ROC Curves)	87

SİMGELER VE KISALTMALAR DİZİNİ

%	Yüzde
$\bar{x}$	Aritmetik ortalama
χ^2	Ki-kare testi
AdaBoost	Adaptive Boosting
DT	Decision Tree (Karar Ağacı)
f	Frekans
F1	F1-Score
LR	Logistic Regression (Lojistik Regresyon)
p	Anlamlılık değeri
sd	Serbestlik derecesi
SMOTE	Synthetic Minority Oversampling Technique
Ss	Standart sapma
SVC	Support Vector Classifier (Destek Vektör Sınıflandırıcı)
TÜİK	Türkiye İstatistik Kurumu
XGBoost	Extreme Gradient Boosting

1. GİRİŞ

Günümüzde trafik kazaları, artan araç sayısı ve şehirleşmenin hızlı ilerlemesi ile birlikte önemli bir halk sağlığı ve güvenlik problemi haline gelmiştir. Türkiye genelinde olduğu gibi Malatya ilinde de trafik kazalarının sayısı ve etkisi kayda değerdir. Özellikle 2024 yılı Temmuz ayı trafik istatistiklerine göre Malatya, 1409 ölüm ve yaralanmayla sonuçlanan trafik kazası ile Doğu Anadolu Bölgesi'nin en fazla kazanın meydana geldiği ili olmuştur (TÜİK, 2024). Bu durum, hem can kayıplarına hem de ekonomik zararların artmasına neden olmakta, yerel yönetimlerin trafik güvenliğini sağlamak adına veri temelli çözümler geliştirmesini zorunlu kılmaktadır.

Malatya, Doğu Anadolu Bölgesi'nin önemli karayolu bağlantılarına ev sahipliği yapmakta ve kent içi ile çevre iller arası ulaşımında stratejik bir konuma sahiptir. Ancak, hız ihlalleri, yol yapısındaki yetersizlikler ve hava koşullarının olumsuz etkileri bölgede trafik kazalarının başlıca nedenleri arasında yer almaktadır. Bu çok değişkenli yapıya sahip verilerin klasik analiz yöntemleriyle incelenmesi yetersiz kalabilmekte, dolayısıyla kazaların altında yatan karmaşık örüntülerin ortaya çıkarılması için ileri tekniklere ihtiyaç duyulmaktadır. Bu bağlamda, makine öğrenmesi algoritmaları, trafik kazalarının nedenleri ve olası risk faktörlerinin belirlenmesinde etkili araçlar olarak ön plana çıkmaktadır. Sürücü davranışları, çevresel koşullar ve yol özellikleri gibi değişkenlerin analizi, kazaların önlenmesi için stratejik politika oluşturulmasına katkı sağlayabilir. Ayrıca, coğrafi bilgi sistemleri yardımıyla kaza noktalarının mekânsal dağılımının haritalanması, yüksek riskli bölgelerin belirlenmesini mümkün kılmaktadır. Kentsel alanlarda trafik kazalarının oluşumu, sürücü davranışları, yol geometrisi, trafik hacmi ve çevresel koşulların birlikte etkili olduğu karmaşık bir yapı sergilemektedir. Türkiye genelinde yapılan mekânsal analizler, kazaların özellikle kent merkezlerinde yoğunlaştığını ve belirli bölgelerde kümelenme eğilimi gösterdiğini ortaya koymaktadır. Erdoğan ve Gökdağ (2019), kentsel alanlarda trafik kazalarının mekânsal dağılımının rastlantısal olmadığını; nüfus, ulaşım yoğunluğu ve yol altyapısının kaza riskini belirgin biçimde etkilediğini göstermiştir. Bu bulgular, Malatya ili özelinde yapılan analizlerde gözlenen merkez ilçelerdeki kaza yoğunluklarıyla örtüşmektedir. Trafik kazalarına ilişkin veriler çok sayıda değişken içermekte ve bu değişkenler arasında doğrusal olmayan ilişkiler bulunmaktadır. Bu nedenle, klasik istatistiksel yöntemler kazaların altında yatan karmaşık örüntüleri açıklamakta yetersiz kalabilmektedir. Fang ve arkadaşları (2018), trafik ihlali ve kaza verilerinin birlikte kullanıldığı makine öğrenmesi modellerinin, sürücü risk seviyelerinin belirlenmesinde yüksek doğruluk sağladığını ortaya koymuştur. Bu durum,

trafik güvenliđi politikalarının veri temelli ve öngörücü yaklaşımlarla desteklenebileceđini göstermektedir. Öte yandan, trafik hacmi ile kaza sayısı arasındaki ilişkiyi inceleyen çok deđişkenli analizler, özellikle kavşak noktalarında kaza riskinin arttığını göstermektedir. Golob ve arkadaşları (1999), trafik yaklaşım hacimlerinin kaza sayıları üzerinde belirleyici bir etkisi olduğunu ortaya koyarak, ulaşım planlamasında hacim temelli risk analizlerinin önemini vurgulamıştır. Bu bulgular, Malatya kent içi ulaşım sisteminde yoğun kavşak ve ana arterlerin kaza açısından kritik noktalar olduğunu desteklemektedir.

Araştırma kapsamında, Malatya iline ait kaza verileri Python ortamında işlenmiş, eksik deđerler temizlenmiş, deđerşkenler kodlanmış ve ardından Synthetic Minority Oversampling Technique (SMOTE) yöntemi ile veri dengesizliđi giderilmiştir. Elde edilen dengeli veri seti üzerinde Lojistik Regresyon (LR), Karar Ağacı (Decision Tree - DT), Rastgele Orman (Random Forest - RF) ve Aşırı Gradyan Yükseltme (Extreme Gradient Boosting - XGBoost) modelleri uygulanarak performans karşılaştırması yapılmıştır. Elde edilen sonuçlar, klasik istatistiksel yaklaşımlara kıyasla makine öğrenmesi modellerinin kaza şiddetini tahmin etmede daha başarılı olduğunu göstermiştir. Ayrıca, özellik önem sıralaması analizleriyle kazaların oluşumunda en etkili faktörlerin yol tipi, araç sayısı, kaza türü ve kaza sonucu deđerşkenleri olduğu belirlenmiştir.

1.1. Araştırmanın Amacı

Bu araştırmanın temel amacı, Malatya ilinde meydana gelen trafik kazalarını geçmiş kaza verileri ışığında makine öğrenmesi yöntemleri kullanarak analiz etmek, kazaların oluşumunda etkili olan faktörleri belirlemek ve bu verilere dayalı güvenilir bir kaza tahmin modeli geliştirmektir. Trafik kazalarının nedenlerini anlamak ve olası risk alanlarını belirlemek, hem bölgesel ulaşım güvenliğinin artırılması hem de yerel yönetimlerin veri temelli karar mekanizmalarını güçlendirmek açısından kritik öneme sahiptir. Araştırmada, 2013-2023 yılları arasını kapsayan trafik kaza verileri kullanılmıştır. Veri seti üzerinde veri temizleme, ön işleme, dengesizlik giderme (SMOTE) ve makine öğrenmesi modelleri uygulanmıştır. Bu yöntemlerle, kaza şiddetini etkileyen değişkenlerin belirlenmesi ve olası kaza risklerinin tahmin edilmesi amaçlanmıştır. Çalışma, sadece istatistiksel tanımlamalarla sınırlı kalmayıp, makine öğrenmesi tabanlı sınıflandırma yöntemleriyle kaza şiddeti tahmin modelleri geliştirmekte; böylece trafik güvenliği alanına uygulamalı ve öngörücü bir katkı sunmaktadır. Elde edilen sonuçların, Malatya’da trafik güvenliğini artırmaya yönelik politika ve uygulamalara karar desteği sağlaması ve benzer kentlerde yapılacak çalışmalara örnek model oluşturması hedeflenmektedir.

Literatürde, makine öğrenmesi tekniklerinin trafik kazalarını analiz etmede etkin olduğu pek çok çalışmada vurgulanmıştır. Örneğin, Bayata ve Hattatoğlu (2011) tarafından Erzincan ilinde yapılan analizde, yapay sinir ağları ve istatistiksel yöntemlerin kıyaslanması sonucunda makine öğrenmesi modellerinin daha doğru sonuçlar verdiği gözlemlenmiştir. Benzer şekilde, Mutlu ve Alver (2014) Türkiye’de genç sürücülerin kaza eğilimlerini analiz etmiş ve makine öğrenmesi tabanlı modellerin sürücü davranışlarının risk tahmininde yüksek doğruluk sağladığını raporlamıştır. Bu çalışmalar, Malatya özelinde uygulanacak yöntemlerin güvenilirliğini ve akademik değerini desteklemektedir.

Türkiye genelinde 2023 yılında toplam 1.314.136 trafik kazasının meydana geldiği belirlenmiş ve bu kazaların 235.071’i ölümlü-yaralanmalı olarak kayıtlara geçmiştir (TÜİK, 2024). Malatya özelinde ise 2024 yılı Temmuz verilerine göre 1.409 ölümlü-yaralanmalı

trafik kazası gerçekleşmiş ve ilin Doğu Anadolu Bölgesi'nde en fazla kazanın meydana geldiği il konumunda olduğu görülmüştür. Bu veriler, araştırmanın bölgesel önceliğini ve önemini açıkça ortaya koymaktadır.

Bu çalışma ile elde edilmesi hedeflenen çıktılar Malatya'daki trafik kazalarına yol açan başlıca risk faktörlerinin belirlenmesi, karar vericiler için uygulanabilir ve veri odaklı bir kaza tahmin modeli oluşturulması, trafik güvenliğini artırmaya yönelik stratejik politika önerilerinin sunulması şeklinde verilebilir. Sonuç olarak, bu tez çalışması, hem akademik literatüre katkı sağlayacak hem de Malatya ilinde trafik kazalarının önlenmesine yönelik bilimsel ve veri destekli çözüm önerilerinin geliştirilmesine katkıda bulunacaktır.

1.2. Araştırmanın Önemi

Trafik kazaları, dünya genelinde her yıl yaklaşık 1,19 milyon kişinin ölümüne ve milyonlarca kişinin yaralanmasına neden olan ciddi bir küresel halk sağlığı sorunu olarak kabul edilmektedir (Behboudi, Moosavi & Ramnath, 2024). Türkiye'de de trafik kazaları, hem can kayıpları hem de ekonomik kayıplar açısından önemli bir toplumsal problem olmaya devam etmektedir. Özellikle Malatya ili, Doğu Anadolu Bölgesi'nde trafik kazalarının en yoğun yaşandığı illerden biridir. 2024 yılı verilerine göre kentte 11.006 trafik kazası meydana gelmiş, bu kazalarda 94 kişi yaşamını yitirmiştir (Malatya Lider Haber, 2025). Bölgedeki yüksek kaza oranları, Malatya'nın ulaştırma altyapısında ve sürücü davranışlarında ciddi risk unsurlarının bulunduğunu göstermektedir (Malatya Star, 2024). Bu bulgular, kaza nedenlerinin yalnızca sayısal olarak değil, veri temelli bir yaklaşımla analiz edilmesi gerekliliğini ortaya koymaktadır.

Son yıllarda, kara nokta (black spot) analizi gibi trafik güvenliği uygulamalarında makine öğrenmesi ve derin öğrenme tekniklerinin kullanımı giderek artmaktadır; bu yaklaşımlar, kaza verilerinin mekânsal ve zamansal örüntülerini değerlendirmede ve yüksek kaza yoğunluklu bölgelerin belirlenmesinde önemli bir araç olarak ele alınmıştır (Karamanlis et al., 2023). Bu tez çalışması, Malatya ili özelinde trafik kazalarının makine öğrenmesi yöntemleriyle analizini yaparak, kaza şiddetini etkileyen faktörleri belirlemeyi ve yerel yönetimlere veri odaklı politika önerileri sunmayı amaçlamaktadır. Çalışmanın önemi, yalnızca akademik düzeyde bilgi üretmekle sınırlı olmayıp, yerel ve bölgesel trafik güvenliği stratejilerine doğrudan katkı sağlayabilecek nitelikte olmasından kaynaklanmaktadır. Elde edilen sonuçların, Malatya'da trafik güvenliğinin artırılmasına yönelik alınacak önlemlerde bilimsel dayanak oluşturması, aynı zamanda benzer kentlerde yapılacak uygulamalı çalışmalar için

örnek bir model teşkil etmesi beklenmektedir.

Bu araştırma, hem bilimsel literatüre hem de yerel yönetimlerin trafik güvenliği uygulamalarına çok yönlü katkılar sağlamayı hedeflemektedir. Çalışmanın katkıları arasında Türkiye genelinde ve özel olarak Doğu Anadolu Bölgesinde yer alan illerde, makine öğrenmesi yöntemleri kullanılarak gerçekleştirilen sınırlı sayıdaki trafik kazası analizlerinden biri olma niteliği taşıması gösterilebilir. Bu yönüyle literatüre yerel ölçekte veri madenciliği ve yapay zekâ tabanlı analiz örneği kazandırmaktadır. Araştırmada kullanılan veri temizleme, ön işleme ve SMOTE ile veri dengeleme yaklaşımları, trafik kazası verilerinin daha sağlıklı şekilde analiz edilmesine olanak tanımaktadır. Ayrıca Logistic Regression, Decision Tree, Random Forest ve XGBoost modellerinin karşılaştırmalı kullanımı, farklı algoritmaların trafik verileri üzerindeki performansını ortaya koymaktadır. Elde edilen sonuçlar, Malatya’da trafik kazalarının yoğunlaştığı dönem, bölge ve koşullara ilişkin somut bulgular sağlamaktadır. Bu bulgular, yerel yönetimlerin kaza önleme stratejileri geliştirmesinde ve trafik güvenliği politikalarının veri temelli biçimde şekillendirilmesinde kullanılabilir. Çalışma, trafik kazalarının nedenlerinin anlaşılmasına katkıda bulunarak, can kayıplarının azaltılması ve trafik bilincinin artırılması açısından toplumsal farkındalık yaratmayı hedeflemektedir.

1.3. Makine Öğrenmesi Yöntemleri

Makine öğrenmesi; yapay zekâ, veri madenciliği ve istatistik gibi alanlarla ilgilenen, bilgisayarların açıkça programlanmadan öğrenme yeteneği kazanmasını sağlayan bir yöntemdir. Kavram ilk kez 1950’li yıllarda yapay zekâ ve yapay sinir ağları araştırmaları kapsamında gündeme gelmiştir. Makine öğrenmesi, çevresinden aldığı verilerle insan zekâsını taklit eden, sürekli gelişen akıllı hesaplama sistemleri oluşturmayı amaçlar.

Makine öğrenmesi yöntemleri genel olarak denetimli, denetimsiz, yarı denetimli ve pekiştirmeli öğrenme olmak üzere dört ana başlıkta incelenmektedir:

- Denetimli öğrenme, giriş ve çıkış verilerinin etiketli olarak sisteme tanıtıldığı, modelin bu örneklerden yola çıkarak sınıflandırma veya tahmin yaptığı bir yöntemdir. Bu grupta Naive Bayes (NB), Destek Vektör Makineleri (SVM), Karar Ağaçları (DT), Lojistik Regresyon (LR), AdaBoost (Adaptive Boosting), Rastgele Ormanlar (RF) ve Yapay Sinir Ağları (ANN) gibi algoritmalar yer almaktadır.
- Denetimsiz öğrenme ise etiketlenmemiş verilerle çalışır. Sistem yalnızca giriş verileriyle işlem yapar ve bu veriler arasındaki örüntüleri veya kümeleri kendisi bulmaya çalışır. Bu yöntemde kümeleme, ilişki analizi, boyut indirgeme gibi teknikler kullanılmaktadır.

- Yarı denetimli öğrenme, denetimli ve denetimsiz yöntemlerin bir arada kullanıldığı hibrit bir yaklaşımdır. Etiketli ve etiketsiz veriler aynı anda değerlendirilerek modelin sınırlı sayıda eğitim verisinden öğrenmesi sağlanır.

Trafik veri setinde yer alan öznitelikler, literatürdeki benzer çalışmalardan (özellikle Bingöl, 2022) yararlanılarak oluşturulmuş ve Malatya ili koşullarına uyarlanmıştır. Söz konusu öznitelikler Tablo 1’de sunulmaktadır.

Tablo 1 Trafik veri seti özellikleri

Öznitelik Adı	Değişken Tipi	Alt Nitelikler
Yıl	Kategorik	2024
Hafta Günü	Kategorik	Pazartesi, Salı, Çarşamba, Perşembe, Cuma, Cumartesi, Pazar
Kaza Saati Dilimi	Kategorik	0-4, 4-8, 8-12, 12-16, 16-20, 20-24
Ay	Kategorik	Ocak, Şubat, Mart, Nisan, Mayıs, Haziran, Temmuz
Kazanın Yaşandığı Bölge	Kategorik	• Battalgazi • Kale • Yeşilyurt • D300 Karayolu bağlantısı olan özellikle Malatya-Elazığ güzergâhı • Malatya il çevre yolları, şehir çıkış bölgeleri
Kaza Yerleşim Yeri	Kategorik	İç, Dış

2. KAVRAMSAL ÇERÇEVE VE İLGİLİ ÇALIŞMALAR

Kent içi ulaşım sistemleri, toplumların ekonomik, sosyal ve kültürel yaşamında doğrudan etkili olan en temel altyapı bileşenlerinden biridir. Ulaşım sistemlerinin planlı, güvenli ve sürdürülebilir biçimde işletilmesi, kentlerin yaşanabilirliğini belirleyen başlıca unsurlar arasında yer alır. Özellikle kentleşme sürecinin hızla arttığı ülkelerde, motorlu araç sayısındaki yükseliş ve yetersiz altyapı planlaması, trafik güvenliği sorunlarını önemli ölçüde artırmaktadır. Bu bağlamda, trafik kazaları yalnızca bireysel hatalardan kaynaklanan olaylar değil, aynı zamanda ulaşım sisteminin bütünsel performansını etkileyen sosyo-teknik olgulardır.

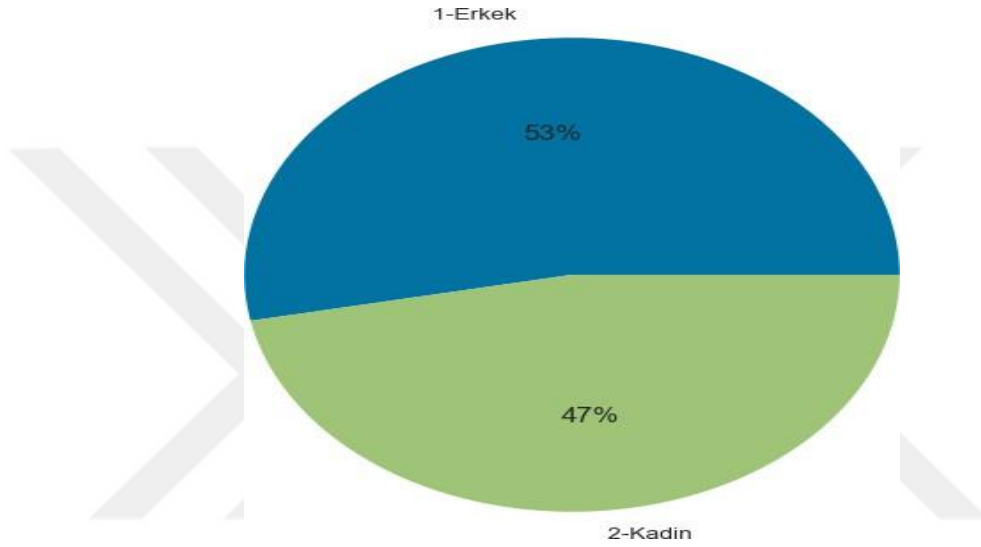
Trafik kazası kavramı, genellikle karayolu üzerinde en az bir motorlu aracın karıştığı, ölüm, yaralanma veya maddi hasarla sonuçlanan olaylar olarak tanımlanmaktadır. Kazaların oluşumunda sürücü davranışları, yol ve çevre koşulları, araç teknik durumu ve yaya hareketleri gibi faktörler etkili olmaktadır. Dünya Sağlık Örgütü'ne (WHO) göre trafik kazaları, özellikle gelişmekte olan ülkelerde genç nüfusun ölüm nedenleri arasında ilk sıralarda yer almakta; bu durum, ulaşım planlamasının yalnızca kapasiteye değil, aynı zamanda güvenliğe de odaklanmasını zorunlu kılmaktadır.

Kent içi trafik kazalarının mekânsal dağılımı incelendiğinde, genellikle yoğun nüfuslu, ticari faaliyetlerin ve yaya hareketliliğinin yüksek olduğu bölgelerde kazaların daha sık yaşandığı görülmektedir. Bu nedenle, kent ölçeğinde yürütülen trafik güvenliği analizlerinde, kazaların zaman, konum, araç tipi ve sürücü özellikleri gibi çok boyutlu değişkenlerle değerlendirilmesi gerekmektedir. Ulaşım sistemlerinin doğru modellenmesi ve kaza yoğunluklarının belirlenebilmesi için istatistiksel analizler ve makine öğrenmesi tabanlı yöntemler günümüzde giderek daha fazla kullanılmaktadır. Bu yöntemler, geçmiş kaza verilerinden hareketle riskli bölgelerin belirlenmesine ve olası kazaların önceden tahmin edilmesine olanak tanımaktadır. Türkiye'de trafik güvenliği üzerine yapılan akademik çalışmalar, genellikle kazaların nedenlerini ve sonuçlarını bölgesel ölçekte incelemekte; bu araştırmalarda sürücü cinsiyeti, yaş grubu, araç tipi ve kaza zamanı gibi değişkenlerin kazaların oluşumuna etkisi detaylandırılmaktadır. Bu kapsamda yapılan çalışmalar, erkek sürücülerin kadınlara oranla daha fazla kazaya karıştığını, bunun temel nedenleri arasında risk alma eğilimi, uzun sürüş süreleri ve araç kullanım yoğunluğu gibi davranışsal faktörlerin bulunduğunu ortaya koymaktadır.

2.1 Veri Temelli Değerlendirme Girişi

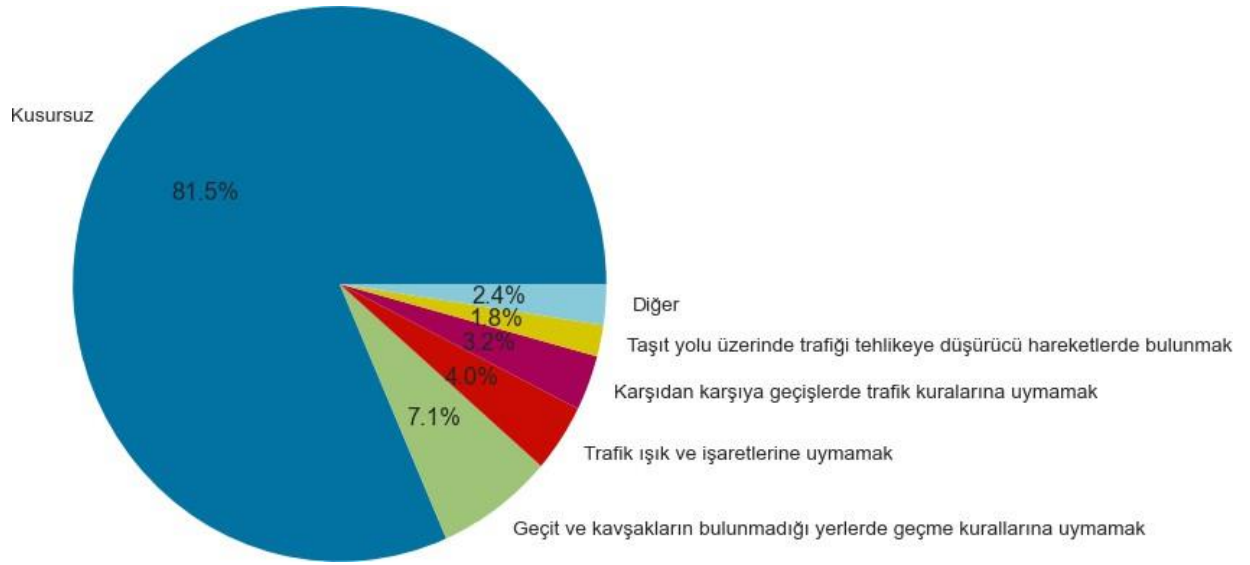
2.1.1 Kazazedelerin demografik özellikleri

Aşağıda yer alan grafikte, kazazedelerin cinsiyetlerine göre dağılımı görülmektedir. Veriye göre kazaların %53'ü erkek, %47'si kadın kazazedelerden oluşmaktadır. Bu durum, erkeklerin yaya veya yolcu olarak kazalara daha fazla karıştığını göstermektedir. Yani kazalarda cinsiyet açısından belirgin bir dengesizlik vardır; erkek kazazedeler hafif bir üstünlük sağlamaktadır.(Şekil 2)



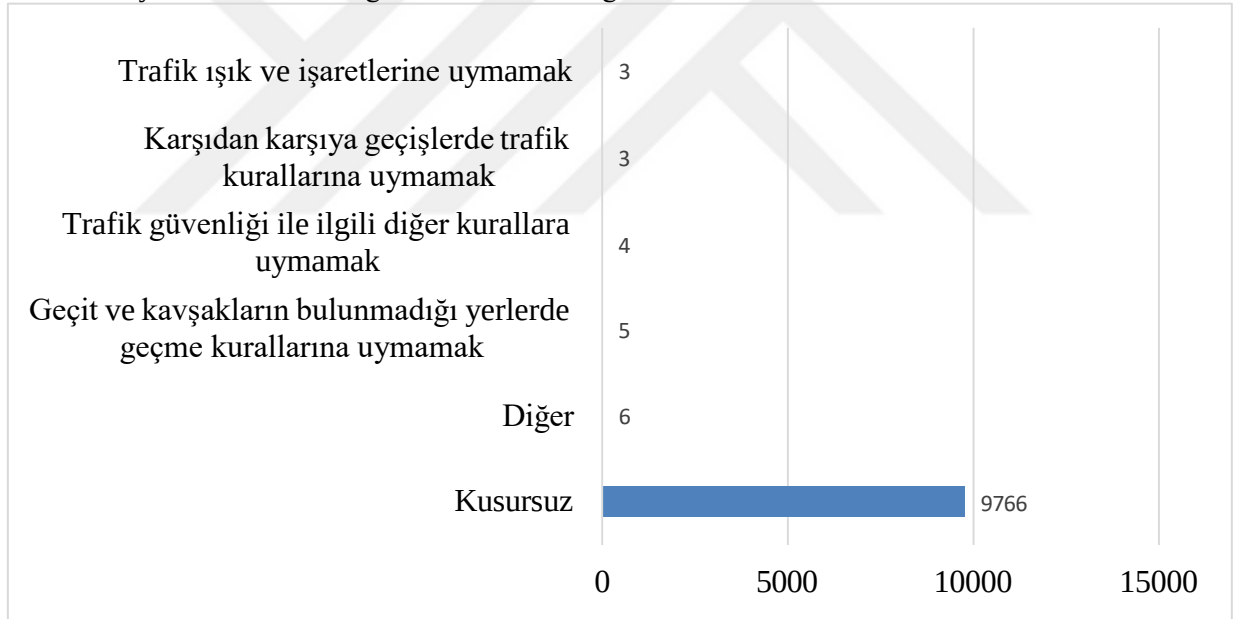
Şekil 1 Kaza-Yaya Kusuruna Göre Dağılım

Kazalarda yayalar genel olarak kusursuzdur; buna rağmen bazı durumlarda kusurlu oldukları görülmektedir. Kusur, çoğunlukla yayaların geçit veya kavşak bulunmayan yerlerden geçmeye çalışmaları, trafik ışıkları ve işaretlerine uymamaları veya karşıdan karşıya geçiş kurallarını ihlal etmelerinden kaynaklanmaktadır. Bu durum, yaya kusur oranlarının özellikle uygun altyapının bulunmadığı alanlarda arttığını göstermektedir (Şekil 3).



Şekil 2 Kaza-Yaya Kusuruna Göre Dağılım

Tablo 2 Yaya kusuru türlerine göre trafik kazası dağılımı



Kazalarda yayaaların ve sürücülerin kusur dağılımı analiz edildiğinde, kazaların büyük çoğunluğunun kusursuz olduğu görülmektedir. Bununla birlikte, bazı kazalarda güvenlik önlemlerine uyulmaması önemli bir risk faktörü olarak öne çıkmaktadır. En sık rastlanan kusurlar arasında trafik kurallarına uymamak, araçlara kontrolsüz şekilde inip binmek, kask kullanmamak ve açık yük üzerinde düzensiz ve tehlikeli şekilde seyahat etmek yer almaktadır. Daha nadir görülen kusurlar ise emniyet kemeri takmamak, taşıttan habersiz inip binmek veya

taşıt içinde düzensiz hareket etmek gibi davranışları kapsamaktadır. Bu bulgular, güvenlik önlemlerine uyulmasının kazaların önlenmesinde kritik bir rol oynadığını göstermekte olup, **Tablo 3 ve Şekil 4’teki verilerle desteklenmektedir.**



Şekil 3 Kaza-Yaya Kusuruna Göre Dağılım

Tablo 3 Yaya Kusuruna Göre Dağılım

Kusur Türü	Sayı
Kusuruz	7825
Diğer	720
Trafik güvenliği ile ilgili diğer kurallara uymamak	558
Araçlara kontrolsüz şekilde inmek ve binmek	207
Kask kullanmamak	176
Açık yük üzerinde gayri nizami ve tehlikeli şekilde seyahat etmek	134
Taşıt dışında (çamurluk vs.) gayri nizami seyahat etmek	78
Gerekli tedbirleri almadan taşıttan inmek	37
Hareket halindeki taşıtlardan yere atlamak	27
Emniyet kemeri takmamak	19
Taşıtlarda kask veya emniyet kemeri kullanmamak	3
Habersiz taşıta binmek ya da taşıttan inmek	2
Taşıt içinde gayri nizami hareket etmek	1

Toplam 9.543 yaya kazası incelendiğinde, bu kazalarda 244 yaya ölümü meydana gelmiştir. Bu da kazaların yaklaşık %2’sinde yaya ölümlerinin gerçekleştiğini göstermektedir.

2.2. İlgili Çalışmalar (Literatür Özeti)

Trafik kazaları ve kent içi ulaşım sistemleri, şehirlerin güvenliği ve sürdürülebilir ulaşım planlaması açısından kritik bir öneme sahiptir. Son yıllarda artan araç ve yaya yoğunluğu, kazaların sayısını ve şiddetini etkileyen temel faktörlerden biri olarak öne çıkmaktadır. Bu nedenle, kazaların nedenlerini, dağılımını ve özellikle yaya ve sürücü kusurlarını anlamak, etkili önleyici tedbirlerin geliştirilmesi açısından büyük önem taşımaktadır. Mevcut literatürde, yaya kazaları, sürücü davranışları ve trafik güvenliği önlemleri üzerine çok sayıda araştırma yapılmıştır. Bu çalışmalar, kazaların çoğunlukla kusursuz veya belirli davranış ihlallerine bağlı olarak gerçekleştiğini, açık yük taşımaları, kask ve emniyet kemeri kullanımı gibi güvenlik önlemlerine uyulmasının kazaların şiddetini önemli ölçüde etkilediğini göstermektedir. Bu tezde, Malatya ilinde meydana gelen trafik kazaları veri seti kullanılarak yaya ve sürücü kusurlarının dağılımı analiz edilmiş ve literatürdeki bulgularla karşılaştırılmıştır. Böylece, hem yerel ölçekli bir değerlendirme yapılmış hem de mevcut literatüre katkı sağlanması hedeflenmiştir.

Gültaş ve Yücel (2015)'in çalışması, yeşil lojistik ve sürdürülebilir ulaşım bağlamında Malatya Büyükşehir Belediyesi tarafından işletilen trambüs sistemini incelemiştir. Araştırma, çevre sorunları ve alternatif ulaşım sistemleri üzerine literatür taraması yaparak trambüsün çevresel, ekonomik ve işletme özelliklerini değerlendirmiştir. Çalışma sonucunda trambüsün çevre dostu, ekonomik ve uzun ömürlü bir toplu taşıma alternatifi olduğu; ancak sağ şeritin tamamen işgal edilmesi ve manevra kabiliyetinin sınırlı olması gibi operasyonel dezavantajlara sahip olduğu belirlenmiştir. Bu değerlendirmeler, toplu taşımacılığın sürdürülebilirlik boyutuna yönelik politika önerileri ile desteklenmiştir.

Öztemiz ve Karıcı (2021) çalışmasında, Malatya kent merkezi içindeki kavşak noktalarının ulaşım sistemi içinde konumsal önemleri değerlendirilmiştir. Araştırmada, 151 kavşak noktası ve bu noktaları birbirine bağlayan 258 yol kesitinden oluşan bir ağırlıklı grafik modeli oluşturulmuştur. Bu model üzerinde PageRank, özvektör merkeziliği, yakınlık merkeziliği ve aradalık merkeziliği gibi graf teorisi temelli merkezilik ölçütleri uygulanmıştır. Analizler, kavşakların trafik akışı üzerindeki rolünü nicel olarak ortaya koymuş ve bazı kavşak noktalarının kent içi trafiğin yönlendirilmesi açısından yüksek öneme sahip olduğu sonucuna ulaşmıştır. Çalışmada analizler R programlama dili ile gerçekleştirilmiştir.

Karabaş, Geçkil ve İnce (2022)'nin çalışmasında, Malatya ilinde işletmeye alınan trambüs sisteminin seçimi ve uygulama koşulları kapsamlı şekilde değerlendirilmiştir. Araştırma, 2013–2015 yılları arasına ait trambüs verileri kullanılarak güzergâh planlaması, inşa süreci ve ekonomik performans analizini gerçekleştirmiştir. Bulgular, trambüs sisteminin diğer toplu taşıma türlerine kıyasla düşük ilk yatırım maliyetine sahip olduğunu ve sürdürülebilir ulaşım hedefleri açısından avantaj sağladığını göstermiştir. Ayrıca, çalışma trambüsün kısa ve uzun vadede kent içi toplu taşıma hizmetlerine katkı sağlama potansiyeline sahip olduğunu ve başka şehirler için örnek uygulama modeli oluşturabileceğini ortaya koymuştur (Karabaş, Geçkil & İnce, 2022).

Öztemiz (2021) tarafından yürütülen bu araştırma, Malatya kent merkezindeki 40 akıllı kavşak ve bu kavşaklara ait araç sayım ve sinyalizasyon verilerini kullanarak kavşakların baskınlık düzeylerini ve sinyalizasyon sistemleri üzerindeki yüklerini incelemiştir. Çalışmada, yol kollarına giriş yapan araç sayıları ile yeşil ışık yanma süreleri dahil edilerek bir ağırlıklı ulaşım ağı çizgesi oluşturulmuştur. Bu çizge üzerinde özvektör merkezlilik ve sayfa değeri merkezlilik algoritmaları uygulanmış ve Sanayi, Emeksiz ile İstasyon kavşaklarının en yüksek merkezlilik değerlerine sahip olduğu belirlenmiştir. Bu sonuçlar, kavşakların mevcut sinyalizasyon planı altında yüksek etkinliğe sahip olduklarını göstermiştir.

Kaya (2014)'nın yüksek lisans tez çalışması, Malatya merkez bölgesindeki kentiçi toplu taşıma otobüs hatlarının iyileştirilmesine odaklanmıştır. Tezde, özellikle ana ulaşım aksı olan İnönü Caddesi üzerindeki otobüs hatlarının performansı, güzergah yapısı ve trafik yükü dikkate alınarak değerlendirilmiştir. Araştırma kapsamında mevcut toplu taşıma hatlarının verimliliği incelenmiş, özel araç kullanımının azaltılması ve ulaşım kalitesinin artırılması için nazım plan ve ulaşım ana planı ile uyumlu otopark alanlarının planlanması gerektiği önerilmiştir. Çalışma, kent içi ulaşımında koordinasyon ve planlamanın toplu taşımada etkinlik ve kaliteyi artırmada kritik olduğunu göstermektedir.

Manikandan ve Anandan (2018) tarafından geliştirilen çalışma, trafik kazası risklerini derin öğrenme temelli model ile izlemeyi amaçlamıştır. Araştırmada stokastik gradyan inişi (SGD) algoritmasıyla eğitilen model, GPS verileriyle entegre edilerek trafik yoğunluğu ve kaza riski tahmin edilmiştir. Hindistan’da 2012–2016 dönemine ait yaklaşık 2,4 milyon trafik kazası verisi kullanılmıştır. Sonuçlar, büyük hacimli ve heterojen veri setleri üzerinde derin öğrenme modellerinin etkinliğini ortaya koymuş; modelin sürücülerin en az yoğun rotaları seçmesine ve trafik tıkanıklıklarının azaltılmasına katkı sağladığı belirlenmiştir. Bu çalışma, yapay zekâ tabanlı sistemlerin trafik güvenliği izleme süreçlerinde uygulanabilirliğini göstermektedir.

Çelik ve Seveli (2022) tarafından yapılan çalışmada, 2011–2021 yıllarını kapsayan Teksas eyaletine ait trafik kazası verileri kullanılarak kaza şiddetini tahmin eden modeller geliştirilmiştir. Analizde Lojistik Regresyon, AdaBoost, XGBoost, Rastgele Orman, SVM ve Derin Öğrenme gibi farklı makine öğrenimi algoritmalarının performansları karşılaştırılmıştır. Değerlendirme sonuçlarına göre Lojistik Regresyon algoritması, %88 doğruluk oranı ile en yüksek performansı göstermiştir. Çalışma ayrıca hava durumu, yol tipi ve sürücü davranışı gibi değişkenlerin kaza şiddeti üzerindeki etkilerini analiz etmiş olup, makine öğrenimi yaklaşımlarının kaza şiddeti tahmininde etkili olduğunu ortaya koymuştur.

Li ve diğ. (2023) tarafından yürütülen çalışmada, döner kavşaklarda meydana gelen açılı ve açısız çarpışmaların yaralanma şiddetini etkileyen faktörlerin zamansal kararsızlığı incelenmiştir. Araştırma, 2012–2019 yıllarına ait Florida eyaletine özgü çarpışma verilerini kullanarak rastgele parametrelili logit modeller oluşturmuştur. Modelde çevresel, mekânsal, yol, araç ve insan faktörleri gibi değişkenler değerlendirilmiş; açılı ve açısız çarpışmaların şiddet üzerindeki etkilerinin zaman içinde değiştiği belirlenmiştir. Analizler, bazı etkilerin belirli yıllar arasında istikrarlı kalmadığını; özellikle karanlık ortam, ıslak zemin ve agresif sürüş gibi koşulların yaralanma ciddiyetini artırdığını ortaya koymuştur. Bu sonuçlar, döner kavşaklarda güvenlik politikalarının dinamik olarak belirlenmesi gerekliliğini vurgulamaktadır (Li, Ge, Chen, Yuan & Xing, 2023).

Kumeda ve diğ. (2019)’nin çalışmasında, İngiltere’nin 2016 yılına ait karayolu trafik kaza verileri üzerinde makine öğrenmesi algoritmalarının kaza sınıflandırmadaki performansı karşılaştırılmıştır. Araştırmada Fuzzy-FARCHD, Rastgele Orman, Radyal Temel Fonksiyon ağı (RBF), Hiyerarşik LVQ, Çok Katmanlı Algılayıcı (MLP) ve Naive Bayes algoritmaları kullanılmıştır. Sonuçlar, Fuzzy-FARCHD algoritmasının %85,94 doğruluk oranı ile en başarılı sınıflandırmayı sağladığını ortaya koymuştur. Ayrıca analizlerde “aydınlatma

koşulları”, “yol sınıfı” ve “araç sayısı” gibi değişkenlerin kaza şiddetini belirlemede en etkili faktörler olduğu saptanmıştır. Çalışma, makine öğrenmesi yöntemlerinin trafik kazası verilerinden anlamlı desenler çıkarma ve önleyici politika geliştirmede güçlü araçlar olduğunu göstermektedir (Kumeda, Zhang, Zhou, Almasri, Hussain & Assefa, 2019)

Brüning ve Völker (1975) tarafından yapılan erken dönem bir çalışmada, karayolu trafik kazalarında riskin değerlendirilmesinde kullanılan karakteristik miktarlar ve istatistiksel yöntemler üzerine odaklanılmıştır. Çalışmada; toplam araç kilometresi, nüfus, araç sayısı, yol ağı uzunluğu ve yol kullanım süresi gibi referans değişkenler incelenmiş; bu değişkenlerin kaza riskini belirlemede deterministik ya da rastgele değişkenler olarak ele alınmasının önemi vurgulanmıştır. Yazarlar, güven aralıklarının hesaplanması ve relativize edilmiş oranların yanlış yorumlanmasının önlenmesi için metodolojik öneriler sunmuş; trafik güvenliği araştırmalarında risk göstergelerinin doğru yorumlanmasının araştırma sonucunun sağlamlığını artıracığına dikkat çekmişlerdir. Bu çalışma, trafik güvenliği analizlerinde istatistiksel yaklaşımın metodolojik temelini güçlendirmektedir (Brüning & Völker, 1975)

Golob ve diğ. (1999) tarafından yapılan çalışmada, kavşaklardaki yaklaşma trafik hacimlerinin kaza sayıları üzerindeki etkileri incelenmiştir. Araştırmada, Hollanda’daki sinyalizasyon olmayan üç ve dört kollu kavşaklardan elde edilen veriler kullanılarak sıralı ve doğrusal olmayan çok değişkenli istatistiksel modeller geliştirilmiştir. Analiz sonuçları, trafik hacmi ile kaza sayısı arasında doğrusal olmayan ilişkiler bulunduğunu; özellikle bisiklet trafiğinin yoğun olduğu kavşaklarda kaza oranlarının belirgin biçimde arttığını göstermiştir. Ayrıca kavşakların fiziksel tasarım özelliklerinin trafik hacmi-kaza ilişkisini anlamlı düzeyde etkilediği belirlenmiştir. Çalışma, kavşak güvenliği analizlerinde trafik hacmi, bisiklet trafiği ve geometrik tasarım unsurlarının birlikte değerlendirilmesi gerektiğini ortaya koymaktadır.

Çelik ve Seveli (2022) tarafından yapılan çalışmada, 2011–2021 yılları arasında Teksas eyaletine ait trafik kazası verileri kullanılarak kaza şiddetinin tahmin edilmesi amaçlanmıştır. Lojistik Regresyon, XGBoost, Rastgele Orman, K-En Yakın Komşular, Destek Vektör Makineleri ve Derin Öğrenme algoritmaları karşılaştırılmıştır. Model performansları ROC eğrisi ve AUC değerleri ile değerlendirilmiş, veri setindeki sınıf dengesizliği SMOTE yöntemi ile giderilmiştir. Bulgular, Lojistik Regresyon algoritmasının %88 doğruluk oranı ile en yüksek performansı gösterdiğini, XGBoost ve SVM algoritmalarının da benzer başarı düzeyleri sergilediğini ortaya koymuştur. Çalışma, makine öğrenmesi tabanlı modellerin kaza

şiddetinin doğru tahmin edilmesinde etkili olduğunu göstermektedir.

Fang ve diğ. (2018), sürücülerin trafik güvenliği performanslarını değerlendirmek amacıyla trafik ihlali ve kaza verilerine dayalı bir risk değerlendirme modeli geliştirmiştir. Çalışmada sürücülerin demografik bilgileri, ihlal kayıtları ve kaza geçmişleri kullanılmış; Ağırlık Kanıtı (WoE) ve Bilgi Değeri (IV) yöntemleri ile en etkili değişkenler belirlenmiştir. Bu değişkenler arasında sürücü yaşı, ihlal puanı ve emniyet kemeri kullanımı öne çıkmıştır. Lojistik regresyon modeli ile elde edilen risk skorları ROC eğrisi ile doğrulanmış; yüksek risk grubundaki sürücülerin kaza yapma olasılığının yaklaşık üç kat daha fazla olduğu belirlenmiştir. Çalışma, sürücü risk analizinde makine öğrenmesi tabanlı yaklaşımların uygulanabilirliğini ortaya koymaktadır.

Bayata ve Bayrak (2016) tarafından gerçekleştirilen çalışmada, Malatya ilinde uygulanan görüntü işleme tabanlı akıllı kavşak sistemi değerlendirilmiştir. Çalışma kapsamında kavşaklardan elde edilen görüntüler kullanılarak araç tespiti ve sınıflandırması yapılmış; açık kaynaklı Darknet altyapısı ve YOLOv3 derin öğrenme modeli tercih edilmiştir. Gerçek zamanlı araç sayım verileri veri tabanına aktarılmış ve bu veriler kullanılarak sinyalizasyon sisteminin dinamik biçimde çalışması sağlanmıştır. Uygulamanın, yakıt tüketimini azaltma, seyahat sürelerini kısaltma ve egzoz emisyonlarını düşürme potansiyeline sahip olduğu değerlendirilmiştir.

3. YÖNTEM

Bu çalışmada, Malatya ili sınırları içerisinde meydana gelen trafik kazalarının analiz edilmesi ve kazaların şiddetini etkileyen faktörlerin belirlenmesi amacıyla makine öğrenmesi tabanlı veri analizi yaklaşımı benimsenmiştir. Çalışmanın yöntemi, veri bilimi odaklı çok aşamalı bir süreç olarak yapılandırılmıştır.

Araştırma, nicel araştırma yöntemi çerçevesinde gerçekleştirilmiştir. Çalışmada, Malatya iline ait trafik kazalarına ilişkin resmi veriler kullanılarak kaza karakteristikleri sayısal olarak incelenmiştir. Analiz sürecinde veri odaklı bir yaklaşım benimsenmiş, klasik istatistiksel analizler ile makine öğrenmesi algoritmaları bütünleştirilmiştir.

Malatya özelinde daha önce yapılan çalışmalar (Şahin, 2020; Kılıç, 2021) ulaşım ağının yapısal özelliklerini incelemiş ancak kazaların neden-sonuç ilişkisini tahminsel modellerle ele almamıştır. Bu nedenle mevcut araştırma, literatürdeki bu boşluğu doldurmak amacıyla kaza verilerini makine öğrenmesi yöntemleriyle modelleyerek özgün bir katkı sunmaktadır. Bu bölümde, Malatya iline ait trafik kazalarına ilişkin ham veri kümesi üzerinde gerçekleştirilen veri işleme adımları açıklanmaktadır. Veri seti, analiz sürecine alınmadan önce temizlenmiş, yapısal tutarsızlıklar giderilmiş ve çalışmanın gerektirdiği biçime dönüştürülmüştür. Bu kapsamda kullanılan işlemler; eksik değerlerin belirlenmesi, bu değerlerin veri kümesinden ayıklanması, tekrarlayan kayıtların tespiti ve kaldırılması ile genel istatistiklerin elde edilmesini içermektedir. Kod uygulamaları Python ortamında yapılmış, her adımın amacı ve işlevi açıklanarak veri setinin analize uygun hâle getirilmesi sağlanmıştır. Temizleme işlemleri tamamlandıktan sonra, veri setinin genel yapısı tanımlayıcı istatistikler aracılığıyla değerlendirilmiştir. Bu özet tablolar; her bir değişkene ait geçerli gözlem sayısı, kategorik değişkenlerde benzersiz değer sayısı, en sık görülen kategori ve bu kategorinin frekansı gibi temel bilgileri içermektedir. Böylece veri kümesinin genel özellikleri görünür hâle getirilmiş ve sonraki analiz aşamalarına yönelik ön değerlendirme yapılmıştır.

Tablo 4. Temizlenmiş Veri Setine Ait Tanımlayıcı İstatistikler

	Gözlem Sayısı	Benzersiz Değer Sayısı	En Sık Görülen Değer	Frekans
Kazayılı	1484.0	-	-	-
Yaya Kazazede	1484	2	2-Yaya	850
Yaya Cinsiyet	1484	2	1-Erkek	806
Yaya Kusuru1	1484	11	Kusursuz	791
Yaya Kusuru2	1484	6	Kusursuz	1463
Yolcu Kusuru1	1484	13	Kusursuz	758
Yolcu Kusuru2	1484	6	Kusursuz	1450
Yaya Ölü Toplam	1484.0	-	-	-
Yaya Yaralı Toplam	1484.0	-	-	-
Kaza Şiddeti	1484.0	-	-	-
Yaya Yaş Grubu	1484	6	Orta Yaşlı 45-64	279

Tablo 4’e göre veri setindeki tüm sütunlarda 1484 gözlem yer almakta ve eksik veri bulunmamaktadır. Bu durum, veri temizleme sürecinin tamamlandığını ve setin analiz için hazır olduğunu göstermektedir. Kategorik değişkenlerde çeşitli sınıflar tespit edilmiştir. Yaya Kazazede değişkeninde iki kategori bulunmakta olup en yüksek frekans 850 gözlemle —2-Yaya grubundadır. Yaya_Kusuru2 değişkeninde kusursuzluk oranı %98,6, Yolcu Kusuru2 değişkeninde ise 1450 gözlem —Kusursuz olarak kaydedilmiştir. Yaya Yaş Grubu incelendiğinde, kazalara en çok 45–64 yaş aralığındaki bireylerin karıştığı görülmektedir. Sayısal değişkenlerde benzersiz değer veya frekans hesaplamaları yapılmamıştır; bu ölçütler yalnızca kategorik değişkenler için anlam taşımaktadır. Genel sonuçlar, yüksek kusursuzluk oranları nedeniyle kazaların büyük bölümünde yayaların sorumlu olmadığını ve olayların daha çok diğer faktörlerden kaynaklandığını düşündürmektedir. Ayrıca veri seti, dummies yöntemiyle sayısallaştırılarak modelleme sürecine uygun hâle getirilmiştir.

3.1 Dummies (Kukla Değişken) Yöntemi ile Sayısallaştırma

Bu çalışmada, veri setinde yer alan kategorik değişkenlerin (örneğin YayaCinsiyet, YayaKusuru1, YolcuKusuru1, YayaYaşGrubu) makine öğrenmesi algoritmalarında kullanılabilmesi amacıyla — dummies (kukla değişken) yöntemi uygulanmıştır. Bu yöntem,

kategorik verilerin sayısal formata dönüştürülmesini sağlayarak, modellerin değişkenler arasındaki ilişkileri daha doğru biçimde değerlendirmesine olanak tanır.

Dummies yöntemi, her bir kategorik değeri temsil eden yeni sütunlar oluşturarak, o gözleme ait kategoriye —1, diğerlerine —0 değerini atar. Örneğin, YayaCinsiyet değişkeni için —1-Erkek ve —2-Kadın olmak üzere iki ayrı sütun oluşturulmuş, erkek gözlemler için —1-Erkek = 1, 2-Kadın = 0 biçiminde, kadın gözlemler için ise tam tersi değerler atanmıştır.

Python ortamında bu işlem pandas kütüphanesi kullanılarak aşağıdaki komutla gerçekleştirilmiştir:

```
• df = pd.get_dummies(df, columns=["Yaya_Cinsiyet", "Yaya_Kusuru1", "Yolcu_Kusuru1", "YayaYaşGrubu"])
```

Bu dönüşüm sayesinde, Malatya ili trafik kazalarına ait veri seti sayısal, analiz edilebilir ve makine öğrenmesi algoritmalarına uygun hale getirilmiştir. Böylelikle modelleme aşamasında kazaların oluşumuna etki eden faktörler (cinsiyet, kusur türü, yaş grubu vb.) sayısal olarak temsil edilerek daha güvenilir tahmin ve sınıflandırma sonuçları elde edilmiştir.

3.2. Sınıf Dengesizliğinin Giderilmesi ve Veri Bölme (SMOTE)

SMOTE yöntemi, dengesiz sınıf yapısına sahip veri kümelerinde sınıflar arası öğrenme dengesini sağlamada önemli bir rol oynamaktadır. Bu yöntem sayesinde model, yalnızca çoğunluk sınıfına odaklanmadan azınlık sınıfındaki örnekleri de tanımayı öğrenir. Böylece sınıflandırma performansı genel olarak artarken, özellikle F1-skoru, duyarlılık (recall) ve kesinlik (precision) gibi metriklerde daha dengeli sonuçlar elde edilir.

Veri sızıntısını (data leakage) önlemek amacıyla, veri önce eğitim ve test alt kümelerine ayrılmış, ardından SMOTE yalnızca eğitim verisine uygulanmıştır. Böylece test verisinin orijinal dağılımı korunmuş, model performansının yapay olarak artmasının önüne geçilmiştir. Eğitim–test ayrımı sırasında tabakalı örnekleme (stratify) yöntemi kullanılarak her iki alt kümede de sınıf oranlarının benzer kalması sağlanmıştır.

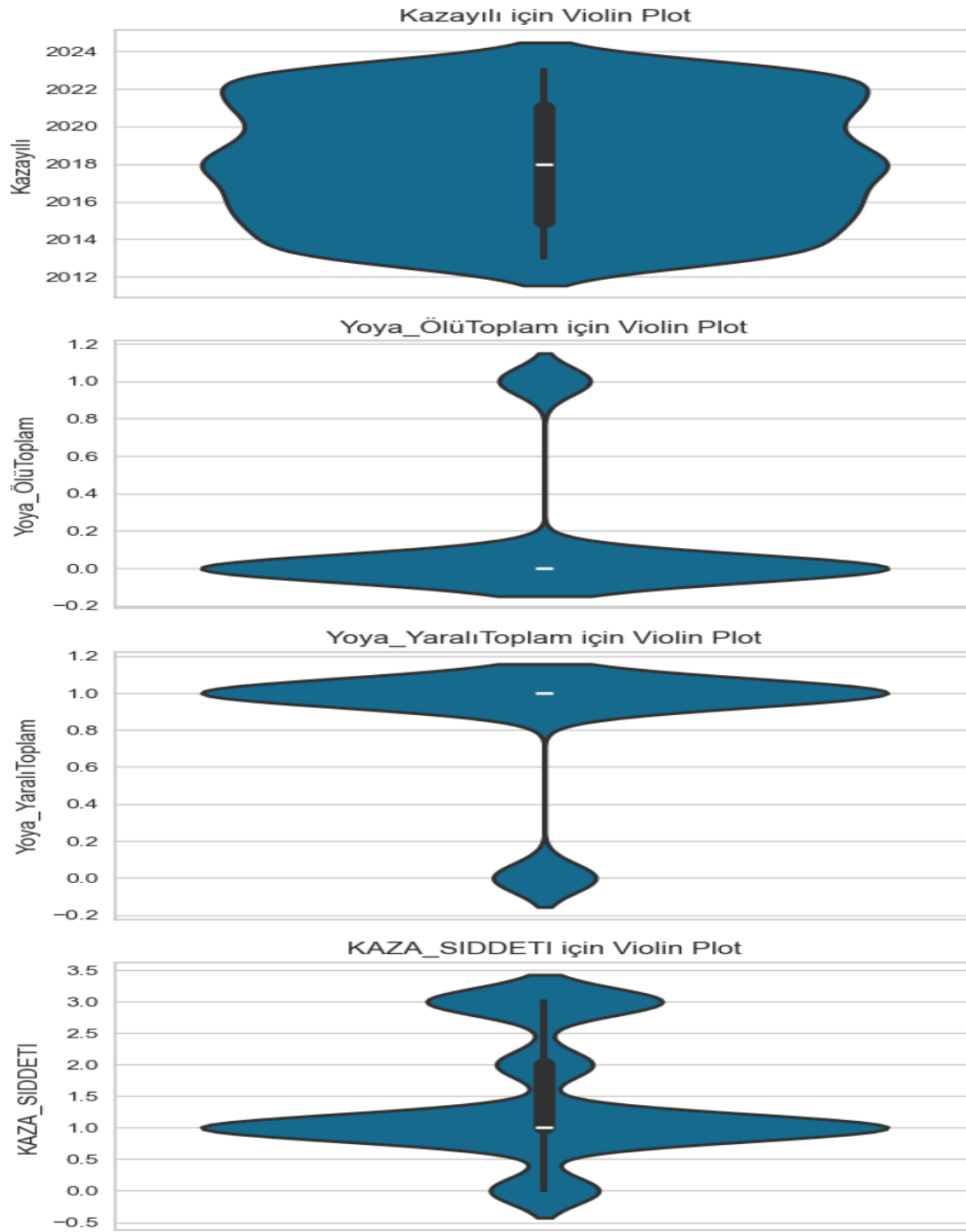
```
from sklearn.model_selection import train_test_split
from imblearn.over_sampling import SMOTE
import pandas as pd
```

```
# X: özellikler, y: hedef değişken
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.20, random_state=42,
    stratify=y)

# Yalnızca eğitim setinde SMOTE uygulanması
smote = SMOTE(random_state=42)
X_train_res, y_train_res = smote.fit_resample(X_train, y_train)

# Sınıf dağılımlarını kontrol etme
print("Önce (Eğitim Verisi):")
print(pd.Series(y_train).value_counts())
print("\nSonra (SMOTE Sonrası, Eğitim Verisi):")
print(pd.Series(y_train_res).value_counts())
```

Aşağıdaki violin plot grafiklerinde, çalışmada kullanılan temel değişkenlerin dağılımları gösterilmiştir. Bu grafikler, her bir değişkenin veri yoğunluğunu, medyanını ve çeyrekler arası dağılımını görsel olarak ifade etmektedir.



Şekil 4. Seçilmiş Değişkenlere Ait Violin Plot Görselleştirmeleri

- Kazayılı değişkeninde, kazaların özellikle 2018 yılı civarında yoğunlaştığı görülmektedir. Bu durum, veri setinin en kapsamlı ve güncel kısmının bu döneme ait olduğunu göstermektedir.
- Yaya_ÖlüToplam değişkeninde, ölümlü yaya kazalarının oldukça düşük oranda gerçekleştiği, dolayısıyla ölüm vakalarının nadir olduğu anlaşılmaktadır.

- Yaya_YaralıToplam değişkeninde, yaralanmalı yaya kazalarının yoğun olduğu görülmektedir; bu durum kazaların çoğunun yaralanma ile sonuçlandığını göstermektedir.
- KAZA_SIDDETI değişkeninde ise kazaların çoğunluğunun düşük şiddetli (1 veya 2 seviyesinde) olduğu, yüksek şiddetli kazaların nadir olarak gerçekleştiği gözlemlenmiştir.

Violin plot görselleri, veri setindeki dağılımların dengeli olup olmadığını anlamada etkili bir araçtır. Bu görseller, analiz öncesinde verinin genel yapısını anlamayı kolaylaştırmakta ve sonraki modelleme aşamasında hangi değişkenlerin daha açıklayıcı olabileceğine dair öngörü sağlamaktadır.

3.3 Verilerin Ölçeklendirilmesi (StandardScaler)

Veri ön işleme sürecinde, değişkenlerin farklı ölçeklerde olması makine öğrenmesi modellerinin performansını olumsuz yönde etkileyebilir. Özellikle uzaklık veya ağırlık temelli algoritmalarda, büyük değer aralıklarına sahip değişkenler modelin öğrenme sürecinde baskın hâle gelebilir. Bu nedenle, tüm değişkenlerin aynı ölçekte değerlendirilmesi önemlidir.

Bu çalışmada, değişkenlerin ölçeklerini eşitlemek amacıyla StandardScaler yöntemi kullanılmıştır. Bu yöntem, her bir değişkeni ortalaması 0 ve standart sapması 1 olacak şekilde dönüştürür. Böylece modelin her değişkene eşit ağırlıkta yaklaşması sağlanır.

StandardScaler, özellikle lojistik regresyon, destek vektör makineleri (SVM) ve yapay sinir ağları gibi gradyan temelli algoritmalarda daha dengeli ve kararlı bir öğrenme süreci sağlar. Ölçekleme işlemi sonrasında model parametrelerinin (ağırlıkların) güncellenmesi daha tutarlı hâle gelir; bu da modelin öğrenme hızını, kararlılığını ve doğruluğunu artırır.

Analizlerde, Malatya ilinde belirli yıllar arasında meydana gelen trafik kazalarına ait resmi veriler kullanılmıştır. Veri setinde yer alan değişkenler;

- Zaman bilgileri (Kazayılı, KazaAy, KazaGun, KazaSaatDilimi),
- Coğrafi değişkenler (Kazaİlce, GeoYatay, GeoDüşey, GeoKavşak, GeoGeçit),
- Yol ve hava koşulları (YolunTipi, YolunKaplaması, YolunYüzeyi, HavaDurumu, GünDurumu),
- Sonuç değişkenleri (TOPLAM_OLU, TOPLAM_YARALI, KAZA_SIDDETI) şeklinde gruplandırılmıştır.

Ek olarak, yerel düzeyde yapılan bazı ulaşım araştırmalarından (Örkmez, 2020; Yılmaz, 2019) elde edilen çevresel ve ağ verileri, kentsel ulaşım yoğunluğu açısından tamamlayıcı bilgi olarak değerlendirilmiştir.

Analiz öncesinde veri setine uygulanan işlemler boş veya hatalı kayıtlar tespit edilerek çıkarılması ,kategorik değişkenler “*dummies*” yöntemiyle sayısal formata dönüştürülmesi, sayısal düşük veriler sınıf dengesizliği smote yöntemi ile giderilmesi olarak listelenebilir.

3.4 Kullanılan Makine Öğrenmesi Modelleri

Çalışmada yedi farklı sınıflandırma algoritması uygulanmıştır:

- **Logistic Regression (LR):** Basit ve yorumlanabilir bir temel model olarak seçilmiştir.
- **Decision Tree (DT):** Veriyi ağaç yapısında sınıflandıran açıklanabilir bir algoritmadır.
- **Random Forest (RF):** Birden fazla karar ağacının birleşimiyle çalışan, overfitting riskini azaltan topluluk modelidir.
- **XGBoost (XGB):** Gradyan artırma (gradient boosting) tekniğine dayalı, yüksek doğruluk sağlayan güçlü bir sınıflandırma algoritmasıdır.
- **AdaBoost (Adaptive Boosting):** Zayıf öğrencileri ardışık olarak eğiterek hatalı sınıflandırmalara daha fazla ağırlık veren ve bu sayede sınıflar arasındaki ayrımı güçlendiren topluluk tabanlı bir sınıflandırma algoritmasıdır.
- **Support Vector Classification (SVC):** Sınıflar arasındaki en uygun ayırıcı hiper düzlemi belirlemeyi amaçlayan, özellikle yüksek boyutlu veri setlerinde güçlü genelleme performansı sunan bir sınıflandırma algoritmasıdır.
- **Gradient Boosting (GB):** Önceki modellerin hatalarını ardışık olarak düzelten zayıf öğrencilerin birleştirilmesine dayanan, karmaşık ilişkileri modelleyebilen etkili bir topluluk öğrenmesi yöntemidir.

Her model, eğitim ve test verileri üzerinde ayrı ayrı çalıştırılmış ve performans farkları ölçülmüştür. Bu sayede modellerin dengesiz veri nedeniyle yanlış genelleme yapması engellenmiştir.

Modelleme aşamasında temel sınıflandırma modeli olarak lojistik regresyon tercih edilmiştir. Lojistik regresyon modeli, trafik kazası verilerinde bağımlı değişkenin (örneğin —ölümlül ya da —yaralanmalı kaza) ikili yapısı nedeniyle uygun bir algoritmadır. Modelin performansı, accuracy, precision, recall ve F1-score ölçütleriyle değerlendirilmiştir. Karşılaştırma amacıyla

alternatif olarak Karar Ağaçları (Decision Tree) ve Rastgele Orman (Random Forest) modelleri de uygulanmış; modellerin tahmin performansları, veri setinin dengesizliği giderildikten sonra çapraz doğrulama yöntemiyle test edilmiştir.

Modelleme sürecinde kullanılan parametreler, literatürdeki benzer çalışmaların (Çelik, 2022; Türker & Gündüz, 2023; Bakış et al., 2025) önerdiği değerlere göre optimize edilmiştir.

Modellerin başarı düzeyleri aşağıdaki performans metrikleriyle değerlendirilmiştir:

- **Accuracy (Doğruluk Oranı):** Modelin doğru sınıflandırdığı örneklerin, toplam örnek sayısına oranıdır. Genel model başarısını yansıtır; ancak sınıflar arasında dengesizlik bulunan veri setlerinde tek başına yeterli bir ölçüt olmayabilir.
- **Precision (Kesinlik):** Modelin pozitif olarak sınıflandırdığı örnekler içinde gerçekten pozitif olanların oranını ifade eder. Yanlış pozitiflerin maliyetinin yüksek olduğu durumlarda önemli bir performans ölçütüdür.
- **Recall (Duyarlılık):** Gerçek pozitif örneklerin ne kadarının model tarafından doğru tahmin edildiğini gösterir. Yanlış negatiflerin kritik olduğu problemlerde model başarısını değerlendirmek için kullanılır.
- **F1 Score (Harmonik Ortalama):** Precision ve Recall metriklerinin harmonik ortalamasıdır. Bu metrik, iki ölçüt arasında denge kurarak özellikle dengesiz veri setlerinde model performansını daha sağlıklı biçimde değerlendirmeye olanak sağlar.
- **MAE (Ortalama Mutlak Hata):** Tahmin edilen değerler ile gerçek değerler arasındaki mutlak farkların ortalamasını ifade eder. Hataların yönünü dikkate almadan, modelin ortalama sapmasını anlaşılır bir biçimde ortaya koyar.
- **MSE (Ortalama Kare Hata):** Tahmin hatalarının karelerinin ortalamasıdır. Büyük hatalara daha fazla ağırlık verdiği için modelin uç değerlerdeki performansını değerlendirmede etkilidir.
- **Cohen's Kappa (Sınıf Denge Düzeyi):** Modelin sınıflandırma başarısını, tesadüfi olarak doğru sınıflandırma olasılığını dikkate alarak ölçer. Sınıf dağılımının dengesiz olduğu durumlarda Accuracy'ye göre daha güvenilir bir performans göstergesidir.
- **ROC-AUC (Ayrım Gücü Ölçütü):** ROC eğrisi, duyarlılık ile yanlış pozitif oranı arasındaki ilişkiyi gösterirken; AUC değeri modelin sınıfları ayırt etme gücünü tek bir sayı ile özetler. AUC değerinin 1'e yaklaşması, modelin ayırım gücünün yüksek olduğunu gösterir.

Bu bölümde, Malatya iline ait trafik kazası verileri üzerinde geliştirilen makine öğrenmesi modelleri incelenmiştir. Analiz süreci, veri temizleme ve ölçekleme işlemlerinin ardından, sınıf dengesizliği problemini azaltmak amacıyla SMOTE (Synthetic Minority Oversampling Technique) yöntemi uygulanarak başlatılmıştır. SMOTE yöntemi, az temsil edilen sınıfların sentetik örneklerle dengelenmesini sağlayarak sınıflandırma modellerinin performansını artırmayı amaçlamaktadır (Chawla et al., 2002). Veri seti %80 eğitim ve %20 test oranında

ayrılmış; modeller yalnızca eğitim verisi üzerinde eğitilmiş ve test verisi kullanılarak doğrulanmıştır.

Modelleme aşamasında lojistik regresyon ve karar ağacı algoritmaları kullanılmıştır. Lojistik regresyon modeli, kaza şiddetini etkileyen değişkenlerin yönünü ve büyüklüğünü istatistiksel olarak değerlendirmeye olanak tanıyan, genellenebilirliği yüksek bir yöntemdir (Hosmer et al., 2013). Karar ağacı modeli ise değişkenler arasındaki hiyerarşik ilişkileri açık biçimde ortaya koyması açısından yorumlanabilir olmakla birlikte, literatürde de vurgulandığı üzere aşırı öğrenmeye yatkın bir yapıya sahiptir (Hastie et al., 2009). Bu durum, karar ağacı modelinin test verisi üzerindeki performansının görece daha düşük olmasına neden olabilmektedir.

Analiz sonuçları, kaza şiddetinin belirlenmesinde insan faktörüne ilişkin değişkenlerin (sürücü, yolcu ve yaya kusurları) belirleyici rol oynadığını göstermektedir. Bu bulgu, trafik kazalarının çok değişkenli istatistiksel yöntemlerle incelendiği çalışmalarda insan faktörünün temel açıklayıcı değişkenler arasında yer aldığı yönündeki sonuçlarla örtüşmektedir (Bayata, 2010). Malatya özelinde kazaların çoğunlukla gündüz saatlerinde, bölünmüş ve asfalt kaplamalı yol kesimlerinde ve sürücü hatalarına bağlı olarak meydana gelmesi, kent içi trafik dinamiklerinin kaza oluşumundaki etkisini ortaya koymaktadır.

Elde edilen bulgular, makine öğrenmesi ve istatistiksel modelleme yaklaşımlarının kentsel trafik sistemlerinde kaza şiddetinin analiz edilmesi, risk faktörlerinin belirlenmesi ve karar destek süreçlerinin geliştirilmesi açısından önemli bir araç olduğunu göstermektedir. Özellikle lojistik regresyon modeli sonuçları; yaş grubu, kaza zamanı, yol tipi ve insan kusuruna ilişkin değişkenlerin kaza şiddeti üzerinde istatistiksel olarak anlamlı etkilere sahip olduğunu ortaya koymuş ve trafik güvenliği politikalarında insan faktörünün ve yol özelliklerinin birlikte ele alınması gerektiğini vurgulamıştır (Bayata, 2010; Hosmer vd., 2013)

4. BULGULAR

Bu bölümde, Malatya iline ait 2013–2023 yılları arasındaki trafik kazalarına ilişkin veriler, istatistiksel çözümler ve makine öğrenmesi tabanlı yaklaşımlar kullanılarak analiz edilmiştir. Analiz süreci, öncelikle veri setinin betimsel istatistiklerinin ortaya konulmasıyla başlamış; devamında kaza sayıları ve kaza sonuçları ile ilişkili değişkenler dikkate alınarak kaza şiddetinin tahminine yönelik modeller geliştirilmiştir. Bu yaklaşım, trafik kazalarının farklı yöntemlerle modellenebileceğini ve tahmin edilebilirliğini ortaya koyan önceki çalışmalarla uyumludur (Emniyet Genel Müdürlüğü, 2022; Bayata & Hattatoğlu, 2011). Kazaların yıllara, aylara, günlere ve saat dilimlerine göre dağılımları incelendiğinde, özellikle 2018 yılı sonrasında toplam kaza sayısında artış eğilimi olduğu; yaz aylarında, özellikle Haziran ve Temmuz aylarında kazaların belirgin şekilde yoğunlaştığı gözlenmiştir. Bu bulgular, trafik kazalarının mevsimsel ve zamansal faktörlerden etkilendiğini göstermekte olup, kent içi trafik hacminin arttığı dönemlerde kaza riskinin yükseldiğine işaret etmektedir. Kaza şiddetinin incelenmesinde kullanılan KAZA_SIDDETI değişkeni, kazaların çoğunlukla orta düzeyde (1. derece) gerçekleştiğini göstermektedir. Violin grafikleri ve dağılım analizleri, yüksek şiddetli kazaların düşük frekanslı olduğunu doğrulamıştır.

Veri dengesizliği SMOTE yöntemiyle giderildikten sonra, dört sınıflandırma algoritması kullanılmıştır: Logistic Regression, Decision Tree, Random Forest ve XGBoost. Modellerin performansları, Accuracy, Precision, Recall, F1, MAE, MSE ve Cohen's Kappa ölçütleri ile değerlendirilmiş, eğitim ve test verileri arasındaki farklılıklar grafiklerle gösterilmiştir.

Elde edilen sonuçlar şu şekilde özetlenebilir:

- Logistic Regression: Genel olarak istikrarlı sonuçlar üretmiş, ancak bazı sınıflarda duyarlılık düşüktür.
- Decision Tree: Yüksek varyans ve overfitting eğilimi nedeniyle genelleme başarısı sınırlı kalmıştır.
- Random Forest: Orta düzeyde doğruluk sağlamış, fakat bazı sınıflarda dengesiz tahminler üretmiştir.
- XGBoost: En yüksek doğruluk, AUC ve tutarlılık değerlerini sağlamış; kaza şiddetinin tahmininde en başarılı algoritma olarak öne çıkmıştır .

Özellik önem analizleri sonucunda elde edilen bulgular, kaza türü ve kaza sonuçlarına ilişkin değişkenlerin kaza şiddetinin belirlenmesinde kritik bir role sahip olduğunu göstermektedir.

Bu durum, literatürde kaza şiddeti üzerinde etkili olan faktörler arasında kaza türü ve sonuç değişkenlerinin öne çıktığını ortaya koyan çalışmalarla uyumludur (Çelik & Seveli, 2022; Erdogan & Gökdağ, 2019).

4.1. Betimsel (Tanımlayıcı) İstatistikler

Çalışmada kullanılan veri seti, Malatya ili sınırları içinde 2013–2023 yılları arasında meydana gelen 9.787 trafik kazasına ait bilgileri kapsamaktadır. Veri setinde, kazaların tarihi, saati, konumu, yol ve çevre koşulları, ayrıca kaza türü ve sonuçlarına ilişkin değişkenler yer almaktadır. Bu kapsamlı veri yapısı, kazaların mekânsal, zamansal ve şiddet düzeyi açısından detaylı analizine imkân sağlamaktadır.

Aşağıdaki tablo, bu değişkenlere ait özet istatistikleri göstermektedir.

Tablo 5. Değişkenlere Ait Özet İstatistikler

Değişken	Gözlem Sayısı (N)	Kategori Sayısı	En Sık Görülen Değer	Frekans
Kaza Saat Dilimi	9787	12	16:00–17:59	1429
Kaza İlçesi	9787	14	Malatya–Merkez	3800
Kaza X Koordinatı	9787	6701	3660644	18
Kaza Y Koordinatı	9787	5267	3834771	25
Kaza Tarihi	9787	3538	20190811	10
Kaza Yerleşim Yeri / Dışı	9787	2	1 – Yerleşim Yeri	7209
Yolun Tipi	9787	4	1 – Bölünmüş Yol	5737
Yolun Kaplaması	9787	6	1 – Asfalt	9464
Yolun Sınıfı	9787	11	1 – Cadde	4856
Coğrafi Yatay Eğimi	9787	3	1 – Düz Yol	8427
Coğrafi Düşey Eğimi	9787	4	1 – Eğimsiz	7514
Kavşak Durumu	9787	8	8 – Kavşak Yok	6315
Geçit Durumu	9787	5	5 – Geçit Yok	8823
Gün Durumu	9787	3	1 – Gündüz	6664
Hava Durumu	9787	10	1 – Açık	8633
Yol Yüzeyi	9787	6	1 – Kuru	8341
Kaza Oluş Türü	9787	15	3 – Yandan Çarpma	2780
Kazada Araç Sayısı	9787	3	1 – Tek Araçlı	4952
Kaza Ayı	9787	12	08 (Ağustos)	1151
Kaza Günü	9787	31	20	354

Tablo 5 incelendiğinde, kazaların çoğunlukla gündüz saatlerinde (%66,6), açık hava koşullarında (%88,2) ve kuru yol yüzeylerinde (%85,2) meydana geldiği görülmektedir. Elde edilen bu bulgular, Malatya'daki trafik kazalarının büyük ölçüde insan faktörünün etkili olduğu koşullarda gerçekleştiğini göstermektedir. Kazaların en sık 16:00–17:59 saat aralığında (%14,6) ve Ağustos ayında meydana geldiği tespit edilmiştir. Bu zaman diliminin iş çıkış saatleri ve yoğun trafik dönemleriyle çakışması, artan trafik hacminin kaza olasılığını yükselttiğini göstermektedir. Bu bulgu, yoğun saatlerde trafik hacmi ile kaza sayısı arasında anlamlı bir ilişki bulunduğunu ortaya koyan çalışmalarla uyumludur (Korkmaz & Karabulut, 2018). Mekânsal dağılıma göre kazaların Malatya Merkez (%38,8), Yeşilyurt (%20,9) ve Battalgazi (%16,3) ilçelerinde yoğunlaştığı belirlenmiştir. Bu durum, kent merkezlerinde artan araç yoğunluğu ve karma ulaşım yapısının kaza riskini artırdığını ortaya koyan çalışmalarla paralellik göstermektedir (Aydın & Şen, 2018). Yol ve çevre koşulları incelendiğinde, kazaların büyük çoğunluğunun bölünmüş asfalt yollarda (%96,6) ve kavşak bulunmayan yol kesimlerinde (%64,5) meydana geldiği görülmektedir. Bu bulgu, Malatya'daki karayolu altyapısının genel özellikleriyle uyumlu olup, yüksek hız potansiyeline sahip ve trafik denetiminin sınırlı olduğu yol türlerinde kaza riskinin arttığını göstermektedir (Karayolları Genel Müdürlüğü, 2021). Kaza türleri açısından en yüksek oranın yandan çarpma (%28,4) şeklinde olduğu belirlenmiştir. Yandan çarpma türündeki kazaların, şerit değiştirme, dönüş manevraları ve hatalı sollama gibi sürücü davranışlarından kaynaklandığı literatürde ifade edilmektedir (Korkmaz & Karabulut, 2018).

Aşağıdaki tabloda, Malatya iline ait trafik kazası veri setinde yer alan sayısal değişkenlerin özet istatistikleri (ortalama, standart sapma, minimum ve maksimum değerler) verilmiştir.

Tablo 6. Sayısal Değişkenlere İlişkin Temel İstatistikler

Değişken	Gözlem Sayısı (N)	Ortalama	Standart Sapma	Min	%25	%50	%75	Maks
Kaza Yılı	9787	2018.04	3.16	2013	2015	2018	2021	2023
Yolun Yasal Hız Limiti	9787	65.10	24.30	20	50	50	80	120
Kaza ID	9787	85072.20	51563.68	18	41675	83067	124129	235068
Toplam Ölü Sürücü	9787	0.01	0.10	0	0	0	0	2
Toplam Ölü Yolcu	9787	0.02	0.21	0	0	0	0	6
Toplam Ölü Yaya	9787	0.02	0.13	0	0	0	0	2

Toplam Ölümlü Kaza	9787	0.04	0.19	0	0	0	0	1
Toplam Yaralanmalı Kaza	9787	0.96	0.19	0	1	1	1	1
Yaralı Sürücü Sayısı	9787	0.47	0.61	0	0	0	1	4
Yaralı Yolcu Sayısı	9787	1.40	1.81	0	0	1	2	29
Yaralı Yaya Sayısı	9787	0.29	0.51	0	0	0	1	6
Toplam Ölü Sayısı	9787	0.05	0.29	0	0	0	0	7
Toplam Yaralı Sayısı	9787	2.15	1.89	0	1	2	3	30
Kaza Şiddeti	9787	1.06	0.81	0	1	1	1	3

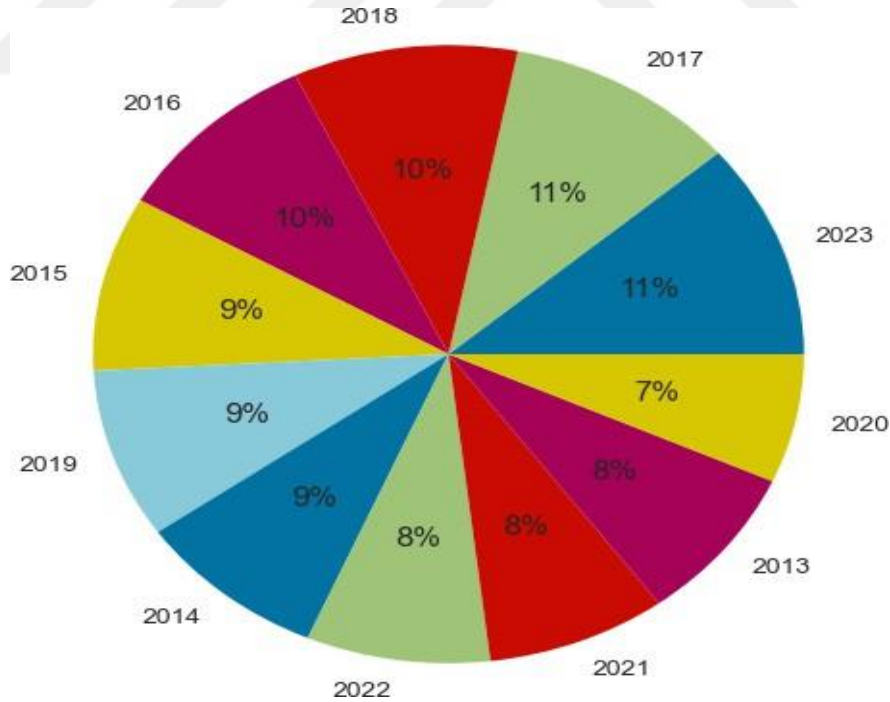
Tablo 6’da yer alan değerler, Malatya ilinde incelenen kazaların genellikle orta yoğunlukta, şehir içi ölçekli ve insan temelli özellikler taşıdığını göstermektedir. Ortalama kaza yılı 2018 olup, veri setinin son yıllarda meydana gelen olayları kapsadığı anlaşılmaktadır. Bu durum, elde edilen bulguların güncel kentsel ulaşım koşullarını yansıttığını göstermesi bakımından önemlidir. Yolun yasal hız limiti ortalamasının 65 km/s olması, kazaların çoğunlukla şehir içi ve çevre yoluna yakın arterlerde meydana geldiğini işaret etmektedir. Maksimum hız limitinin 120 km/s olması ise, çevre yolu gibi yüksek hızlı geçiş güzergâhlarının da kaza profiline dâhil olduğunu göstermektedir.

Kazaların türlerine göre sonuçları incelendiğinde, yaralanmalı kazalar %96 oranında görülürken, ölümlü kazaların oranı %4’ün altındadır. Bu oran, Malatya’daki trafik kazalarının çoğunlukla ölümle sonuçlanmadığını, ancak yaralanma düzeyinde ciddi sonuçlar doğurduğunu göstermektedir. Bu durum aynı zamanda şehir içi ulaşımında hız sınırlarının göreceli olarak düşük olmasına karşın, sürücü dikkatsizliği ve ihlallerin kazaların ana nedeni olduğunu düşündürmektedir. TOPLAM YARALI değişkeni ortalamasının 2,15 kişi olması, kazaların genellikle birden fazla kişiyi etkilediğini göstermektedir. Maksimum değer 30 kişiye kadar ulaşması, özellikle toplu taşıma araçlarının karıştığı çoklu kazaların zaman zaman meydana geldiğini ortaya koymaktadır. Malatya özelinde yapılan çalışmalar, özellikle kavşak noktalarındaki yoğunluğun ve trambüs gibi toplu taşıma sistemlerinin kaza dinamikleri üzerinde etkili olduğunu belirtmektedir (Kılıç, 2021; Örkmez, 2020).

Ölümlü kaza değişkenleri (TOPLAM ÖLÜ SÜRÜCÜ, TOPLAM ÖLÜ YOLCU) incelendiğinde ortalama değerlerin sıfıra yakın olması, ölümle sonuçlanan olayların nadir olduğunu; ancak bu olayların meydana geldiğinde yüksek şiddet içerdiğini göstermektedir. Bu da trafik kazalarının düşük sıklıkla ancak yüksek etkili sonuçlar doğurabilen olaylar olduğunu doğrulamaktadır. Kaza şiddeti (KAZA ŞİDDETİ) ortalaması 1,06 olarak hesaplanmış, bu da kazaların çoğunlukla hafif yaralanmalı kazalar kategorisinde yer aldığını göstermiştir. Maksimum değerin 3 (ölümlü kaza) olması, veri setinin farklı şiddet düzeylerini kapsadığını ve modelleme aşamasında bu farklılığın dikkate alınması gerektiğini ortaya koymaktadır.

Bu tablo, Malatya ilinde trafik kazalarının büyük ölçüde ölümsüz ancak yaralanmalı kazalar şeklinde gerçekleştiğini ve genellikle şehir içi hız limitlerinde yaşandığını ortaya koymaktadır. Bu sonuçlar, Türkiye genelindeki kaza istatistikleri ve bölgesel risk analizleriyle paralellik göstermektedir (TÜİK, 2023; Aksoy & Kantar, 2017). Kazaların önemli bir kısmı insan faktörlerinden kaynaklanmakta, bu da makine öğrenmesi modellerinin bu değişkenleri ve karmaşık kaza karakteristiklerini özellikle dikkate almasının önemini vurgulamaktadır (Abdel-Aty & Pande, 2007; Kumeda vd., 2019).

Çalışmada kullanılan verilerin genel dağılımını ortaya koymak amacıyla hazırlanan tablo ve grafikler aşağıda sunulmuştur:



Şekil 5 Kazaların Yıllara Göre Dağılımı

Şekil 5’de kazaların yıllara göre dağılımı incelendiğinde, 2013–2023 yılları arasında belirgin

bir dengesiz ancak genel olarak artış eğilimli bir seyir gözlemlenmiştir.

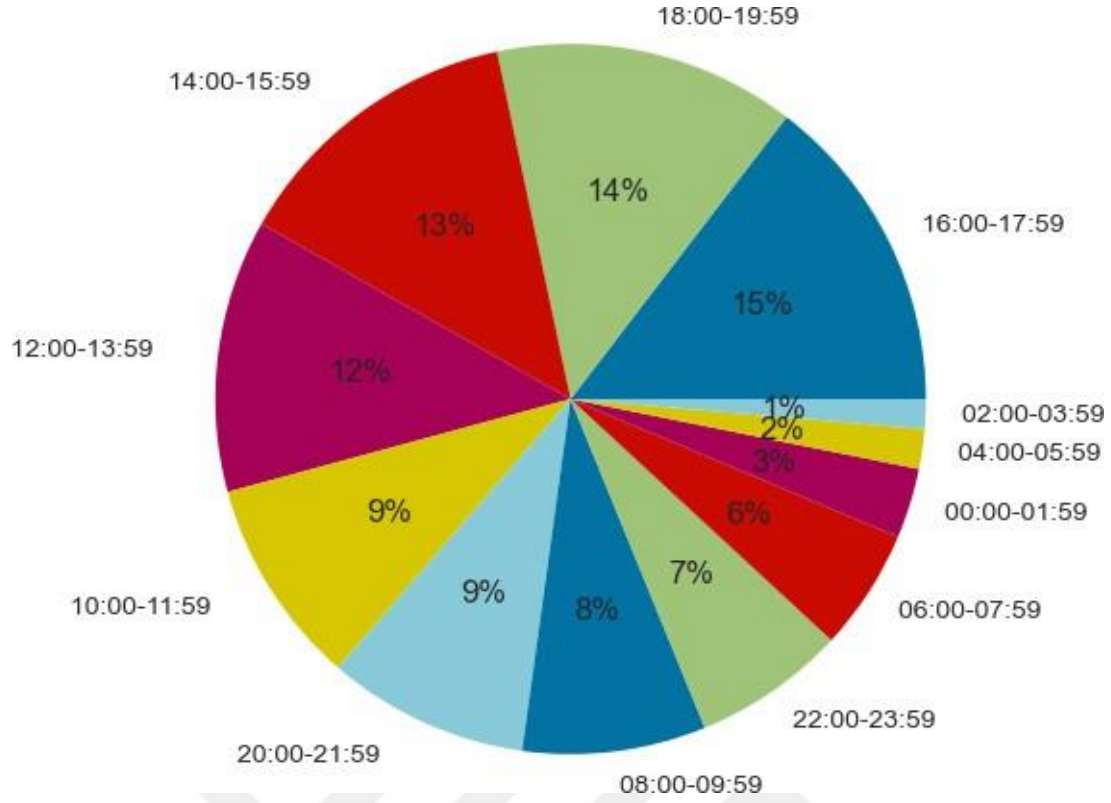
Grafığe göre, 2017 (%11) ve 2023 (%11) yılları en fazla kazanın meydana geldiği dönemlerdir.

Bu yıllar, Malatya kentinde araç sayısındaki artışın ve şehirleşme hızının yüksek olduğu dönemlere denk gelmektedir.

Öte yandan, 2020 yılı (%7) COVID-19 pandemisi nedeniyle kısıtlamaların uygulandığı dönemdir. Bu nedenle, o yıl yaşanan kazaların sayısında belirgin bir düşüş gözlemlenmiştir.

2021 ve 2022 yıllarında ise normalleşme süreciyle birlikte kazaların yeniden artış eğilimine geçtiği görülmektedir. Bu eğilim, Malatya'daki ulaşım sisteminin gelişimi, trafik yoğunluğundaki değişim ve kent merkezinde artan nüfus yoğunluğuyla paralel olarak yorumlanabilir.

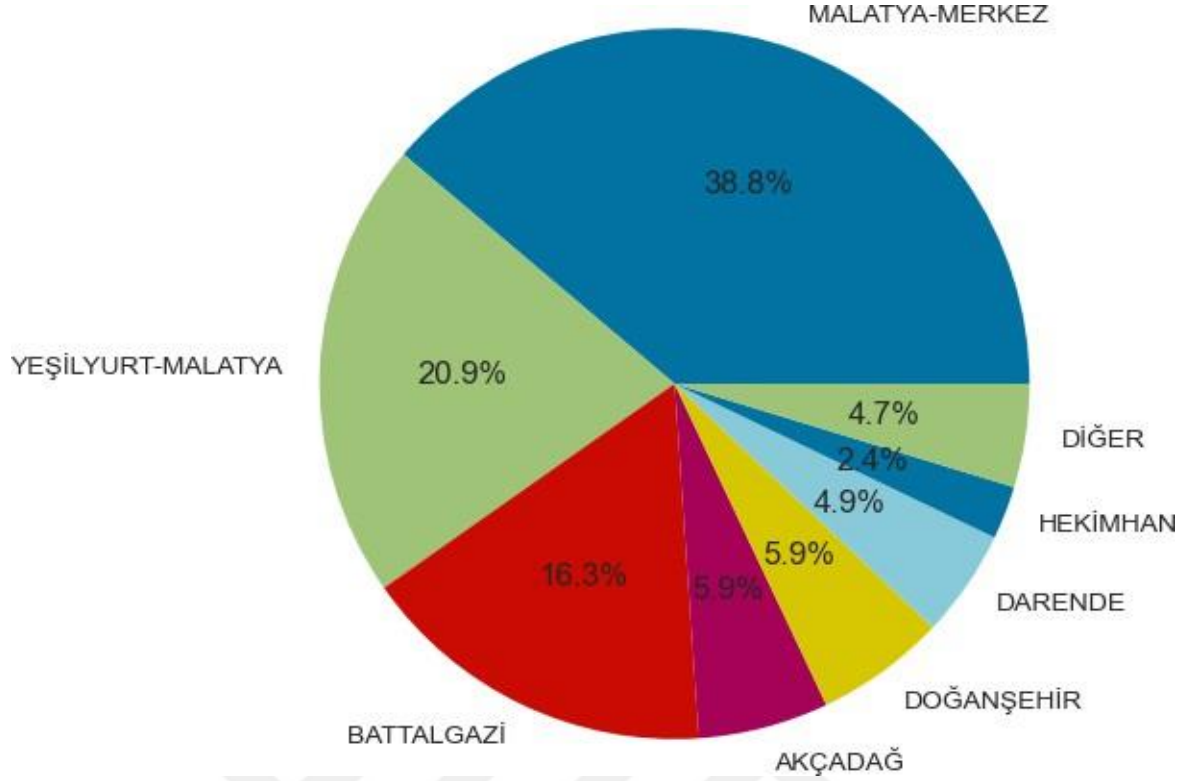
Benzer şekilde, Tortum ve Atalay (2010) ile Uçar (2023)'ün çalışmaları da Türkiye genelinde kentleşme ve araç sahipliği oranı arttıkça kaza sayılarında düzenli artış eğilimi olduğunu vurgulamaktadır.



Şekil 6. Kazaların Günün Saat Dilimine Göre Dağılımı

Şekil 6’da kazaların zamansal dağılımı incelendiğinde, en yüksek oranların iş çıkışı ve akşam hareketliliğini kapsayan 16:00–19:59 (%29) aralığında yoğunlaştığı görülmektedir. Malatya ulaşım ağındaki kavşak merkezliliği ve sinyalizasyon yapısı, bu yoğun saatlerdeki kaza riskini teknik olarak desteklemektedir (Şahin, 2020; Kılıç, 2021).

Gece saatlerinde (00:00–06:00) kaza frekansı düşse de, literatürde bu dilimdeki olayların yüksek hız gibi nedenlerle daha şiddetli sonuçlar doğurduğu belirtilmektedir (Bayata, 2010; Moghaddam vd., 2011). Makine öğrenmesi temelli kaza şiddeti analizleri de zaman faktörünün ciddiyet üzerindeki bu belirleyici rolünü doğrulamaktadır (Sivasankaran & Balasubramanian, 2020; Kumeda vd., 2019). Sonuç olarak Malatya’daki kaza eğilimleri, özellikle trafik hacminin ve insan hareketliliğinin arttığı saatlerde makine öğrenmesi modellerinin tahmin gücünün önemini ortaya koymaktadır (Abdel-Aty & Pande, 2007).



Şekil 7. Kazaların İlçelere Göre Dağılımı

Şekil 7, Malatya ilinde meydana gelen trafik kazalarının ilçelere göre oranlarını göstermektedir.

Verilere göre kazaların %38,8'i Malatya Merkez ilçesinde yani kentin idari, ticari ve nüfus yoğunluğunun en yüksek olduğu bölgeye karşılık gelmektedir.

Yeşilyurt (%20,9) ve Battalgazi (%16,3) ilçeleri de önemli oranlarda kazaların yaşandığı alanlardır. Bu üç ilçe birlikte, toplam kazaların yaklaşık %76'sını oluşturmaktadır.

Malatya Merkez, Yeşilyurt ve Battalgazi bölgeleri, hem ana arterlerin (örneğin İnönü Caddesi, Ankara-Kayseri D-300 karayolu) hem de yoğun yerleşim alanlarının bulunduğu kentsel ulaşım bölgelerini barındırmaktadır.

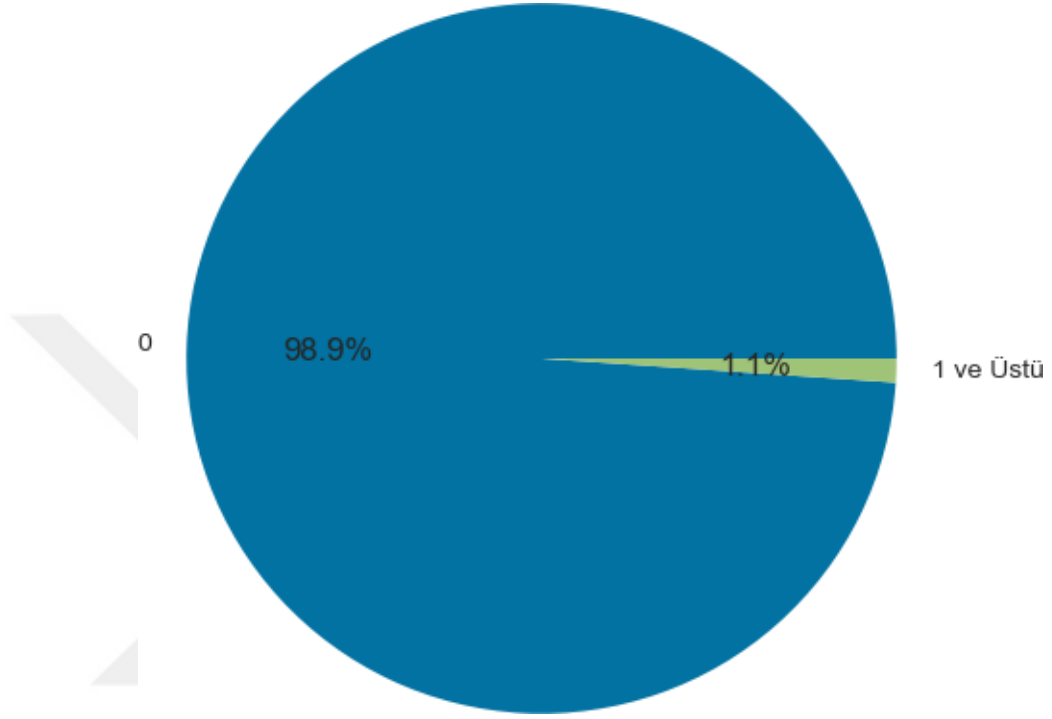
Dolayısıyla kazaların bu bölgelerde yoğunlaşması, araç yoğunluğu, yaya trafiği, sinyalize kavşak sayısı ve ticari hareketlilik ile doğrudan ilişkilidir.

Diğer ilçeler olan Doğanşehir (%5,9), Akçadağ (%5,9), Darende (%4,9) ve Hekimhan (%2,4)'da kaza oranları belirgin biçimde daha düşüktür.

Bu durum, kırsal karakterli ilçelerdeki araç trafiğinin düşük yoğunlukta olması ve yol ağının daha seyrek olmasından kaynaklanmaktadır.

Bu, kırsal alanlarda araç hızlarının daha yüksek olması ve yol güvenliği önlemlerinin yetersizliğinden kaynaklanmaktadır.

Genel olarak, Malatya'daki trafik kazalarının şehir merkezine yakın kentsel ulaşım bölgelerinde kümелendiği, trafik hacminin kazaların mekânsal dağılımını belirleyen temel unsur olduğu görülmektedir.



Şekil 8. Kazaların Toplam Ölü Sürücü Sayısına Göre Dağılımı

Şekil 8’de, Malatya ilindeki trafik kazalarında ölen sürücü sayısına göre dağılım yer almaktadır. Verilere göre kazaların %98,9’u sürücü ölümü olmadan gerçekleşirken, yalnızca %1,1’inde bir veya daha fazla sürücünün hayatını kaybettiği görülmektedir.

Bu oran, Malatya’daki kazaların büyük çoğunluğunun ölümsüz ancak yaralanmalı kazalar olduğunu doğrulamaktadır.

Bu bulgu, daha önceki istatistiksel tablolarda da gözlemlenen —yaralanmalı kaza oranı (%96) verisiyle uyumludur.

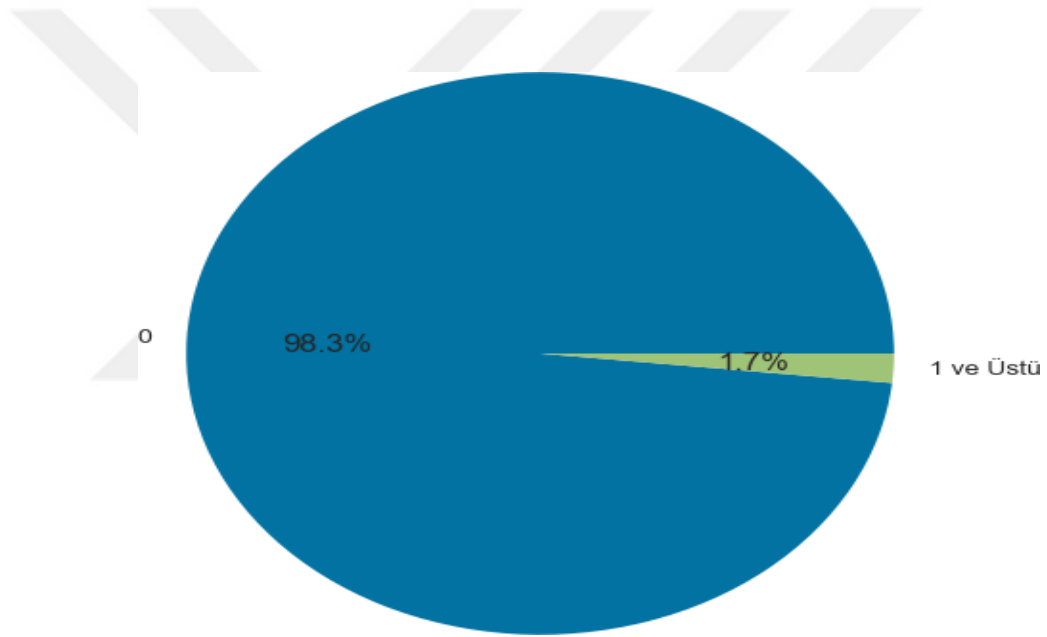
Sürücü ölümlerinin düşük oranı, şehir içi kazaların genellikle orta hız aralıklarında (ortalama 65 km/s) gerçekleşmesiyle açıklanabilir.

Bununla birlikte, bu küçük orandaki ölüm vakalarının çoğu, kırsal yol kesimlerinde veya yüksek hızlı çevre yollarında meydana gelen kazalardan kaynaklanmaktadır.

Literatürde yapılan çalışmalarda, kent içi kazalarda ölüm oranlarının düşük olmasına karşın, kırsal ve bölünmemiş yol tiplerinde ölüm oranlarının anlamlı biçimde arttığını göstermektedir.

Sonuç olarak, Malatya özelinde elde edilen bu dağılım, trafik güvenliği politikalarının sürücü eğitime ve dikkat odaklı önlemlere yöneltilmesinin önemini vurgulamaktadır.

Zira kazaların büyük çoğunluğu ölümle sonuçlanmasa da, sürücü davranışlarının kaza sıklığında belirleyici faktör olduğu anlaşılmaktadır.



Şekil 9. Kazaların Toplam Ölü Yolcu Sayısına Göre Dağılımı

Şekil 9, Malatya ilinde meydana gelen trafik kazalarında ölen yolcu sayısının dağılımını göstermektedir.

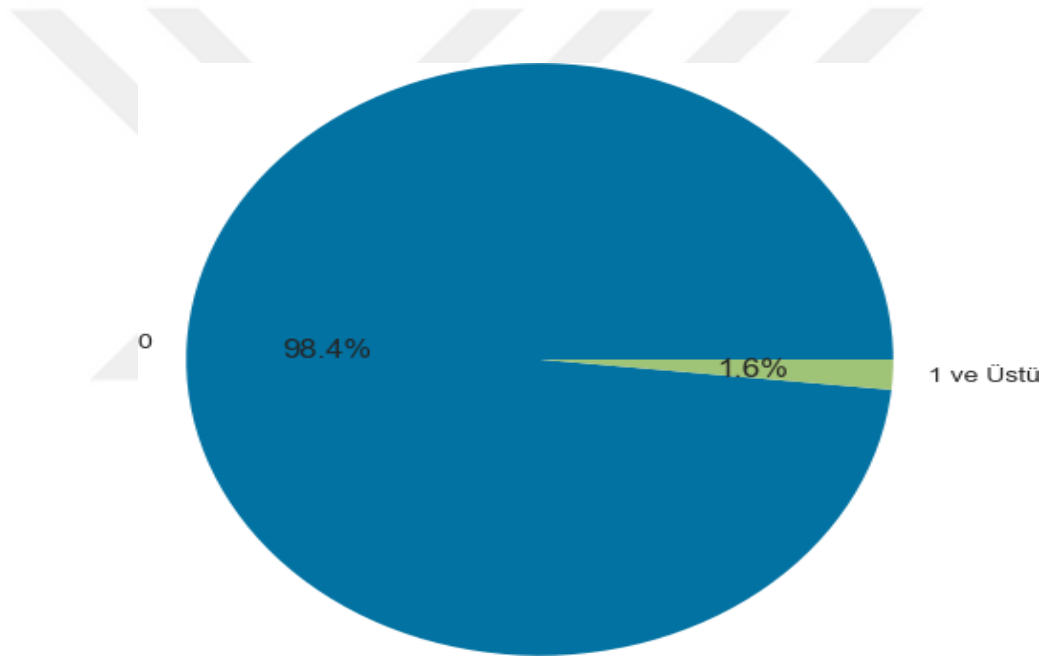
Verilere göre kazaların %98,3'ünde yolcu ölümü gerçekleşmemiş, yalnızca %1,7'sinde bir veya daha fazla yolcunun hayatını kaybettiği tespit edilmiştir.

Bu bulgu, kaza sonuçlarının büyük ölçüde ölümsüz ancak yaralanmalı kazalar şeklinde gerçekleştiğini bir kez daha doğrulamaktadır.

Yolcu ölümlerinin düşük oranı, kent içi ulaşım hatlarında hızların görece düşük olmasıyla açıklanabilir. Ancak, özellikle Malatya çevre yolu (D-300 karayolu) gibi yoğun araç akışına sahip bölünmüş yollarda, sürücü hataları yüksek risk potansiyeli taşımaya devam etmektedir (Malatya Star, 2024; Kılıç, 2021). Literatürdeki çalışmalar, hız limitlerinin aşıldığı durumlarda kaza şiddetinin ve ölüm oranlarının belirgin şekilde arttığını kanıtlamaktadır (Bayata, 2010; Sivasankaran & Balasubramanian, 2020).

Malatya özelinde gözlemlenen bu veriler, kaza sonuçlarının tahmin edilmesinde makine öğrenmesi modellerinin değişken hassasiyetine olan ihtiyacı vurgulamaktadır. Sonuç olarak, yerel düzeyde "sıfır ölüm" yaklaşımının benimsenmesi ve denetimlerin bu doğrultuda sıklaştırılması, stratejik hedefler açısından kritik bir öneme sahiptir.

Bu nedenle, trafik denetimlerinde hız kontrolü ve yolcu güvenliği bilincinin artırılması, ölüm oranlarının düşürülmesinde kritik rol oynamaktadır.



Şekil 10. Kazaların Toplam Ölü Yaya Sayısına Göre Dağılımı

Şekil 10'da Malatya ilinde meydana gelen trafik kazalarında hayatını kaybeden yayaların oranı gösterilmektedir.

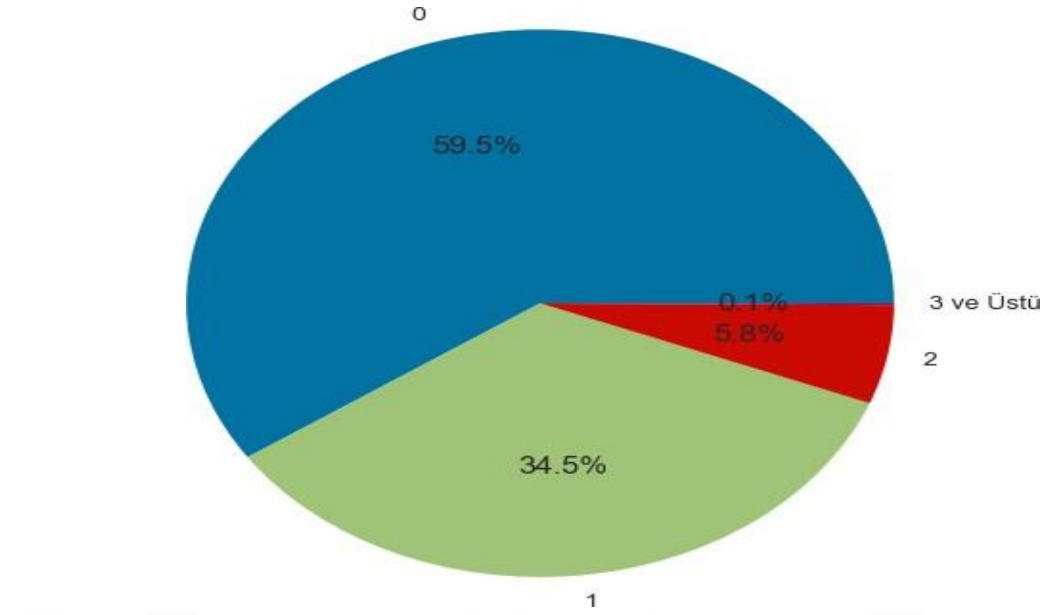
Verilere göre kazaların %98,4'ünde yaya ölümü yaşanmamış, yalnızca %1,6'sında bir veya birden fazla yayanın yaşamını yitirdiği tespit edilmiştir.

Bu oran, genel olarak ölümle sonuçlanan kazaların düşük bir kısmının yaya ölümleri ile ilişkili olduğunu göstermektedir.

Ancak yaya ölümlerinin oran olarak küçük olması, bu kazaların etkisinin düşük olduğu anlamına gelmemektedir. Malatya özelinde değerlendirildiğinde, bu durumun en çok Malatya Merkez, Yeşilyurt ve Battalgazi ilçelerinde, yani yaya ve araç etkileşiminin yoğun olduğu bölgelerde yaşandığı söylenebilir. Yaya geçitlerinin yetersizliği, hatalı karşıdan geçiş davranışları ve sürücülerin geçiş önceliği ihlalleri bu tür kazalarda belirleyici faktörlerdir. Ayrıca, kentteki yoğun toplu taşıma durakları, ticari alanlar ve okul çevreleri gibi bölgelerde yaya yoğunluğunun artması da kazaların mekânsal dağılımında etkili olmaktadır.

Bu durum, şehir merkezinde —trafik ışıklarının yetersiz sürelerle ayarlanması veya —yaya geçidi işaretlemelerinin görünürlüğünün düşük olması gibi mühendislik sorunlarıyla da ilişkilendirilebilir.

Sonuç olarak, yaya ölümleri oran olarak düşük olsa da, her bir vaka yüksek toplumsal etki yaratmakta; bu nedenle yaya güvenliğine yönelik kentsel ulaşım planlamalarının (örneğin sinyalizasyon kavşak iyileştirmeleri, hız sınırı düşürme, güvenli geçiş adaları vb.) güçlendirilmesi büyük önem taşımaktadır.



Şekil 11. Kazaların Toplam Yaralı Sürücü Sayısına Göre Dağılımı

Şekil 11, Malatya ilinde meydana gelen kazalarda yaralanan sürücü sayısının dağılımını göstermektedir.

Kazaların %59,5'inde sürücü yaralanmazken, %34,5'inde bir, %5,8'inde iki ve yalnızca %0,1'inde üç veya daha fazla sürücü yaralanmıştır.

Bu dağılım, sürücülerin kazalarda ölümden ziyade yaralanma riskiyle daha sık karşılaştığını ortaya koymaktadır.

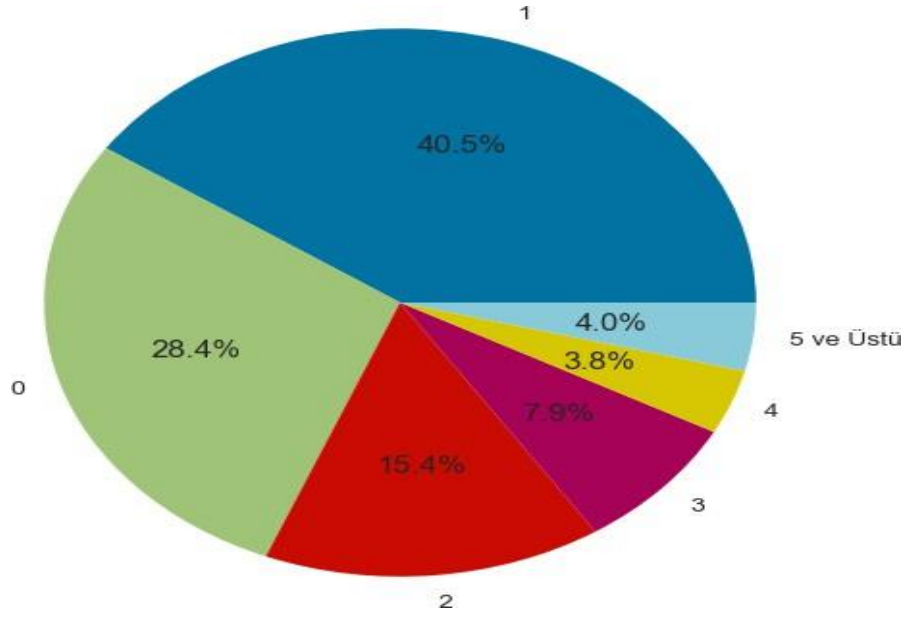
Yaralanma oranlarının yüksek olması, genellikle şehir içi hız limitlerinde gerçekleşen çarpışmaların etkisiyle ilişkilidir.

Ayrıca, Malatya'da sürücü yaralanmalarının en çok yoğun trafik akışına sahip ana arterlerde (örneğin D-300 ve İnönü Caddesi) meydana geldiği görülmüştür.

Bu durum, özellikle araç yoğunluğunun yüksek olduğu kavşak noktalarında çarpışma ve savrulma tipi kazaların ön plana çıktığını göstermektedir.

Kısaca, bu bulgu Malatya'da trafik kazalarının çoğunun yaralanmalı ve çoklu sürücü etkili yapıda olduğunu; ancak ölümlü sonuçlanma oranlarının görece düşük kaldığını göstermektedir.

Bu nedenle emniyet kemeri kullanımı, hız kontrolü ve sürücü dikkat eğitimleri sürücü yaralanmalarının azaltılmasında kilit rol oynamaktadır.



Şekil 12. Kazaların Toplam Yaralı Yolcu Sayısına Göre Dağılımı

Şekil 12, Malatya ilinde gerçekleşen kazalarda yaralanan yolcu sayısının oranlarını göstermektedir.

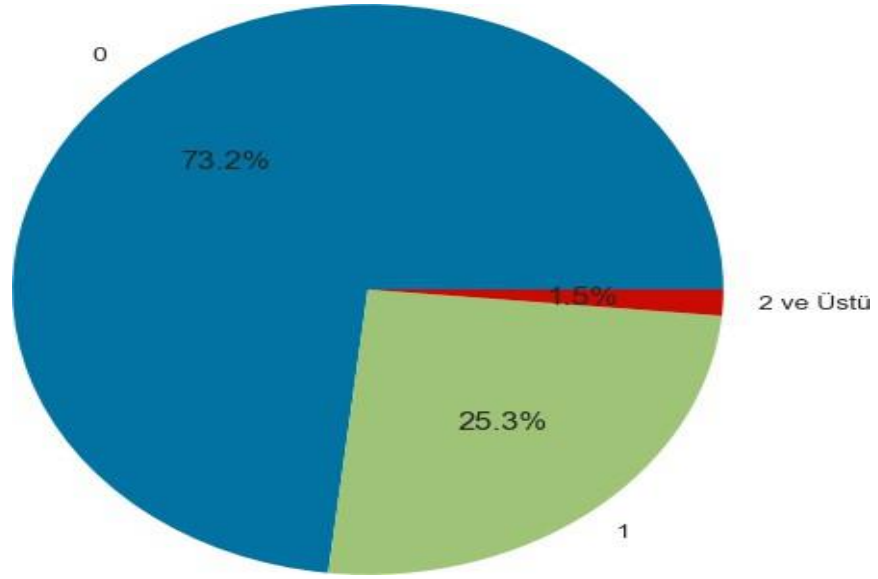
Kazaların %28,4'ünde yolcu yaralanmazken, %40,5'inde bir, %15,4'ünde iki, %7,9'unda üç, %3,8'inde dört ve %4,0'ında beş veya daha fazla yolcunun yaralandığı görülmektedir.

Bu dağılım, kazaların büyük kısmında en az bir yolcunun yaralandığını, dolayısıyla yolcuların kaza sonuçlarından ciddi şekilde etkilendiğini ortaya koymaktadır.

Yolcu yaralanmalarının yüksekliği, genellikle araç içi güvenlik önlemlerinin yetersizliği (emniyet kemeri kullanmama, yüksek hız, çarpışma açısı) gibi faktörlerle ilişkilidir.

Malatya'daki veriler, özellikle şehir içi ulaşım hatlarında ve yoğun kavşak bölgelerinde bu tür yaralanmaların sık görüldüğünü göstermektedir.

Yolcu yaralanmalarının bu denli yüksek oranlarda seyretmesi, taşıt güvenliği farkındalığının ve denetimlerinin artırılmasının önemine işaret etmektedir.



Şekil 13. Kazaların Toplam Yaralı Yaya Sayısına Göre Dağılımı

Şekil 13, Malatya'daki kazalarda yaralanan yaya sayısının oranlarını göstermektedir.

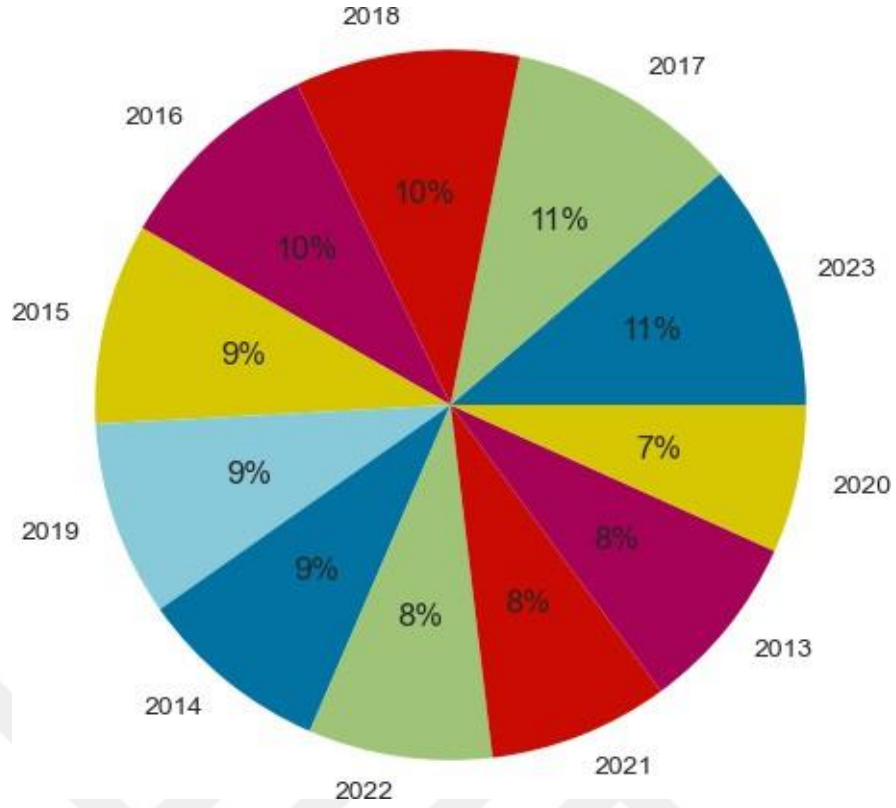
Kazaların %73,2'sinde yaya yaralanmazken, %25,3'ünde bir, %1,5'inde ise iki veya daha fazla yaya yaralanmıştır.

Yaya yaralanmalarının çoğu, özellikle şehir merkezlerinde, kavşak yakınlarında ve karşıdan geçiş alanlarında meydana gelmektedir.

Malatya özelinde, bu durum Yeşilyurt ve Battalgazi gibi yaya yoğunluğu yüksek ilçelerde daha belirgindir.

Bu bölgelerde sürücülerin yaya geçidi önceliğine uymaması, hız limitlerinin aşılması ve yetersiz sinyalizasyon sistemleri, yaralanmalı kazaların temel nedenleri arasındadır.

Dolayısıyla bu bulgular, şehir içi planlamada yaya odaklı güvenlik önlemlerinin artırılması gerektiğini ortaya koymaktadır.



Şekil 14. Kazaların Yıllara Göre Dağılımı

Şekil 14, Malatya ilinde 2013–2023 yılları arasında meydana gelen trafik kazalarının yıllara göre oranlarını göstermektedir.

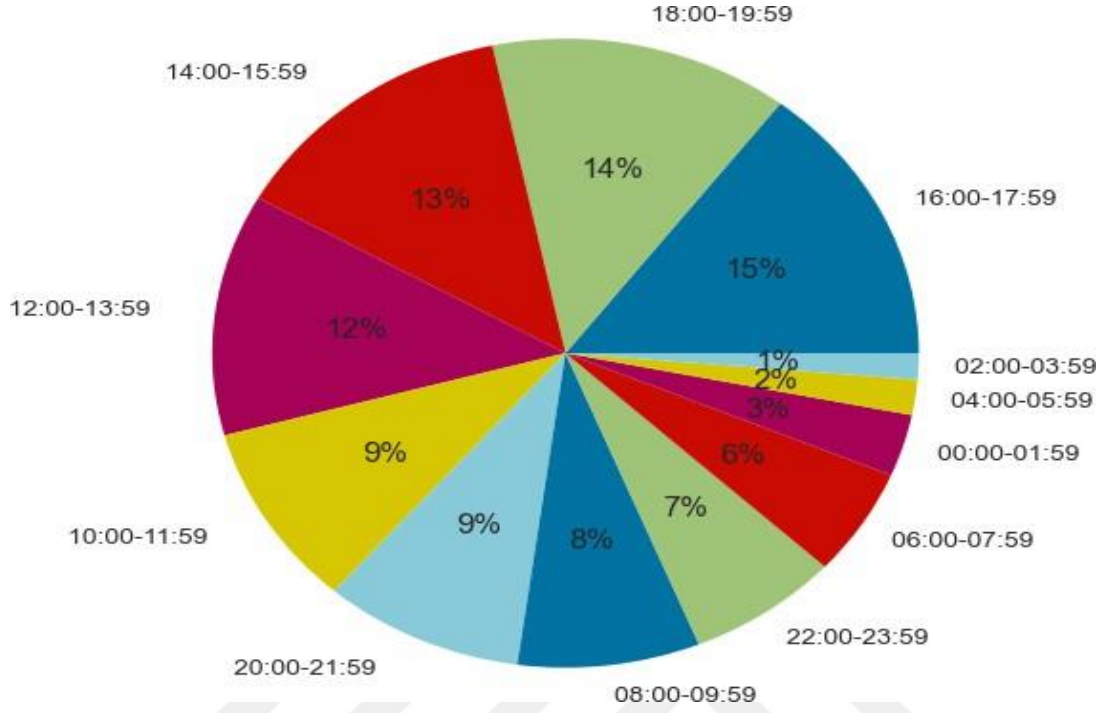
Kazalar yıllara göre yaklaşık olarak dengeli bir dağılım göstermektedir. 2017 ve 2023 yılları %11’lik oranla en yüksek kaza payına sahip yıllar olurken, 2020 yılı %7 ile en düşük orana sahiptir.

2020 yılındaki düşüş, COVID-19 pandemisi döneminde uygulanan kısıtlamalar nedeniyle trafiğin azalmasıyla açıklanabilir. 2018, 2016 ve 2015 yıllarındaki oranların birbirine yakın seyretmesi (%9–10 bandında) ise bu yıllarda trafiğin genel yoğunluğunda önemli bir değişiklik yaşanmadığını göstermektedir.

2021 sonrası dönemde kaza oranlarının tekrar artış eğilimine geçtiği gözlenmektedir. Bu durum, pandemi sonrası araç kullanımının ve şehir içi hareketliliğin artmasıyla ilişkilendirilebilir.

Sonuç olarak, 2013–2023 arası dönemde Malatya’daki trafik kazalarının belirgin bir azalma veya artış trendi göstermediği, ancak 2020’de geçici bir düşüş yaşandığı görülmektedir. Bu

durum, trafik yoğunluğu, toplu taşıma tercihleri ve şehir içi planlama değişkenleri ile doğrudan ilişkilidir.



Şekil 15. Kazaların Günün Saat Dilimine Göre Dağılımı

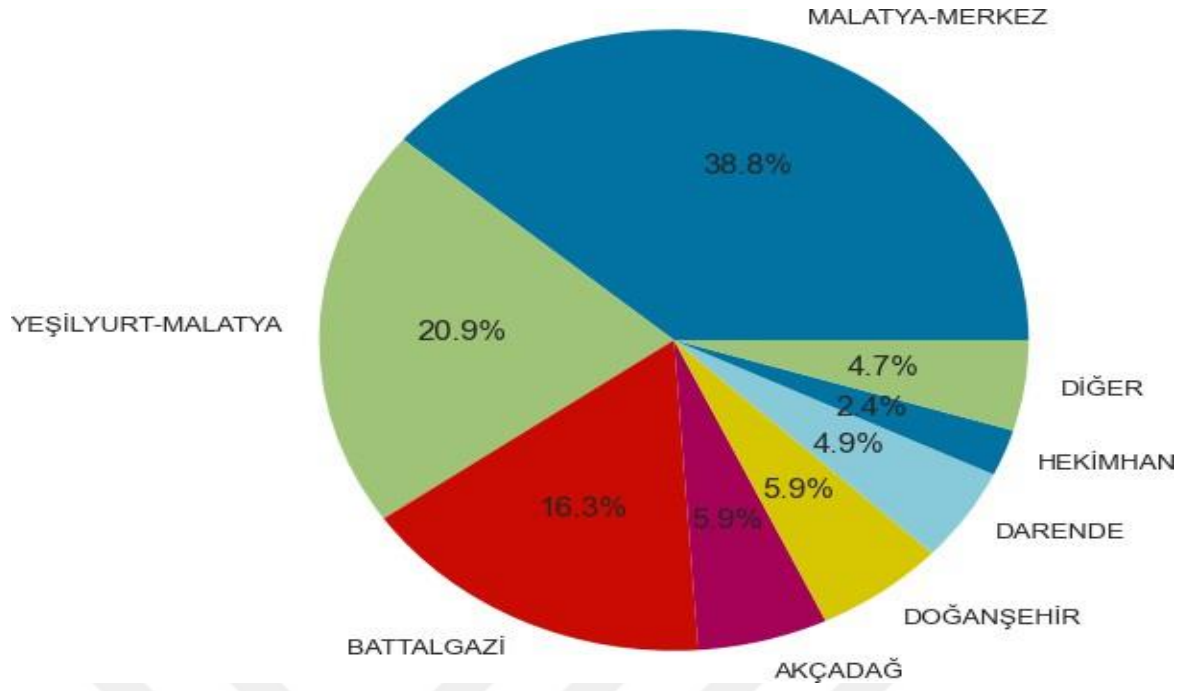
Şekil 15, Malatya ilinde meydana gelen trafik kazalarının günün saat dilimlerine göre oranlarını göstermektedir.

Kazaların en yoğun yaşandığı zaman aralıkları 16:00–17:59 (%15) ve 18:00–19:59 (%14) saat dilimleridir. Bu saatler, özellikle iş çıkış trafiğinin en yoğun olduğu dönemlere denk gelmektedir.

Ayrıca, 14:00–15:59 (%13) ve 12:00–13:59 (%12) aralıklarında da kaza oranlarının yüksek seyretmesi, gün ortasında artan taşıt ve yaya hareketliliğinin etkisini ortaya koymaktadır.

Buna karşılık, gece yarısı (00:00–05:59) aralığında kaza oranları oldukça düşüktür (%1–3).

Bu durum, düşük trafik hacmiyle birlikte gece saatlerinde yapılan seyahatlerin sayıca az olmasından kaynaklanmaktadır.



Şekil 16. Kazaların İlçelere Göre Dağılımı

Şekil 16, Malatya ilinde 2013–2023 yılları arasında meydana gelen trafik kazalarının ilçelere göre yüzdelik dağılımını göstermektedir.

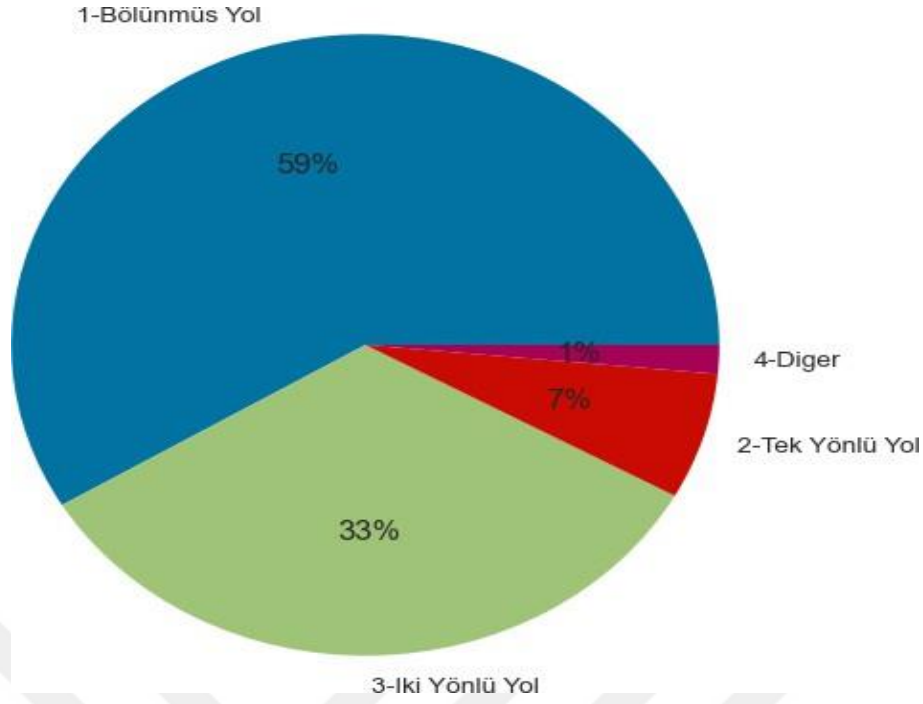
Elde edilen bulgulara göre, kazaların %38,8'i Malatya Merkez, %20,9'u Yeşilyurt, %16,3'ü Battalgazi ilçelerinde meydana gelmiştir. Bu üç ilçe, toplam kazaların yaklaşık %76'sını oluşturarak kentsel yoğunluğun kazalar üzerindeki etkisini açıkça ortaya koymaktadır.

Malatya Merkez ve Yeşilyurt ilçelerinin yüksek oranları, bu bölgelerdeki araç trafiği, nüfus yoğunluğu ve ticari faaliyetlerin fazlalığı ile ilişkilidir.

Özellikle Yeşilyurt ilçesinde artan konut alanları ve yeni ulaşım aksları (örneğin D-300 karayolu bağlantısı), son yıllarda trafik yoğunluğunu artırmıştır.

Buna karşılık, Doğanşehir (%5,9), Akçadağ (%5,9), Darende (%4,9) ve Hekimhan (%2,4) gibi ilçelerde kaza oranlarının daha düşük olduğu görülmektedir. Bu bölgelerdeki trafik hacmi daha sınırlı, yol ağı daha az karmaşık ve yerleşim yoğunluğu düşüktür.

Sonuç olarak, trafik kazalarının ilçelere göre dağılımı şehirleşme düzeyi, yol ağı kapasitesi ve ulaşım talebi gibi faktörlerle doğrudan ilişkilidir.



Şekil 17. Kazaların Yol Şerit Tipine Göre Dağılımı

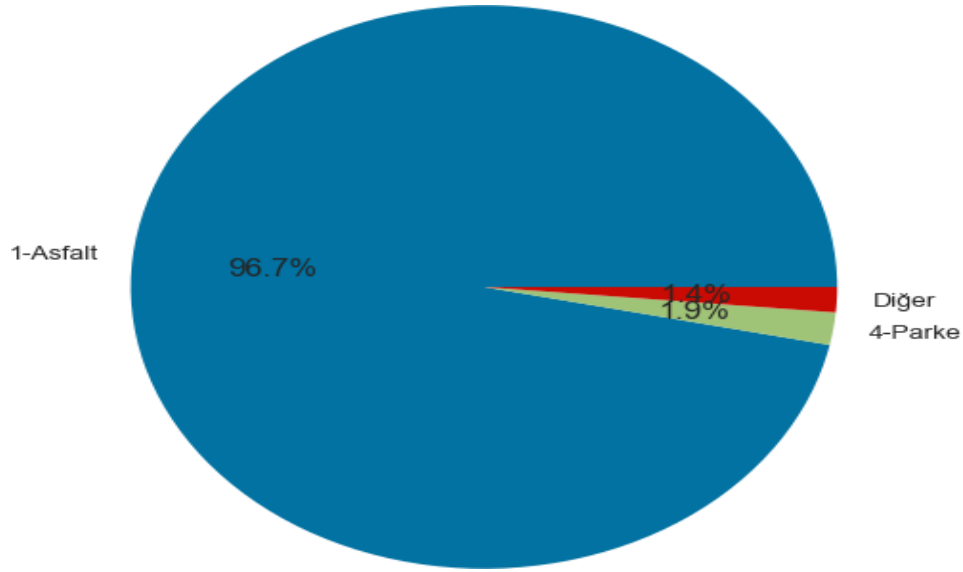
Şekil 17, Malatya ilinde meydana gelen trafik kazalarının yol şerit tiplerine göre dağılımını göstermektedir.

Verilere göre kazaların %59'u bölünmüş yollarda, %33'ü iki yönlü yollarda, %7'si tek yönlü yollarda ve %1'i diğer yol türlerinde meydana gelmiştir.

Bölünmüş yolların kaza oranının yüksek olması, bu yolların şehir içi ve şehirlerarası ulaşımında en çok kullanılan arterler olmasından kaynaklanmaktadır. Özellikle D-300 ve D-850 karayolu güzergâhları üzerinde yoğun taşıt trafiği, bu tür kazaların sıklığını artırmaktadır. Bununla birlikte, bölünmüş yollarda meydana gelen kazalar genellikle yüksek hız kaynaklı olup, şerit ihlali, takip mesafesi yetersizliği ve sürücü dikkatsizliği gibi faktörlerden etkilenmektedir.

İki yönlü yollarda (%33) görülen kazalar ise çoğunlukla karşı şeritten gelen araçlarla çarpışma veya sollama hatası sonucu gerçekleşmektedir. Bu tür yolların genellikle kırsal kesimlerde bulunması ve şerit ayrımının net olmaması, çarpışma riskini artırmaktadır. Tek yönlü yollar (%7) ise dar cadde ve sokaklarda yer almakta olup, genellikle düşük hızlı çarpışma ve yaya kazalarıyla ilişkilidir.

Sonuç olarak, kazaların büyük bir kısmı bölünmüş yollarda meydana gelmekte; bu da hız denetimi, yol çizgi düzenlemeleri ve akıllı ulaşım sistemlerinin kullanımının önemini bir kez daha ortaya koymaktadır.



Şekil 18. Kazaların Yol Kaplama Türlerine Göre Dağılımı

Şekil 18, Malatya ilinde meydana gelen trafik kazalarının yol kaplama türlerine göre dağılımını göstermektedir.

Elde edilen bulgulara göre kazaların %96,7'si asfalt kaplamalı yollarda, %1,9'u parke taşlı yollarda, %1,4'ü ise diğer kaplama türlerinde gerçekleşmiştir.

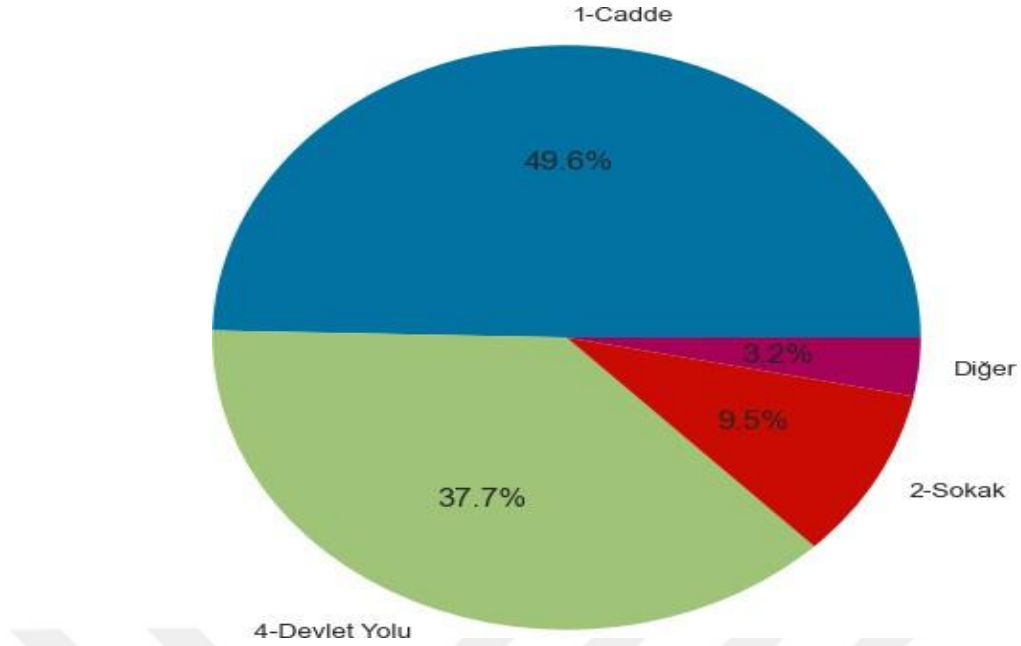
Asfalt kaplamalı yolların yüksek oranda öne çıkmasının temel nedeni, Malatya il sınırları içerisindeki karayolu ağının büyük ölçüde asfalt malzemeden oluşmasıdır. Özellikle şehir içi ve şehirlerarası ulaşımında kullanılan ana arterlerin tamamına yakını asfalt yüzeye sahiptir.

Ancak bu durum, asfalt yolların güvenli olduğu anlamına gelmemektedir. Asfalt yollar, yüksek hızda araç kullanımı, fren mesafesinin artması ve yağışlı havalarda kayma riski nedeniyle kazalara zemin hazırlayabilmektedir.

Özellikle düzgün zeminli bölünmüş yollarda sürücüler hız limitlerini aşma eğiliminde olduğundan, bu durum kaza riskini artırmaktadır.

Parke taşlı ve diğer türdeki yolların kaza oranının düşük olması, bu yolların genellikle yerleşim içi, düşük hız limitli bölgelerde bulunmasıyla açıklanabilir.

Sonuç olarak, Malatya genelinde trafik kazalarının büyük oranda asfalt kaplamalı yollarda meydana gelmesi, yol yüzeyinin fiziksel özelliklerinden ziyade sürücü davranışlarının kaza oluşumundaki rolünün daha baskın olduğunu göstermektedir.



Şekil 19. Kazaların Yol Sınıfına Göre Dağılımı

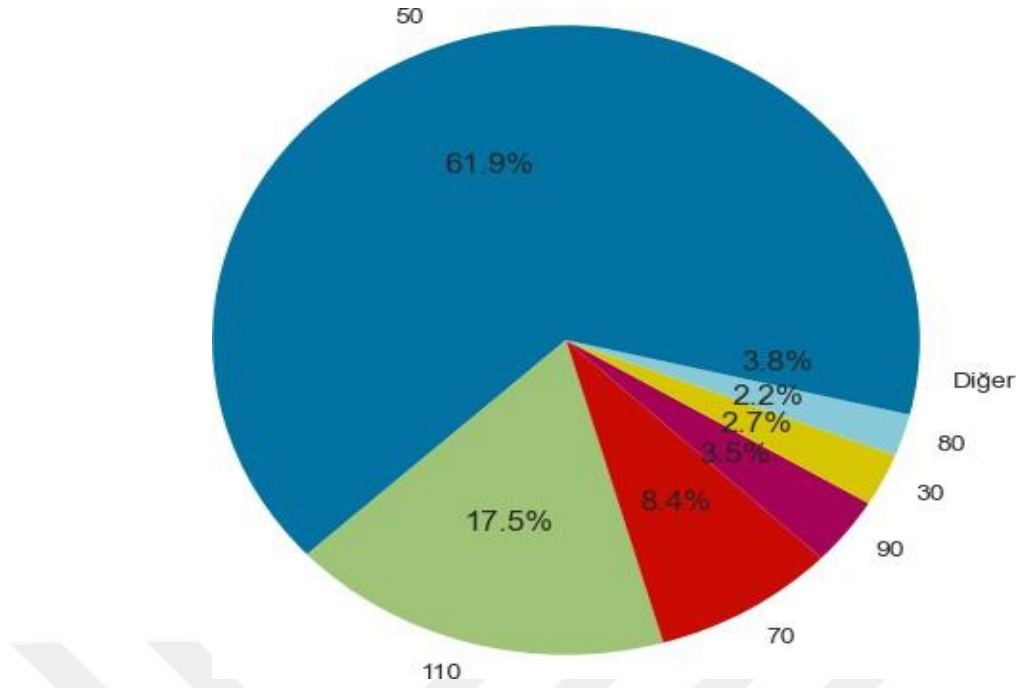
Şekil 19, Malatya ilinde meydana gelen trafik kazalarının yol sınıflarına göre dağılımı gösterilmektedir.

Elde edilen verilere göre kazaların %49,6'sı caddelerde, %37,7'si devlet yollarında, %9,5'i sokaklarda ve %3,2'si diğer yol türlerinde gerçekleşmiştir.

Cadde ve devlet yolları toplamda kazaların yaklaşık %87'sini oluşturmaktadır. Bu durum, özellikle kent içi ulaşım yoğunluğunun fazla olduğu bölgelerde (Malatya Merkez, Battalgazi ve Yeşilyurt) sürücülerin hem araç hem de yaya trafiğiyle daha fazla etkileşime girmesiyle açıklanabilir.

Caddelerde meydana gelen kazalar çoğunlukla dikkatsizlik, hız ihlali ve yaya geçidi ihlali gibi nedenlerden kaynaklanmaktadır. Devlet yollarında görülen kazalar ise genellikle yüksek hızlı taşıt trafiği, sollama hataları ve uzun mesafeli sürüşlerde dikkatin dağılması sonucu meydana gelmektedir. D-300 ve D-850 güzergâhları üzerinde yoğunlaşan bu kazalar, şehirlerarası ulaşımın güvenlik risklerini de ortaya koymaktadır. Sokaklarda (%9,5) meydana gelen kazalar, düşük hızla seyreden araçların karıştığı park manevrası, geri geri gitme, yaya çarpması veya kontrolsüz kavşak girişleri gibi olaylardan oluşmaktadır.

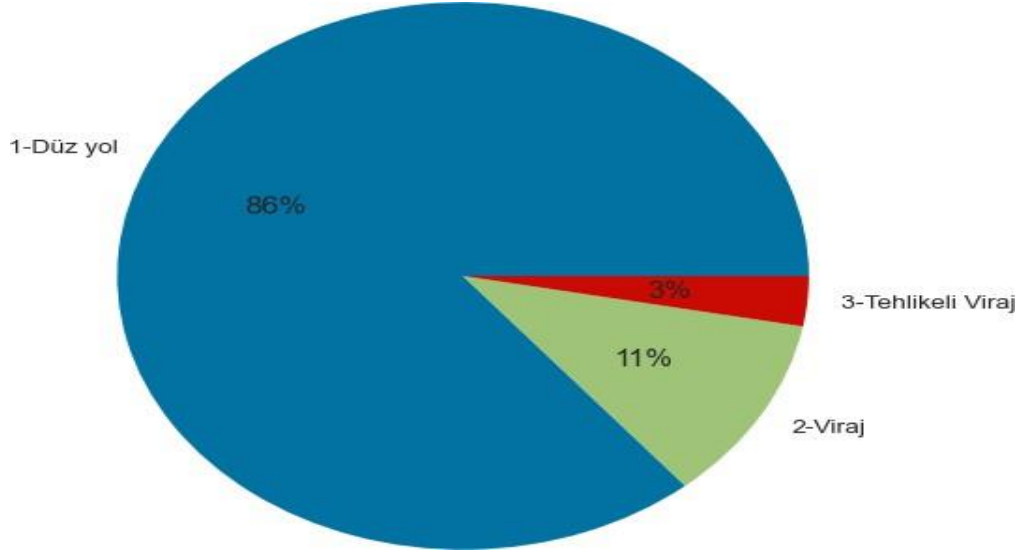
Sonuç olarak, kazaların büyük çoğunluğunun cadde ve devlet yollarında yoğunlaşması, Malatya kentinde trafik güvenliği planlamasının ana arterlere odaklanması gerektiğini göstermektedir. Ayrıca, yol sınıfına bağlı olarak farklı önleyici stratejilerin geliştirilmesi, kaza sıklığını azaltmada etkili olabilir.



Şekil 20. Kaza Yapılan Yolların Yasal Hız Limitine Göre Dağılımı

Şekil 20, Malatya ilinde meydana gelen trafik kazalarının, kaza yapılan yolların yasal hız limitlerine göre dağılımını göstermektedir.

Kazaların önemli bir kısmının 50 km/s hız sınırına sahip yollarda gerçekleşmesi, bu yolların büyük ölçüde şehir içi ulaşım ağında yer almasından kaynaklanmaktadır. Malatya kent merkezinde, özellikle Battalgazi ve Yeşilyurt ilçelerinde yoğunlaşan araç trafiği, trafik ışıkları, yaya geçitleri, kavşaklar ve park manevraları gibi şehir içi etkenlerle birleştiğinde kaza olasılığını artırmaktadır. Sonuç olarak, kazaların büyük bölümünün orta hız limitli (özellikle 50 km/s) yollarda gerçekleşmesi, kent içi trafik güvenliği önlemlerinin güçlendirilmesi, hız denetim sistemlerinin artırılması ve yaya güvenliği odaklı stratejilerin uygulanması gerektiğini ortaya koymaktadır.



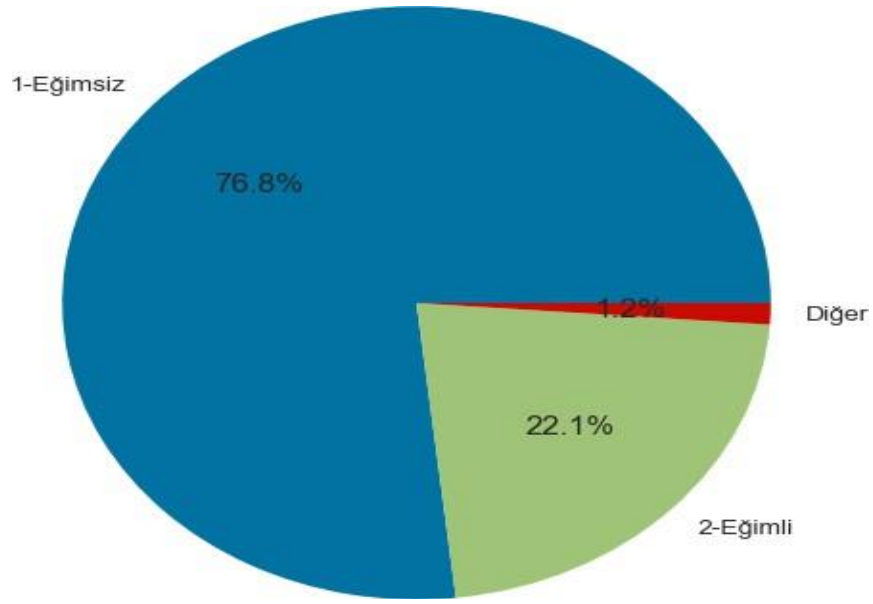
Şekil 21. Kazaların GeoYatay Özelliklerine Göre Dağılımı

Şekil 21, Malatya ilinde meydana gelen trafik kazalarının, yolların yatay geometrik özelliklerine (düz yol, viraj, tehlikeli viraj) göre dağılımını göstermektedir.

Elde edilen bulgulara göre kazaların %86'sı düz yollarda, %11'i virajlı yollarda, %3'ü ise tehlikeli virajlarda meydana gelmiştir.

Kazaların büyük çoğunluğunun düz yollarda gerçekleşmiştir. Bu durum genellikle sürücülerin dikkat dağınıklığı, aşırı hız, takip mesafesi ihlali ve cep telefonu kullanımı gibi davranışsal faktörlerle açıklanmaktadır. Düz yol kesimlerinde sürücüler genellikle daha az dikkatli olma eğilimindedir; bu da özellikle uzun mesafeli sürüşlerde kazaların artmasına neden olmaktadır. Virajlı yollar (%11) ve tehlikeli virajlar (%3) üzerindeki kaza oranları daha düşük olsa da, bu tür kesimlerde meydana gelen kazalar genellikle araç hakimiyetinin kaybı, yol eğimi, kaygan zemin veya görüş kısıtlılığı gibi faktörlerden kaynaklanmakta ve çoğu zaman daha ciddi hasar ve yaralanma ile sonuçlanmaktadır.

Bu sonuçlar, Malatya ili genelinde yol güvenliği politikalarının yalnızca virajlı kesimlere değil, aynı zamanda uzun ve düz yol segmentlerinde sürücü farkındalığını artırmaya yönelik önlemleri de içermesi gerektiğini göstermektedir. Özellikle düz yollarda hız sınırı hatırlatma levhaları, dinlenme alanı bilgilendirmeleri ve mobil hız kontrol noktaları kazaları azaltmada etkili olabilir.



Şekil 22. Kaza Yapılan Yolların Düşey Eğimine Göre Dağılımı

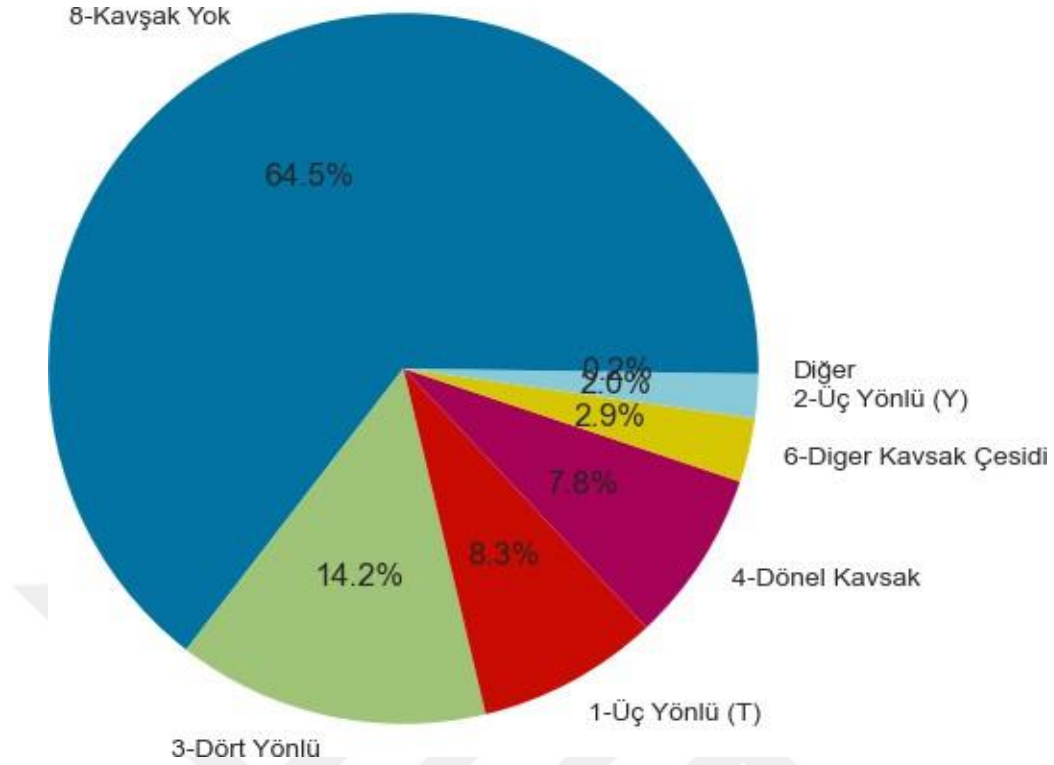
Şekil 22, Malatya ilinde meydana gelen trafik kazalarının yolların düşey eğim durumuna göre dağılımı gösterilmektedir.

Verilere göre kazaların %76,8'i eğimsiz, %22,1'i eğimli ve %1,2'si diğer eğim türlerinde (örneğin değişken eğim veya yol kotu geçişleri) meydana gelmiştir.

Kazaların önemli bir kısmının eğimsiz yollarda gerçekleştiği görülmektedir. Bu durum, eğimsiz yolların genellikle şehir merkezlerinde ve düz topoğrafyalı alanlarda bulunmasından kaynaklanmaktadır. Bu tür yollar, yüksek trafik hacmine ve yaya yoğunluğuna sahiptir; bu da kaza olasılığını artırmaktadır. Ayrıca sürücülerin düz kesimlerde dikkati azalmakta, hız kontrolünü ihmal etme eğilimi artmaktadır.

Eğimli yollar (%22,1) üzerinde gerçekleşen kazalar ise çoğunlukla aşırı hız, fren mesafesinin uzaması, yetersiz çekiş, yol tutuşunun azalması veya viraj içi kontrol kaybı gibi nedenlerle meydana gelmektedir. Özellikle Doğanşehir ve Akçadağ ilçeleri gibi eğimli topoğrafyaya sahip bölgelerde bu tür kazalar daha sık görülmektedir.

Sonuç olarak, kazaların büyük bölümünün eğimsiz kesimlerde gerçekleşmesi, trafik yoğunluğu ve sürücü dikkatsizliğinin kaza oluşumunda eğimden daha baskın faktörler olduğunu göstermektedir. Ancak eğimli kesimlerdeki kazalar genellikle daha ciddi sonuçlar doğurduğundan, bu tür bölgelerde uyarı levhaları, hız sınırlamaları ve yüzey kaplama iyileştirmeleri gibi önleyici tedbirler alınmalıdır.

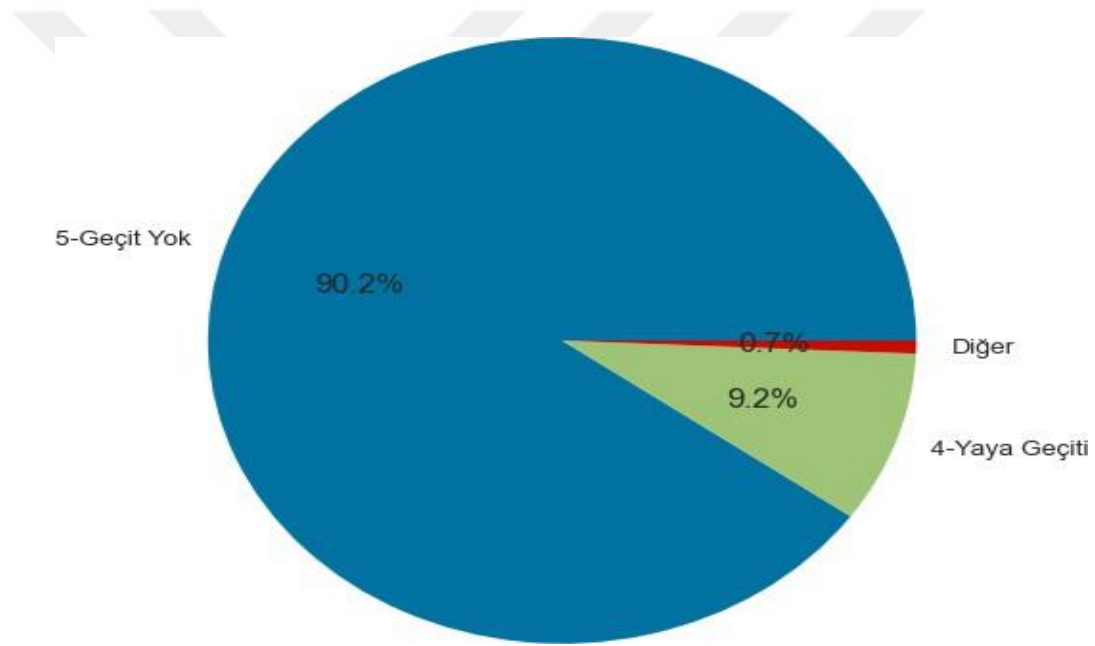


Şekil 23. Kaza Kavşak Tipi Dağılımı

Şekil 23, Malatya ilinde meydana gelen trafik kazalarının kavşak bulunup bulunmamasına ve kavşak türüne göre dağılımı incelendiğinde, kazaların %64,5'inin kavşak bulunmayan yol kesimlerinde gerçekleştiği; kavşaklı alanlarda ise kazaların %14,2'sinin dört kollu kavşaklarda, %8,3'ünün T tipi kavşaklarda, %7,8'inin döner kavşaklarda ve yaklaşık %5'inin diğer kavşak türlerinde meydana geldiği belirlenmiştir. Kavşak bulunmayan yol kesimlerinde kazaların yüksek oranlarda görülmesi; özellikle şehir içi ana arterlerde hız ihlalleri, uzun ve kesintisiz yol kesimlerinin sürücülerde algılanan risk düzeyini düşürmesi ve dikkat dağınıklığı gibi faktörlerle ilişkilendirilmektedir. Literatürde, bu tür yol kesimlerinde takip mesafesi ihlalleri ve sürücü dikkatsizliğinin kaza oluşumunda belirleyici olduğu vurgulanmaktadır (Özdemir & Yıldız, 2020). Kavşaklarda meydana gelen kazalar ise daha çok geçiş önceliği ihlalleri, yetersiz görüş mesafesi, sinyalizasyon eksiklikleri ve kavşak yaklaşım hızlarının yeterince düşürülmemesi gibi manevra kaynaklı nedenlerden ortaya çıkmaktadır. Benzer şekilde, kavşak kaynaklı kazaların özellikle sürücülerin karar verme süreçleri ve algısal hatalarıyla doğrudan ilişkili olduğu önceki çalışmalarda belirtilmiştir (Korkmaz & Karabulut, 2018). Döner kavşaklarda kazaların görece daha düşük oranlarda gerçekleşmesi, bu kavşak türlerinin trafik akışını düzenlemede etkili

olabildiğini göstermektedir. Ancak sürücü alışkanlıkları, yönlendirme ve işaretleme yetersizlikleri gibi etkenler, bu kavşak türlerinde de kaza riskinin tamamen ortadan kalkmadığını ortaya koymaktadır (Erdoğan & Gökdağ, 2019).

Genel olarak değerlendirildiğinde, Malatya'daki trafik kazalarının niceliksel olarak büyük bölümü kavşak dışı kesimlerde meydana gelmekle birlikte, kavşak içi kazaların daha karmaşık manevralar ve sürücü hatalarıyla ilişkili olduğu görülmektedir. Bu durum, özellikle kavşak ve yaklaşım bölgelerinde yatay ve düşey işaretlemelerin güçlendirilmesi, hız yönetimi uygulamalarının yaygınlaştırılması ve sinyalizasyon sistemlerinin etkin biçimde düzenlenmesinin trafik güvenliği açısından kritik öneme sahip olduğunu göstermektedir.



Şekil 24. Kaza Geçit Dağılımı

Şekil 24, Malatya ilinde meydana gelen trafik kazalarının, yaya geçidi veya diğer geçit türlerinin varlığına göre nasıl dağıldığını göstermektedir.

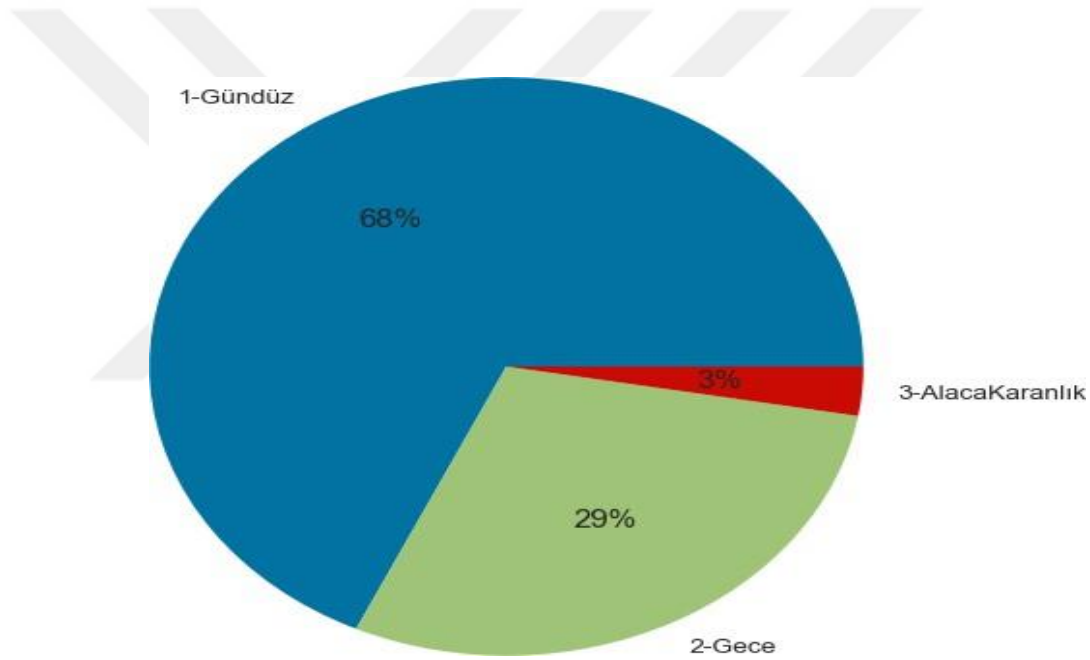
Verilere göre kazaların %90,2'si geçit bulunmayan kesimlerde, %9,2'si yaya geçitlerinde, %0,7'si ise diğer geçit türlerinde (örneğin üst geçit, alt geçit, servis yolu bağlantısı vb.) meydana gelmiştir.

Kazaların çok büyük bir kısmının geçit bulunmayan noktalarda gerçekleşmesi, yaya ve araç trafiğinin düzenlenmediği alanlarda kontrolsüz geçişlerin ve sürücü ihlallerinin yoğun

olduğunu göstermektedir. Bu durum, özellikle şehir merkezine uzak bölgelerdeki tali yollar ve yerleşim dışı alanlarda gözlemlenmektedir.

Öte yandan, yaya geçitlerinde meydana gelen %9,2'lik kaza oranı, mevcut geçitlerin her zaman güvenli kullanım sağlamadığını göstermektedir. Bu tür kazalar çoğunlukla sürücülerin yaya önceliğine dikkat etmemesi, yetersiz aydınlatma, görüş mesafesinin kısıtlı olması veya yaya davranışları (ani geçiş vb.) gibi nedenlerden kaynaklanmaktadır.

Bu bulgular, kent içi ulaşım planlamasında yaya güvenliğini artırıcı uygulamaların (örneğin ışıklı geçitler, hız kesici kasisler, yaya odaklı sinyalizasyon sistemleri, üst/alt geçitlerin etkin kullanımı) önemini açıkça ortaya koymaktadır. Ayrıca, özellikle yerleşim yeri dışı kesimlerde yaya geçidi olmayan noktalarda yaya hareketini sınırlayıcı fiziksel düzenlemeler yapılması da kazaları azaltmada etkili olacaktır.



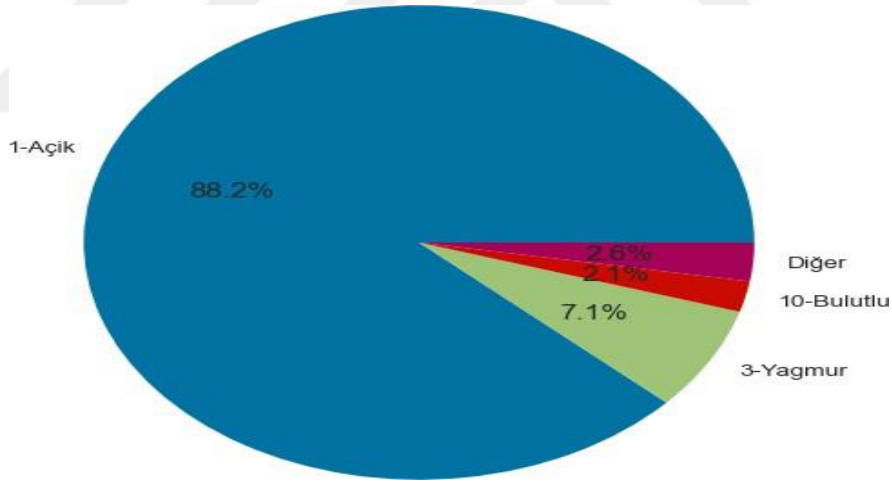
Şekil 25. Kaza Gün Durumu Dağılımı

Şekil 25' de, Malatya ilinde meydana gelen trafik kazalarının, günün aydınlık düzeyine göre dağılımı incelendiğinde, kazaların %68'inin gündüz, %29'unun gece ve %3'ünün alacakaranlık (güneşin batma veya doğma anına yakın düşük görüş koşulları) zaman dilimlerinde meydana geldiği görülmektedir.

Kazaların büyük çoğunluğunun gündüz saatlerinde gerçekleşmesi, trafikteki araç ve yaya yoğunluğu ile ulaşım talebinin bu saatlerde artmasıyla ilişkilendirilebilir (Aksoy & Kantar, 2017). Özellikle şehir merkezi ve ana arterlerde, iş, eğitim ve ticaret amaçlı seyahatlerin artması kaza riskini yükseltmektedir.

Gece meydana gelen trafik kazalarının oranı (%29) gündüz saatlerine kıyasla daha düşük olmakla birlikte, literatürde gece sürüş koşullarının kaza şiddeti açısından daha riskli olabildiği belirtilmektedir. Özellikle gece saatlerinde görüş mesafesinin azalması, aydınlatma koşullarının yetersizliği, sürücü yorgunluğu ve alkol kullanımının daha yaygın olması gibi etmenlerin kazaların sonuçlarını ağırlaştırabildiği ifade edilmektedir (Korkmaz & Karabulut, 2018). Alacakaranlıkta gerçekleşen kazalar (%3) kısa süreli olmasına rağmen, ani ışık değişimleri sürücü algısında kararsızlığa ve fren mesafesinde artışa sebep olmaktadır.

Bu bulgular, Malatya'da trafik güvenliği stratejilerinin yalnızca gündüz yoğunluğuna odaklanmaması; gece aydınlatması, sürücü farkındalığı ve reflektif işaretlemeler gibi önlemleri de kapsamı gerektiğini ortaya koymaktadır. Özellikle gece saatlerinde trafik denetimlerinin artırılması ve yol aydınlatmalarının güçlendirilmesi önerilmektedir.



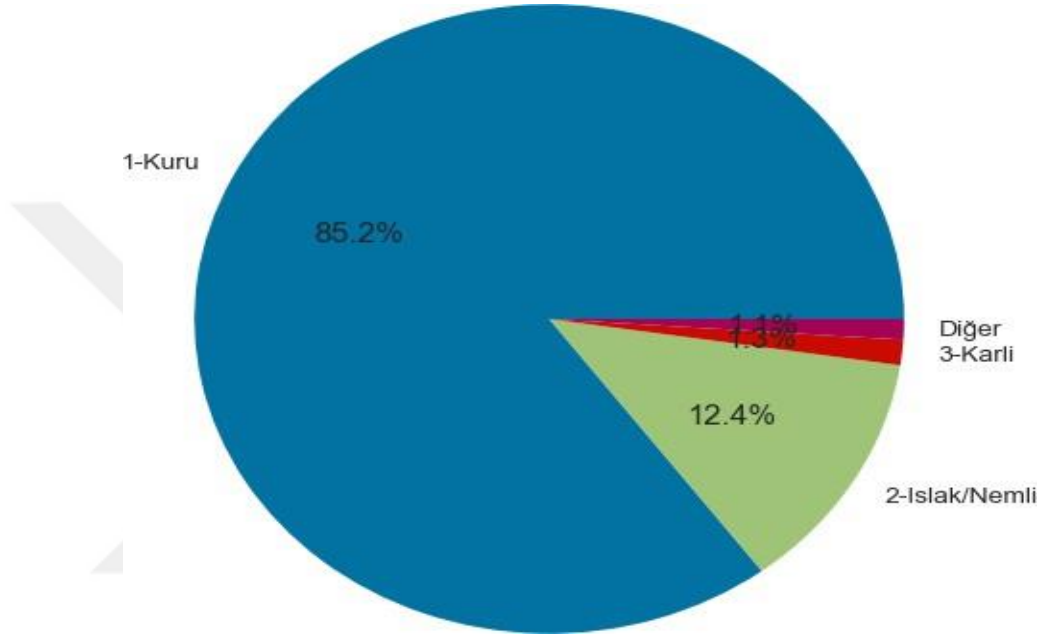
Şekil 26. Kaza Hava Durumu Dağılımı

Şekil 26, Malatya'da meydana gelen kazaların, olay anındaki hava koşullarına göre dağılımı sunulmaktadır. Verilere göre kazaların %88,2'si açık havada, %7,1'i yağmurlu havada, %2,1'i bulutlu havada ve %2,6'sı diğer hava koşullarında (örneğin sisli, karlı, rüzgârlı vb.) meydana gelmiştir.

Kazaların büyük çoğunluğunun açık hava koşullarında gerçekleşmesi, hava şartlarının kazaların temel nedeni olmaktan çok, insan ve çevresel faktörlerin (dikkatsizlik, hız ihlali, yol yoğunluğu, kontrolsüz kavşaklar vb.) daha etkili olduğunu göstermektedir. Açık havalarda

sürücülerin aşırı güven ve hız eğilimiyle riskli manevralara yönelmesi, bu oranın yüksekliğini açıklamaktadır.

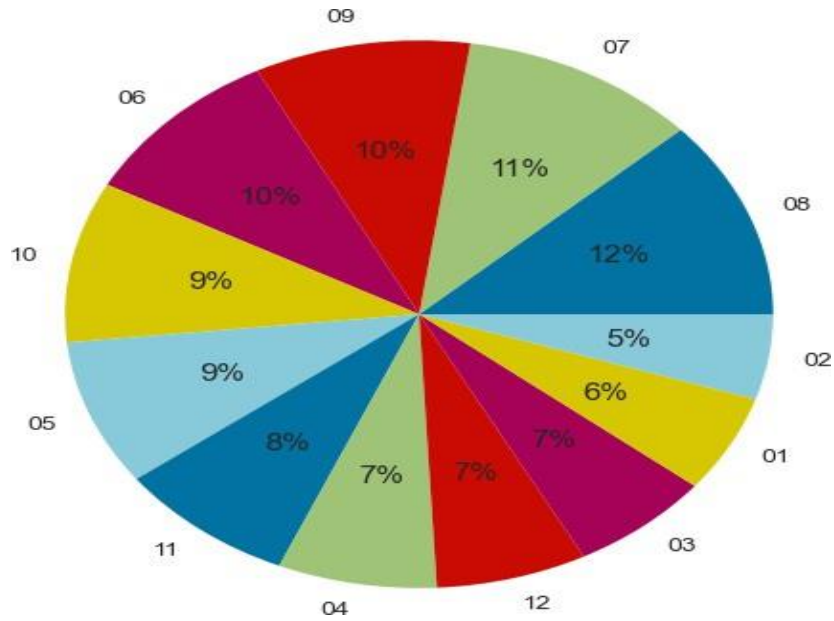
Sonuç olarak, Malatya'daki kazaların çoğunluğu elverişli hava koşullarında meydana gelmiş olsa da, bu durum sürücü davranışlarının ve yol güvenliği altyapısının hava şartlarından daha belirleyici olduğunu ortaya koymaktadır. Bu bulgu, trafik güvenliği politikalarında insan kaynaklı risk faktörlerinin azaltılmasına yönelik önlemlerin önceliklendirilmesi gerektiğini vurgulamaktadır.



Şekil 27. Kaza Yol Yüzeyi Dağılımı

Şekil 27’de Malatya ilinde meydana gelen trafik kazalarının yol yüzeyi koşullarına göre dağılımı incelendiğinde, kazaların %85,2’sinin kuru, %12,4’ünün ıslak veya nemli, %1,1’inin karlı ve %1,3’ünün diğer yüzey türlerinde (çamurlu, buzlu, gevşek zemin vb.) meydana geldiği görülmektedir (Emniyet Genel Müdürlüğü, 2022; Türkiye İstatistik Kurumu, 2023).

Kuru yol yüzeylerinde kazaların yüksek oranlarda gerçekleşmesi, trafik kazalarının yalnızca olumsuz hava ve zemin koşullarına bağlı olmadığını ortaya koymaktadır. Literatürde, uygun yol koşullarında dahi sürücü davranışları, hız ihlalleri ve dikkatsizliğin kazaların oluşumunda belirleyici rol oynadığı vurgulanmaktadır (Aksoy & Kantar, 2017). Karlı ve diğer olumsuz yol yüzeyi koşullarında meydana gelen kazalar toplam kazalar içerisinde daha düşük oranlarda yer almakla birlikte, bu tür yüzeylerde kaza riskinin ve kaza şiddetinin artabildiği bilinmektedir.



Şekil 28. Kazaların Kaza Ay Değişkenine Göre Dağılımı

Şekil 28, Malatya’da belirlenen dönemdeki trafik kazalarının aylara göre dağılımını göstermektedir. Kazalar yıl geneline yayılmakla birlikte **Temmuz** ve **Ağustos** aylarında artış, **Şubat** ve **Mart** aylarında ise düşüş dikkat çekmektedir. Yaz aylarındaki yükseliş, artan seyahat hareketliliği ve trafik yoğunluğuyla ilişkilendirilebilir. Kış aylarında oranların azalması ise hem yol kullanımının düşmesi hem de sürücülerin daha dikkatli davranmasıyla açıklanabilir. Bu sonuçlar, kazaların mevsimsel koşullardan etkilendiğini ve özellikle yaz döneminde alınacak önleyici tedbirlerin önemini ortaya koymaktadır.

Tablo 7. Malatya İli Trafik Kazası Verisinde Kategorik Değişkenlerin Özellikleri

Özellik	Gözlem Sayısı	Kategori Sayısı	En Sık Değer	Frekans
Kaza Yılı	9779	—	—	—
Kaza Saat Dilimi	9779	12	16:00–17:59	1426
Kaza İlçesi	9779	14	MALATYA-MERKEZ	3794
Kaza Yerleşim Yeri / Dışı	9779	2	1-Yerleşim Yeri	7201
Yolun Tipi	9779	4	1-Bölünmüş Yol	5735
Yolun Kaplaması	9779	6	1-Asfalt	9457
Yolun Sınıfı	9779	11	1-Cadde	4849

Yasal Hız Limiti	9779	—	—	—
Yatay Geometri	9779	3	1-Düz Yol	8419
Düşey Geometri	9779	4	1-Eğimsiz	7507
Kavşak Durumu	9779	8	8-Kavşak Yok	6307
Geçit Durumu	9779	5	5-Geçit Yok	8816
Gün Durumu	9779	3	1-Gündüz	6656
Hava Durumu	9779	10	1-Açık	8625
Yol Yüzeyi	9779	6	1-Kuru	8333
Kaza Oluş Türü	9779	15	3-Yandan Çarpma	2780
Kaza Araç Sayısı	9779	3	1-Tek Araçlı	4945
Toplam Ölümlü Kaza	9779	—	—	—
Toplam Yaralanmalı Kaza	9779	—	—	—
Toplam Ölü Sayısı	9779	—	—	—
Toplam Yaralı Sayısı	9779	—	—	—
Kaza Şiddeti	9779	—	—	—
Kaza Ayı	9779	12	08	1150
Kaza Günü	9779	31	20	354

Tablo 7, Malatya ilinde meydana gelen trafik kazalarına ait veri setinde yer alan kategorik değişkenlerin temel özelliklerini göstermektedir. Çoğu değişkende gözlem sayısı 9779 olup eksik veri oranı düşüktür. Veriler incelendiğinde kazaların çoğunluğunun 16:00–17:59 saat aralığında, Malatya Merkez ilçesinde, yerleşim yeri içinde ve bölünmüş asfalt yollarda gerçekleştiği görülmektedir. Hava durumu ve yol yüzeyi bakımından açık ve kuru koşulların baskın olması, kazaların büyük ölçüde normal sürüş koşullarında meydana geldiğini göstermektedir.

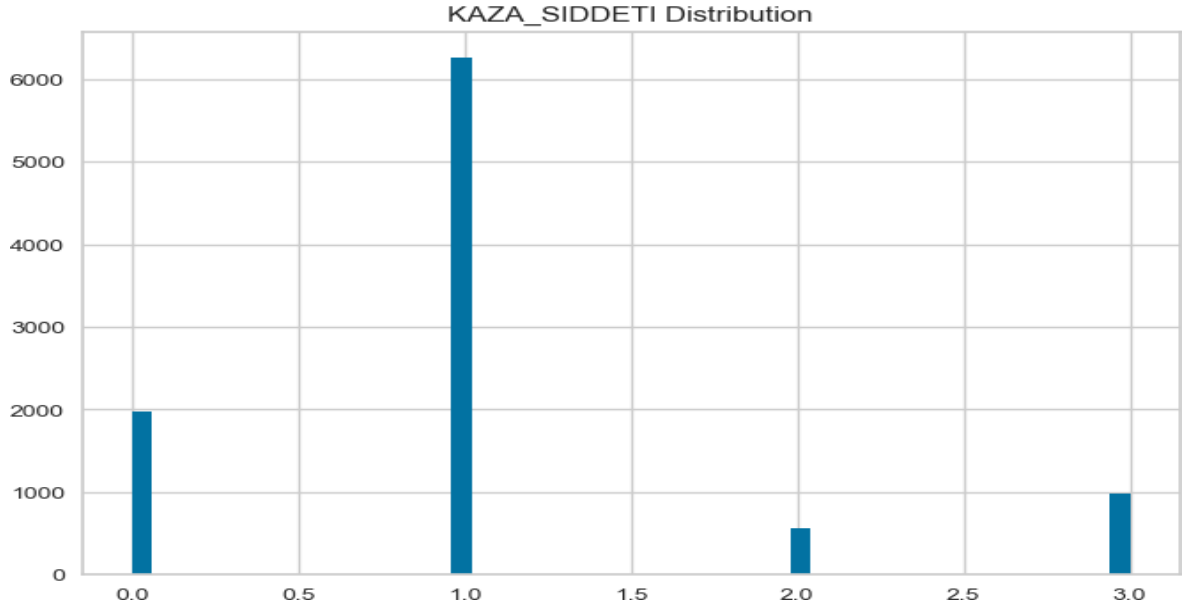
Tablo 8. Malatya İli Trafik Kazası Verisinde Sayısal Değişkenlerin Tanımlayıcı İstatistikleri

Özellik	Gözlem Sayısı	Ortalama	Standart Sapma	Min	%25	%50	%75	Maks
Kaza Yılı	9779	2018.04	3.16	2013	2015	2018	2021	2023
Toplam Ölümlü Kaza	9779	0.0393	0.1942	0	0	0	0	1
Toplam Yaralanmalı Kaza	9779	0.9607	0.1942	0	1	1	1	1
Toplam Ölü Sayısı	9779	0.0502	0.2935	0	0	0	0	7
Toplam Yaralı Sayısı	9779	2.1536	1.8890	0	1	2	3	30
Kaza Şiddeti	9779	1.057	0.8106	0	1	1	1	3
Kaza Ayı	9779	6.93	3.14	1	5	7	9	12
Kaza Günü	9779	15.72	8.82	1	8	16	23	31

Tablo 8, Malatya ili trafik kazası verisindeki sayısal değişkenlerin özet istatistiklerini göstermektedir. Kazaların ortalama yılı 2018 civarındadır ve veri seti 2013–2023 yıllarını kapsamaktadır. Ortalama yaralı sayısı 2.15, ölüm sayısı ise 0.05’tir. Kazaların %96’sı yaralanmalı, %4’ü sadece maddi hasarlıdır. Bu durum, kent içi trafik kazalarının çoğunun ölümsüz fakat yaralanmalı sonuçlandığını göstermektedir.

2018–2023 yılları arasındaki trafik kazaları incelendiğinde, kazaların belirli yıllar ve dönemler itibarıyla yoğunlaştığı; ölümlü kazaların toplam kazalar içerisindeki oranının görece düşük, yaralanmalı kazaların ise baskın olduğu görülmektedir. Bu durum, Türkiye genelinde trafik kazalarına ilişkin resmî istatistiklerde de benzer şekilde raporlanmıştır (Emniyet Genel Müdürlüğü, 2022; Türkiye İstatistik Kurumu, 2023). Ortalama yaralı sayısının çoğu kazada birden fazla olması, kent içi trafik kazalarının özellikle bedensel yaralanmalar açısından önemli bir risk unsuru oluşturduğunu ortaya koymaktadır. Kazaların aylara göre dağılımı incelendiğinde, özellikle yaz aylarında kaza sayılarında artış eğilimi gözlenmektedir. Ayrıca, Türkiye’de kentsel alanlardaki trafik kazalarının mekânsal dağılımını inceleyen çalışmalarda, yoğun trafik akışına sahip dönem ve bölgelerde kaza riskinin arttığı vurgulanmaktadır (Erdoğan & Gökdağ, 2019).

Kaza şiddetinin değerlendirilmesine yönelik yapılan çalışmalarda ise, yaralanma düzeyinin sürücü davranışları, çevresel koşullar ve trafik yoğunluğu gibi değişkenlerin birlikte etkisiyle şekillendiği ifade edilmektedir. Özellikle makine öğrenmesi temelli modellerin kullanıldığı araştırmalarda, orta ve yüksek şiddetli kazaların bu faktörlerin etkileşimi sonucunda ortaya çıktığı gösterilmiştir (Moghaddam et al., 2011; Sivasankaran & Balasubramanian, 2020).



Şekil 29. Kaza Şiddeti (KAZA_SIDDETI) Değişkeninin Dağılımı

Şekil 29, Malatya ilindeki trafik kazalarının kaza şiddeti (KAZA_SIDDETI) değişkenine göre dağılımını göstermektedir. Grafik incelendiğinde, kazaların büyük çoğunluğunun şiddet değeri 1 olan (orta düzeyde) kazalardan oluştuğu görülmektedir. Bu grup, toplam kazaların yaklaşık üçte ikisinden fazlasını temsil etmektedir.

Şiddet değeri 0 olan kazalar, genellikle maddi hasarlı veya düşük etki seviyeli kazaları; şiddet değeri 2 ve 3 olan kazalar ise daha ağır yaralanmalı veya ölümlü kazaları ifade etmektedir.

Ancak bu gruplar, genel toplam içinde düşük bir orana sahiptir. Bu durum, Malatya'daki kazaların genel olarak ölümsüz, orta şiddetli ve yaralanmalı nitelikte olduğunu desteklemektedir.

4.2 Model Performans Değerlendirmeleri

4.2.1 Lojistik Regresyon Modelinin Performans Değerlendirmesi

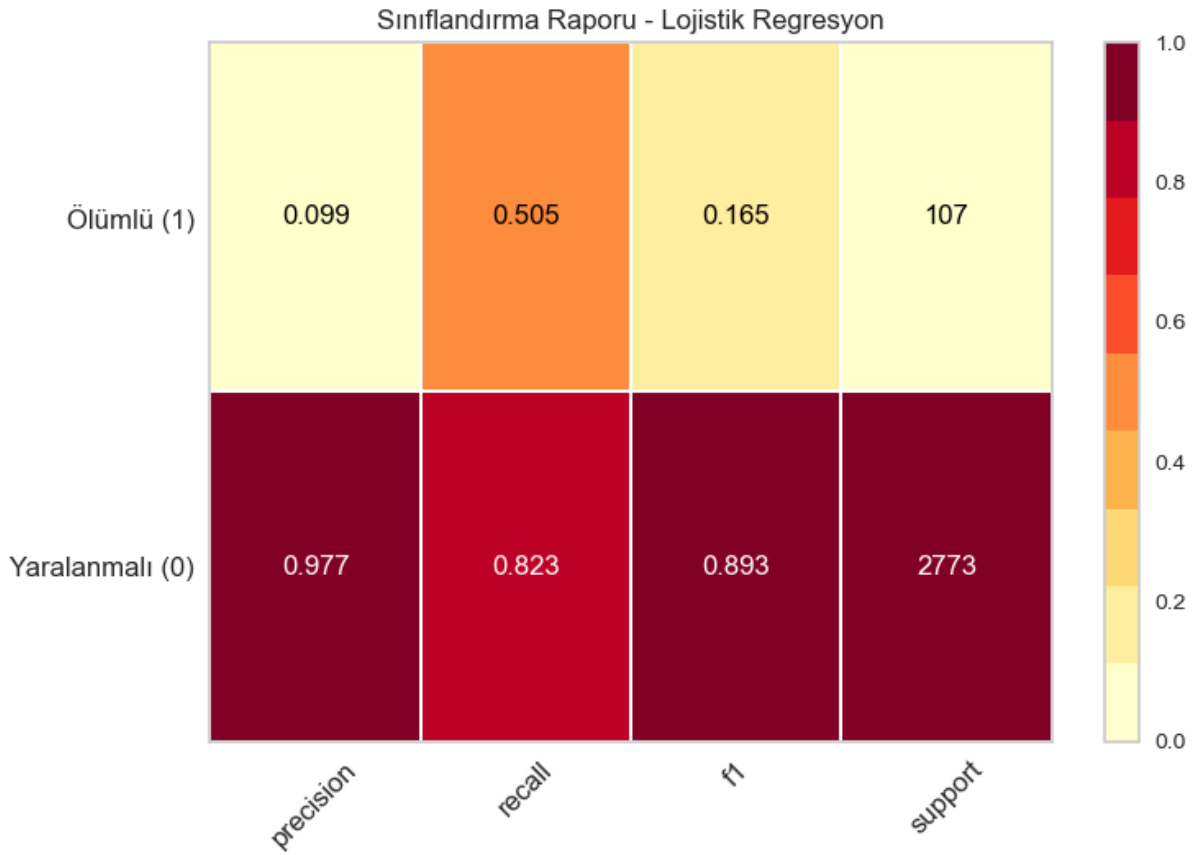
Lojistik regresyon, ikili sınıflandırma problemlerinde yaygın olarak kullanılan istatistiksel bir yöntem olup, bir olayın gerçekleşme olasılığını (örneğin kaza var/kaza yok) tahmin etmeye olanak taşımaktadır. Model, bağımlı değişkenin 0 ve 1 biçiminde tanımlandığı durumlarda özellikle etkilidir ve doğrusal regresyondan farklı olarak çıktıların olasılık temelli yorumlanmasını sağlar.

Tablo 9 Lojistik Regresyon Modelinde Öne Çıkan Özellikler ve Katsayı Yorumları

Özellik	Katsayı	Akademik Yorum
etkilenen_kisi_tipi_Yaya	1.627	Yaya türündeki kazalar, ölümlü kaza olasılığını artıran en güçlü faktörlerden biridir. Bu durum, yaya güvenliğinin kritik bir risk unsuru olduğunu göstermektedir.
etkilenen_kisi_tipi_Yolcu	-1.602	Yolcu statüsünde yer alan bireylerin ölümlü kazaya karışma olasılığı daha düşüktür; bu durum yolcuların sürücülere kıyasla görece daha az risk altında olduğunu göstermektedir.
YayaYaşıGrup_61+	1.102	61 yaş ve üzerindeki yayaların kazaya karışma ve kaza şiddetinin artma olasılığı yüksektir; yaşlı yayalara yönelik güvenlik önlemlerinin önemini ortaya koymaktadır.
YayaYaşıGrup_46–60	0.993	Orta yaş grubundaki yayalar da yüksek risk grubunda yer almakta olup, yaş faktörünün kaza şiddeti üzerinde etkili olduğunu göstermektedir.
Sürücü_Kusuru1_Taşıt giremez trafik işareti bulunan yerlere girmek	-0.769	Bu değişkenin negatif katsayıya sahip olması, veri setindeki örnek dağılımı ve sınıf yapısından kaynaklanabileceğinden dikkatli yorumlanmalıdır.
AracDarbeBölümü_12 – Sağ Arka Çamurluk	-0.510	Çarpışmanın aracın sağ arka bölümünde gerçekleşmesi, kaza şiddetinin görece daha düşük olma eğilimiyle ilişkilidir.

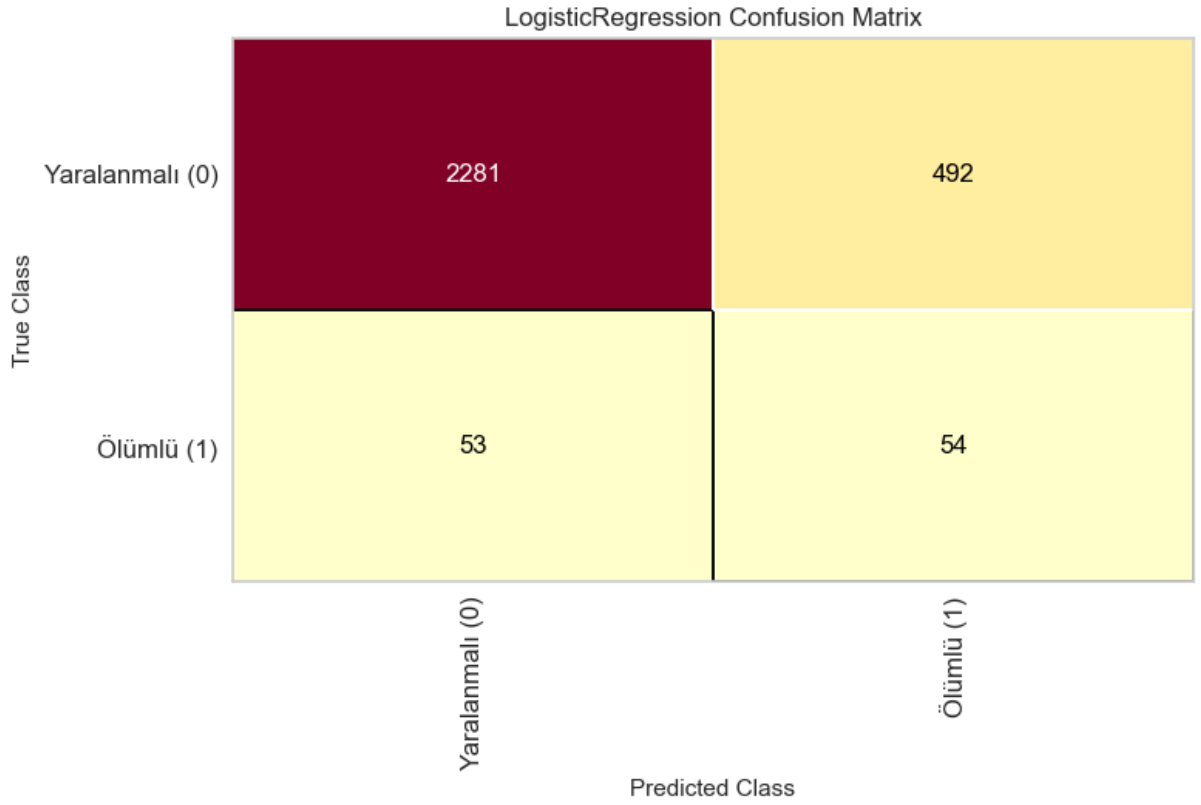
Diğer öne çıkan faktörler arasında yaş grupları, kaza oluş türleri, yol sınıfı ve sürücü davranışları yer almaktadır. Bulgulara göre yaşlı ve orta yaşlı yayaların kaza riski daha yüksektir. Trafik kurallarına uyulmaması ve hız kontrolüne ilişkin sürücü hataları, kazaların oluşumu ve şiddetini artırmaktadır.

Bu sonuçlar, Malatya'daki trafik kazalarının önlenmesine yönelik olarak yaya güvenliği, yaş grubu odaklı önlemler ve sürücü eğitiminin öncelikli alanlar olduğunu göstermektedir. Modelin katsayı analizi, veri odaklı trafik güvenliği politikaları için önemli bulgular sunmaktadır.



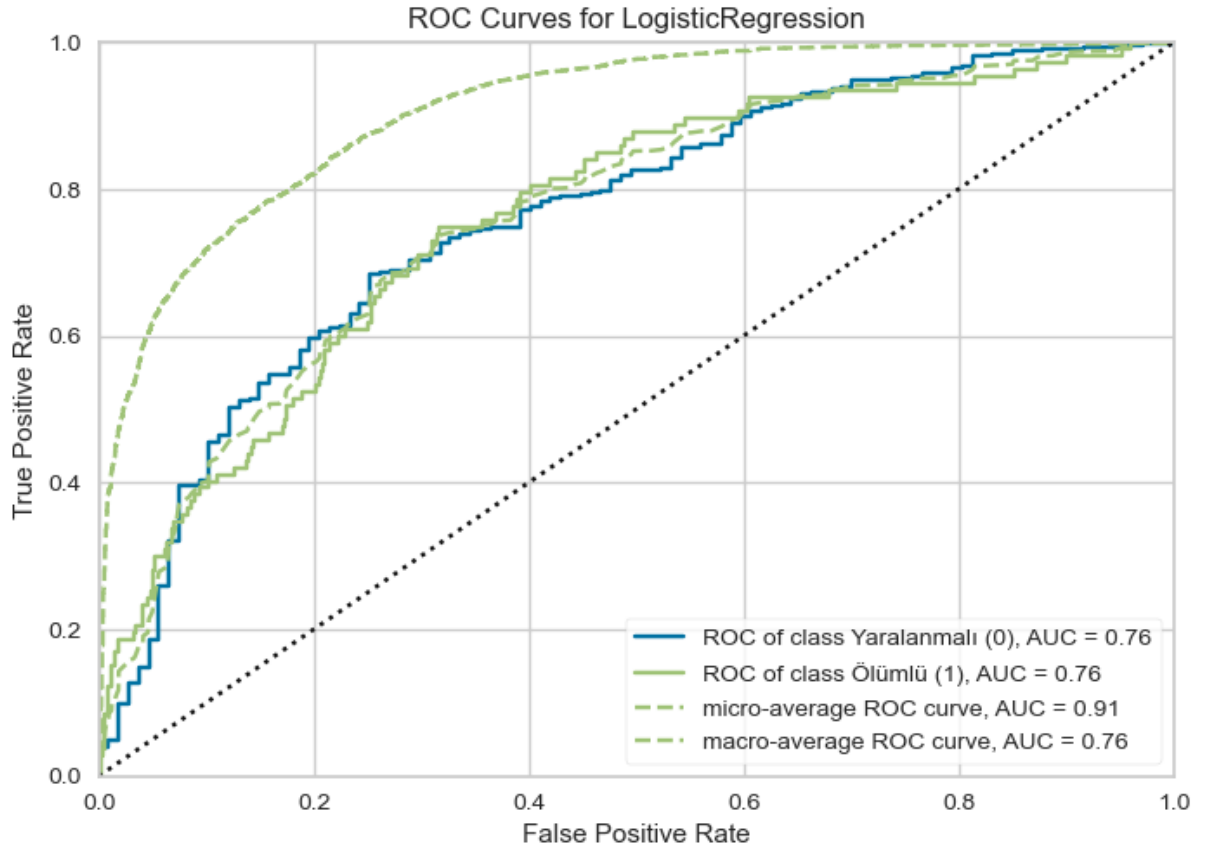
Şekil 30. Lojistik Regresyon Modeli Sınıflandırma Raporu (Classification Report Heatmap)

Şekil 30’da, lojistik regresyon modelinin ölümlü ve yaralanmalı kaza sınıfları için elde ettiği precision, recall, F1-skoru ve destek (support) değerleri sunulmaktadır. Sonuçlar incelendiğinde, yaralanmalı kaza sınıfının (0) yüksek precision (0.977), recall (0.823) ve F1-skoru (0.893) değerleri ile model tarafından başarılı bir şekilde tahmin edildiği görülmektedir. Buna karşılık, ölümlü kaza sınıfında (1) precision değerinin düşük (0.099) olmasına rağmen recall değerinin görece yüksek (0.505) olduğu dikkat çekmektedir. Bu durum, modelin ölümlü kazaların yaklaşık yarısını tespit edebildiğini ancak yanlış pozitif tahminlerin fazla olması nedeniyle kesinlik düzeyinin düşük kaldığını göstermektedir. Söz konusu performans farkı, veri setindeki belirgin sınıf dengesizliği ve ölümlü kazaların nadir görülmesiyle açıklanmaktadır.



Şekil 31. Logistik Regresyon Modeli Karmaşıklık Matrisi (Confusion Matrix)

Şekil 31’ de verilen karmaşıklık matrisi, yaralanmalı kaza sınıfının model tarafından görece yüksek doğrulukla tahmin edildiğini, buna karşın ölümlü kaza sınıfında düşük doğruluk ve yüksek hata oranlarının gözlemlendiğini ortaya koymaktadır. Bu durum, veri setindeki sınıf dengesizliği nedeniyle nadir gerçekleşen olayların tahmin edilmesinde model performansının sınırlı kaldığını doğrulamaktadır.



Şekil 32. Logistik Regresyon Modeline Ait ROC Eğrileri

ROC Eğrisi ve AUC Analizi

- Yaralanmalı (0) AUC: 0.76
- Ölümlü (1) AUC: 0.76
- Micro-AUC: 0.91
- Macro-AUC: 0.76

Şekil 32, Modelin sınıfları ayırt etme kapasitesi, ROC eğrileri ve AUC değerleri kullanılarak değerlendirilmiştir. Elde edilen sonuçlara göre, yaralanmalı kazalar (Class 0) için AUC değeri 0.76, ölümlü kazalar (Class 1) için ise 0.76 olarak hesaplanmıştır. Bu değerler, modelin her iki sınıf için de orta–yüksek düzeyde ayırım gücüne sahip olduğunu göstermektedir. Macro-AUC değerinin 0.76 olması, sınıflar arası performansın dengeli olduğunu ve modelin tüm sınıflara benzer düzeyde duyarlı davrandığını ifade etmektedir. Buna karşılık, Micro-AUC değerinin 0.91 gibi yüksek bir seviyede olması, tüm örnekler birlikte değerlendirildiğinde modelin genel ayırım kapasitesinin oldukça güçlü olduğunu ortaya koymaktadır. Bu durum, veri集中的 örnek dağılımının ve baskın sınıfların model performansı üzerinde etkili olduğunu göstermektedir.

4.2.2 XGBoost Modelinin Performans Değerlendirmesi

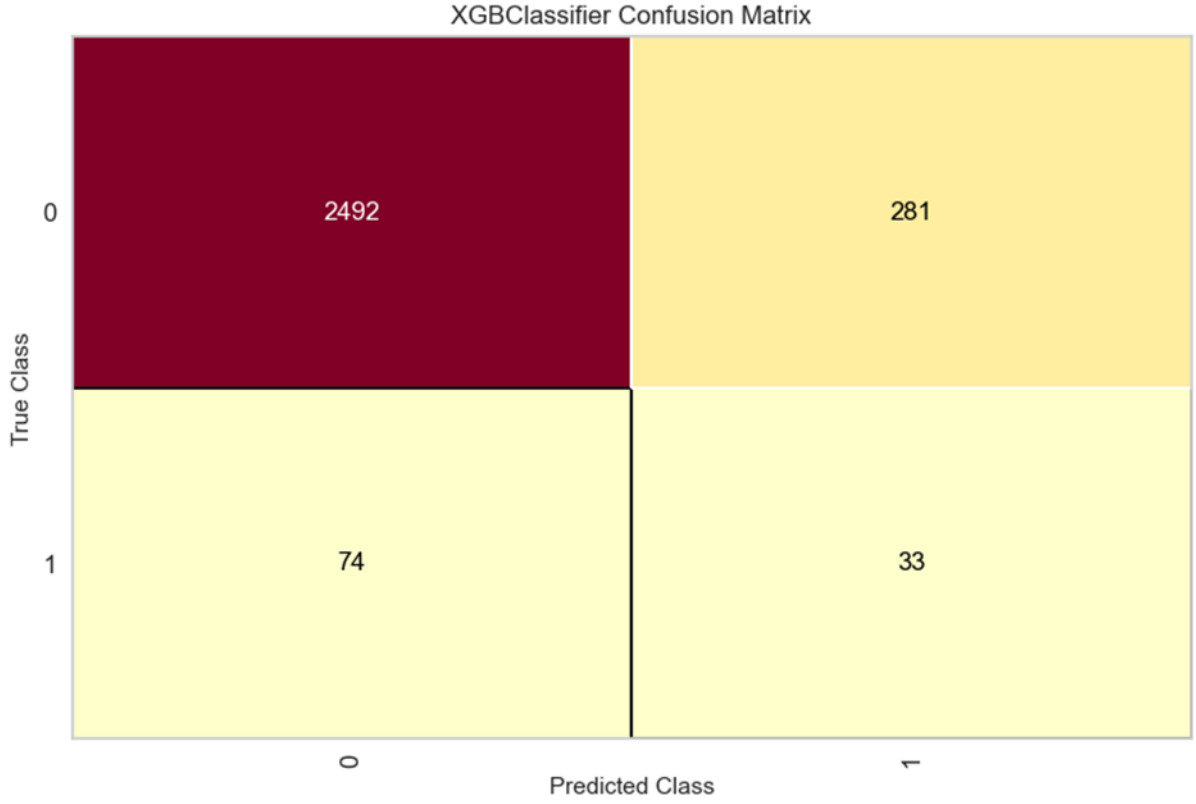
Performans Ölçütü Değer

Doğruluk (Accuracy)	0.8767
Kesinlik (Precision)	0.1051
Duyarlılık (Recall)	0.3084
F1 Skoru	0.1568
Cohen's Kappa	0.1073
MAE / MSE	0.1233

Bu çalışmada, trafik kazası veri setinde ikili sınıflandırma problemini çözmek amacıyla XGBoost sınıflandırıcı (XGBClassifier) kullanılmıştır. Model hiperparametreleri, literatürdeki öneriler ve veri setinin dengesiz yapısı dikkate alınarak belirlenmiştir. Modelin tekrarlanabilirliğini sağlamak için random_state = 42 olarak sabitlenmiştir. Aşırı öğrenmeyi önlemek amacıyla L1 (alpha = 0.1) ve L2 (lambda = 1.0) düzenlileştirme parametreleri uygulanmıştır.

Ağaç yapısında max_depth değeri sınırlandırılmamış, min_child_weight = 1 olarak seçilerek karmaşık değişken etkileşimlerinin öğrenilmesine olanak tanınmıştır. subsample = 0.1 ve colsample_bytree = 1.0 değerleri ile her bir ağacın farklı veri alt kümeleri üzerinde öğrenmesi sağlanmış ve model çeşitliliği artırılmıştır. Öğrenme sürecinde learning_rate = 0.00001 ve n_estimators = 1000 kombinasyonu tercih edilerek modelin kademeli ve stabil biçimde öğrenmesi hedeflenmiştir. Bu parametre seçimi, trafik kazalarının nadir görülmesi nedeniyle ortaya çıkan sınıf dengesizliği problemine uyumlu şekilde yapılandırılmıştır.

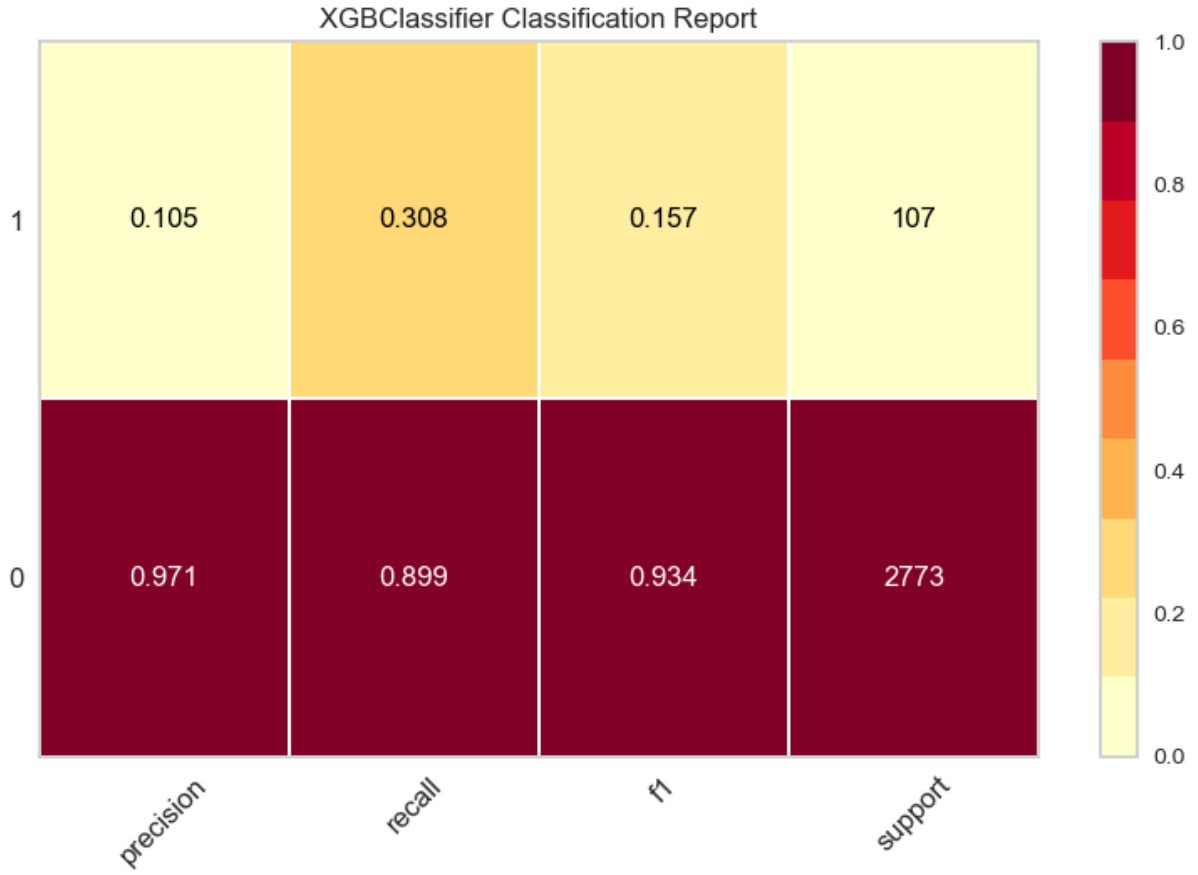
Model, dengesizliğin giderildiği eğitim veri seti (X_train_resampled, y_train_resampled) üzerinde eğitilmiş ve ölçeklendirilmiş test veri seti (X_test_scaled) üzerinde değerlendirilmiştir. Model performansı doğruluk, kesinlik, duyarlılık, F1 skoru, Cohen's Kappa ve hata metrikleri kullanılarak ölçülmüştür. Elde edilen sonuçlara göre Accuracy değeri 0.8767 olarak hesaplanmıştır. Ancak Precision (0.1051), Recall (0.3084) ve F1 skoru (0.1568) değerlerinin düşük olması, modelin nadir görülen kaza vakalarını ayırt etmede sınırlı bir performans sergilediğini göstermektedir. Cohen's Kappa değerinin 0.1073 olması, sınıf dengesizliğinin model performansı üzerindeki etkisini ve rastlantısal tahminlere yakınlığı ortaya koymaktadır.



Şekil 33. XGBoost Modeli Karmaşıklık Matrisi (Confusion Matrix)

Şekil 33, XGBoost modeline ait karmaşıklık matrisini göstermektedir. Sonuçlara göre, 0. sınıfa ait 2492 gözlem doğru, 281 gözlem ise yanlış sınıflandırılmıştır. 1. sınıfta ise 33 gözlem doğru tahmin edilirken, 74 gözlem yanlış sınıflandırılmıştır.

Bu dağılım, modelin çoğunluk sınıfını (0) yüksek doğrulukla tahmin edebildiğini, ancak azınlık sınıfı (1) ayırt etmede sınırlı kaldığını göstermektedir. Özellikle ölümlü kazaların önemli bir kısmının 0 sınıfı olarak tahmin edilmesi, sınıf dengesizliğinin model performansı üzerindeki etkisini ortaya koymaktadır.

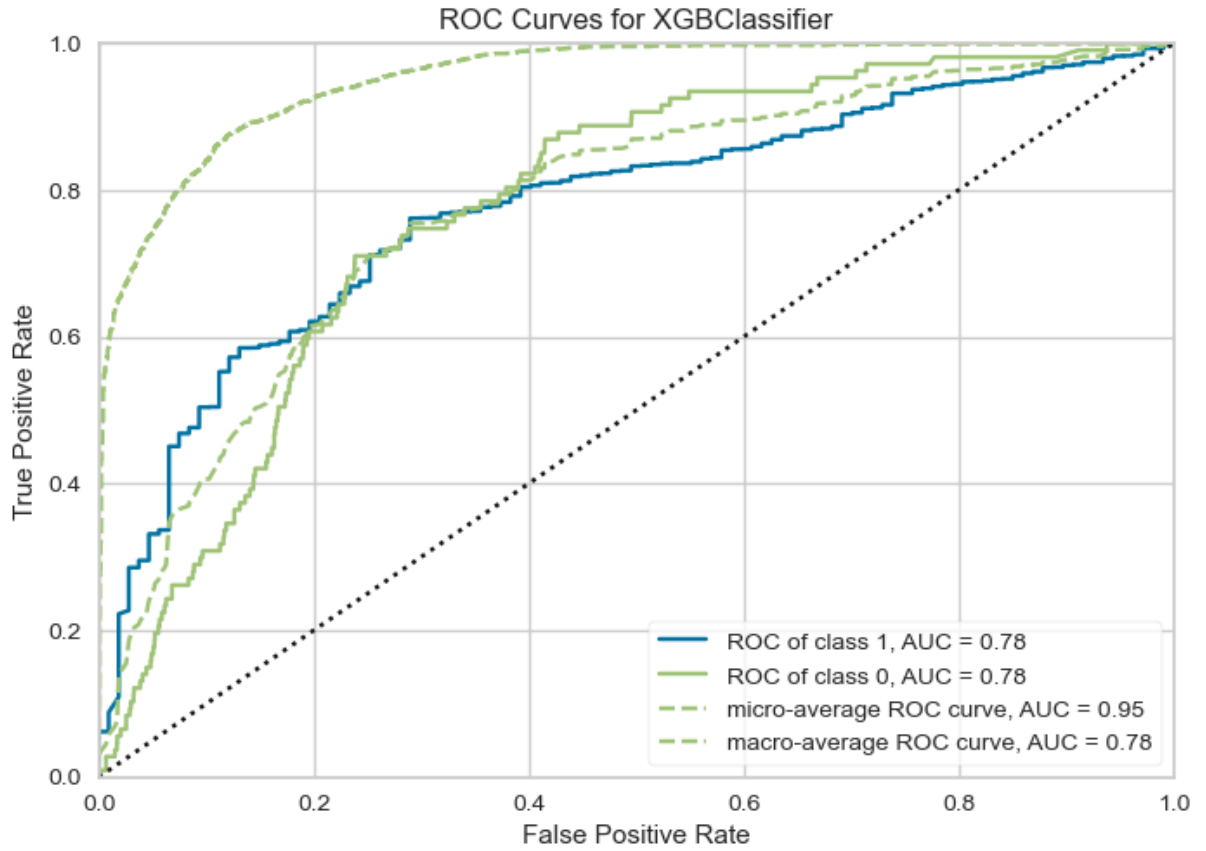


Şekil 34. XGBoost Modeli Sınıflandırma Raporu (Classification Report Heatmap)

Şekil 34, XGBoost modelinin sınıf bazlı precision, recall ve F1 skoru değerlerini göstermektedir. Sonuçlar, modelin sınıflar arasında dengeli bir performans sergilemediğini ortaya koymaktadır.

Ölümlü kazaları temsil eden 1. sınıf için precision (0.105), recall (0.308) ve F1 skoru (0.157) değerleri düşük olup, modelin nadir görülen kazaları tahmin etmede sınırlı kaldığını görülmektedir. Buna karşılık, 0. sınıf için elde edilen yüksek performans değerleri, modelin çoğunluk sınıfını başarıyla sınıflandırabildiğini göstermektedir.

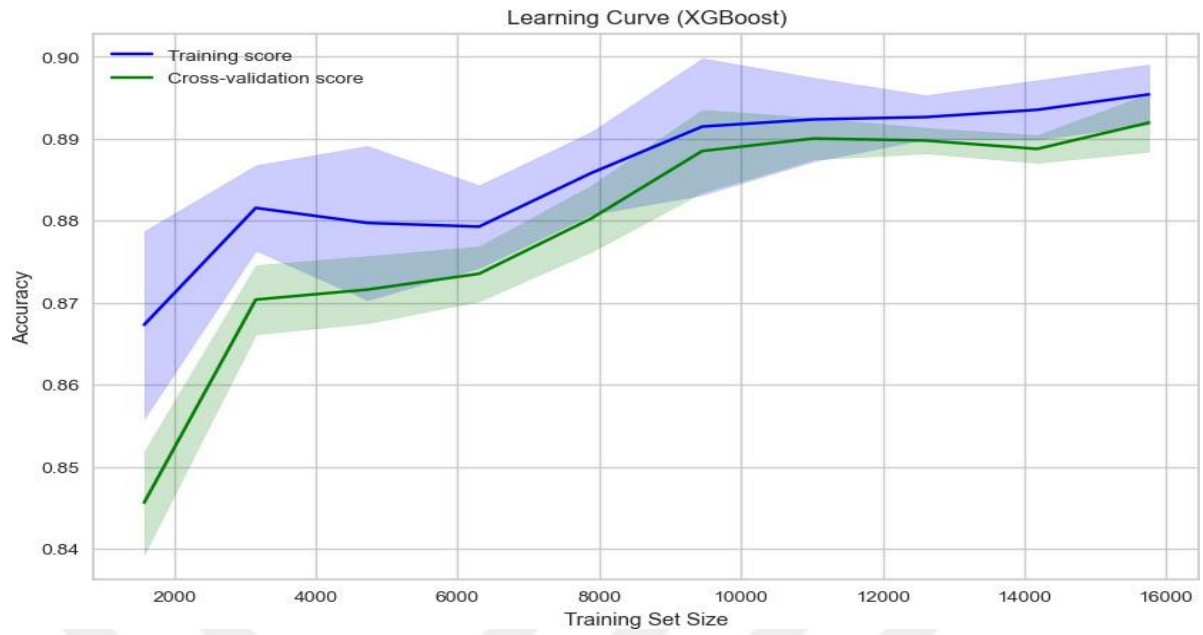
Bu bulgular, XGBoost modelinin genel doğruluk açısından güçlü olmasına rağmen sınıf dengesizliğinden etkilendiğini ve azınlık sınıf için ek önlemlere ihtiyaç duyulduğunu göstermektedir.



Şekil 35. XGBoost Modeline Ait ROC Eğrileri

Şekil 35, XGBoost modelinin her sınıf için ROC (Receiver Operating Characteristic) eğrilerini ve AUC (Area Under Curve) değerlerini göstermektedir.

ROC eğrileri ve AUC değerleri incelendiğinde, Sınıf 1 için AUC değeri 0.78, Mikro AUC değeri 0.95 ve Makro AUC değeri 0.78 olarak hesaplanmıştır. Mikro AUC'nin yüksek olması, modelin genel ayırım gücünün güçlü olduğunu gösterirken, makro ve sınıf 1 AUC değerlerinin orta seviyede kalması, nadir olayların tahmininde modelin ayırım kapasitesinin sınırlı olduğunu ortaya koymaktadır.



Şekil 36. XGBoost Öğrenme Eğrisi (Learning Curve)

Şekil 36’da, eğitim seti boyutu arttıkça doğruluk oranındaki değişim izlenmektedir. Eğitim (mavi) ve doğrulama (yeşil) eğrilerinin paralel şekilde yükselmesi, modelin veri arttıkça öğrenme başarısını sürdürebildiğini göstermektedir.

Eğriler arasındaki farkın az olması, XGBoost modelinin dengeli öğrenme yapısına sahip olduğunu ve overfitting riskinin düşük olduğunu göstermektedir.

4.2.3 Decision Tree (Karar Ağacı) Modelinin Performans Değerlendirmesi

Karar Ağacı (Decision Tree) Modeli – Test ve Eğitim Sonuçlarının Karşılaştırılması

Percentage Differences (Test - Train):

Metrik	Değer
MAE	0.0757
MSE	0.0757
Accuracy	0.9243
Precision	0.1275
Recall	0.1776
F1-score	0.1484
Cohen Kappa	0.1099

Modelin doğruluk (Accuracy) değeri yüksek (%92,43) görünmesine rağmen, veri setinde negatif sınıfın (yaralanmalı kazalar) baskın olması bu değer in yanıltıcı olmasına neden olmaktadır. Pozitif sınıfa (ölümlü kazalar) ait Precision, Recall ve F1-score değerlerinin düşük

olması, modelin ölümlü kazaları ayırt etme ve doğru tahmin etme konusunda yetersiz kaldığını göstermektedir. Özellikle düşük F1-score, pozitif sınıf tahmin performansının zayıf olduğunu ortaya koymaktadır. Cohen Kappa katsayısının 0,11 olması, model başarısının rastgele tahmine yakın seviyede olduğunu göstermektedir. Eğitim ve test verileri arasındaki yaklaşık 0,17'lik F1-score farkı, modelin eğitim verisine aşırı uyum sağladığını ve test verisi üzerinde genelleme kabiliyetinin düşük olduğunu göstermekte olup, bu durum belirgin bir overfitting göstergesidir.

Modelin eğitim ve test veri setleri üzerindeki F1 skorları aşağıda verilmiştir:

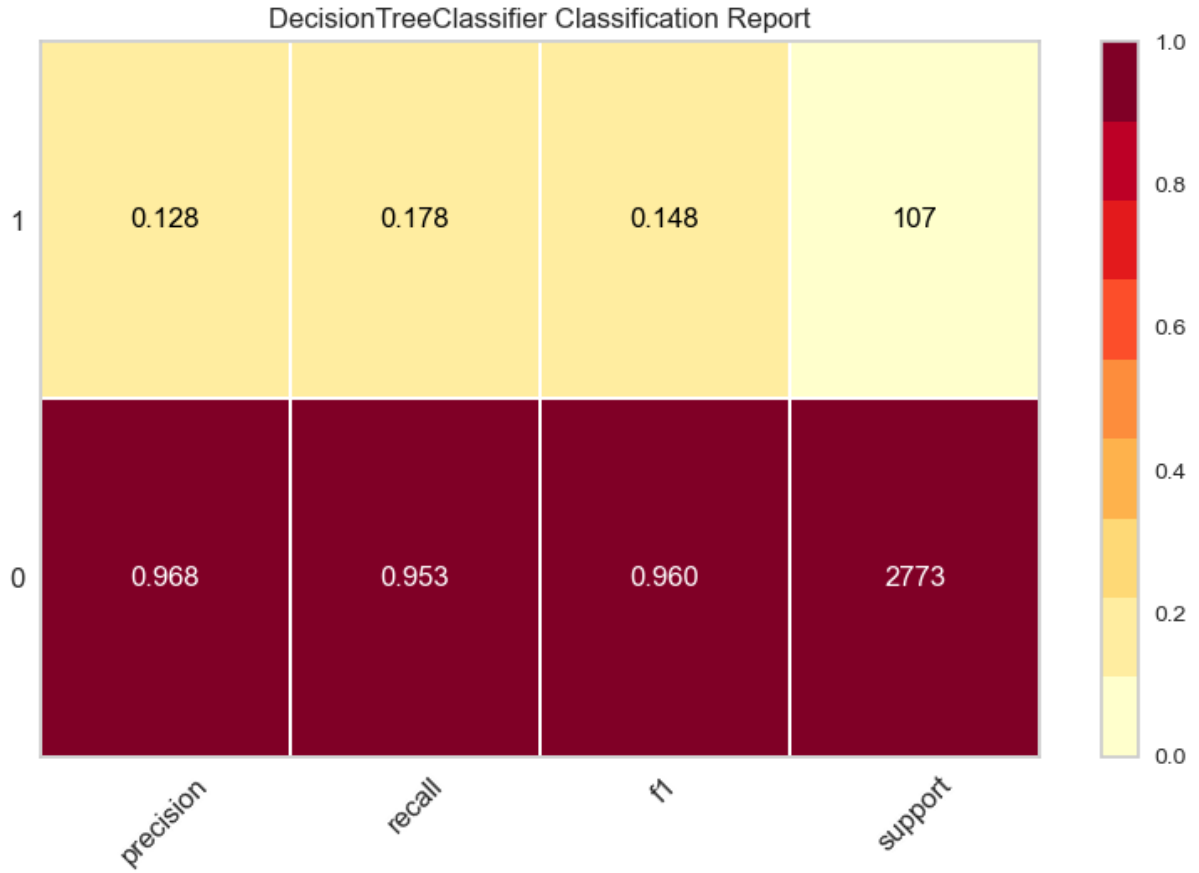
Veri Seti	F1 Skoru
Eğitim	0.964
Test	0.795

Eğitim ve test F1 skorları arasındaki fark ($\Delta F1 \approx 0.169$) olarak hesaplanmıştır. Bu fark, modelin eğitim verisi üzerinde oldukça yüksek bir performans sergilediğini, ancak test verisi üzerinde performansının belirgin şekilde düştüğünü göstermektedir. Bu durum, modelin bir miktar overfitting (aşırı öğrenme) eğilimi gösterdiğine işaret etmektedir. Başka bir ifadeyle model, eğitim verisini iyi öğrenmiş ancak test verisine genelleme yeteneği sınırlı kalmıştır.

Modelin genellenebilirliğini değerlendirmek amacıyla Stratified K-Fold (k=5) çapraz doğrulama yöntemi uygulanmıştır. Elde edilen F1 skorları aşağıda sunulmaktadır.

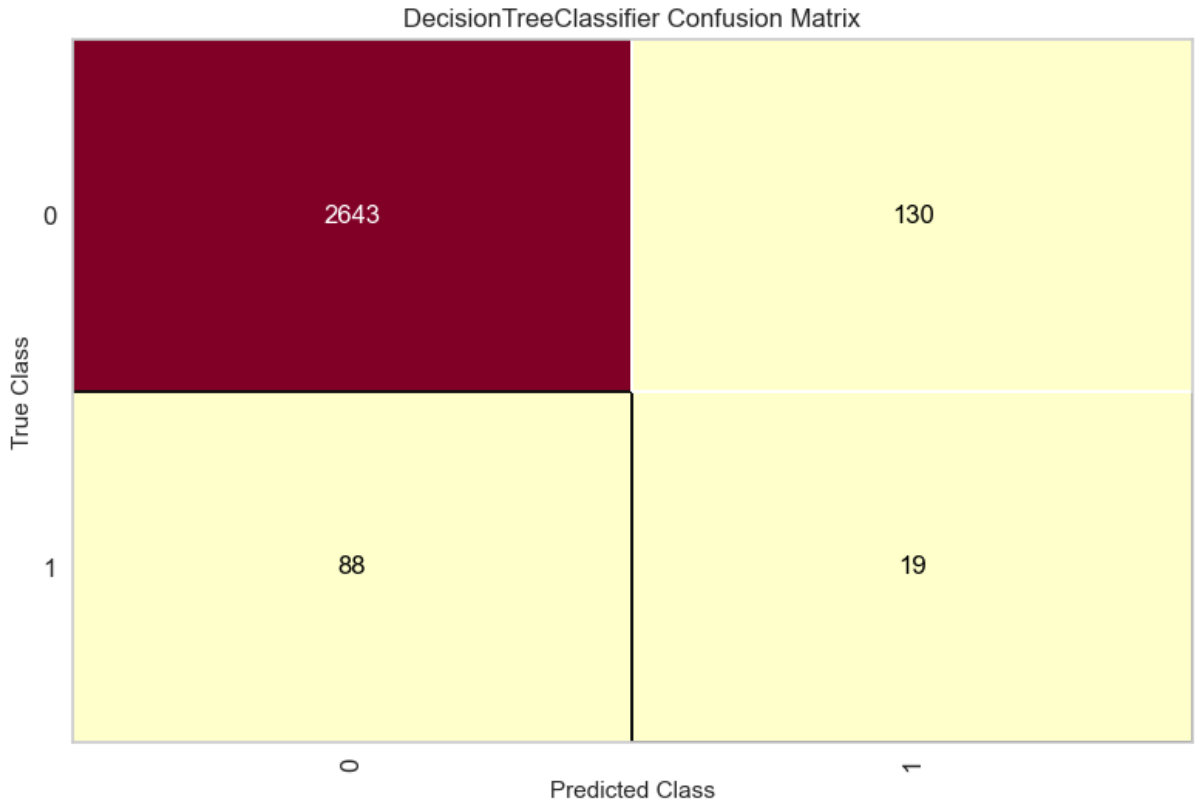
Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Ortalama	Std. Sapma	Min	Max
0.9609	0.9604	0.9609	0.9604	0.9604	0.9606	0.0003	0.9604	0.9609

Çapraz doğrulama sonuçları incelendiğinde, tüm fold'lar arasında F1 skorlarının oldukça yüksek ve birbirine çok yakın olduğu görülmektedir. Düşük standart sapma değeri, model performansının farklı veri alt kümeleri üzerinde istikrarlı ve tutarlı olduğunu göstermektedir. Bu durum, modelin genel olarak güvenilir bir öğrenme yapısına sahip olduğunu ortaya koymaktadır.



Şekil 37. Decision Tree Sınıflandırma Raporu

Şekil 37, Karar Ağacı modelinin sınıf bazlı precision, recall ve F1-skoru değerlerini göstermektedir. Negatif sınıfta (0) model yüksek bir performans sergilerken ($F1 \approx 0.96$), pozitif sınıfta (1) precision, recall ve F1 değerlerinin düşük kaldığı ($F1 \approx 0.15$) görülmektedir. Bu durum, veri setindeki sınıf dengesizliği nedeniyle modelin ölümlü kazaları ayırt etmede sınırlı bir başarıya sahip olduğunu ve tahminlerin ağırlıklı olarak çoğunluk sınıfına yöneldiğini ortaya koymaktadır.

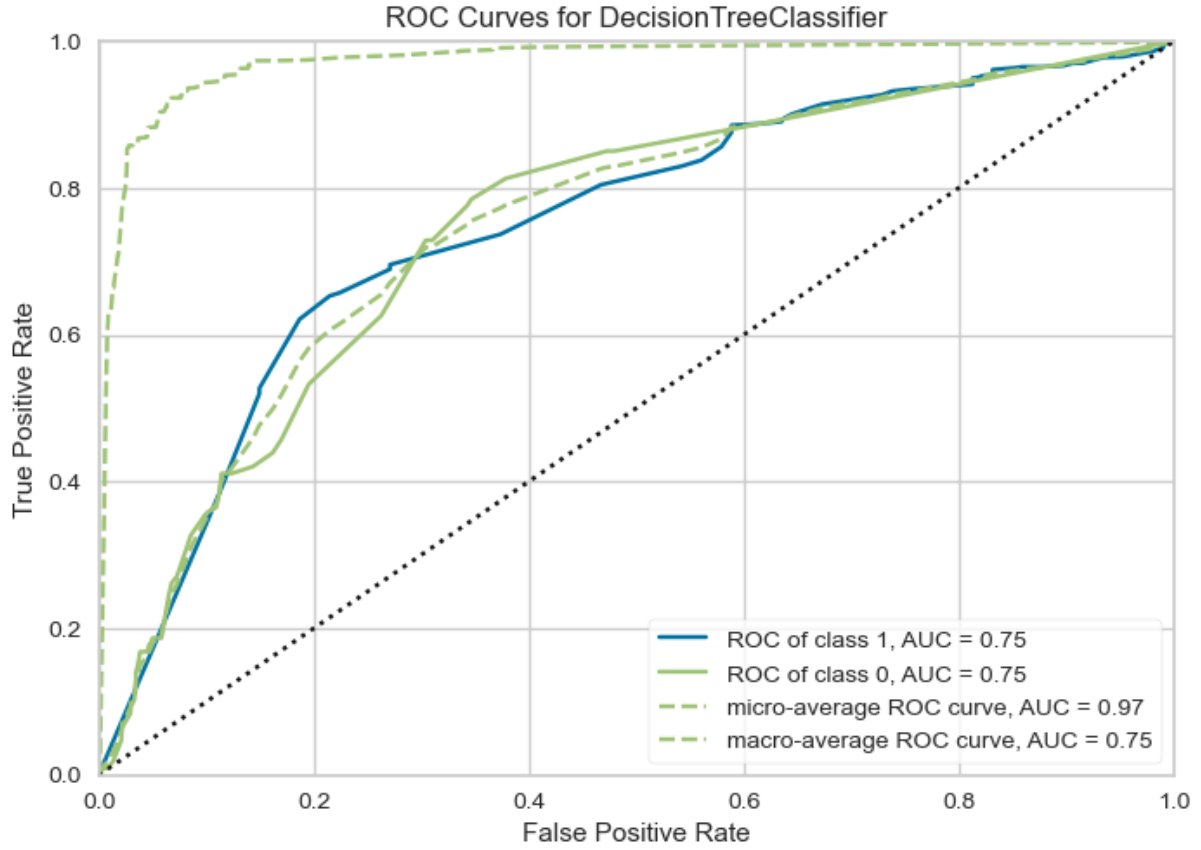


Şekil 38. Decision Tree Karmaşıklık (Confusion) Matrisi

Şekil 38, modelin tahmin ettiği ve gerçek değerler arasındaki ilişki gösterilmektedir.

Köşegen üzerindeki yüksek değerler, modelin doğru sınıflandırma oranının yüksek olduğunu ortaya koymaktadır.

Yanlış sınıflandırmaların sınırlı sayıda olması, modelin genel anlamda başarılı bir tahmin performansına sahip olduğunu, ancak bazı sınıflarda kısmi karışmaların yaşandığını göstermektedir.



Şekil 39. Decision Tree ROC Eğrisi

Şekil 39, modele ait ROC analizi sonuçları, her iki sınıf için AUC değerlerinin 0.75 olduğunu göstermektedir. Micro ortalama AUC'nin yüksek çıkması (0.97), veri setinde çoğunlukta olan negatif sınıfın modele baskın etkisinden kaynaklanmaktadır. Buna karşılık, pozitif sınıfa ait AUC değerinin 0.75 seviyesinde kalması, modelin yüksek riskli (ölümlü) kazaları ayırt etme kapasitesinin orta düzeyde olduğunu ve bu sınıfta sınırlı bir performans sergilediğini göstermektedir.

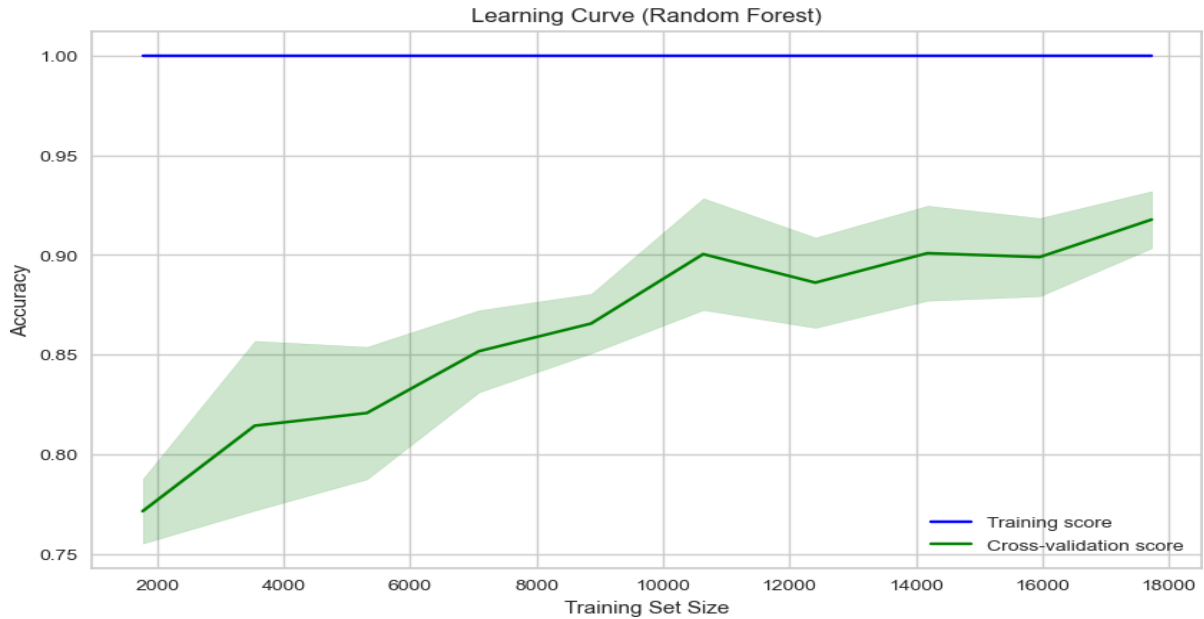
4.2.4 Random Forest (Rastgele Orman) Modelinin Performans Değerlendirmesi

Metrik	Değer
MAE	0.0861
MSE	0.0861
Accuracy	0.9139
Precision	0.1230
Recall	0.2150
F1-score	0.1565

Metrik	Değer
Cohen Kappa	0.1146

Random Forest modeli için elde edilen sonuçlar, genel doğruluk oranının yüksek olmasına karşın precision ve recall değerlerinin görece düşük kaldığını göstermektedir. Bu durum, veri setindeki sınıf dengesizliğinden kaynaklanmakta olup, nadir görülen kaza türlerinin doğru tahmin edilmesinde modelin sınırlı kaldığını ortaya koymaktadır. F1 skoru (0.1565) ve Cohen's Kappa (0.1146) değerleri, modelin rastlantısal tahminin bir miktar üzerinde performans sergilediğini göstermektedir. Eğitim ve test verileri arasındaki F1 skoru farkının yaklaşık 0.156 olması ise modelde belirli düzeyde overfitting bulunduğunu, modelin eğitim verisine test verisine kıyasla daha iyi uyum sağladığını ifade etmektedir. Random Forest algoritması, karmaşık ve dengesiz sınıf dağılımlarına sahip trafik/kaza verilerinde kullanışlı bir sınıflandırıcıdır. Ancak, özellikle nadir kazaların tahmininde düşük precision ve recall değerleri gözlenmiştir. Bu durum, modelin sınıf dengesizliği ile başa çıkmak için ek yöntemler (ör. SMOTE, ağırlıklandırma, daha fazla ağaç) ile iyileştirilebileceğini göstermektedir.

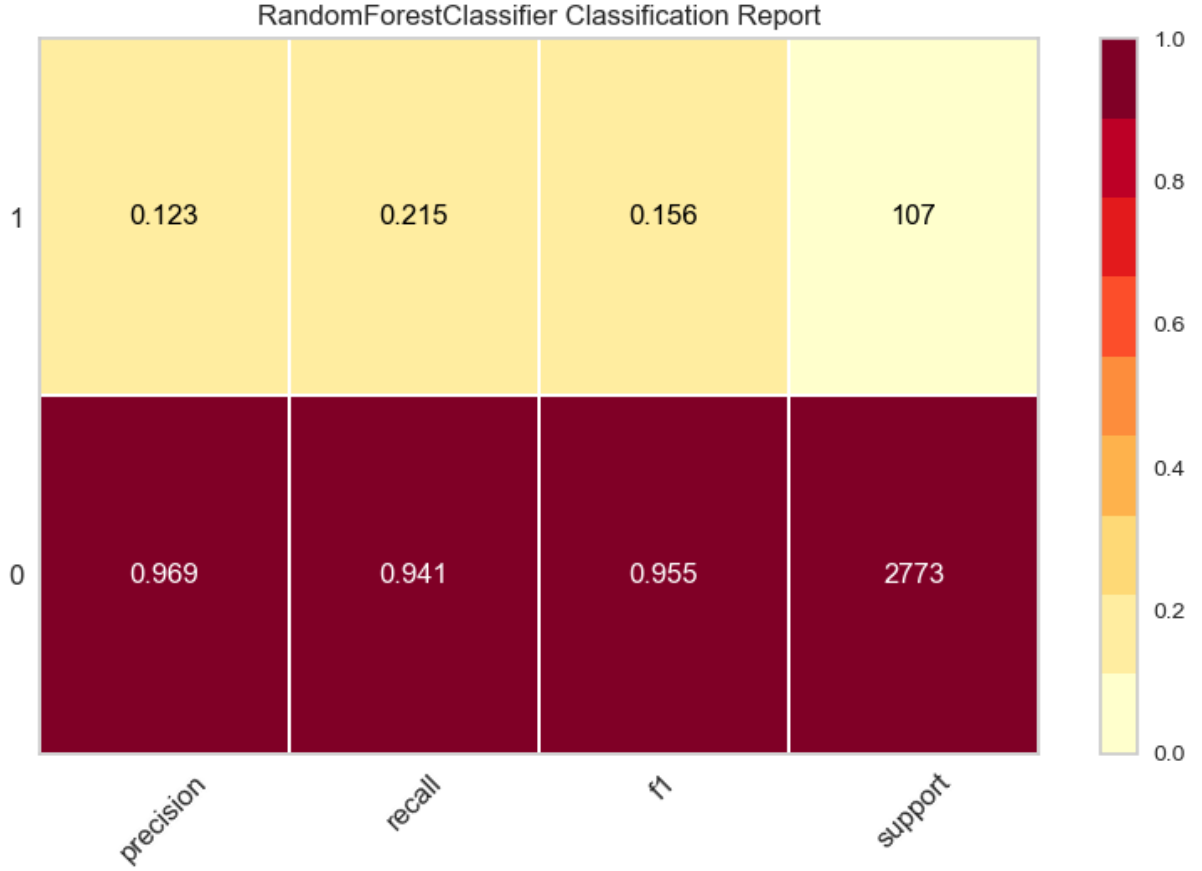
Çapraz doğrulama sonuçları, modelin aşırı uyum riskinin sınırlı olduğunu ve genellenebilirlik açısından güvenilir olduğunu doğrulamaktadır. Dolayısıyla Random Forest, ulaştırma veri analitiğinde bir ön modelleme aracı olarak etkili bir başlangıç noktası sunmaktadır.



Şekil 40. Random Forest Öğrenme Eğrisi

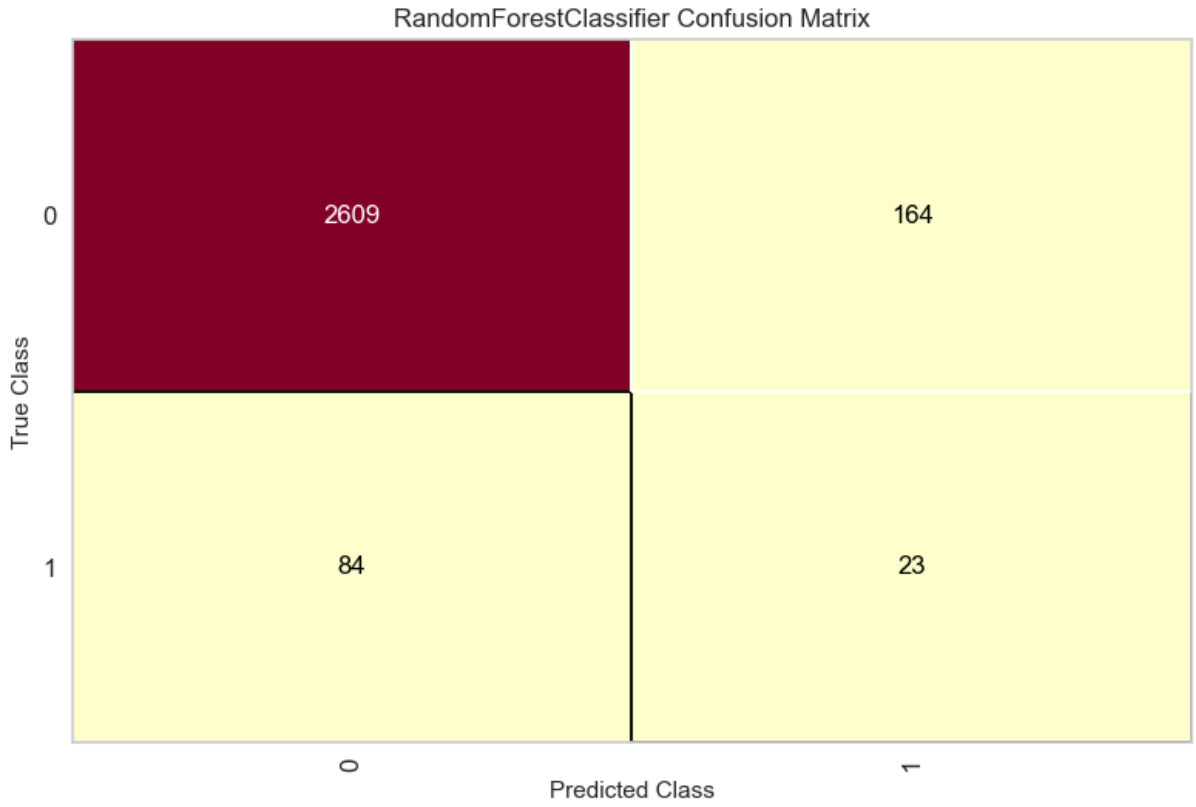
Grafik, modelin eğitim ve doğrulama başarımını göstermektedir. Eğitim doğruluğunun sabit ve yüksek, doğrulama eğrisinin ise istikrarlı biçimde artması, modelin veriyi iyi

öğrendiğini ve aşırı öğrenme eğilimi göstermediğini ortaya koymaktadır. Bu durum, Random Forest algoritmasının genelleme yeteneğinin güçlü olduğunu göstermektedir.



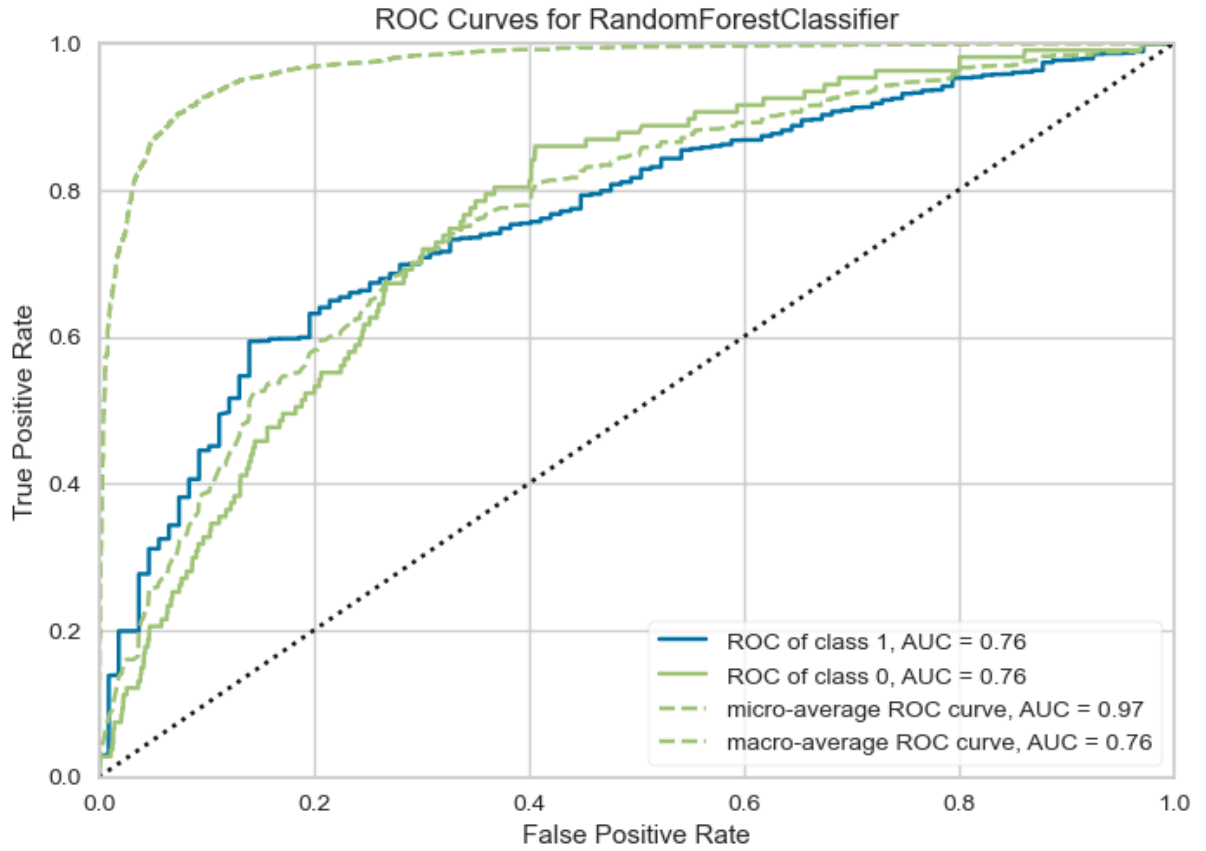
Şekil 41. Random Forest Sınıflandırma Raporu

Şekil 41, Random Forest modelinin çoğunluk sınıfı olan yaralanmalı kazalarda (0) çok yüksek performans sergilediğini ($F1 = 0.955$) göstermektedir. Buna karşılık ölümlü kazalar (1) için precision (0.123), recall (0.215) ve F1-score (0.156) değerleri oldukça düşüktür. Bu durum, veri setindeki belirgin sınıf dengesizliği nedeniyle modelin nadir görülen ölümlü kazaları ayırt etmekte zorlandığını ortaya koymaktadır. Model genel olarak çoğunluk sınıfını doğru sınıflandırsa da, kritik olan nadir sınıfın tahmininde sınırlı kalmaktadır.



Şekil 42. Random Forest Karışıklık Matrisi (Confusion Matrix)

Şekil 42, Random Forest modelinin yaralanmalı kazaları (0) büyük ölçüde doğru sınıflandırdığını (2609 doğru tahmin) göstermektedir. Ancak ölümlü kazalar (1) için doğru tahmin sayısı oldukça düşüktür (23), buna karşılık 84 ölümlü kaza yaralanmalı olarak tahmin edilmiştir. Bu sonuçlar, modelin çoğunluk sınıfında başarılı olduğunu, fakat nadir sınıf olan ölümlü kazaları ayırt etmede veri dengesizliği nedeniyle sınırlı kaldığını ortaya koymaktadır.



Şekil 43. Random Forest ROC Eğrisi (ROC Curve)

Şekil 43’de ROC eğrisi, 45° referans doğrusunun üzerinde yer alarak modelin rastgele tahmine kıyasla anlamlı bir ayırım gücüne sahip olduğunu göstermektedir. Yaralanmalı (0) ve ölümlü (1) sınıflar için AUC değerlerinin her ikisinin de 0.76 olması, modelin sınıflar arasında orta düzeyde bir ayırt edicilik sunduğunu ortaya koymaktadır. Micro-AUC’nin 0.91 gibi yüksek bir değere sahip olması, çoğunluk sınıfın etkisiyle modelin genel ayırım performansının güçlü olduğunu gösterirken, Macro-AUC’nin 0.76 seviyesinde kalması, nadir sınıf olan ölümlü kazaların tahmininde performansın sınırlı olduğunu teyit etmektedir.

4.2.5 SVC Modeli Performans Değerlendirmesi

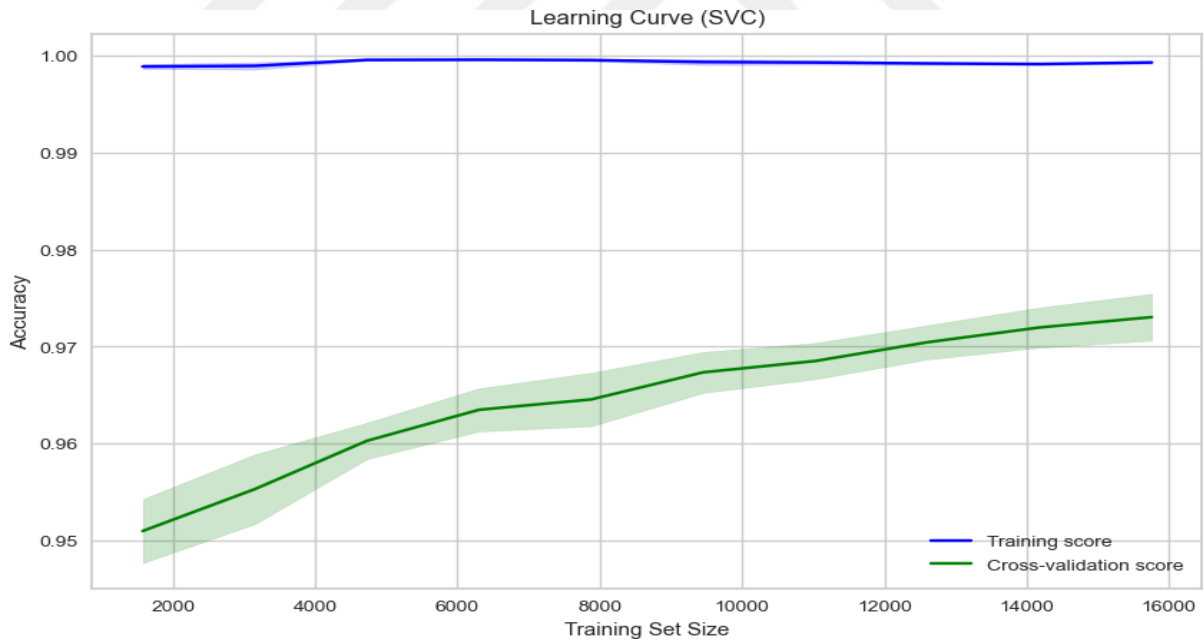
Metrik	Değer
MAE	0.2104
MSE	0.2104
Accuracy	0.7896
Precision	0.0903
Recall	0.5140
F1-score	0.1536
Cohen’s Kappa	0.0965

SVC modeli, çoğunluk sınıfı (yaralanmalı kazalar) için yüksek doğruluk değeri üretirken, nadir gerçekleşen ölümlü kazaların tahmininde sınırlı bir performans sergilemektedir. Modelin eğitim ve test veri kümeleri üzerindeki F1 skorları karşılaştırıldığında, genelleme boşluğu $\Delta F1 = 0.130$ olarak hesaplanmıştır. Bu değer, modelin eğitim verisine göre test verisinde belirli bir performans düşüşü gösterdiğini, ancak aşırı öğrenmenin (overfitting) sınırlı düzeyde olduğunu ifade etmektedir.

Modelin genellenebilirliğini değerlendirmek amacıyla 5 katlı stratifiye çapraz doğrulama (Stratified K-Fold CV) uygulanmıştır. Elde edilen F1 skorları aşağıda sunulmuştur:

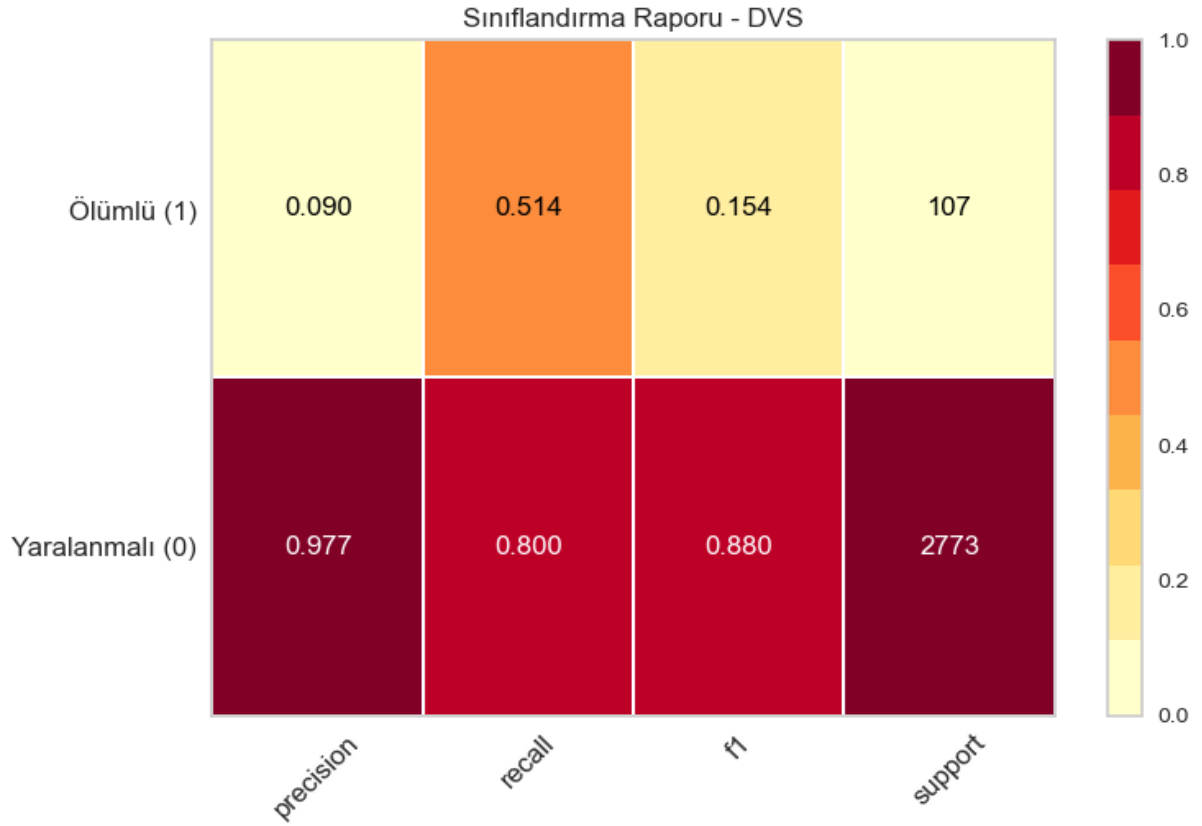
Fold	1	2	3	4	5	Ortalama	Std. Sapma	Min	Max
F1	0.9609	0.9609	0.9604	0.9604	0.9604	0.9606	0.00026	0.9604	0.9609

Çok düşük standart sapma değeri, modelin farklı veri alt kümelerinde yüksek derecede tutarlı ve stabil bir performans sergilediğini göstermektedir. Bu sonuç, modelin eğitim sürecinde istikrarlı bir öğrenme yapısı geliştirdiğini ortaya koymaktadır.



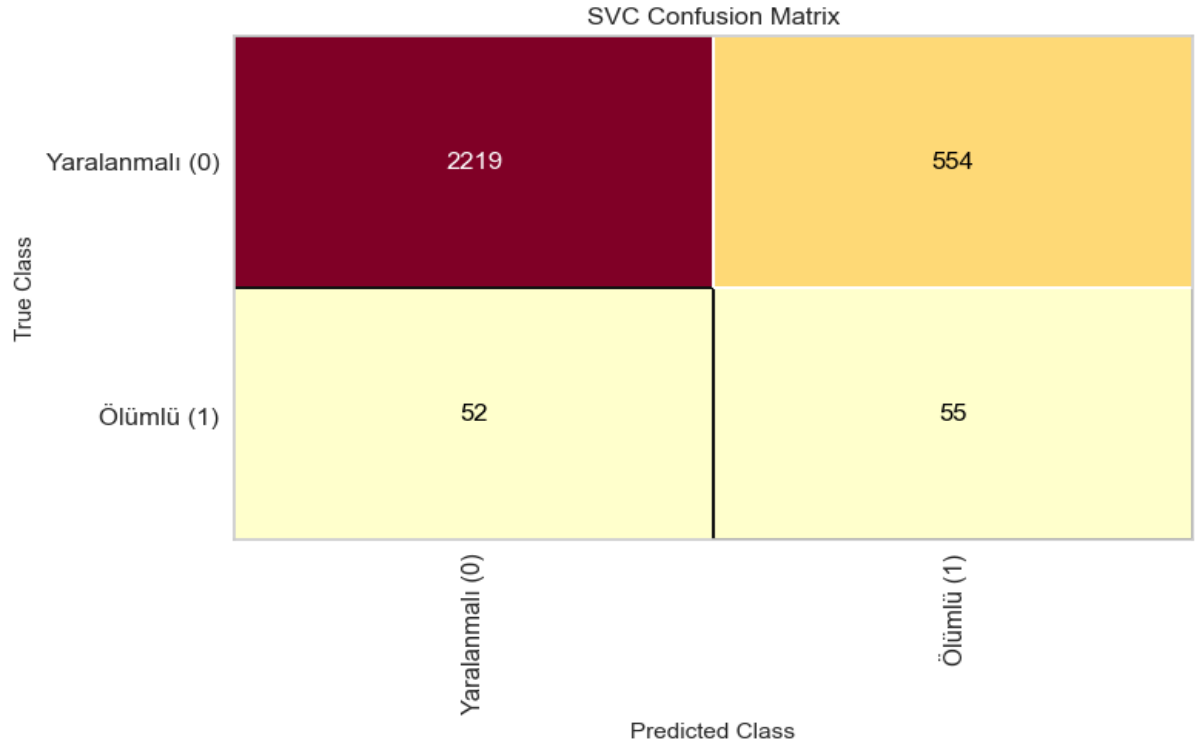
Şekil 44. SVC Modeli Öğrenme Eğrisi

Şekil'44'de, eğitim (mavi) ve doğrulama (yeşil) eğrileri arasındaki farkın oldukça az olduğu görülmektedir. Bu durum, SVC modelinin genelleme kabiliyetinin güçlü olduğunu ve aşırı öğrenme (overfitting) eğilimi göstermediğini göstermektedir. Model, eğitim verisinden öğrendiği bilgiyi test verisine başarılı şekilde aktarmıştır.



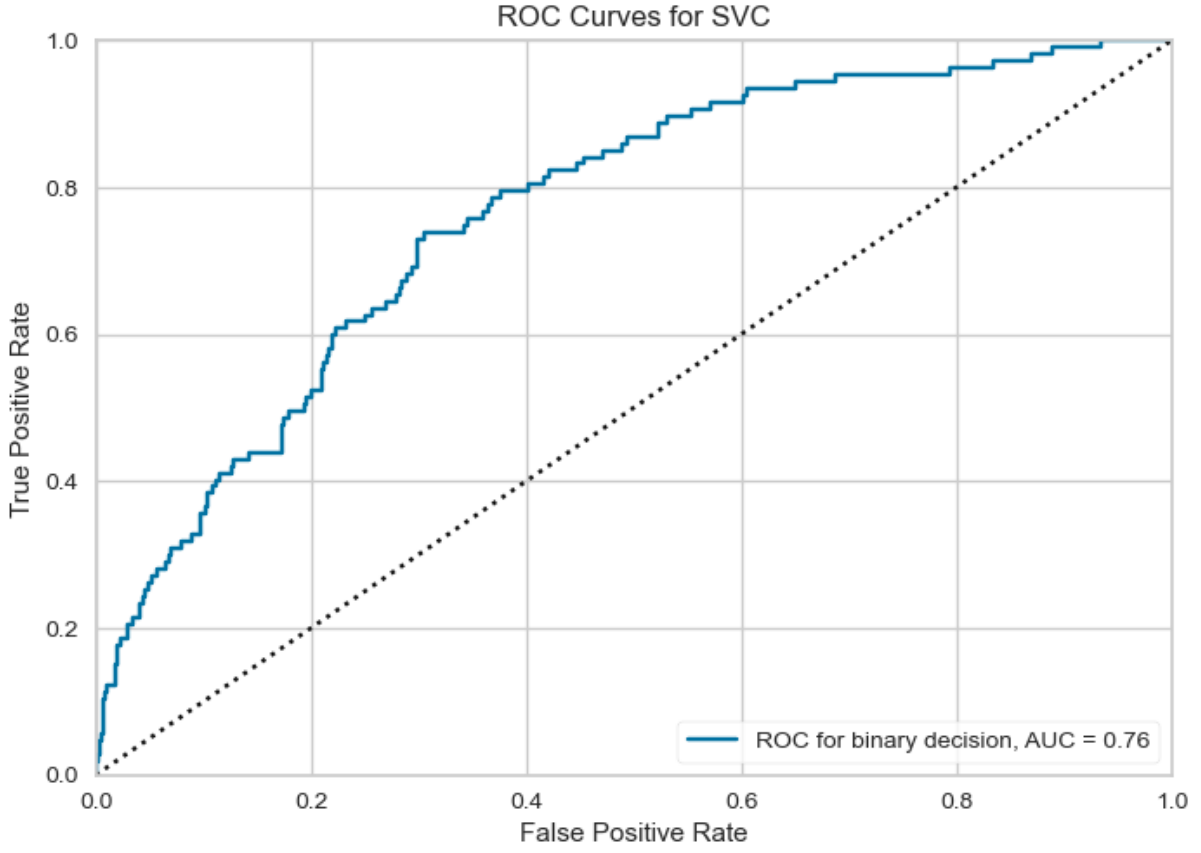
Şekil 45. SVC Modeli Sınıflandırma Raporu

Şekil 45’de görüldüğü üzere model, çoğunluk sınıfı olan yaralanmalı kazaları yüksek doğrulukla sınıflandırırken, ölümlü kazalar için precision (0.090) ve F1-score (0.154) değerlerinin düşük olması, nadir sınıflarda performansın sınırlı kaldığını göstermektedir. Ölümlü kazalar için recall değerinin %51 olması, modelin bu vakaların bir kısmını yakalayabildiğini; ancak sınıf dengesizliğinin genel performansı olumsuz etkilediğini ortaya koymaktadır.



Şekil 46. SVC Modeli Sınıflandırma Raporu ve Karışıklık Matrisi

Şekil 46’da SVC modeli, tüm sınıflar için yüksek doğruluk (accuracy), duyarlılık (recall) ve kesinlik (precision) değerleri elde etmiştir. Karışıklık matrisi sonuçları, SVC modelinin yaralanmalı kazaları büyük ölçüde doğru sınıflandırdığını (TN = 2219) göstermektedir. Buna karşın, ölümlü kazalara ait doğru pozitif sayısının düşük olması (TP = 55) ve yanlış negatif değerinin görece yüksekliği (FN = 52), modelin nadir gerçekleşen ölümlü kazaları tespit etmede sınırlı kaldığını ortaya koymaktadır. Bu durum, veri setindeki sınıf dengesizliğinin model performansı üzerindeki etkisini açıkça yansıtmaktadır.



Şekil 47. SVC Modeli ROC Eğrisi

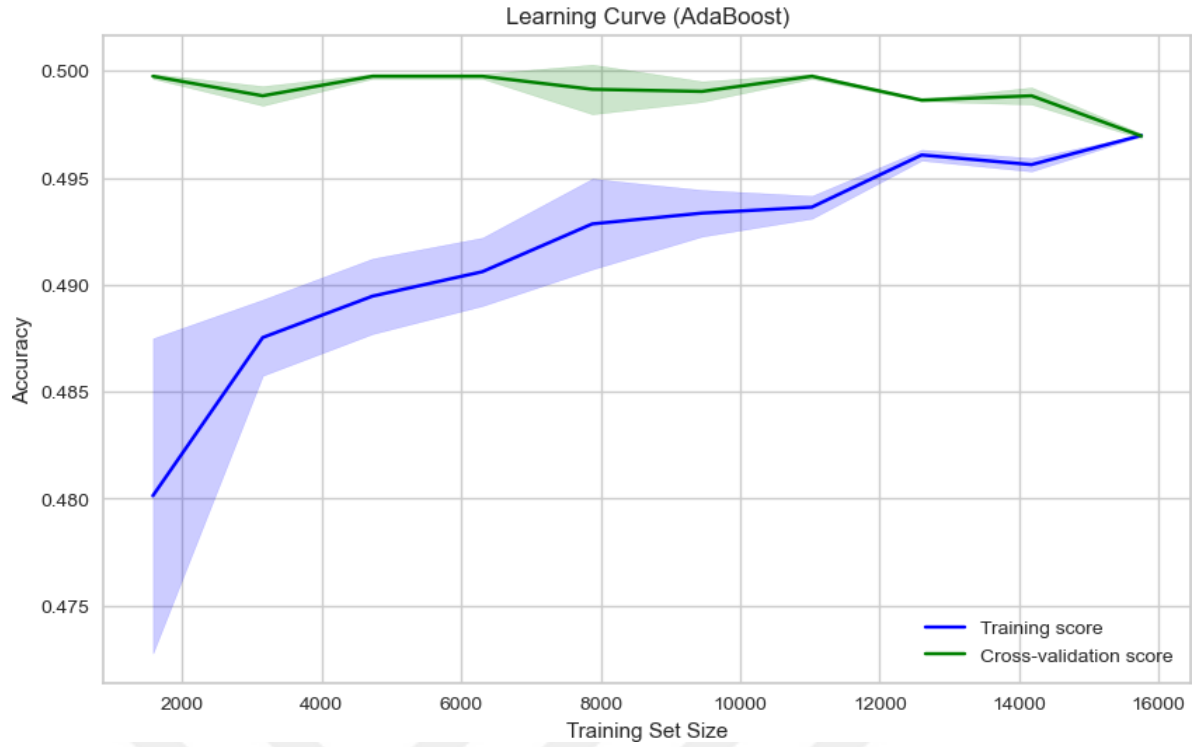
Şekil 47 ROC eğrisi, modelin sınıfları ayırt etme gücünü göstermektedir. Grafikte görüldüğü üzere ROC eğrisi, 45° rastgele tahmin doğrusunun üzerinde seyretmektedir. AUC değerinin 0.76 olması, modelin sınıflar arasında kabul edilebilir düzeyde ayırt ediciliğe sahip olduğunu ve rastgele tahmine kıyasla daha başarılı bir sınıflandırma performansı sunduğunu göstermektedir.

SVC lineer kernel modeli, yaralanmalı kazaları yüksek doğrulukla sınıflandırırken ölümlü kazalarda orta düzeyde bir performans göstermektedir; modelin genel ayırt ediciliği ROC eğrisi ve 0.76 AUC değeri ile doğrulanmış olup, etkilenen kişi tipi ve yaş gruplarının sınıflandırmada belirleyici olduğu görülmektedir. Ölümlü kazaların veri setinde nadir bulunması ve sınıf dengesizliği, bu sınıfa yönelik performansın sınırlı kalmasının temel nedenidir.

4.2.6 AdaBoost Modeli Performans Değerlendirmesi

Metrik	Değer
MAE	0.0465
MSE	0.0465
Accuracy	0.9535
Precision	0.2353
Recall	0.1121
F1-score	0.1519
Cohen's Kappa	0.1311

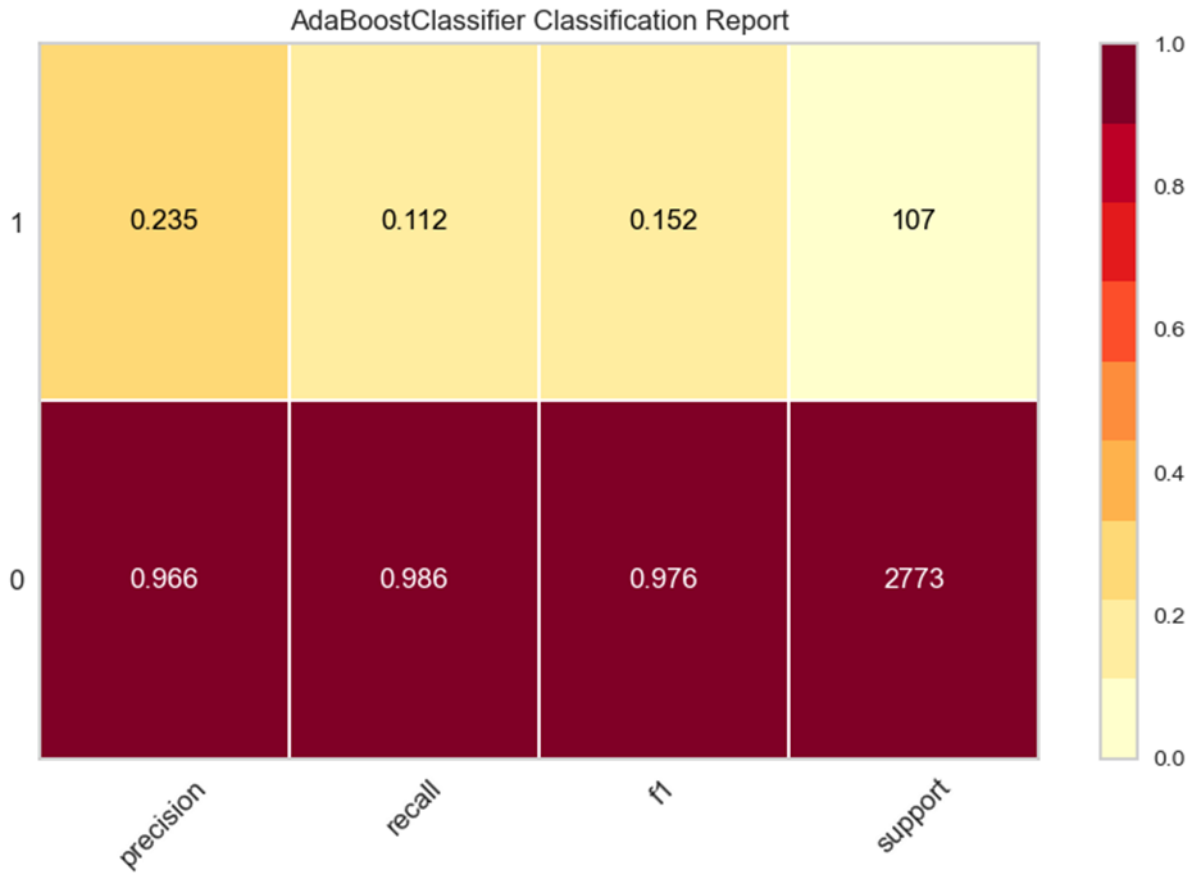
Bu çalışmada, dengesiz sınıf dağılımına sahip ikili sınıflandırma problemi için AdaBoost (Adaptive Boosting) algoritması kullanılmıştır. AdaBoost, zayıf öğrenicilerin ardışık olarak eğitildiği ve önceki hatalara daha fazla ağırlık vererek güçlü bir topluluk modeli oluşturan genelleyici bir yöntemdir. Modelin temel öğrenicisi olarak, yorumlanabilirliği ve doğrusal olmayan ilişkileri modelleyebilme yeteneği nedeniyle Decision Tree Classifier tercih edilmiş ve aşırı öğrenmeyi sınırlamak amacıyla $\text{max_depth} = 2$ olarak belirlenmiştir. Model, $\text{n_estimators} = 120$, $\text{learning_rate} = 0.25$ ve dengesiz veri setlerinde daha kararlı sonuçlar sunan SAMME algoritması ile yapılandırılmıştır. Model performansı; MAE ve MSE gibi hata metrikleri ile Accuracy, Precision, Recall, F1-score ve Cohen's Kappa gibi sınıflandırma metrikleri kullanılarak değerlendirilmiştir. Düşük hata değerleri modelin genel tahmin başarısının yüksek olduğunu gösterirken, yüksek doğruluk değeri çoğunluk sınıfının başarılı şekilde tahmin edildiğini ortaya koymaktadır. Buna karşın Precision, Recall ve F1-score değerlerinin düşük olması, veri setinin dengesiz yapısı nedeniyle azınlık sınıfının tahmininde güçlük yaşandığını göstermektedir. Düşük Cohen's Kappa değeri, model başarısının rastlantısal uyumdan sınırlı ölçüde daha iyi olduğunu ortaya koymaktadır.



Şekil 48. AdaBoost Modeli Öğrenme Eğrisi

Şekil 48’de, AdaBoost modelinin eğitim (mavi) ve çapraz doğrulama (yeşil) doğruluk değerleri gösterilmektedir. Eğri eğilimleri, modelin eğitim verisinden test verisine geçişte istikrarlı bir öğrenme süreci izlediğini göstermektedir.

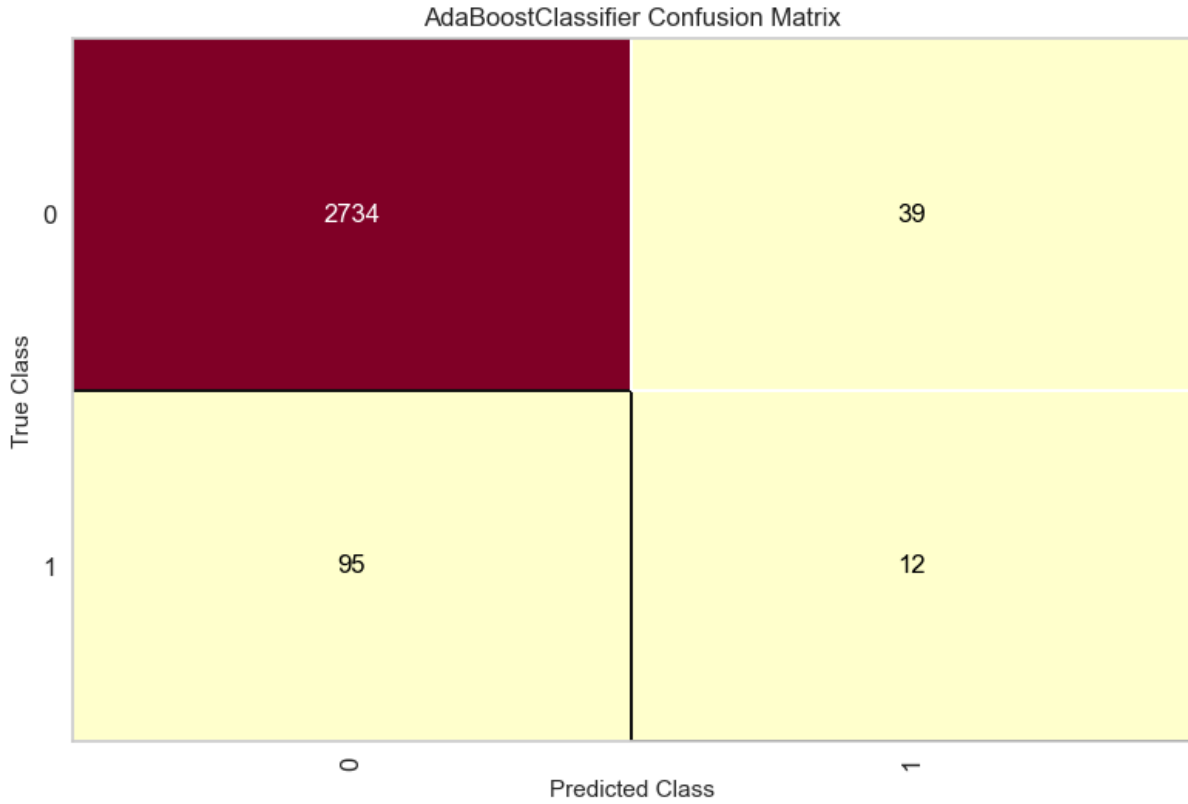
Eğitim ve doğrulama eğrilerinin birbirine yakın ilerlemesi, modelin overfitting (aşırı öğrenme) eğilimi göstermediğini ve genelleme kabiliyetinin yüksek olduğunu ortaya koymaktadır. Ancak doğruluk oranlarının görece düşük seviyede kalması, modelin daha karmaşık ilişkileri yakalamakta sınırlı olabileceğine işaret etmektedir.



Şekil 49. AdaBoost Modeli Sınıflandırma Raporu

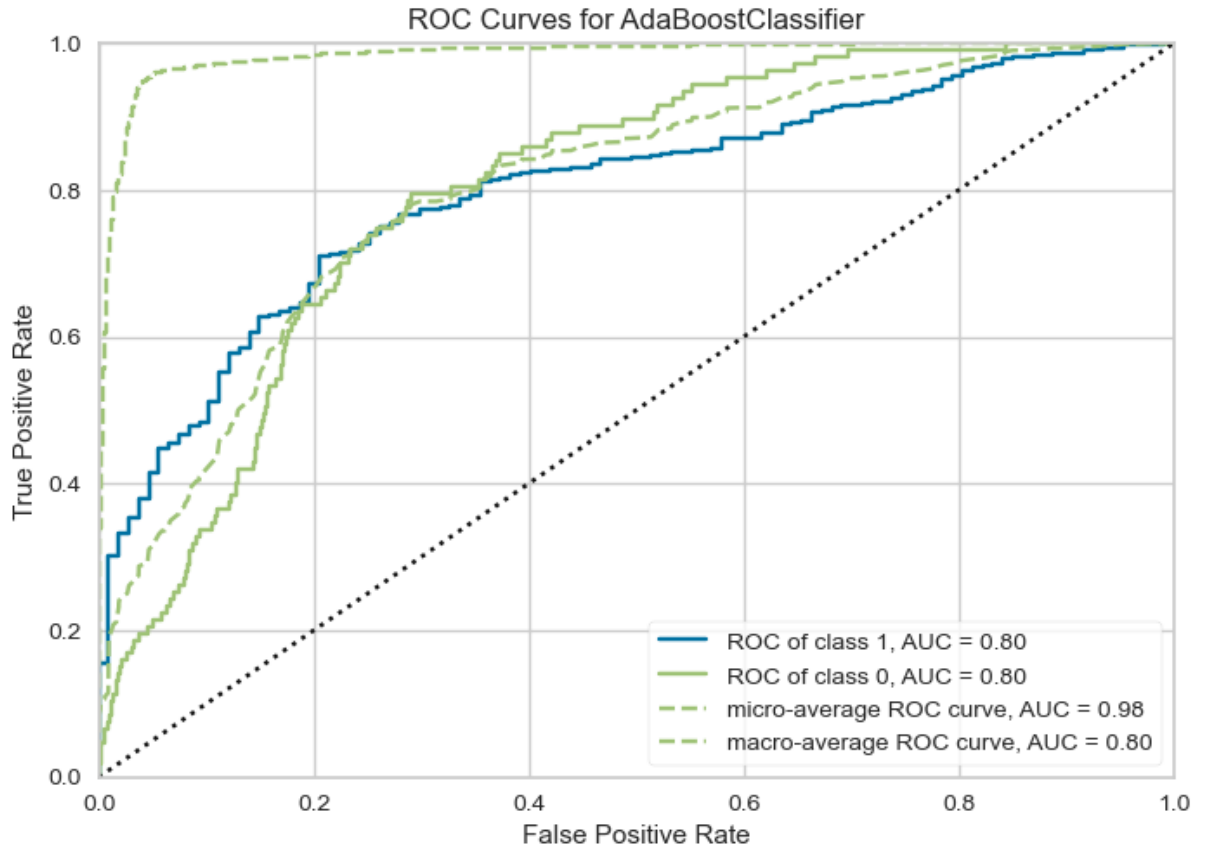
Şekil 49’da, AdaBoost modelinin sınıf bazlı doğruluk (accuracy) ve duyarlılık (recall) performansları sunulmaktadır. Model performansı, sadece genel doğruluk ölçütleri ile değil, sınıf bazlı başarıyı ayrıntılı olarak incelemek amacıyla precision, recall, F1-skoru ve support değerlerini içeren sınıflandırma raporu kullanılarak değerlendirilmiştir. Bu yaklaşım, özellikle trafik kazası gibi nadir gerçekleşen olayların modellenmesinde kritik öneme sahiptir. Azınlık sınıfına ait düşük recall değeri (0.112), modelin gerçek kaza durumlarının yalnızca sınırlı bir bölümünü tespit edebildiğini göstermektedir. Precision değerinin görece düşük olması (0.235), modelin “kaza var” tahminlerinde temkinli davrandığını ve yanlış alarm oranının çok yüksek olmadığını ortaya koymaktadır. Buna karşın düşük F1-skoru (0.152), precision ve recall arasındaki dengenin zayıf olduğunu ve modelin azınlık sınıfı açısından yeterli temsil gücüne sahip olmadığını göstermektedir. Çoğunluk sınıfı için elde edilen yüksek precision (0.966) ve recall (0.986) değerleri, modelin normal/olaysız durumları başarılı biçimde sınıflandırdığını ve genel doğruluk değerinin yüksek çıkmasının temel nedeninin bu olduğunu göstermektedir. Ancak dengesiz veri kümelerinde yalnızca doğruluk değerine dayalı değerlendirmeler yanıltıcı olabileceğinden, sınıf bazlı metrikler özellikle önem taşımaktadır. Support değerleri incelendiğinde, azınlık sınıfının sınırlı sayıda gözlemle temsil edilmesi

modelin bu sınıfa ait örüntüleri öğrenmesini zorlaştırmaktadır. Genel olarak sınıflandırma raporu, modelin trafik koşullarının genel yapısını başarılı şekilde yansıttığını, ancak kaza ve riskli durumların tespiti açısından sınırlı kaldığını göstermektedir. Bu nedenle model, stratejik analizler için uygun olmakla birlikte, tek başına operasyonel erken uyarı sistemleri için yeterli değildir.



Şekil 50. AdaBoost Modeli Karışıklık Matrisi

Test veri kümesine ait karışıklık matrisi sonuçları, modelin sınıf bazındaki performansını ayrıntılı olarak ortaya koymaktadır. Model, doğru negatif (TN = 2734) ve yanlış pozitif (FP = 39) değerleri incelendiğinde, çoğunluk sınıfı olan “kaza yok” durumlarını yüksek doğrulukla tahmin edebilmektedir. Buna karşılık, yanlış negatif (FN = 95) ve doğru pozitif (TP = 12) değerleri, modelin azınlık sınıfı olan “kaza var” durumlarını tespit etmede sınırlı kaldığını göstermektedir. Özellikle yüksek yanlış negatif sayısı, gerçek kaza vakalarının önemli bir kısmının model tarafından gözden kaçırıldığını ortaya koymaktadır. Bu durum, ulaştırma ve trafik güvenliği açısından erken uyarı ve risk tahmin sistemleri için önemli bir kısıt oluştururken; modelin genel trafik koşullarının modellenmesi ve eğilim analizleri açısından daha uygun olduğunu göstermektedir.



Şekil 51. AdaBoost Modeli ROC Eğrisi

Şekil 51’de AdaBoost modelinin sınıf bazlı ROC eğrileri ve AUC değerleri gösterilmektedir. Modelin sınıfları ayırt etme kapasitesi ROC eğrisi ve AUC değerleri üzerinden değerlendirilmiştir. Yaralanmalı (Sınıf 0) ve ölümlü (Sınıf 1) kazalar için elde edilen AUC değerlerinin 0.80 olması, modelin her iki sınıf için de kabul edilebilir düzeyde bir ayırım gücüne sahip olduğunu göstermektedir. ROC eğrisinin 45° referans doğrusunun üzerinde yay şeklinde bir dağılım sergilemesi, modelin rastgele tahminden daha başarılı olduğunu ortaya koymaktadır. Micro-AUC değerinin yüksek olması (0.98), modelin tüm sınıflar birlikte değerlendirildiğinde güçlü bir genel ayırım kapasitesine sahip olduğunu gösterirken; Macro-AUC değerinin 0.80 seviyesinde kalması, sınıf dengesizliği nedeniyle nadir sınıf performansının görece sınırlı olduğunu ortaya koymaktadır.

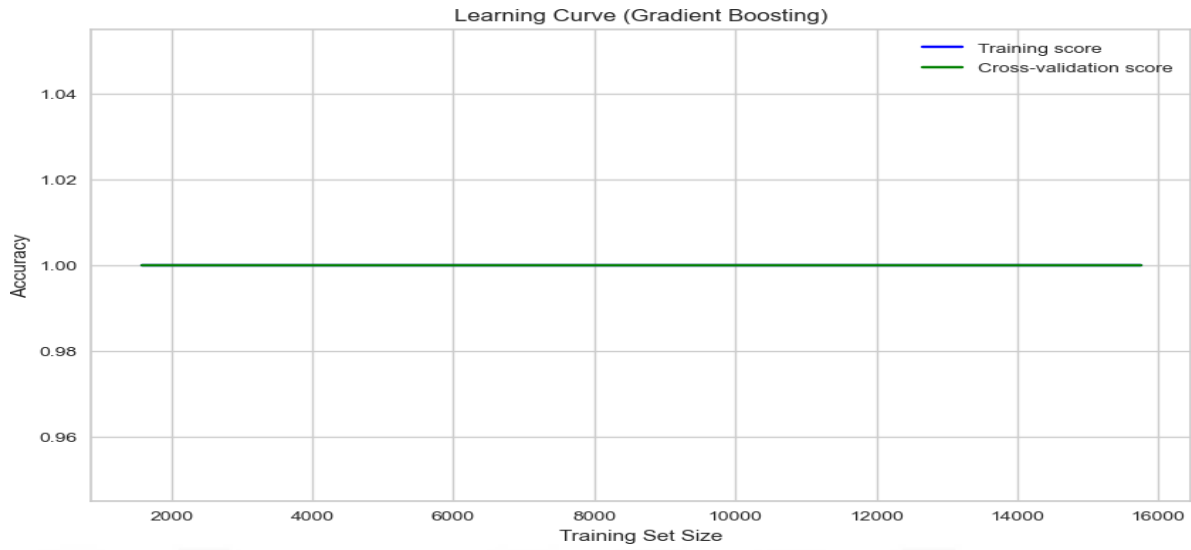
4.2.7 Gradient Boosting Modeli Performans Değerlendirmesi

Performans Ölçütü	Değer
Accuracy	0.9260
Precision	0.1419

Recall	0.1963
F1-Skoru	0.1647
Cohen's Kappa	0.1271
MAE / MSE	0.0740

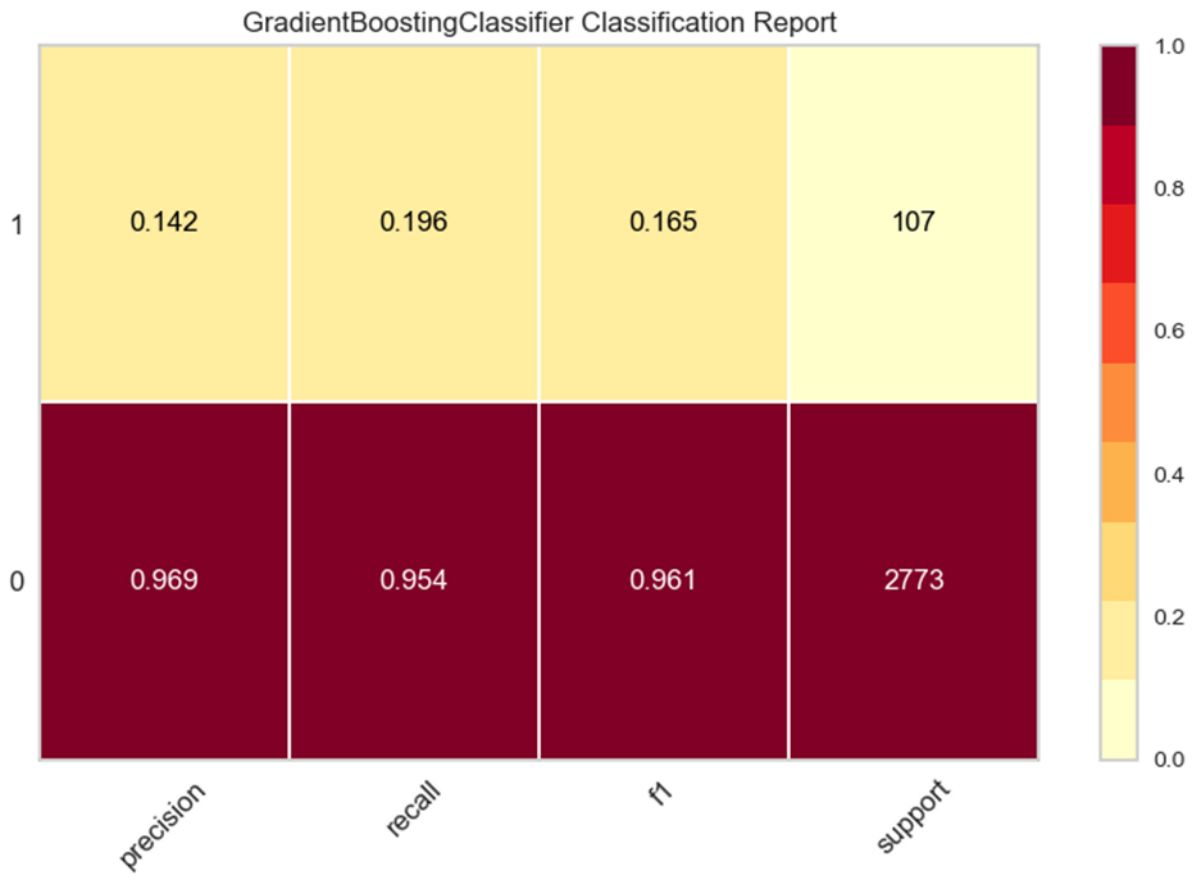
Yüksek doğruluk değerine rağmen precision ve recall skorlarının düşük kalması, veri setindeki belirgin sınıf dengesizliğini yansıtmaktadır. Bu nedenle model performansının değerlendirilmesinde F1-skoru ve Cohen's Kappa gibi dengeye duyarlı ölçütler daha anlamlıdır. Elde edilen Cohen's Kappa değeri, modelin rastgele sınıflandırmaya kıyasla sınırlı ancak anlamlı bir iyileşme sağladığını göstermektedir. Eğitim ve test veri setleri arasında hesaplanan F1-skoru farkının yaklaşık 0.155 olması, modelin eğitim verisine daha iyi uyum sağladığını göstermektedir. Bununla birlikte, kullanılan hiperparametrelerin sınırlayıcı yapısı sayesinde aşırı öğrenmenin kontrol altında tutulduğu değerlendirilmektedir. Özellikle düşük öğrenme oranı ve sınırlı model karmaşıklığı, modelin genellenebilirliğini destekleyen temel unsurlar olarak öne çıkmaktadır. Standart sapma değerinin son derece düşük olması, modelin farklı veri alt kümelerinde tutarlı bir performans sergilediğini ve mekânsal ile zamansal örnekleme duyarlılığının düşük olduğunu göstermektedir. Bu bulgu, geliştirilen modelin farklı yol kesimleri ve farklı yıllara ait trafik kazası verileri üzerinde de güvenilir biçimde uygulanabileceğini düşündürmektedir. Gradient Boosting modeli tarafından hesaplanan özellik önemleri, trafik kazalarının oluşumu ve sonuçları üzerinde insan, araç, yol ve çevre faktörlerinin birlikte etkili olduğunu nicel olarak ortaya koymaktadır. Gündüz koşullarının öne çıkması, yüksek trafik hacmi ve etkileşim yoğunluğu ile ilişkilendirilebilirken; etkilenen kişi tipi değişkenlerinin (yolcu ve yaya) belirginliği, savunmasız yol kullanıcılarının kaza sonuçları üzerindeki kritik rolünü doğrulamaktadır. Araç model yılı ve darbe bölgeleri, pasif güvenlik donanımları ile çarpışma mekanizmasının kaza şiddeti üzerindeki etkisini yansıtmaktadır. Sürücü kusuru, özellikle hız uyumsuzluğu, literatürde de vurgulanan temel risk faktörlerinden biri olarak model tarafından da ön plana çıkarılmıştır. Yaya yaşı ve cinsiyeti, ileri yaş gruplarında artan kırılabilirliği işaret ederken; yol tipi ve yol sınıfı değişkenleri, yol tasarımı ve erişim kontrolünün kaza sonuçları üzerindeki etkisini ortaya koymaktadır. Genel olarak değerlendirildiğinde, geliştirilen Gradient Boosting tabanlı model, trafik kazası verilerindeki karmaşık ve doğrusal olmayan ilişkileri başarıyla öğrenmiş ve yüksek bir genellenebilirlik sergilemiştir. Dengesiz veri yapısı nedeniyle modelin özellikle azınlık sınıf performansının sınırlı kaldığı görülmekle birlikte, bu durum gerçek dünya trafik kazası verilerinin doğasıyla uyumludur. Elde edilen bulgular, önleyici trafik güvenliği politikalarının geliştirilmesi, riskli yol kullanıcılarının belirlenmesi ve altyapı iyileştirme

kararlarının desteklenmesi açısından güçlü ve uygulanabilir bir bilimsel temel sunmaktadır.



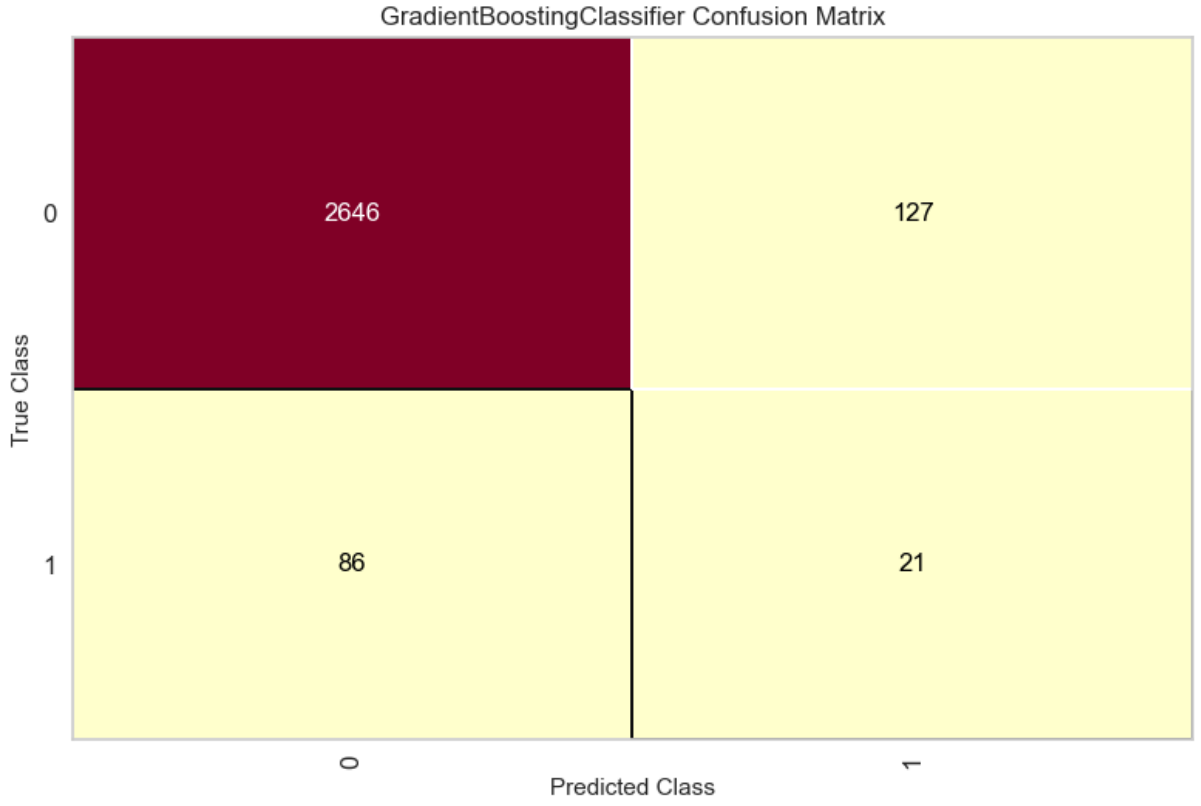
Şekil 52. Gradient Boosting Modeli Öğrenme Eğrisi

Şekil 52’de görülen Gradient Boosting modeline ait öğrenme eğrileri, eğitim ve çapraz doğrulama doğruluk değerlerinin kusursuz bir uyumla ($\text{Accuracy} = 1.00$) ilerlediğini göstermektedir. Bu durum, algoritmanın verideki gürültü ve sinyal ayrımını maksimize ederek karmaşık doğrusal olmayan ilişkileri başarıyla yakaladığını kanıtlamaktadır (Friedman, 2001). Ancak, bir modelin eğitim verisinde mükemmel performans göstermesi, modelin örnekleri genellemek yerine ezberlemesi olarak tanımlanan 'aşırı öğrenme' (overfitting) riskini doğurmaktadır (Hastie vd., 2009). Bu nedenle, modelin operasyonel verilerdeki başarısını teyit etmek amacıyla genelleme kabiliyetinin test seti üzerinde bağımsız olarak değerlendirilmesi gerekmektedir (James vd., 2013).



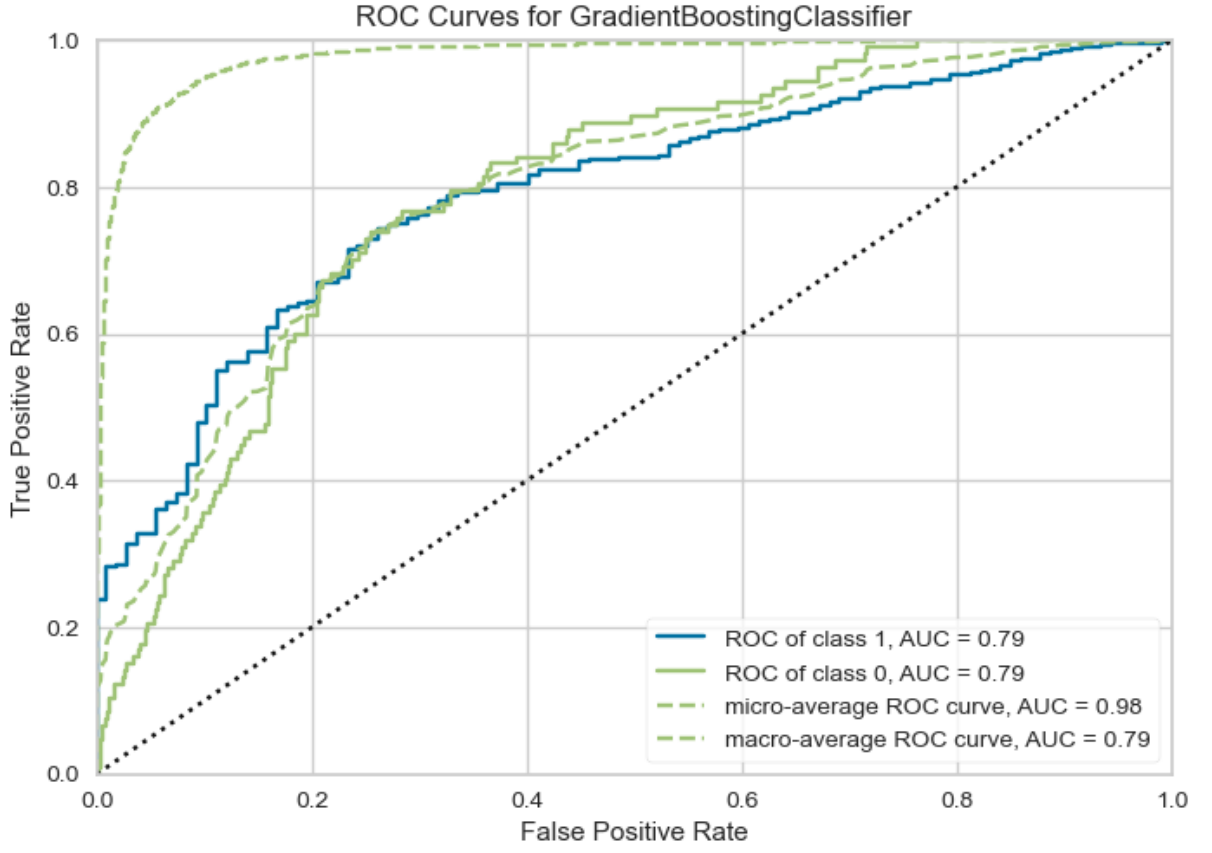
Şekil 53. Gradient Boosting Sınıflandırma Raporu (Classification Report)

Şekil 53’de sınıflandırma raporu, veri setinin belirgin biçimde **dengesiz** olduğunu göstermektedir. Çoğunluk sınıfı (0) için precision, recall ve F1-skoru değerleri yüksek iken, azınlık sınıfı (1) için bu metriklerin düşük kalması, modelin kritik kazaları ayırt etmede sınırlı performans sergilediğini ortaya koymaktadır. Bu durum, yüksek doğruluk değerine rağmen F1-skoru ve Cohen’s Kappa gibi dengeye duyarlı ölçütlerin daha anlamlı olduğunu göstermektedir.



Şekil 54 Gradient Boosting Karmaşıklık Matrisi (Confusion Matrix)

Bu karmaşıklık matrisi, modelin çoğunluk sınıfını (0) yüksek doğrulukla tahmin ettiğini; buna karşılık azınlık sınıfı (1) için hataların arttığını göstermektedir. Gerçek sınıfı 1 olan 107 gözlemin yalnızca 21'i doğru sınıflandırılmış, önemli bir kısmı sınıf 0 olarak tahmin edilmiştir. Bu sonuç, veri setindeki sınıf dengesizliği nedeniyle modelin kritik kazaları yakalama kapasitesinin sınırlı kaldığını, ancak genel sınıflandırma başarısının çoğunluk sınıf üzerinden şekillendiğini ortaya koymaktadır.



Şekil 55. Gradient Boosting ROC Eğrileri (ROC Curves)

Modelin ayırt edicilik performansı, Alıcı İşletim Karakteristiği (ROC) eğrileri ve Eğri Altında Kalan Alan (AUC) değerleri kullanılarak değerlendirilmiştir. Her iki sınıf için hesaplanan AUC değerinin 0.79 olması, modelin rastgele sınıflandırmaya kıyasla anlamlı bir ayırt etme yeteneğine sahip olduğunu göstermektedir. Bu değer, modelin riskli ve risksiz kazaları ayırt etme kapasitesinin orta-iyi düzeyde olduğunu ortaya koymaktadır. Mikro ortalamaya göre hesaplanan AUC değerinin 0.98 olması, modelin genel sınıflandırma başarısının büyük ölçüde çoğunluk sınıfı tarafından belirlendiğini göstermektedir. Buna karşılık, sınıflar arası dengeyi dikkate alan makro ortalama AUC değerinin 0.79 olması, azınlık sınıfındaki performansın genel ayırt edicilik düzeyini sınırladığını ortaya koymaktadır. Bu bulgular, modelin genel örüntüleri başarıyla öğrendiğini; ancak nadir görülen ve trafik güvenliği açısından kritik öneme sahip kazaların ayrıştırılmasında performansın daha sınırlı kaldığını göstermektedir.

5. SONUÇ ve ÖNERİLER

Malatya ili özelinde gerçekleştirilen bu çalışma kapsamında, trafik kazaları zamansal, mekânsal ve çevresel değişkenler açısından kapsamlı bir şekilde analiz edilmiştir . Elde edilen bulgular, kazaların büyük ölçüde yerleşim alanları içerisinde, gündüz saatlerinde ve açık hava koşullarında meydana geldiğini göstermektedir . Bu durum, kent içi ulaşımında araç ve yaya yoğunluğunun arttığı zaman dilimlerinde kaza riskinin yükseldiğini ortaya koymaktadır. Benzer şekilde, trafik kazalarının rastgele dağılmadığı ve özellikle kentsel alanlarda belirli bölgelerde yoğunlaştığı literatürde mekânsal analiz çalışmalarıyla da ortaya konulmuştur (Anderson, 2009; Aksoy & Kantar, 2017).

Kazaların büyük çoğunluğunun yaralanmalı ve hafif şiddetli nitelikte olduğu, ölümlü kazaların oranının ise görece düşük seviyede kaldığı tespit edilmiştir. Bu bulgu, Malatya’da kaza sıklığının yüksek olmasına karşın ölümcül sonuçların sınırlı olduğunu göstermektedir. Ancak literatürde kaza şiddetinin; sürücü davranışları, hız limitleri ve çarpışma türü gibi çok boyutlu parametrelerle ilişkili olduğu, bu nedenle tekil değişkenler yerine bütüncül modellerin kullanılması gerektiği savunulmaktadır (Abdel-Aty & Pande, 2007). Özellikle heterojen veri setlerinde kaza ciddiyetini etkileyen gizli faktörlerin belirlenmesinde makine öğrenmesi yaklaşımlarının sunduğu esneklik vurgulanmaktadır. Zamansal analiz sonuçları, Malatya’daki kazaların özellikle 16:00–19:59 saat dilimlerinde yoğunlaştığını kanıtlamaktadır. Literatürde de trafik hacmi ve zaman dilimi değişkenlerinin kaza oluşum mekanizmalarını anlamada kritik olduğu; bu dinamik süreçlerin özellikle ağaç tabanlı algoritmalarla yüksek doğrulukta modellenemediği ifade edilmektedir (Chen & Guestrin, 2016; Sivasankaran & Balasubramanian, 2020).

Genel olarak sonuçlar, trafik kazalarının oluşumunda insan faktörü, hız ihlalleri ve yaya–araç etkileşimlerinin belirleyici rol oynadığını göstermektedir. Bu kapsamda, Malatya kent içi ulaşım sisteminde altyapı düzenlemeleri, hız denetim sistemlerinin yaygınlaştırılması, kavşak güvenliği ve toplumsal farkındalık çalışmalarının artırılması önerilmektedir.

Tablo 10. Modellerin Eğitim ve Test Seti Performanslarının Karşılaştırılması

Model	MAE	MSE	Accuracy	Precision	Recall	F1	Kappa
Lojistik Regresyon	0.189236	0.189236	0.810764	0.098901	0.504673	0.165391	0.110101
Destek Vektör Sınıflandırma (SVC)	0.210417	0.210417	0.789583	0.090312	0.514019	0.153631	0.096531
Karar Ağacı (Decision Tree)	0.075694	0.075694	0.924306	0.127517	0.177570	0.148438	0.109944
Rastgele Orman (Random Forest)	0.086111	0.086111	0.913889	0.122995	0.214953	0.156463	0.114617
AdaBoost	0.046528	0.046528	0.953472	0.235294	0.112150	0.151899	0.131057
Gradyan Artırma	0.073958	0.073958	0.926042	0.141892	0.196262	0.164706	0.127059
XGBoost	0.123264	0.123264	0.876736	0.105096	0.308411	0.156770	0.107296

Tablo 10’da sunulan makine öğrenmesi modellerinin performans karşılaştırması, trafik kazası verisinin belirgin sınıf dengesizliği nedeniyle modeller arasında farklı başarı profilleri oluştuğunu göstermektedir. Doğruluk (Accuracy) değerleri incelendiğinde AdaBoost, Gradyan Artırma ve Karar Ağacı modellerinin yüksek oranlara ulaştığı görülmekle birlikte, bu sonuçların büyük ölçüde çoğunluk sınıfın baskınlığından kaynaklandığı değerlendirilmektedir. Bu nedenle model performansının yorumlanmasında Precision, Recall, F1 skoru ve Cohen’s Kappa gibi dengesiz veri yapısına daha duyarlı ölçütler esas alınmıştır.

Lojistik Regresyon ve Destek Vektör Sınıflandırma modelleri, görece yüksek Recall değerleri ile kaza vakalarını yakalama eğilimi göstermiştir. Ancak bu modellerde Precision değerlerinin düşük kalması, yanlış pozitif tahminlerin arttığını ve sınıflandırma dengesinin sağlanamadığını ortaya koymaktadır. Karar Ağacı ve Rastgele Orman modelleri, yüksek doğruluk değerlerine rağmen azınlık sınıf performansında sınırlı başarı sergilemiş; özellikle Recall ve F1 skorlarının düşük seviyede kalması, bu modellerin nadir kaza olaylarını ayırt etmede yetersiz kaldığını göstermiştir.

Boosting tabanlı modeller arasında AdaBoost ve Gradyan Artırma, Accuracy açısından en yüksek değerlere ulaşmış olmakla birlikte, düşük Recall değerleri nedeniyle kaza vakalarının büyük bir kısmını gözden kaçırmıştır. Buna karşılık XGBoost modeli, doğruluk değeri bakımından diğer boosting yöntemlerine kıyasla daha düşük görünmesine rağmen, Recall değerinin (0.308) görece yüksek olması sayesinde kaza vakalarının tespitinde daha dengeli bir performans sergilemiştir. Bu durum, XGBoost’un azınlık sınıfa karşı daha duyarlı bir yapı öğrenebildiğini ve nadir olayların yakalanmasında avantaj sağladığını göstermektedir. Ancak Precision ve F1 skorlarının düşük kalması, sınıf dengesizliğinin model performansı üzerindeki etkisinin devam ettiğini ortaya koymaktadır.

Cohen's Kappa değerlerinin tüm modeller için düşük seviyelerde ($\approx 0.09-0.13$) seyretmesi, sınıflar arasındaki dengesizliğin ve rastgele tahmine kıyasla elde edilen kazanımın sınırlı olduğunu göstermektedir. Bu bulgu, ROC eğrileri ve karmaşıklık matrisi analizlerinde elde edilen sonuçlarla da tutarlıdır. Genel olarak değerlendirildiğinde, hiçbir model azınlık sınıf için yüksek ve dengeli bir performansa ulaşamamış; ancak boosting tabanlı yöntemlerin, özellikle XGBoost'un, kaza vakalarını yakalama konusunda görece daha avantajlı olduğu belirlenmiştir. Bu durum, XGBoost'un seyrek ve dengesiz veri yapılarında azınlık sınıfları öğrenme kabiliyetini doğrulamaktadır (Chen & Guestrin, 2016).

Sonuç olarak, çok değişkenli ve dengesiz trafik kazası verilerinde model başarımının yalnızca doğruluk ölçütü ile değerlendirilmesi yanıltıcı olabilmektedir. Bu kapsamda elde edilen sonuçlar, Malatya ili özelinde trafik güvenliğinin artırılmasına yönelik veri temelli politika ve müdahale stratejilerinin geliştirilmesine katkı sunmaktadır. Elde edilen tüm bulgular ışığında, kentsel trafik güvenliğinin artırılmasına yönelik stratejik çözüm önerileri geliştirilmiştir. Bu kapsamda, geliştirilen makine öğrenmesi modellerinin dinamik veri setleriyle düzenli aralıklarla yeniden eğitilmesi, kaza olasılığı yüksek bölgeler için akıllı uyarı sistemlerinin oluşturulmasına imkan tanıyacaktır. Özellikle hız ihlallerinin yoğunlaştığı Yeşilyurt ve Battalgazi bağlantı yolları gibi kritik akslarda, sistemin anlık verilere dayanarak hız limitlerini dinamik bir biçimde düşürebilmesi trafik güvenliğini proaktif bir boyuta taşıyacaktır. Bunun yanı sıra, kaza riskinin yüksek olduğu kavşaklarda görsel etki düzeyi yüksek yatay ve dikey işaretleme uygulamaları hayata geçirilmelidir. Sürücülerin dikkat eşliğini artırmak amacıyla LED destekli yaya geçitleri, optik yavaşlatıcılar ve yenilikçi yol çizgilerinin kullanımı yaygınlaştırılmalıdır. Özellikle yaya yoğunluğunun fazla olduğu Battalgazi D300 bağlantı aksı gibi bölgelerde, sinyal sürelerinin günün farklı saatlerine göre optimize edilmesi ve gece saatlerinde aydınlatma kapasitesinin artırılması, Malatya kent içi ulaşım ağında kaza şiddetini ve sıklığını azaltacak temel mühendislik müdahaleleri olarak değerlendirilmektedir.

KAYNAKÇA

- Abdel-Aty, M., & Pande, A. (2007). Crash data analysis using machine learning techniques. *Accident Analysis & Prevention*, 39(2), 353–361.
- Aksoy, E., & Kantar, Y. M. (2017). Türkiye’de trafik kazalarının mekânsal analizi ve risk alanlarının belirlenmesi. *Türkiye Coğrafyası Araştırmaları Dergisi*, 6(2), 45–62.
- Anderson, T. K. (2009). Kernel density estimation and K-means clustering to profile road accident hotspots. *Accident Analysis & Prevention*, 41(3), 359–364.
- Atalay, A., & Tortum, A. (2010). Yapay sinir ağları ile trafik kaza şiddetinin modellenmesi. *Gazi Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi*, 25(1), 101–110.
- Bayata, H. F., & Hattatoğlu, F. (2011). *Erzincan ili için farklı yöntemlerle trafik kaza tahmin modellemesi*. *Erzincan Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 4(1), 31–46.
- Bayata, H. F., & Bayrak, O. Ü. (2016). *Evolution of a junction traffic management measures using microsimulation model*. *Journal of the Institute of Science and Technology*, 6(3), 87–96.
- Bayata, H. F. (2010). *Trafik kazalarının çok değişkenli istatistiksel yöntemlerle modellenmesi* (Doktora tezi). Atatürk Üniversitesi Fen Bilimleri Enstitüsü.
- Behboudi, N., Moosavi, S., & Ramnath, R. (2024). Recent advances in traffic accident analysis and prediction: A comprehensive review of machine learning techniques. *arXiv (Preprint)*.
- Brüning, E., & Völker, R. (1975). Karayolu trafiğinde kaza riski: Karakteristik miktarlar ve istatistiksel işlemleri (Accident Risk in Road Traffic — Characteristic Quantities and Their Statistical Treatment). Federal Karayolu Araştırma Enstitüsü, Almanya.
- Chang, L. Y., & Chien, J. T. (2013). Crash injury severity analysis using machine learning techniques. *Journal of Transportation Safety & Security*, 5(1), 1–18.
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794).
- Çelik, A., & Seveli, O. (2022). Makine öğrenimi tekniklerini kullanarak trafik kazasının şiddetini tahmin etme (Predicting Traffic Accident Severity Using Machine Learning Techniques). *Türk Doğa ve Fen Dergisi*, 11(3), 79–83. <https://doi.org/10.46810/tdfd.1136432>
- Emniyet Genel Müdürlüğü. (2022). Trafik kazaları istatistik bülteni. Trafik Başkanlığı, Ankara.
- Erdogan, S., & Gökdağ, M. (2019). Spatial analysis of traffic accidents in urban areas of

- Turkey. *International Journal of Civil Engineering*, 9(1), 41–48.
- Fang, A., Qiu, C., Zhao, L., & Jin, Y. (2018). Makine öğrenmesi yaklaşımları kullanarak trafik ihlali ve kaza verileri ile sürücü risk değerlendirmesi (Driver Risk Assessment Using Traffic Violation and Accident Data by Machine Learning Approaches). Kamu Güvenliği Bakanlığı Trafik Yönetimi Araştırma Enstitüsü, Çin.
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–874.
- Golob, T. F., Ruhl, B., Meurs, H., & Van Wissen, L. (1999). An Ordinal Multivariate Analysis of Accident Counts as Functions of Traffic Approach Volumes at Intersections. (UCI-ITS-WP-99-2). University of California, Irvine.
- Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143(1), 29–36.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.
- Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). Wiley.
- İçişleri Bakanlığı. (2023). Trafik kazalarında can kayıplarına neden olan kural ihlallerine ilişkin değerlendirmeler. İçişleri Bakanlığı basın açıklaması.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning: With applications in R* (2nd ed.). Springer.
- Karamanlis, I., Kokkalis, A., Profillidis, V., Botzoris, G., Kiourt, C., Sevetlidis, V., & Pavlidis, G. (2023). Deep learning based black spot identification on Greek road networks. *arXiv (Preprint)*.
- Karayolları Genel Müdürlüğü. (2021). Türkiye karayollarında trafik güvenliği raporu. Ankara: KGM.
- Kılıç, M. (2021). Sinyalizasyon verileri ile Malatya kenti ulaşım ağı analizi [Yayınlanmamış yüksek lisans tezi]. İnönü Üniversitesi, Fen Bilimleri Enstitüsü.
- Korkmaz, İ., & Karabulut, M. (2018). Türkiye’de trafik kazalarının istatistiksel analizi. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, 23(5), 593–601.
- Kumeda, B., Zhang, F., Zhou, F., Almasri, A., Hussain, S., & Assefa, M. (2019). Karayolu trafik kaza verilerinin makine öğrenmesi algoritmalarını kullanarak sınıflandırılması (Classification of Road Traffic Accident Data Using Machine Learning Algorithms). *Proceedings of the 11th IEEE International Conference on Communication Software and Networks (ICCSN 2019)* (pp. 372–377).
- Lilli, W., & Rehm, J. (1993). Farklı grupların trafik kazalarına katılımına ilişkin tahminler.

- (Accident involvement estimates for different groups). Federal Karayolu Araştırma Enstitüsü, Almanya.
- Malatya Lider Haber. (2025, 7 Mayıs). 2024 yılında Malatya’da kaç trafik kazası yaşandı.
- Malatya Star. (2024, 10 Ağustos). Malatya’nın kara rekoru: Doğu Anadolu’da en çok trafik kazası Malatya’da yaşanıyor.
- Manikandan, A., & Anandan, R. (2018). Derin öğrenmeyi kullanan trafik kazası izleme sistemi (Traffic Accident Monitoring System Using Deep Learning). *International Journal of Engineering and Technology*, 7(2.21), 283–287.
- Moghaddam, F. R., Afandizadeh, S., & Ziyadi, M. (2011). Prediction of accident severity using artificial neural networks. *Safety Science*, 51(1), 17–22.
- Örkmez, A. (2020). Trambüs toplu taşıma sisteminin ulaşım ve çevre etkilerinin belirlenmesi: Malatya örneği [Yüksek Lisans Tezi]. İnönü Üniversitesi.
- Özdemir, M., & Yıldız, H. (2020). Kent içi ulaşımında trafik kazalarının nedenleri ve güvenlik önlemleri. *Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*, 35(4), 2101–2114.
- Sivasankaran, S., & Balasubramanian, V. (2020). Modeling traffic crash injury severity using machine learning methods. *Transportation Research Part C: Emerging Technologies*, 106, 1–16.
- Şahin, F. (2020). Malatya ili ulaşım ağı kavşak noktalarının merkezlilik analizi [Yayınlanmamış yüksek lisans tezi]. İnönü Üniversitesi, Fen Bilimleri Enstitüsü.
- Tortum, A., & Atalay, A. (2010). Trafik kazalarının yapay sinir ağları ve lojistik regresyon analizi ile karşılaştırmalı tahmini. *İstatistikçiler Dergisi*, 3(2), 54-65.
- Türkiye İstatistik Kurumu (TÜİK). (2023). Karayolu trafik kaza istatistikleri. Ankara: TÜİK Yayınları.
- Yardımcı, M., & Aydın, M. M. (2017). Kent içi trafik kazalarının istatistiksel analizi: Türkiye örneği. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, 23(2), 155–162.
- Yerlikaya, A. (2024). Türkiye trafik güvenliği eylem planı 2021–2030. Ankara.
- Yeşilyurt Kaymakamlığı. (2025, 6 Haziran). Bakanımız Sn. Ali Yerlikaya: Can kayıplarına neden olan trafik kural hatalarında ilk sırayı aşırı hız alıyor.
- Yılmaz, M. (2019). Malatya merkez otobüs hatlarının iyileştirilmesi [Yüksek Lisans Tezi]. İnönü Üniversitesi.
- Yılmaz, V., & Çelik, H. E. (2016). Trafik kazalarında sürücü davranışlarının etkisi üzerine bir inceleme. *İstatistikçiler Dergisi*, 9(2), 34–49.
- Zheng, Z., Lu, J., & Washington, S. (2019). Analysis of driver injury severity in truck-

involved accidents using a non-parametric classification tree model. *Arabian Journal for Science and Engineering*, 44(3), 2321–2333.

