

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

TRAFİK KAZALARININ MAKİNE ÖĞRENMEYİLE MEKÂNSAL ANALİZİ

Alparslan GENÇ

Danışman: Doç. Dr. Halim Ferit BAYATA

TEZ JÜRİ ÜYELERİ

Doç. Dr. Halim Ferit BAYATA

Doç. Dr. Osman Ünsal BAYRAK

Dr. Öğr. Üyesi Ömer Faruk KELEŞ

YÜKSEK LİSANS TEZİ

ERZİNCAN, 2026

© 2026 [Alparslan GENÇ]. Tüm hakları saklıdır.

## Kabul ve Onay Sayfası

Doç. Dr. Halim Ferit BAYATA danışmanlığında, Alparslan GENÇ tarafından hazırlanan bu çalışma ..... tarihinde aşağıdaki jüri tarafından ..... Anabilim Dalı ..... Bilim Dalı'nda Yüksek Lisans/Doktora Tezi olarak kabul oybirliği/oy çokluğu (.../...) ile kabul edilmiştir.

Başkan	:	Unvan Ad SOYAD	İmza:
Üye	:	Unvan Ad SOYAD	İmza:
Üye	:	Unvan Ad SOYAD	İmza:
Üye	:	Unvan Ad SOYAD	İmza:
Üye	:	Unvan Ad SOYAD	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**

“Trafik Kazalarının Makine Öğrenmesiyle Mekânsal Analizi” isimli “Yüksek Lisans” tezim tarafımda intihal tespit programı ile incelenmiştir. Buna göre tezimde bilimsel etik ihlali ve intihal olarak nitelendirilebilecek herhangi bir durum olmadığını taahhüt ederim.

Bu çalışmadaki tüm bilgilerin, akademik ve etik kurallara uygun bir biçimde elde edildiğini; aynı zamanda bu kural ve davranışların gerektirdiğı gibi, bu çalışmanın özünde olmayan tüm materyal ve sonuçları tam olarak aktardığımı ve referans gösterdiğimi beyan ederim. 10/06/2026

(İmza)

**Alparslan GENÇ**

## ÖZET

### TRAFİK KAZALARININ MAKİNE ÖĞRENME Sİ İYLE MEKÂNSAL ANALİZİ

Alparslan GENÇ

Yüksek Lisans Tezi

Erzincan Binali Yıldırım Üniversitesi, Fen Bilimleri Enstitüsü,

İnşaat Mühendisliği Anabilim Dalı

Danışman: Doç. Dr. Halim Ferit BAYATA

2026, 107 sayfa

Trabzon ilinde 2014–2023 yılları arasında meydana gelen trafik kazalarının mekânsal dağılımı ve kaza şiddeti, Coğrafi Bilgi Sistemleri (CBS) ve makine öğrenmesi yöntemleri kullanılarak analiz edilmiştir. Çalışmada veri seti eksik ve yinelenen kayıtlardan arındırılmış, kategorik değişkenler sayısallaştırılmış ve sınıf dengesizliği SMOTE yöntemi ile giderilmiştir. CBS analizleri kapsamında kaza verileri haritalandırılmış, noktasal dağılım ve Kernel Yoğunluk Tahmini (KDE) analizleri gerçekleştirilmiştir. Sonuçlar, trafik kazalarının özellikle şehir merkezi, ana ulaşım arterleri ve yoğun trafik bölgelerinde kümелendiğini göstermiştir. İlçe bazında yapılan değerlendirmelerde Ortahisar ve Maçka ilçelerinde belirgin yoğunlaşmalar tespit edilmiştir. Makine öğrenmesi aşamasında Lojistik Regresyon, Karar Ağacı, Rastgele Orman, AdaBoost, Gradient Boosting ve XGBoost algoritmaları uygulanmıştır. Modeller doğruluk, kesinlik, duyarlılık, F1 skoru ve Cohen's Kappa katsayısı gibi performans ölçütleri kullanılarak değerlendirilmiştir. AdaBoost modeli en yüksek doğruluk oranına (%97,06) ulaşmasına rağmen düşük Recall ve F1 skorları nedeniyle ölümcül kazaların tahmininde sınırlı performans göstermiştir. Lojistik Regresyon modeli yüksek Recall değeri ile öne çıkarken, Karar Ağacı ve Rastgele Orman modellerinde aşırı öğrenme eğilimi gözlenmiştir. Tüm performans ölçütleri birlikte değerlendirildiğinde Gradient Boosting modeli %96,05 doğruluk, %21,43 F1 skoru, 0,1941 Cohen's Kappa katsayısı ve en düşük aşırı öğrenme düzeyi ile en başarılı model olarak belirlenmiştir. Sonuçlar, CBS ve makine öğrenmesi yöntemlerinin birlikte kullanımının trafik kazalarının mekânsal örüntülerinin belirlenmesi ve kaza şiddetinin tahmin edilmesinde etkili bir yaklaşım sunduğunu göstermektedir.

**Anahtar Kelimeler:** Trafik kazaları, Makine öğrenmesi, Mekânsal analiz, Coğrafi bilgi sistemleri, Kernel yoğunluk analizi

## ABSTRACT

### SPATIAL ANALYSIS OF TRAFFIC ACCIDENTS USING MACHINE LEARNING

Alparslan GENÇ

Master's Thesis

Erzincan Binali Yıldırım University, Institute of Science and Technology,

Department of Civil Engineering

Advisor: Assoc. Prof. Dr. Halim Ferit BAYATA

2026, 107 pages

The spatial distribution and severity of traffic accidents that occurred in Trabzon Province between 2014 and 2023 were analyzed using Geographic Information Systems (GIS) and machine learning techniques. The dataset was preprocessed by removing missing and duplicate records, converting categorical variables into numerical formats, and addressing class imbalance through the Synthetic Minority Over-sampling Technique (SMOTE). Within the scope of GIS analyses, accident data were mapped using coordinate information, and point pattern as well as Kernel Density Estimation (KDE) analyses were conducted. The results revealed that traffic accidents were predominantly clustered in the city center, along major transportation corridors, and in areas with high traffic density. District-level analyses indicated significant concentrations of accidents in Ortahisar and Maçka districts. In the machine learning phase, Logistic Regression, Decision Tree, Random Forest, AdaBoost, Gradient Boosting, and XGBoost algorithms were implemented. Model performances were evaluated using Accuracy, Precision, Recall, F1-score, and Cohen's Kappa coefficient. Although the AdaBoost model achieved the highest accuracy rate (97.06%), its low Recall and F1-score values indicated limited effectiveness in predicting fatal accidents. Logistic Regression demonstrated a high Recall value, whereas Decision Tree and Random Forest models exhibited overfitting tendencies. Considering all performance metrics collectively, the Gradient Boosting model emerged as the most successful approach, achieving 96.05% accuracy, a 21.43% F1-score, a Cohen's Kappa coefficient of 0.1941, and the lowest level of overfitting among the evaluated models. Overall, the findings demonstrate that the integrated use of GIS and machine learning provides an effective framework for identifying spatial traffic accident patterns and predicting accident severity.

**Keywords:** Traffic accidents, Machine learning, Spatial analysis, Geographic Information Systems (GIS), Kernel density analysis,

## TEŞEKKÜR

Bu çalışmanın yürütülmesi süresince bilgi birikimi ve tecrübeleriyle bana rehberlik eden, akademik gelişimime önemli katkılar sunan danışmanım Doç. Dr. Halim Ferit BAYATA'ya içtenlikle teşekkür ederim. Çalışma sürecinde kıymetli görüş ve önerileriyle destek veren Doç. Dr. Fatih İrfan BAŞ'a ve Dr. Öğr. Üyesi M. Ali ÇOLAK'a teşekkürlerimi sunarım. Ayrıca, tez çalışmamın değerlendirilmesi sürecinde değerli görüş, öneri ve katkılarıyla çalışmama yön veren jüri üyelerim Doç. Dr. Osman Ünsal BAYRAK'a ve Dr. Öğr. Üyesi Ömer Faruk KELEŞ'e teşekkürlerimi sunarım. Tez çalışmasının yürütülmesi sürecinde değerli katkıları, destekleri ve yol gösterici önerileriyle çalışmaya önemli katkılar sunan Dr. Fatih Ahmet DENİZ'e teşekkürlerimi sunarım. Ayrıca araştırma kapsamında veri temini ve katkılarıyla destek sağlayan Trabzon'daki ilgili kurum ve kuruluşlara şükranlarımı ifade ederim. Çalışmanın CBS ile yürütülen analiz aşamalarında katkı ve desteklerini esirgemeyen kıymetli kardeşim Abdullah Emin KOCATÜRK'e teşekkür ederim. Yüksek lisans sürecim boyunca her türlü desteği sağlayan ve yol gösteren değerli arkadaşım Yüksek Biyomedikal Mühendisi Safiye Betül BULUT'a teşekkür ederim. Son olarak, bu süreç boyunca her zaman yanımda olan, maddi ve manevi desteklerini esirgemeyen anneme ve babama sonsuz teşekkür ederim.

Alparslan GENÇ

Haziran, 2026

# İÇİNDEKİLER

ÖZET .....	i
ABSTRACT .....	ii
TEŞEKKÜR .....	iii
İÇİNDEKİLER.....	iv
TABLOLAR DİZİNİ.....	vii
ŞEKİLLER DİZİNİ .....	viii
SİMGELER VE KISALTMALAR DİZİNİ .....	xi
1.GİRİŞ.....	1
1.1. Araştırmanın Amacı .....	2
1.2. Araştırmanın Önemi .....	2
1.3. Varsayımlar .....	3
1.4. Sınırlılıklar.....	3
2. KAVRAMSAL ÇERÇEVE VE İLGİLİ ÇALIŞMALAR .....	4
2.1. Trafik Kazalarında Risk ve Şiddet Kavramı.....	5
2.1.1. Trafik kazalarını etkileyen faktörler .....	5
2.1.1.1. Demografik faktörler .....	6
2.1.1.2. Yol ve çevresel faktörler.....	6
2.1.1.3. Araç ve davranışsal faktörler.....	6
2.2. Makine Öğrenmesi Yaklaşımları.....	7
2.2.1. Denetimli öğrenme .....	7
2.2.2. Denetimsiz öğrenme .....	8
2.2.3. Pekiştirmeli öğrenme.....	8
2.2.4. Derin öğrenme .....	9
2.3. Coğrafi Bilgi Sistemleri ve Mekânsal Analiz.....	9
2.3.1. Noktasal mekânsal veri yapısı .....	10
2.3.2 Yoğunluk analizi .....	10
3. YÖNTEM.....	13
3.1. Çalışma Bölgesi.....	14
3.1.1. Nüfus ve motorlu araç istatistikleri .....	15
3.2. Veri Setinin Tanımı .....	16
3.2.1. Veri setinin genel özellikleri.....	16
3.2.2. Değişkenlerin tanımı .....	17

3.2.3. Hedef deęişken (Kaza Şiddeti) .....	17
3.3. Veri Ön İşleme Süreci .....	18
3.3.1. Eksik verilerin analizi ve temizlenmesi .....	19
3.3.2. Yinelenen verilerin kaldırılması .....	19
3.4. Deęişken Dönüştürme ve Gruplandırma .....	20
3.4.1. Araç model yılı gruplandırması .....	20
3.4.2. Yaya yaş gruplarının oluşturulması .....	21
3.4.3. Sürücü yaş gruplarının oluşturulması .....	21
3.5. Kategorik Deęişkenlerin Sayısallaştırılması .....	22
3.6. Sınıf Dengesizliği Problemi ve SMOTE Yöntemi .....	22
3.7. Mekânsal Analiz ve CBS Uygulaması .....	23
3.7.1. Verilerin CBS ortamına aktarılması ve hazırlanması .....	23
3.7.2. Kaza verilerinin haritalandırılması .....	27
3.8. Eğitim ve Test Veri Setlerinin Oluşturulması .....	30
3.9. Kullanılan Makine Öğrenmesi Yöntemleri .....	31
3.9.1. Lojistik regresyon .....	32
3.9.2. Karar ağacı.....	33
3.9.3. Rastgele orman .....	34
3.9.4. Adaboost.....	35
3.9.5. Gradient Boosting.....	36
3.9.6. XGBoost.....	37
3.10. Model Performans Deęerlendirme Ölçütleri .....	39
3.10.1. Accuracy (Doęruluk).....	39
3.10.2. Precision (Kesinlik) .....	40
3.10.3. Recall (Duyarlılık).....	40
3.10.4. F1 Skoru .....	41
3.10.5. Cohen's Kappa .....	41
3.10.6. MAE ve MSE .....	42
3.11. Confusion Matrix Analizi.....	43
3.12. ROC Eğrisi ve AUC Analizi .....	43
3.13. Çapraz Doğrulama (Cross Validation) .....	44
3.14. Özellik Önem Analizi.....	45
4. BULGULAR .....	47
4.1. Mekânsal Analiz Bulguları.....	47

4.1.1. Trafik kazalarının mekânsal noktasal dağılımının analizi .....	47
4.1.2. Trafik kazalarının yoğunluk analizi bulguları .....	50
4.1.3. İlçe düzeyinde trafik kazası yoğunluklarının değerlendirilmesi .....	53
4.2. Tanımlayıcı İstatistik Bulguları .....	55
4.2.1. Araç model yılı gruplarının dağılımı .....	55
4.2.2. Yaya yaş gruplarının dağılımı .....	57
4.2.3. Sürücü yaş gruplarının dağılımı .....	58
4.2.4. Kaza şiddeti dağılımı .....	60
4.2.5. Kaza özelliklerinin il geneli ve ilçe bazında karşılaştırılmalı analizi .....	61
4.2.5.1. Yol tipine göre dağılım .....	62
4.2.5.2. Kavşak durumuna göre dağılım .....	64
4.2.5.3. Gündüz/gece durumuna göre dağılım .....	65
4.3. Makine Öğrenmesi Modellerinin Sonuçları .....	66
4.3.1. Karar ağacı sonuçları .....	67
4.3.2. Rastgele orman sonuçları .....	72
4.3.3. Adaboost sonuçları .....	77
4.3.4. Gradient Boosting sonuçları .....	82
4.3.5. Lojistik Regresyon sonuçları .....	87
4.3.6. XGBoost sonuçları .....	92
4.4. Modellerin Karşılaştırılması .....	97
5. TARTIŞMA VE SONUÇ .....	102
KAYNAKÇA .....	105

## TABLolar DİZİNİ

Tablo 1. Trabzon ili yıllara göre nüfus ve motorlu kara taşıtları sayıları.....	15
Tablo 2. Çalışmada kullanılan değişkenler.....	17
Tablo 3. Makine öğrenmesi modellerinde kullanılacak parametreler. ....	32
Tablo 4. Makine öğrenmesi modellerinin performans karşılaştırması .....	97
Tablo 5. Precision değerlerine göre model sıralaması.....	98
Tablo 6. Recall değerlerine göre model sıralaması .....	98
Tablo 7. F1 skorlarına göre model sıralaması .....	99
Tablo 8. Cohen's Kappa katsayılarına göre model sıralaması .....	99
Tablo 9. Modellerin hata ölçütlerinin karşılaştırılması .....	100
Tablo 10. Modellerin aşırı öğrenme düzeylerinin karşılaştırılması.....	100
Tablo 11. Performans ölçütlerine göre en başarılı modeller .....	101

## ŞEKİLLER DİZİNİ

Şekil 1. Tez çalışmasına ait akış diyagramı.....	13
Şekil 2. Trabzon'un Türkiye'deki konumu .....	14
Şekil 3. Trabzon ve ilçeleri.....	14
Şekil 4. Trafik kazası verilerinin Excel ortamındaki görünümü .....	24
Şekil 5. Kaza verilerinin XY koordinatları kullanılarak CBS ortamına aktarılması .....	25
Şekil 6. Noktasal kaza verilerinin harita üzerindeki dağılımı .....	25
Şekil 7. Kaza verilerine ait koordinat sistemi (WGS 84) ayarları .....	26
Şekil 8. Trabzon ilçe sınırlarının CBS ortamında oluşturulması.....	26
Şekil 9. Trabzon il sınırının CBS ortamında vektör olarak oluşturulması .....	27
Şekil 10. Trafik kazası verilerinin CBS ortamında noktasal olarak gösterimi .....	28
Şekil 11. Kernel yoğunluk analizi parametre ayarlarının uygulanması.....	29
Şekil 12. Kernel yoğunluk analizi sonucunda elde edilen yoğunluk yüzeyinin CBS ortamında gösterimi.....	30
Şekil 13. Trabzon il geneline ait trafik kazalarının noktasal dağılımı.....	48
Şekil 14. Maçka ilçesine ait trafik kazalarının noktasal dağılımı.....	49
Şekil 15. Ortahisar ilçesine ait trafik kazalarının noktasal dağılımı.....	50
Şekil 16. Trabzon il geneline ait trafik kazası Kernel yoğunluk haritası .....	51
Şekil 17. Maçka ilçesine ait trafik kazası Kernel yoğunluk haritası .....	52
Şekil 18. Ortahisar ilçesine ait trafik kazası Kernel yoğunluk haritası .....	53
Şekil 19. Trabzon iline ait ilçe bazlı trafik kazası yoğunluk haritası .....	54
Şekil 20. Kazaya karışan araçların model yılı gruplarına göre dağılımı. ....	56
Şekil 21. Kazaya karışan yayaların yaş gruplarına göre dağılımı. ....	57
Şekil 22. Kazaya karışan sürücülerin yaş gruplarına göre dağılımı. ....	59
Şekil 23. Kaza şiddeti sınıflarının dağılımı. ....	60
Şekil 24. Trabzon il genelinde yol tipine göre kaza dağılımı.....	62
Şekil 25. Arsin ilçesinde yol tipine göre kaza dağılımı.....	63
Şekil 26. Şalpazarı ilçesinde yol tipine göre kaza dağılımı.....	64
Şekil 27. Trabzon il genelinde kavşak durumuna göre kaza dağılımı.....	65
Şekil 28. Trabzon il genelinde gündüz/gece durumuna göre kaza dağılımı.....	66
Şekil 29. Karar ağacı modeli sınıflandırma raporu.....	68
Şekil 30. Karar ağacı modeli karışıklık matrisi .....	69
Şekil 31. Karar ağacı modeli ROC eğrisi ve AUC değeri .....	70

Şekil 32. Karar ağacı modeli için özellik önem düzeyleri.....	71
Şekil 33. Rastgele orman modeli sınıflandırma raporu .....	73
Şekil 34. Rastgele orman modeli karışıklık matrisi.....	74
Şekil 35. Rastgele orman modeli ROC eğrisi ve AUC değeri.....	75
Şekil 36. Rastgele orman modeli için özellik önem düzeyleri .....	76
Şekil 37. AdaBoost modeli sınıflandırma raporu .....	78
Şekil 38. AdaBoost modeli karışıklık matrisi.....	79
Şekil 39. AdaBoost modeli ROC eğrisi ve AUC değeri.....	80
Şekil 40. AdaBoost modeli için özellik önem düzeyleri .....	81
Şekil 41. Gradient Boosting modeli sınıflandırma raporu.....	83
Şekil 42. Gradient Boosting modeli karışıklık matrisi .....	84
Şekil 43. Gradient Boosting modeli ROC eğrisi ve AUC değeri .....	85
Şekil 44. Gradient Boosting modeli için özellik önem düzeyleri.....	86
Şekil 45. Lojistik regresyon modeli sınıflandırma raporu.....	88
Şekil 46. Lojistik regresyon modeli karışıklık matrisi .....	89
Şekil 47. Lojistik regresyon modeli ROC eğrisi ve AUC değeri .....	90
Şekil 48. Lojistik regresyon modeli için özellik önem düzeyleri .....	91
Şekil 49. XGBoost modeli sınıflandırma raporu .....	93
Şekil 50. XGBoost modeli karışıklık matrisi .....	94
Şekil 51. XGBoost modeli ROC eğrisi ve AUC değeri .....	95
Şekil 52. XGBoost modeli için özellik önem düzeyleri .....	96

## SİMGELER VE KISALTMALAR DİZİNİ

AUC	Area Under the Curve
AdaBoost	Adaptive Boosting
ArcGIS	Arc Geographic Information System
ATM	Automatic Teller Machine
CBS	Coğrafi Bilgi Sistemleri
CNN	Convolutional Neural Network
EGM	Emniyet Genel Müdürlüğü
ESC	Electronic Stability Control
KDE	Kernel Density Estimation
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MSE	Mean Squared Error
NaN	Not a Number
NHTSA	National Highway Traffic Safety Administration
PCA	Principal Component Analysis
QCIS	Quantum Computing Information System
ROC	Receiver Operating Characteristic
SAMME	Stagewise Additive Modeling using a Multi-class Exponential Loss Function
SMOTE	Synthetic Minority Over-sampling Technique
WHO	World Health Organization
XGBoost	Extreme Gradient Boosting

## 1. GİRİŞ

Kamunun kullanımına açık bir yol veya caddede, en az bir hareket hâlindeki aracın karıştığı ve bir ya da daha fazla kişinin yaralanması veya ölümüyle sonuçlanan olaylar karayolu trafik kazası olarak tanımlanmaktadır. Bu kapsamda araçların birbirleriyle çarpışması, araç ile yaya ya da hayvan çarpışmaları ve araçların sabit coğrafi veya yapısal unsurlara çarpması da karayolu trafik kazası kapsamında değerlendirilmektedir.

Karayolu trafik kazaları yalnızca teknik ya da istatistiksel bir sorun değil, aynı zamanda önemli bir toplumsal problemdir. Meydana gelen her kaza; can kaybı, yaralanma, aile bütünlüğünün zedelenmesi ve ekonomik kayıplar gibi çok yönlü sonuçlar doğurmaktadır. Özellikle genç nüfusta görülen erken ölümler ve iş gücü kayıpları, kazaların sosyoekonomik etkilerini daha da ağırlaştırmaktadır. Artan nüfus, kentleşmenin hızlanması ve motorlu araç sayısındaki sürekli yükseliş, ulaşım sistemleri üzerindeki yükü artırmakta; bu durum trafik güvenliğini hem ulusal hem de küresel ölçekte öncelikli bir mesele haline getirmektedir.

Doğu Karadeniz Bölgesi'nin önemli lojistik ve sosyo-ekonomik merkezlerinden biri olan Trabzon, coğrafi ve fiziksel özellikleri bakımından trafik güvenliği açısından dikkat çekici bir yapıya sahiptir. Engebeli arazi yapısı, kıyı hattına paralel biçimde uzanan ve yoğun şekilde kullanılan ana arterler ile yıl boyunca etkili olan yağışlı ve değişken hava koşulları, kentteki trafik akışını doğrudan etkilemektedir. Bunun yanı sıra dar ve virajlı yol kesimleri, eğim farklılıkları ve artan araç yoğunluğu da kazaların oluşumunda rol oynayan unsurlar arasında yer almaktadır.

Trafik kazalarının belirli noktalarda yoğunlaştığını ortaya koyan geleneksel istatistiksel yöntemler, riskli alanların tespitinde önemli katkılar sağlamaktadır. Bununla birlikte, kazaların ortaya çıkışında etkili olan çok sayıda değişkenin birbiriyle olan karmaşık ilişkisini açıklamada bu yöntemlerin zaman zaman yetersiz kaldığı görülmektedir. Yol geometrisi, trafik hacmi, arazi kullanımı, nüfus yoğunluğu ve iklimsel koşullar gibi birçok faktörün etkileşimi sonucunda şekillenen bu olayların daha bütüncül ve mekânsal bir bakış açısıyla ele alınması gerekliliği giderek daha fazla önem kazanmaktadır. Bu çerçevede, trafik güvenliği çalışmalarında veri temelli ve mekânsal analiz yaklaşımlarının kullanımı kaçınılmaz hale gelmiştir.

## 1.1. Araştırmanın Amacı

Bu tez çalışmasında, Trabzon ilinde 2014–2023 yılları arasında meydana gelen toplam 9.466 trafik kazasının mekânsal ve tahmine dayalı analizi amaçlanmaktadır. Çalışmada kullanılan veriler, Emniyet Genel Müdürlüğü Trafik Kaza Veritabanı'ndan elde edilmiştir.

Araştırma kapsamında trafik kazalarının dağılımı Coğrafi Bilgi Sistemleri (CBS) kullanılarak analiz edilmiş olup, noktasal kaza haritaları, yoğunluk analizi ve ilçe bazlı yoğunluk haritaları oluşturulmuştur.

Bununla birlikte, kazaların tahminine yönelik olarak Lojistik Regresyon, Karar Ağacı, Random Forest, AdaBoost ve Gradient Boosting algoritmaları uygulanmıştır. Modellerin performansı Accuracy, Precision, Recall, F1 Skoru, ROC-AUC ve Cohen's Kappa ölçütleriyle değerlendirilmiştir.

Bu doğrultuda çalışma, makine öğrenmesi ve CBS tabanlı mekânsal analizleri bir arada kullanarak Trabzon ilindeki trafik kazalarının risk örüntüsünü ortaya koymayı amaçlamaktadır.

## 1.2. Araştırmanın Önemi

Trafik kazaları, günümüzde hem halk sağlığı hem de sürdürülebilir kalkınma açısından küresel ölçekte ciddi bir sorun olarak karşımıza çıkmaktadır. Karayolu trafik yaralanmaları dünya genelinde en önde gelen ölüm nedenleri arasında yer almakta; özellikle genç nüfus için büyük bir risk oluşturmaktadır ('Global Status Report On Road Safety 2018', WHO). Bu tablo, trafik güvenliğinin yalnızca ulaşım politikalarının teknik bir başlığı olmadığını; aynı zamanda sağlık, ekonomi ve sosyal politika alanlarını doğrudan ilgilendiren hayati bir konu olduğunu göstermektedir.

Trafik kazalarının ekonomik boyutu da en az insani boyutu kadar çarpıcıdır. World Bank tarafından yayımlanan raporlara göre, trafik kazaları ülkelerin gayri safi yurtiçi hasılasının yaklaşık %3'üne kadar ekonomik kayba neden olabilmektedir ('The High Toll of Traffic Injuries: Unacceptable and Preventable', 2017). Bu kayıplar yalnızca sağlık harcamalarıyla sınırlı değildir; iş gücü kaybı, üretkenliğin azalması ve uzun süreli bakım maliyetleri gibi hem doğrudan hem de dolaylı etkileri kapsamaktadır. Başka bir ifadeyle, trafik kazaları yalnızca bireyleri ve aileleri değil, ülke ekonomilerini de derinden etkilemektedir.

Ulusal veriler ise sorunun yerel düzeydeki somut yansımalarını açık biçimde ortaya koymaktadır. Emniyet Genel Müdürlüğü Trafik Başkanlığı'nın 2025 yılı istatistiklerine göre Trabzon ilinde yıl boyunca toplam 6.045 trafik kazası meydana gelmiştir. Bu kazaların 2.539'u ölüm veya yaralanmayla sonuçlanmış, 3.506'sı ise yalnızca maddi hasarla sonuçlanmıştır. Aynı yıl içerisinde 19 kişi hayatını kaybetmiş, 3.608 kişi yaralanmıştır bu rakamlar, trafik kazalarının Trabzon ili açısından da önemli bir toplumsal güvenlik ve kamu sağlığı sorunu olduğunu açıkça göstermektedir.

### **1.3. Varsayımlar**

Bu çalışmada kullanılan trafik kazası verilerinin ilgili kurum tarafından doğru, eksiksiz ve güvenilir biçimde kaydedildiği varsayılmıştır. Veri setinde yer alan değişkenlerin kaza şiddetini açıklamada yeterli olduğu kabul edilmiştir. Makine öğrenmesi modellerinin geçmiş veriler üzerinden öğrendiği örüntülerin benzer gelecekteki kazalara genellenebilir olduğu varsayılmıştır.

Modelleme sürecinde gözlemlerin birbirinden bağımsız olduğu kabul edilmiştir. Her bir kaza kaydının diğer kayıtlardan bağımsız gerçekleştiği varsayımı altında analizler yürütülmüştür.

SMOTE yöntemi ile oluşturulan sentetik örneklerin gerçek veri dağılımını makul düzeyde temsil ettiği kabul edilmiştir. Ayrıca mekânsal analizlerde kullanılan koordinat bilgilerinin doğruluğu esas alınmıştır.

### **1.4. Sınırlılıklar**

Bu araştırma Trabzon ili sınırları içerisinde meydana gelen trafik kazaları ile sınırlıdır. Çalışmada kullanılan veri seti belirli bir zaman aralığını kapsamaktadır. Veri setinde yer almayan sürücü davranışları, psikolojik faktörler veya anlık karar süreçleri modele dahil edilememiştir.

SMOTE yöntemi sınıf dengesizliğini azaltmak amacıyla kullanılmış olsa da sentetik veriler gerçek veri çeşitliliğini tam olarak yansıtmayabilir. Mekânsal analizler ilçe düzeyinde gerçekleştirilmiş olup daha detaylı mikro ölçekli konumsal analizler çalışma kapsamı dışında bırakılmıştır.

## 2. KAVRAMSAL ÇERÇEVE VE İLGİLİ ÇALIŞMALAR

Trafik kazaları, günümüzde artan motorlu araç sayısı, hızlı kentleşme ve karmaşık ulaşım sistemlerinin etkisiyle küresel ölçekte önemli bir halk sağlığı ve ulaşım güvenliği problemi olmaya devam etmektedir. Son yıllarda yapılan çalışmalar, trafik kazalarının her yıl milyonlarca insanın yaşamını etkilediğini ve özellikle genç nüfus üzerinde ciddi bir risk oluşturduğunu ortaya koymaktadır (Behboudi, Moosavi, & Ramnath, 2024). Artan araç yoğunluğu, mevcut ulaşım altyapısı üzerinde ciddi bir baskı oluşturmakta ve bu durum kazaların hem sıklığını hem de şiddetini artırmaktadır. Ayrıca yayalar, bisikletliler ve motosiklet kullanıcıları gibi kırılgan yol kullanıcıları, trafik kazalarından orantısız biçimde etkilenmektedir (Amiri, Afshari, & Soltani, 2025). Bu durum, trafik güvenliği politikalarının yeniden değerlendirilmesini gerekli kılmaktadır.

Trafik kazalarının yalnızca birincil etkileri değil, kazalar sonrasında meydana gelen ikincil kazalar ve trafik akışı üzerindeki olumsuz etkileri de önemli bir sorun oluşturmaktadır. Yoğun trafik ağlarında meydana gelen kazalar, gecikmelere, ekonomik kayıplara ve ek güvenlik risklerine yol açmaktadır. Bu nedenle, trafik kazalarının önlenmesi ve etkilerinin azaltılması, sürdürülebilir ulaşım sistemlerinin geliştirilmesinde kritik bir öncelik hâline gelmiştir (Berhanu, Schröder, Wodajo, & Alemayehu, 2024).

Son yıllarda veri madenciliği, makine öğrenmesi ve derin öğrenme teknikleri, trafik kazalarının analizi ve tahmini konusunda önemli ilerlemeler sağlamıştır. Güncel çalışmalar, geleneksel istatistiksel yöntemlerin ötesine geçerek karmaşık ve çok boyutlu trafik verilerindeki doğrusal olmayan ilişkileri modelleyebilen yapay zekâ tabanlı yaklaşımların daha başarılı sonuçlar verdiğini göstermektedir (Cunha, Silva, Ribeiro, & Gomes, 2025). Özellikle Random Forest, Yapay Sinir Ağları ve derin öğrenme tabanlı modeller, kaza oluşumunu etkileyen faktörlerin belirlenmesi ve kaza şiddetinin tahmin edilmesinde yaygın olarak kullanılmaktadır (Yumak, Hengirmen Tercan, Colak, & Ozcanan, 2025). Bunun yanında, CNN, LSTM ve Grafik Sinir Ağları gibi gelişmiş mimariler, trafik verilerinin zamansal ve mekânsal özelliklerini birlikte analiz ederek daha doğru tahminler yapılmasına olanak tanımaktadır (Li & Chen, 2025).

Güncel araştırmalar, makine öğrenmesi modellerinin yalnızca kaza tahmini değil, aynı zamanda riskli bölgelerin belirlenmesi, güvenli rota planlama ve trafik yönetim sistemlerinin geliştirilmesi gibi alanlarda da etkili olduğunu ortaya koymaktadır (Berhanu et al., 2024).

Büyük veri setleri üzerinde geliştirilen yapay zekâ tabanlı modeller, insan, araç ve çevre faktörlerini birlikte değerlendirerek yüksek doğruluk oranlarına ulaşabilmektedir (Mostafa et al., 2025). Ayrıca veri ön işleme, eksik veri tamamlama ve boyut indirgeme teknikleri, model performansını artırmada önemli rol oynamaktadır.

Makine öğrenmesi ve derin öğrenme modellerinin performansı; doğruluk (accuracy), duyarlılık (recall), F1-skoru, AUC ve RMSE gibi ölçütler kullanılarak değerlendirilmektedir. Son yıllarda yapılan kapsamlı inceleme çalışmaları, uygun veri dengeleme yöntemleri ve algoritma seçimi ile bu modellerin yüksek başarı oranlarına ulaşabildiğini göstermektedir (Kotsyubynska vd., 2026). Bu bağlamda, yapay zekâ temelli yaklaşımlar, trafik kazalarının önlenmesi, yol güvenliğinin artırılması ve ulaşım sistemlerinin daha güvenli hâle getirilmesi açısından önemli katkılar sunmaktadır.

## **2.1 Trafik Kazalarında Risk ve Şiddet Kavramı**

Trafik kazalarında risk ve şiddet kavramları, kazaların hem ortaya çıkma ihtimalini hem de sonuçlarının ne ölçüde ağır olacağını açıklayan temel kavramlardır. Risk, belirli koşullar altında bir kazanın meydana gelme olasılığını ve olası sonuçlarını ifade ederken; şiddet, gerçekleşen kazanın yol açtığı yaralanma, can kaybı ve maddi hasarın düzeyini göstermektedir. Özellikle hız, dikkatsizlik ve olumsuz çevre koşulları gibi etkenler hem kazaya karışma ihtimalini artırmakta hem de kazanın daha ağır sonuçlar doğurmasına neden olabilmektedir. Bu nedenle trafik güvenliğine yönelik çalışmaların, yalnızca kazaların sayısını azaltmaya değil, aynı zamanda meydana gelen kazaların daha hafif sonuçlar doğurmasını sağlamaya odaklanması önem taşımaktadır.

### **2.1.1 Trafik kazalarını etkileyen faktörler**

Trafik kazaları; insan, araç ve çevre unsurlarının bir araya gelmesiyle ortaya çıkan çok boyutlu olaylardır. Literatürde kazaların tek bir nedene bağlı olarak değil, çoğunlukla birden fazla risk faktörünün etkileşimi sonucunda meydana geldiği vurgulanmaktadır. Bu çerçevede trafik güvenliği çalışmalarında faktörler genel olarak demografik özellikler, yol ve çevresel koşullar ile araç ve sürücü davranışlarına bağlı unsurlar şeklinde sınıflandırılmaktadır.

### **2.1.1.1 Demografik faktörler**

Demografik faktörlerin trafik kazaları üzerindeki etkisi literatürde kapsamlı biçimde incelenmiştir. Özellikle yaş ve cinsiyet değişkenlerinin kazaya karışma olasılığı ve kaza sorumluluğu üzerinde belirleyici olduğu belirtilmektedir. Mccarty vd. (2024), genç sürücülerin risk alma eğilimlerinin ve deneyim eksikliğinin kaza riskini artırdığını, ileri yaş grubunda ise bilişsel ve fiziksel gerilemeye bağlı olarak risk düzeyinin yükseldiğini ortaya koymuştur. Benzer şekilde Williams vd, (2003) erkek sürücülerin kadın sürücülere kıyasla daha yüksek oranda ölümcül kazalara karıştığını ve sorumluluk oranlarının daha fazla olduğunu belirtmiştir. Ayrıca eğitim düzeyi ve sosyo-ekonomik durum gibi değişkenlerin trafik güvenliği davranışlarını etkilediği ve düşük eğitim ile gelir seviyelerinin kaza riskini artırabildiği Borhan vd. (Borhan, Ibrahim, Aziz, & Yazid, 2018) tarafından vurgulanmaktadır. Bu bulgular, demografik özelliklerin trafik kazalarının oluşumunda önemli bir risk grubu oluşturduğunu göstermektedir.

### **2.1.1.2 Yol ve çevresel faktörler**

Trafik kazalarının oluşumunda sürücü davranışlarının yanı sıra yol ve çevresel faktörler de önemli bir rol oynamaktadır. Özellikle yolun geometrik özellikleri, görüş koşulları ve fiziksel tasarım unsurları kaza olasılığı ve kaza şiddeti üzerinde belirleyici etkilere sahiptir.

Yol geometrisinin kaza oluşumu ve şiddeti üzerindeki etkisi farklı yöntemlerle yapılan çalışmalarda da ortaya konmuştur. Jima ve Sapos (2022) (Jima & Sapos, 2022), çoklu lojistik regresyon ve yapay sinir ağı modelleri kullanarak geometrik tasarım unsurlarının kaza olasılığı ve şiddeti üzerinde istatistiksel olarak anlamlı etkiler oluşturduğunu belirlemiştir. Benzer şekilde Alpan (2025), devlet yollarında meydana gelen tek taşıtlı kazalar üzerinden yaptığı analizde, özellikle küçük yarıçaplı yatay dönemeçler, yüksek boyuna eğimler ve yatay–düşey geometrik elemanların birlikte bulunduğu kesimlerin kaza riskini artırdığını ortaya koymuştur. Ayrıca yatay kurp yarıçapı ve boyuna eğim değişkenlerinin kilometre başına düşen yıllık kaza sayısını yüksek oranda açıkladığı tespit edilmiştir. Bu bulgular, yol geometrik tasarımının trafik güvenliği üzerindeki belirleyici rolünü açık biçimde ortaya koymaktadır.

### **2.1.1.3 Araç ve davranışsal faktörler**

Trafik kazalarının oluşumunda araç kaynaklı teknik faktörler ile sürücü davranışlarına bağlı değişkenlerin birlikte değerlendirilmesi gerektiği literatürde güçlü biçimde vurgulanmaktadır.

Araç güvenliği alanında yapılan çalışmalar, özellikle aktif güvenlik sistemlerinin kaza şiddetini azaltmada önemli rol oynadığını göstermektedir. Elektronik stabilite kontrolü (ESC) sisteminin tek araçlı ve savrulma kaynaklı ölümlü kazaları anlamlı düzeyde azalttığı rapor edilmiştir (National Highway Traffic Safety Administration, 2015). Küresel ölçekte ise araç güvenlik standartlarının yetersiz olduğu ülkelerde trafik kaynaklı ölüm oranlarının daha yüksek olduğu belirtilmektedir (World Health Organization, 2018). Bu bulgular, araç tasarımı ve güvenlik donanımlarının kazaların sonuçları üzerindeki belirleyici etkisini ortaya koymaktadır. Raporun yer alan istatistiksel analizlere göre, ESC sistemi binek araçlarda tek araçlı ölümlü kazaları yaklaşık %30–35 oranında, SUV tipi araçlarda ise daha yüksek oranlarda azaltmaktadır (National Highway Traffic Safety Administration, 2015)

## **2.2 Makine Öğrenmesi Yaklaşımları**

Makine öğrenmesi, verilerden otomatik olarak öğrenebilen ve yeni örnekler üzerinde tahminler yapabilen algoritmaların geliştirilmesini amaçlayan yapay zekâ alt disiplindir. Son yıllarda makine öğrenmesi, büyük veri ve hesaplama gücündeki artışla birlikte sağlık, finans, otonom sistemler ve doğal dil işleme gibi alanlarda yaygın kullanım bulmuştur (Akanksha Mishra, 2024).

Makine öğrenmesi yaklaşımları genel olarak denetimli öğrenme (supervised learning), denetimsiz öğrenme (unsupervised learning), pekiştirmeli öğrenme (reinforcement learning) ve yarı-denetimli/metaya öğrenme gibi daha yeni paradigmlar altında incelenmektedir. Bu sınıflandırma, algoritmaların veri ile etkileşim biçimine ve öğrenme hedeflerine göre yapılır (Razzaq & Shah, 2025).

### **2.2.1 Denetimli öğrenme**

Denetimli öğrenme, girdiler ile çıktılar arasındaki ilişkilerin öğrenildiği yöntemdir. Bu yaklaşımda model, etiketli veri üzerinden eğitilir ve yeni örneklerde çıktıları tahmin etmeyi öğrenir. Makine öğrenmesi araştırmalarında denetimli öğrenme, sınıflandırma (örneğin görüntü tanıma) ve regresyon (örneğin fiyat tahmini) problemlerinde yaygın olarak kullanılan temel yaklaşımdır (International Business Machines, 2025).

Son dönemde denetimli öğrenme üzerine yapılan çalışmalar, örneğin LightGBM ve XGBoost gibi ağaç tabanlı algoritmaların yüksek performanslı sınıflandırma sonuçları verdiğini

göstermiş; ayrıca derin öğrenme tabanlı sınıflandırıcılar metin ve görsel veri üzerinde ileri düzey başarı sağlamıştır (Kahraman vd., 2025).

Bu çalışmada trafik kaza analizi problemi denetimli öğrenme yaklaşımı ile ele alınmıştır. Çünkü trafik kazalarına ilişkin veri setlerinde genellikle kaza şiddeti, yaralanma durumu veya maddi hasar seviyesi gibi önceden tanımlanmış hedef değişkenler bulunmaktadır. Denetimli öğrenme algoritmaları, hava durumu, yol tipi, sürücü yaşı, araç türü ve zaman bilgisi gibi bağımsız değişkenler ile kaza sonucu arasındaki karmaşık ilişkileri modelleyebilme kapasitesine sahiptir. Ayrıca trafik kaza verileri çoğunlukla tablo formatında ve sayısal/kategorik değişkenlerden oluştuğu için özellikle ağaç tabanlı topluluk yöntemleri bu tür veri yapılarında yüksek performans göstermektedir. Bu nedenle, kaza şiddetinin sınıflandırılması veya kaza riskinin tahmin edilmesi amacıyla denetimli öğrenme yöntemlerinin tercih edilmesi metodolojik olarak uygun ve literatürle uyumlu bir yaklaşımdır.

### **2.2.2 Denetimsiz Öğrenme**

Denetimsiz öğrenme, verilerdeki gizli örüntü ve yapıları keşfetmek için kullanılır. Bu yaklaşımda model, girdide etiket bulunmadan verideki ilişkileri anlamaya çalışır. Kümeleme (örneğin k-ortalamlar) ve boyut indirgeme (örneğin PCA) gibi yöntemler denetimsiz öğrenmenin temel araçlarından (Razzaq & Shah, 2025)

Güncel literatürde, denetimsiz öğrenme aynı zamanda öztemsil (self-representation) ve öz denetimli öğrenme (self-supervised learning) gibi daha gelişmiş paradigmalara desteklenmekte; bu sayede etiketlenmemiş veri üzerinde derin özellik çıkarımı yapılabilmektedir (Zhang, Yang, Sajjadi Mohammadabadi, & Yan, 2025).

### **2.2.3 Pekiştirmeli Öğrenme**

Pekiştirmeli öğrenme, bir ajanın çevresiyle etkileşime girerek ödül sinyalleri aracılığıyla öğrenmesini temel alan bir makine öğrenmesi yaklaşımıdır. Amaç, uzun vadeli toplam ödülü maksimize eden en uygun politika fonksiyonunun elde edilmesidir. Bu süreç genellikle Markov Karar Süreci çerçevesinde modellenmektedir (Sutton & Barto, 2018).

Derin pekiştirmeli öğrenme yöntemleri, politika ve değer fonksiyonlarını derin sinir ağları ile yaklaşık olarak modelleyerek yüksek boyutlu durum uzaylarında etkili sonuçlar üretmektedir.

Deep Q-Network yöntemi bu alandaki temel çalışmalardan biridir (Mnih vd., 2015). Proximal Policy Optimization algoritması ise politika gradyanı yöntemlerinde daha kararlı öğrenme sağlamıştır (Schulman, Wolski, Dhariwal, Radford, & Openai, 2017). Güncel derleme çalışmalarında pekiştirmeli öğrenmenin örnek verimliliği, ölçeklenebilirlik ve gerçek dünya uygulamaları açısından önemli gelişmeler kaydettiği belirtilmektedir ((Levine vd.,2020); (Prudencio, Maximo, & Colombini, 2023)).

#### **2.2.4 Derin Öğrenme**

Derin öğrenme, çok katmanlı yapay sinir ağlarına dayanan ve veriden hiyerarşik özellik temsilleri öğrenebilen bir makine öğrenmesi alt alanıdır. Geleneksel makine öğrenmesi yöntemlerinde özellik çıkarımı çoğunlukla manuel olarak yapılırken, derin öğrenme modelleri ham veriden otomatik olarak anlamlı temsiller öğrenebilmektedir (LeCun vd., 2015). Bu durum, özellikle büyük ve karmaşık veri kümelerinde önemli bir avantaj sağlamaktadır.

Derin öğrenmenin gelişiminde Convolutional Neural Network (CNN) mimarileri önemli bir rol oynamıştır. Görüntü sınıflandırma alanında derin CNN yapılarının başarısı ilk olarak AlexNet modeli ile gösterilmiştir (Krizhevsky vd., 2012). Daha sonra önerilen ResNet mimarisi, artık bağlantılar sayesinde çok daha derin ağların eğitilmesini mümkün kılmış ve gradyan kaybolması problemini azaltarak performans artışı sağlamıştır (He vd., 2016).

Doğal dil işleme alanında ise dikkat mekanizmasına dayalı Transformer mimarisi önemli bir dönüm noktası olmuştur (Vaswani vd., 2017). Bu mimari, uzun bağımlılıkları daha etkili biçimde modelleyerek dil modelleme performansını artırmıştır. Büyük ölçekli dil modellerinin öz-denetimli öğrenme ile ön-eğitim yaparak farklı görevlere uyarlanabildiği gösterilmiştir (Brown vd., 2020). Ayrıca foundation models yaklaşımının, geniş veri kümeleri üzerinde eğitilen modellerin transfer öğrenme yoluyla birçok farklı problemde kullanılmasını mümkün kıldığı belirtilmektedir (Bommasani vd., 2021).

### **2.3 Coğrafi Bilgi Sistemleri ve Mekânsal Analiz**

Coğrafi Bilgi Sistemleri (CBS), mekâna ilişkin verilerin toplanması, depolanması, analiz edilmesi ve görselleştirilmesini sağlayan donanım, yazılım ve insan bileşenlerinden oluşan bütüncül bir sistemdir. CBS, özellikle mekânsal örüntülerin belirlenmesi, dağılımların analiz edilmesi ve karar destek süreçlerinin yürütülmesinde etkin bir araç olarak kullanılmaktadır (Yomralıoğlu, 2000).

Mekânsal analiz ise, coğrafi veriler arasındaki konumsal ilişkilerin matematiksel ve istatistiksel yöntemlerle incelenmesini ifade eder. Mekânsal analiz yöntemleri; dağılım analizi, yoğunluk analizi, komşuluk ilişkileri, mesafe hesaplamaları ve yüzey modellemeleri gibi çeşitli teknikleri kapsamaktadır. Bu analizler sayesinde mekânsal olayların kümelenme eğilimleri, yoğunluk merkezleri ve mekânsal bağımlılık düzeyleri ortaya konulabilmektedir (Anselin, 1995).

CBS tabanlı mekânsal analizler özellikle suç analizleri, sağlık coğrafyası, afet yönetimi, ulaşım planlaması ve kentsel gelişim çalışmalarında yaygın biçimde kullanılmaktadır. Günümüzde ArcGIS, QGIS ve benzeri yazılımlar mekânsal analizlerin uygulanmasında temel araçlar olarak öne çıkmaktadır.

### **2.3.1 Noktasal mekânsal veri yapısı**

Noktasal mekânsal veri yapısı, belirli bir konumu temsil eden tekil koordinat çiftlerinden (x, y) oluşan vektör veri türüdür. Bu veri yapısı genellikle olayların gerçekleştiği spesifik konumları (suç noktaları, kaza yerleri, hastalık vakaları, ATM konumları vb.) göstermek amacıyla kullanılmaktadır.

Vektör veri modelleri üç temel geometrik yapıdan oluşur: nokta, çizgi ve poligon. Nokta veri yapısı, alansal veya doğrusal bir yayılım göstermeyen, belirli bir konumda meydana gelen olayların gösterimi için uygundur (Burrough, Mcdonnell, & Lloyd, n.d.)

Noktasal veriler üzerinde gerçekleştirilen mekânsal analizler, olayların rastgele mi dağıldığını yoksa belirli alanlarda kümelenme eğilimi mi gösterdiğini ortaya koymayı amaçlar. Bu kapsamda En Yakın Komşu Analizi (Nearest Neighbor Analysis), Moran's I ve Getis-Ord  $G_i^*$  gibi istatistiksel yöntemler sıklıkla kullanılmaktadır.

Ancak noktasal veriler, olayın etki alanını doğrudan göstermemekte; yalnızca gerçekleştiği konumu temsil etmektedir. Bu nedenle yoğunluk temelli yüzey analizleri ile desteklenmesi gerekmektedir.

### **2.3.2 Yoğunluk analizi**

CBS kapsamında yoğunluk analizi, mekânsal verilerin belirli bir alan üzerindeki dağılımını incelemek amacıyla kullanılan temel analiz yöntemlerinden biridir. Bu yöntem, noktasal, çizgisel veya alansal verilerin birim alan başına düşen yoğunluğunu hesaplayarak mekânsal

örüntülerin belirlenmesini sağlar ((O'Sullivan & Unwin, 2010); (Wang, 2006). CBS tabanlı analizlerde yoğunluk hesaplamaları, mekânsal karar verme süreçlerinde önemli bir rol oynamaktadır (Güneş, Nihan, & Sönmez, 2026).

Yoğunluk analizinin kuramsal temeli, mekânda gözlemlenen olayların dağılım biçimini inceleyen mekânsal istatistik yaklaşımlarına dayanmaktadır. Bu kapsamda kullanılan yöntemlerden biri olan Kernel Yoğunluk Tahmini (Kernel Density Estimation- KDE), her bir veri noktasına belirli bir etki alanı tanımlayarak bu etkilerin üst üste bindirilmesiyle sürekli bir yoğunluk yüzeyi oluşturur (Xie & Yan, 2013). Kernel yoğunluk analizinde, belirli bir yarıçap içerisinde kalan noktaların yoğunluğu dikkate alınmakta ve bu yoğunluk mesafe arttıkça azalmaktadır. Bu özellik, yöntemin mekânsal dağılımları daha gerçekçi şekilde modellemesini sağlamaktadır.

CBS'de yoğunluk analizleri genel olarak noktasal, çizgisel ve kernel yoğunluk analizleri olmak üzere üç başlık altında incelenmektedir. Noktasal yoğunluk analizinde belirli bir alana düşen nokta sayısı hesaplanırken, kernel yoğunluk analizinde mesafeye bağlı ağırlıklandırma uygulanarak daha yumuşak geçişli bir yoğunluk yüzeyi elde edilir. Bu nedenle kernel yoğunluk yöntemi, özellikle kümelenme ve yoğunlaşma alanlarının belirlenmesinde daha etkin sonuçlar sunmaktadır.

Yoğunluk analizinin önemli kuramsal dayanaklarından biri de mekânsal otokorelasyon kavramıdır. Mekânsal otokorelasyon, coğrafi olarak yakın konumda bulunan olayların birbirine daha fazla benzerlik gösterdiğini ifade eder. Bu durum, Tobler'in Birinci Coğrafya Yasası ile açıklanmaktadır (Tobler, 1970). Güncel çalışmalarda da bu ilke, mekânsal analizlerin temel varsayımlarından biri olarak kabul edilmektedir (Anselin, 2007).

CBS ortamında gerçekleştirilen yoğunluk analizleri çoğunlukla raster veri yapısı kullanılarak yapılmaktadır. Bu süreçte hücre boyutu, arama yarıçapı (bandwidth) ve ağırlıklandırma fonksiyonları gibi parametreler analiz sonuçlarını doğrudan etkilemektedir. Özellikle bant genişliği seçimi, elde edilen yoğunluk yüzeyinin doğruluğu açısından kritik bir öneme sahiptir (Xie & Yan, 2013). Ayrıca uygulamalı çalışmalarda kernel yoğunluk yönteminin nüfus, suç ve çevresel olayların dağılımını analiz etmede yaygın olarak kullanıldığı görülmektedir (Aydoğdu & Bakırcı, 2021).

Sonuç olarak yoğunluk analizi, CBS tabanlı mekânsal analizlerin en önemli bileşenlerinden biri olup, mekânsal dağılımın anlaşılması, kümelenme alanlarının belirlenmesi ve karar destek süreçlerinin geliştirilmesinde etkin bir araçtır. Güncel çalışmalar, bu yöntemin hem teorik hem de uygulamalı açıdan geniş bir kullanım alanına sahip olduğunu göstermektedir.

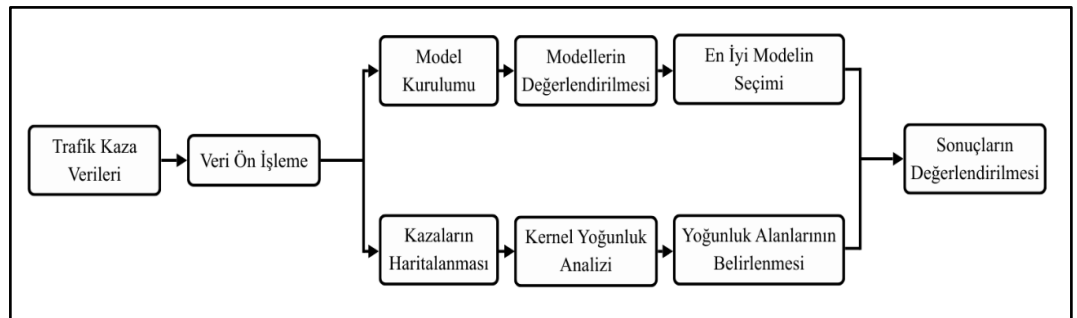
### 3. YÖNTEM

Bu çalışma nicel araştırma yaklaşımı çerçevesinde tasarlanmıştır. Araştırmada Trabzon ili sınırları içerisinde meydana gelen trafik kazalarına ait ikincil veri seti kullanılmıştır. Çalışmanın temel amacı, trafik kazalarının şiddet düzeyini makine öğrenmesi yöntemleri ile tahmin etmek ve kazaların mekânsal dağılımını Coğrafi Bilgi Sistemleri (CBS) aracılığıyla analiz etmektir.

Araştırma tasarımı üç ana aşamadan oluşmaktadır. İlk aşamada, trafik kazalarına ilişkin veri seti temizlenmiş, ön işleme adımlarından geçirilmiş ve sınıflandırma modellerine uygun hale getirilmiştir.

İkinci aşamada, farklı makine öğrenmesi algoritmaları kullanılarak kaza şiddeti tahmin modelleri oluşturulmuş ve model performansları karşılaştırılmıştır. Model başarısı çeşitli performans ölçütleri aracılığıyla değerlendirilmiştir. Böylece en uygun model belirlenmiştir.

Araştırmanın son aşamasında ise mekânsal analiz gerçekleştirilmiştir. Kazalara ait koordinat verileri CBS ortamına aktarılmış, noktasal dağılım haritaları oluşturulmuş ve Kernel yoğunluk analizi uygulanarak kazaların yoğunlaştığı alanlar belirlenmiştir. Ayrıca ilçe düzeyinde yoğunluk analizleri yapılarak mekânsal örüntüler değerlendirilmiştir. Bu araştırma tasarımı, makine öğrenmesi ile elde edilen tahmin sonuçları ile mekânsal analiz bulgularının birlikte değerlendirilmesine olanak sağlamaktadır. Böylece hem istatistiksel hem de mekânsal boyutta bütüncül bir analiz gerçekleştirilmiştir. Şekil 1’ de tez çalışmasına ait akış diyagramı verilmiştir.



Şekil 1. Tez çalışmasına ait akış diyagramı

### 3.1. Çalışma Bölgesi

Bu araştırmanın çalışma bölgesi Trabzon ilidir. Trabzon, Türkiye'nin kuzeydoğusunda, Karadeniz Bölgesi'nde yer almakta olup doğusunda Rize, batısında Giresun, güneyinde Gümüşhane ve Bayburt illeri ile çevrilidir. Kuzey sınırı boyunca Karadeniz'e kıyısı bulunmaktadır. İlin yüzölçümü yaklaşık 4.685 km<sup>2</sup> olup 18 ilçeden oluşmaktadır. Trabzon'un 2024 yılı itibarıyla nüfusu 833.323'tür (TÜİK, 2026).



Şekil 2. Trabzon'un Türkiye'deki konumu (Wikipedia, 2009)

Trabzon, Karadeniz Sahil Yolu üzerinde yer alması ve Doğu Karadeniz Bölgesi'nin önemli ulaşım merkezlerinden biri olması nedeniyle yoğun trafik akışına sahiptir. İl genelinde hem şehir içi hem de şehirlerarası ulaşım hareketliliği yüksektir. Özellikle sahil yolu, bölgesel bağlantı yolları ve kırsal alanlardaki eğimli arazi yapısı trafik güvenliği açısından risk oluşturan unsurlar arasında yer almaktadır.



Şekil 3. Trabzon ve ilçeleri (Trabzon Büyükşehir Belediyesi, 2025)

İlin topoğrafik yapısı genel olarak engebeli olup kıyı kesimlerinden iç kesimlere doğru yükselti artmaktadır. Bu durum yol geometrisi, görüş mesafesi ve sürüş güvenliği üzerinde etkili olabilmektedir. Ayrıca bölgenin yağışlı iklim yapısı, yol yüzeyi koşullarını etkileyerek trafik kazalarının oluşumunda önemli bir çevresel faktör olarak değerlendirilmektedir.

### 3.1.1. Nüfus ve motorlu araç istatistikleri

Trabzon ili nüfus ve motorlu kara taşıtları istatistikleri incelendiğinde, 2007–2025 döneminde her iki değişkenin de genel olarak artış eğiliminde olduğu görülmektedir. İncelenen dönemde nüfus artışı sınırlı ve dalgalı bir seyir izlerken, motorlu kara taşıt sayısındaki artışın daha hızlı ve istikrarlı olduğu dikkat çekmektedir. Bu durum, taşıt sahipliğinin nüfus artış hızının üzerinde bir büyüme gösterdiğini ortaya koymaktadır

Tablo 1. Trabzon ili yıllara göre nüfus ve motorlu kara taşıtları sayıları (TÜİK, 2025)

Yıl	Nüfus	Motorlu Kara Taşıtı Sayısı
2007	740.569	90.144
2008	748.982	94.937
2009	765.127	100.638
2010	763.714	108.178
2011	757.353	117.654
2012	757.898	127.663
2013	758.237	137.560
2014	766.782	146.956
2015	768.417	158.674
2016	779.379	172.343
2017	786.326	185.389
2018	807.903	193.376
2019	808.974	197.471
2020	811.901	209.014
2021	816.684	219.926
2022	818.023	230.979
2023	824.352	249.175
2024	822.270	269.950
2025	823.323	292.063

Yıllar itibarıyla nüfus ve motorlu kara taşıt sayılarında meydana gelen artış, trafik yoğunluğunun artmasına neden olabilmektedir. Trafik yoğunluğundaki bu artış ise trafik kazalarının sayısını etkileyebilecek önemli faktörlerden biri olarak değerlendirilmektedir. Nüfus artışı ve araç sayısındaki yükseliş, özellikle şehir içi ulaşım talebini artırarak kaza riskini dolaylı biçimde etkileyebilmektedir. Bu çerçevede nüfus ve motorlu araç sayılarındaki

değişimin trafik kazaları üzerindeki olası etkisi, ilerleyen bölümde yapılacak istatistiksel analiz kapsamında incelenecektir.

## **3.2. Veri Setinin Tanımı**

### **3.2.1. Veri setinin genel özellikleri**

Bu çalışmada kullanılan veri seti, Trabzon ilinde meydana gelen trafik kazalarına ait kayıtlardan oluşmaktadır. Veri seti kazaya karışan sürücüler, yayalar, araçlar ve yol özelliklerine ilişkin çeşitli değişkenleri içermektedir. Çalışmada kullanılan veri seti toplam 9286 gözlem ve 19 değişkenden oluşmaktadır. Bu değişkenler trafik kazalarının oluşumunda etkili olabilecek farklı faktörleri temsil etmektedir.

Veri setinde yer alan değişkenler genel olarak sürücü özellikleri, yaya özellikleri, yol ve çevre koşulları ile araç özelliklerine ilişkin bilgilerden oluşmaktadır. Sürücü ve yaya özellikleri kapsamında sürücü yaşı, sürücü cinsiyeti, yaya yaşı ve yaya cinsiyeti gibi demografik değişkenler yer almaktadır. Bu değişkenler trafik kazalarında insan faktörünün etkisini incelemek açısından önem taşımaktadır.

Veri setinde ayrıca kazanın meydana geldiği yolun özelliklerine ilişkin değişkenler de bulunmaktadır. Bu kapsamda yolun tipi, yolun sınıfı, yol kaplama durumu, yol yüzeyi, yatay ve düşey geometrik özellikler ile kavşak durumu gibi değişkenler yer almaktadır. Bunun yanı sıra kazanın gerçekleştiği andaki çevresel koşulları ifade eden gündüz/gece durumu ve hava durumu gibi değişkenler de veri setinde bulunmaktadır.

Araç özelliklerine ilişkin olarak ise araç model yılı, kazaya karışan araç sayısı ve araç darbe bölgesi gibi değişkenler veri setinde yer almaktadır. Bu değişkenler kazanın oluş biçimi ve kazanın şiddeti hakkında önemli bilgiler sağlamaktadır.

Çalışmada kullanılan veri setinde yer alan değişkenler trafik kazalarının oluşumunu etkileyebilecek farklı faktörleri temsil etmekte olup, bu değişkenler kullanılarak kazaların şiddeti üzerinde etkili olan faktörlerin belirlenmesi amaçlanmaktadır.

### 3.2.2. Değişkenlerin tanımı

Bu çalışmada kullanılan veri setinde yer alan değişkenler trafik kazalarının oluşumuna ve kaza şiddetine etki edebilecek farklı faktörleri temsil etmektedir. Veri setindeki değişkenler sürücü özellikleri, yaya özellikleri, yol ve çevre koşulları ile araç özelliklerine ilişkin bilgileri içermektedir. Bu değişkenler kullanılarak trafik kazalarının oluşumunu etkileyen faktörlerin ve kaza şiddeti ile ilişkili değişkenlerin belirlenmesi amaçlanmıştır. Çalışmada kullanılan başlıca değişkenler ve bu değişkenlere ilişkin açıklamalar Tablo 2’de sunulmuştur.

Tablo 2. Çalışmada kullanılan değişkenler

Değişken Adı	Açıklama
YayaYaşı	Kazaya karışan yayaların yaşı
Yaya_Cinsiyet	Kazaya karışan yayaların cinsiyeti
Sürücü_Yası	Kazaya karışan sürücünün yaşı
Sürücü_Cinsiyet	Kazaya karışan sürücünün cinsiyeti
YolunTipi	Kazanın meydana geldiği yolun tipi
YolunKaplaması	Yolun kaplama türü
YolunSınıfı	Yolun sınıfı (devlet yolu vb.)
GeoYatay	Yolun yatay geometrik durumu
GeoDüşey	Yolun düşey geometrik durumu
GeoKavşak	Kazanın kavşakta meydana gelip gelmediği
Gündurumu	Kazanın gerçekleştiği zaman dilimi (gündüz/gece)
Havadurumu	Kazanın gerçekleştiği andaki hava durumu
YolunYüzeyi	Yol yüzeyinin durumu
KazaOlusTürü	Kazanın oluş biçimi
KazaAracSayısı	Kazaya karışan araç sayısı
AracDarbeBölümü	Araçta darbenin meydana geldiği bölüm
AracModelYılı	Kazaya karışan aracın model yılı
EtkilenenKisiTipi	Kazadan etkilenen kişi türü
KazaSiddeti	Kazanın şiddetini gösteren hedef değişken

### 3.2.3. Hedef değişken (Kaza Şiddeti)

Bu çalışmada analiz edilen temel değişken kaza şiddeti olup modelleme sürecinde hedef değişken olarak kullanılmıştır. Kaza şiddeti, meydana gelen trafik kazasının sonucunun ne derece ciddi olduğunu ifade etmektedir ve trafik güvenliği çalışmalarında önemli bir

göstermektedir. Bu nedenle kazaların yalnızca oluşum sıklığının değil, aynı zamanda kazaların sonuçlarının da incelenmesi trafik kazalarının nedenlerinin daha iyi anlaşılmasına katkı sağlamaktadır.

Çalışmada kullanılan veri setinde kaza şiddeti “KAZA\_SIDDETI” değişkeni ile temsil edilmektedir. Bu değişken kazaların sonuçlarına göre sınıflandırılmış olup kazaların yaralanmalı veya ölümlü olma durumunu göstermektedir. Bu nedenle çalışmada kullanılan hedef değişken ikili (binary) bir sınıflandırma yapısına sahiptir.

Veri setindeki kaza kayıtları incelendiğinde kazaların büyük bölümünün yaralanmalı kazalardan oluştuğu görülmektedir. Ölümlü kazaların sayısı ise yaralanmalı kazalara göre oldukça düşüktür. Bu durum veri setinde sınıflar arasında bir dengesizlik bulunduğunu göstermektedir. Bu nedenle modelleme sürecinde bu durum dikkate alınmıştır.

Bu çalışmada kaza şiddeti bağımlı değişken olarak ele alınmış, veri setinde yer alan diğer değişkenler ise bağımsız değişkenler olarak modele dahil edilmiştir. Sürücü özellikleri, yaya özellikleri, yol ve çevre koşulları ile araç özelliklerine ilişkin değişkenlerin kaza şiddeti üzerindeki etkileri makine öğrenmesi yöntemleri kullanılarak analiz edilmiştir.

### **3.3. Veri Ön İşleme Süreci**

Makine öğrenmesi çalışmalarında veri setinin analiz ve modelleme sürecine uygun hale getirilmesi önemli bir aşamadır. Ham veri setleri çoğu zaman eksik değerler, hatalı kayıtlar veya tekrar eden gözlemler içerebilmektedir. Bu tür durumlar model performansını olumsuz etkileyebileceği için modelleme aşamasına geçmeden önce veri seti üzerinde çeşitli veri ön işleme işlemleri gerçekleştirilmiştir.

Bu çalışmada veri seti Python programlama dili kullanılarak analiz edilmiş ve veri ön işleme işlemleri “Pandas” ve “NumPy” kütüphaneleri yardımıyla gerçekleştirilmiştir. Veri seti Excel formatından çalışma ortamına aktarılmış ve gerekli düzenlemeler yapılmıştır. Başlangıçta veri seti 9787 gözlem ve 19 değişkenden oluşmaktadır. Veri ön işleme sürecinde eksik veriler incelenmiş, daha sonra veri setinde tekrar eden kayıtlar kontrol edilmiştir. Yapılan bu işlemler sonucunda veri seti analiz için daha uygun ve tutarlı bir yapıya getirilmiştir.

### **3.3.1. Eksik verilerin analizi ve temizlenmesi**

Veri setinin analiz edilmesinden önce veri setinde eksik değerlerin bulunup bulunmadığı incelenmiştir. Bu amaçla veri setindeki tüm değişkenlerde eksik veri olup olmadığı kontrol edilmiştir. Yapılan inceleme sonucunda bazı değişkenlerde eksik gözlemler bulunduğu tespit edilmiştir.

Eksik verilerin modelleme sürecini olumsuz etkilememesi amacıyla veri setinde bulunan boş hücreler öncelikle NaN (Not a Number) değerine dönüştürülmüştür. Daha sonra veri setinde herhangi bir değişkeninde eksik değer bulunan satırlar veri setinden çıkarılmıştır. Bu işlem Python ortamında “Pandas” kütüphanesinde bulunan “dropna()” fonksiyonu kullanılarak gerçekleştirilmiştir.

Bu işlem sonucunda veri setinde eksik veri içeren gözlemler temizlenmiş ve veri setinin tüm değişkenlerinde eksik değer bulunmadığı görülmüştür. Böylece veri seti analiz ve modelleme süreci için daha güvenilir bir yapıya kavuşturulmuştur.

### **3.3.2. Yinelenen verilerin kaldırılması**

Veri setlerinde bazı kayıtların birden fazla kez bulunması mümkün olabilmektedir. Bu tür yinelenen kayıtlar veri analizinde hatalı sonuçlara yol açabileceği için veri seti tekrar eden gözlemler açısından da incelenmiştir.

Bu amaçla veri setinde yinelenen kayıtların olup olmadığı “Pandas” kütüphanesinde bulunan “duplicated()” fonksiyonu kullanılarak kontrol edilmiştir. Yapılan inceleme sonucunda veri setinde tekrar eden kayıtlar belirlenmiş ve bu kayıtlar veri setinden çıkarılmıştır. Bu işlem “drop\_duplicates()” fonksiyonu kullanılarak gerçekleştirilmiştir.

Veri ön işleme sürecinde gerçekleştirilen eksik veri temizleme ve yinelenen veri kaldırma işlemleri sonucunda başlangıçta 9787 gözlemden oluşan veri seti 9286 gözleme düşmüştür. Başka bir ifadeyle veri temizleme işlemleri sonucunda veri setinden toplam 501 gözlem çıkarılmıştır. Böylece veri seti modelleme sürecinde kullanılacak daha tutarlı ve güvenilir bir yapıya getirilmiştir.

### 3.4. Değişken Dönüştürme ve Gruplandırma

Veri analizinde bazı değişkenlerin daha anlamlı ve yorumlanabilir hale getirilmesi amacıyla veri seti üzerinde çeşitli dönüştürme ve gruplandırma işlemleri uygulanmıştır. Özellikle sürekli değişkenlerin belirli aralıklara ayrılması, veri setinin daha düzenli bir yapıya kavuşmasını sağlamakta ve modelleme sürecinde değişkenlerin etkisinin daha kolay yorumlanmasına olanak tanımaktadır.

Bu çalışmada araç model yılı, yaya yaşı ve sürücü yaşı değişkenleri belirli aralıklara ayrılarak yeni kategorik değişkenler oluşturulmuştur. Bu işlem Python programlama dili kullanılarak “Pandas” kütüphanesinde bulunan “cut()” fonksiyonu yardımıyla gerçekleştirilmiştir. Gruplandırma işlemleri sonucunda oluşturulan yeni değişkenler veri setine eklenmiş, daha sonra ise orijinal değişkenler veri setinden çıkarılmıştır.

#### 3.4.1. Araç model yılı gruplandırması

Araçların model yılı trafik güvenliği açısından önemli bir değişken olarak değerlendirilmektedir. Araçların üretim yılı arttıkça araç güvenlik sistemleri ve teknolojik özellikleri de gelişmektedir. Bu nedenle araç model yılı değişkeni analizde daha anlamlı sonuçlar elde edebilmek amacıyla belirli yıl aralıklarına ayrılmıştır.

Bu çalışmada araç model yılı değişkeni 10 yıllık aralıklara bölünerek gruplandırılmıştır. Gruplandırma işlemi sonucunda araç model yılları aşağıdaki kategoriler altında toplanmıştır:

- 1970–1979
- 1980–1989
- 1990–1999
- 2000–2009
- 2010–2023

Gruplandırma işlemi Python ortamında “Pandas” kütüphanesindeki “cut()” fonksiyonu kullanılarak gerçekleştirilmiş ve oluşturulan yeni değişken “AracModelGrup” adıyla veri setine eklenmiştir. Daha sonra analizde tekrar kullanılmaması amacıyla orijinal “AracModelYılı” değişkeni veri setinden çıkarılmıştır.

### 3.4.2. Yaya yaş gruplarının oluşturulması

Yaya yaşı trafik kazalarında risk düzeyini etkileyebilen önemli bir faktördür. Özellikle çocuklar ve ileri yaş gruplarındaki bireyler trafik kazalarından daha fazla etkilenebilmektedir. Bu nedenle yaya yaşı değişkeni analiz sürecinde belirli yaş aralıklarına ayrılarak gruplandırılmıştır.

Bu çalışmada yaya yaşı değişkeni aşağıdaki yaş aralıklarına göre sınıflandırılmıştır:

- 0–18
- 19–30
- 31–45
- 46–60
- 61 ve üzeri

Bu işlem sonucunda “YayaYaşıGrup” adlı yeni bir değişken oluşturulmuştur. Gruplandırma işlemi Python ortamında “Pandas cut()” fonksiyonu kullanılarak gerçekleştirilmiştir. Daha sonra veri setinde tekrar kullanılmaması amacıyla orijinal “YayaYaşı” değişkeni veri setinden çıkarılmıştır.

### 3.4.3. Sürücü yaş gruplarının oluşturulması

Sürücü yaşı da trafik kazalarının oluşumunda önemli rol oynayan faktörlerden biridir. Farklı yaş gruplarındaki sürücülerin sürüş deneyimi, refleksleri ve risk algıları farklılık gösterebilmektedir. Bu nedenle sürücü yaşı değişkeni de analizde daha anlamlı sonuçlar elde edebilmek amacıyla belirli yaş aralıklarına ayrılmıştır.

Bu çalışmada sürücü yaşı değişkeni aşağıdaki yaş gruplarına göre sınıflandırılmıştır:

- 0–18
- 19–30
- 31–45
- 46–60
- 61 ve üzeri

Bu işlem sonucunda “Sürücü\_YasıGrup” adlı yeni bir değişken oluşturulmuştur. Gruplandırma işlemi Python ortamında “Pandas cut()” fonksiyonu kullanılarak gerçekleştirilmiştir. Daha

sonra analizde tekrar kullanılmaması amacıyla orijinal “Sürücü\_Yası” değişkeni veri setinden çıkarılmıştır.

### **3.5. Kategorik Değişkenlerin Sayısallaştırılması**

Makine öğrenmesi algoritmaları genellikle sayısal verilerle çalıştığı için veri setinde bulunan kategorik değişkenlerin sayısal formata dönüştürülmesi gerekmektedir. Bu nedenle çalışmada veri setinde yer alan kategorik değişkenler analiz sürecinde sayısallaştırılmıştır.

Bu işlem Python programlama dili kullanılarak “Pandas” kütüphanesinde bulunan “get\_dummies()” fonksiyonu yardımıyla gerçekleştirilmiştir. One-Hot Encoding yöntemi kullanılarak kategorik değişkenlerde yer alan her bir kategori için yeni bir değişken oluşturulmuştur. Böylece kategorik değişkenler ikili (0 ve 1) değerler alacak şekilde sayısal forma dönüştürülmüştür.

Bu işlem kapsamında veri setinde yer alan yaya türü, yaya cinsiyeti, yaya kusurları, yolcu kusurları, yol özellikleri, hava durumu, yol yüzeyi, araç özellikleri ve sürücü özellikleri gibi kategorik değişkenler sayısal değişkenlere dönüştürülmüştür. One-Hot Encoding işlemi sonucunda veri setindeki değişken sayısı artmış ve modelleme sürecinde kullanılacak sayısal bir veri yapısı elde edilmiştir.

Kategorik değişkenlerin sayısallaştırılması sonucunda veri seti makine öğrenmesi algoritmalarının kullanımı için uygun hale getirilmiştir. Böylece veri seti sınıflandırma modellerinin uygulanabileceği bir yapıya dönüştürülmüştür.

### **3.6. Sınıf Dengesizliği Problemi ve SMOTE Yöntemi**

Makine öğrenmesi çalışmalarında veri setlerinde sıklıkla karşılaşılan problemlerden biri sınıf dengesizliğidir. Sınıf dengesizliği, veri setinde yer alan sınıflardan birinin diğerine göre çok daha fazla gözleme sahip olması durumunda ortaya çıkmaktadır. Bu durum sınıflandırma modellerinin çoğunluk sınıfına yönelmesine neden olabilmekte ve model performansını olumsuz etkileyebilmektedir.

Bu çalışmada kullanılan veri setinde hedef değişken olan kaza şiddeti “KAZA\_SIDDETI” incelendiğinde yaralanmalı kazaların sayısının ölümlü kazalara göre oldukça fazla olduğu görülmüştür. Bu durum veri setinde sınıflar arasında dengesizlik bulunduğunu göstermektedir.

Sınıf dengesizliği problemini azaltmak amacıyla veri setine SMOTE (Synthetic Minority Over-sampling Technique) yöntemi uygulanmıştır. SMOTE yöntemi, azınlık sınıfına ait yeni örnekler oluşturarak veri setindeki sınıfların daha dengeli hale gelmesini sağlamaktadır. Bu yöntem mevcut azınlık sınıfı örnekleri arasındaki benzerlikleri kullanarak sentetik veri üretmekte ve böylece veri setindeki sınıf dağılımını dengelemektedir.

Bu çalışmada SMOTE işlemi Python programlama dili kullanılarak imblearn kütüphanesinde bulunan SMOTE sınıfı yardımıyla gerçekleştirilmiştir. SMOTE uygulanmadan önce veri seti bağımlı değişken “KAZA\_SIDDETI” ve bağımsız değişkenler olarak iki ayrı veri yapısına ayrılmıştır. Daha sonra SMOTE yöntemi uygulanarak azınlık sınıfına ait yeni gözlemler üretilmiş ve veri setindeki sınıf dağılımı dengelenmiştir.

SMOTE uygulaması sonucunda veri setinde yer alan sınıflar dengeli hale getirilmiş ve böylece makine öğrenmesi modellerinin daha sağlıklı sonuçlar üretmesi amaçlanmıştır. Bu işlem modelleme sürecinde sınıflandırma algoritmalarının performansını artırmaya katkı sağlamaktadır.

### **3.7. Mekânsal Analiz ve CBS Uygulaması**

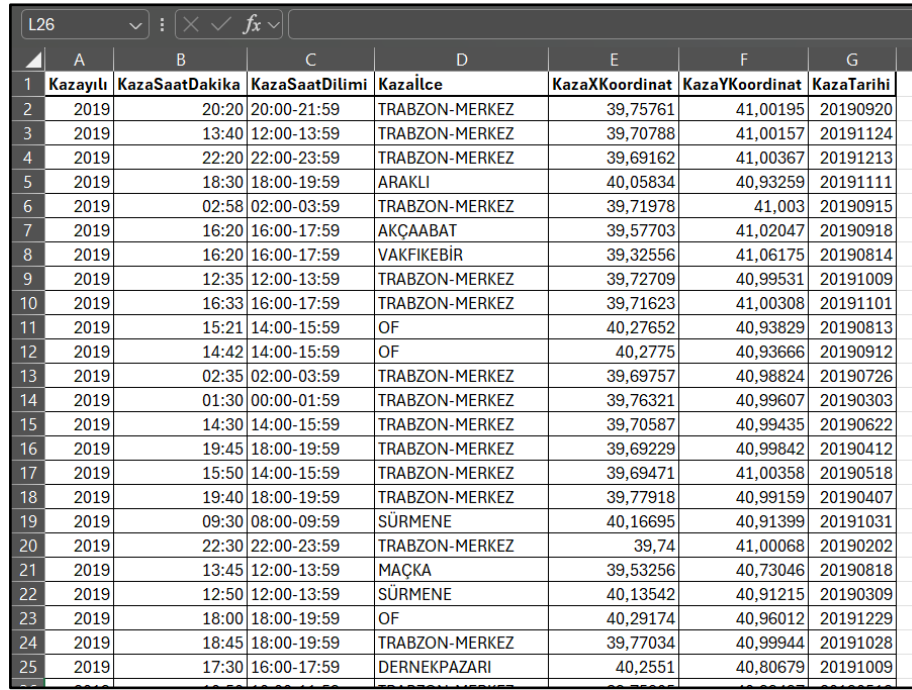
Bu çalışmada trafik kazalarının mekânsal boyutunun incelenmesi amacıyla Coğrafi Bilgi Sistemleri (CBS) kullanılmıştır. Çalışmada kullanılan veriler, Emniyet Genel Müdürlüğü Trafik Kaza Veritabanı’ndan temin edilmiş olup 2018–2023 yılları arasında kapsamaktadır. Elde edilen veriler CBS ortamına aktarılmış, uygun koordinat sisteminde düzenlenmiş ve mekânsal analizlere hazır hale getirilmiştir. Bu süreçte, kaza verilerinin haritalandırılması ve mekânsal dağılımının görselleştirilmesi sağlanmıştır.

#### **3.7.1. Verilerin CBS ortamına aktarılması ve hazırlanması**

Bu çalışmada kullanılan trafik kazası verileri, Emniyet Genel Müdürlüğü Trafik Kaza Veritabanı’ndan temin edilmiş olup 2018–2023 yılları arasında kapsamaktadır. Elde edilen veri seti başlangıçta toplam 6769 kayıttan oluşmaktadır. Ancak veri ön işleme aşamasında gerçekleştirilen kontroller sonucunda, bazı kayıtların hatalı veya eksik koordinat bilgisi içerdiği tespit edilmiştir. Özellikle X ve Y koordinat değerlerinde eksiklik veya gerçek dışı konum bilgisi içeren veriler, mekânsal analizlerin doğruluğunu olumsuz etkileyeceği için veri setinden

çıkarılmıştır. Bu işlemler sonucunda analizlerde kullanılmak üzere toplam veri sayısı 6467'ye düşürülmüştür.

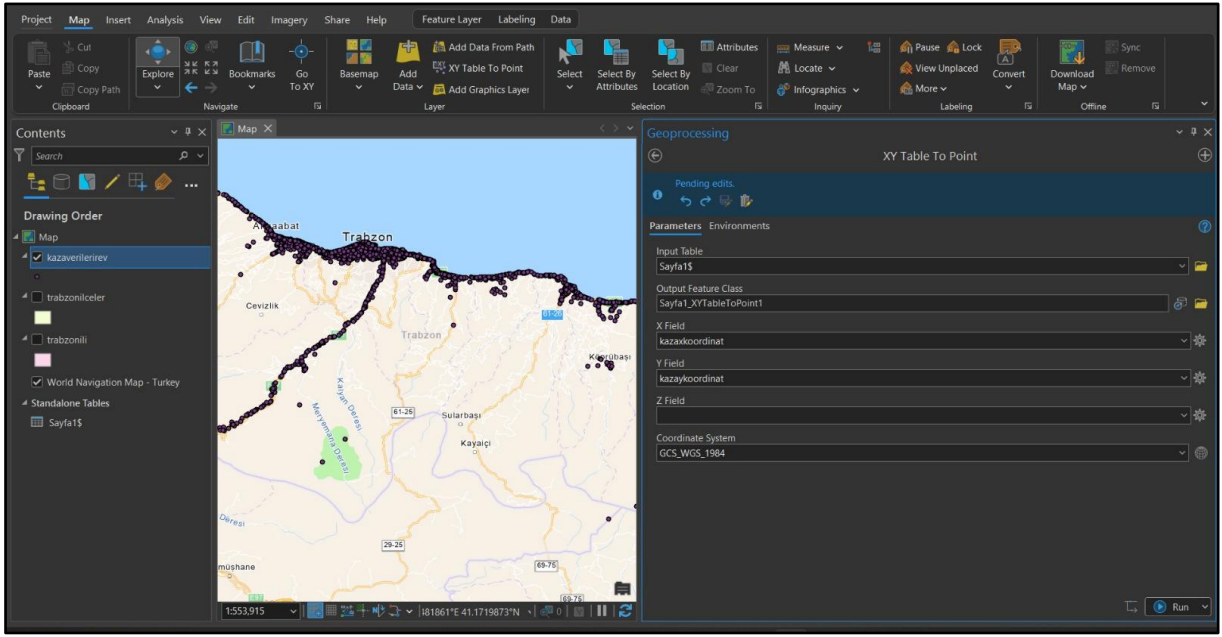
Veri hazırlama sürecinde, öncelikle kaza verileri Microsoft Excel ortamında düzenlenmiş ve CBS yazılımlarında kullanılabilir formata getirilmiştir. Veri setinde kaza yılı, saat bilgisi, ilçe bilgisi, zaman aralığı ve coğrafi konum bilgilerini temsil eden X ve Y koordinatları gibi değişkenler yer almaktadır. Excel ortamında düzenlenen veri setine ait örnek görünüm Şekil 4'te sunulmuştur.



	A	B	C	D	E	F	G
1	Kazayı	KazaSaatDakika	KazaSaatDilimi	Kazalce	KazaXKoordinat	KazaYKoordinat	KazaTarihi
2	2019	20:20	20:00-21:59	TRABZON-MERKEZ	39,75761	41,00195	20190920
3	2019	13:40	12:00-13:59	TRABZON-MERKEZ	39,70788	41,00157	20191124
4	2019	22:20	22:00-23:59	TRABZON-MERKEZ	39,69162	41,00367	20191213
5	2019	18:30	18:00-19:59	ARAKLI	40,05834	40,93259	20191111
6	2019	02:58	02:00-03:59	TRABZON-MERKEZ	39,71978	41,003	20190915
7	2019	16:20	16:00-17:59	AKÇAABAT	39,57703	41,02047	20190918
8	2019	16:20	16:00-17:59	VAKFIKEBİR	39,32556	41,06175	20190814
9	2019	12:35	12:00-13:59	TRABZON-MERKEZ	39,72709	40,99531	20191009
10	2019	16:33	16:00-17:59	TRABZON-MERKEZ	39,71623	41,00308	20191101
11	2019	15:21	14:00-15:59	OF	40,27652	40,93829	20190813
12	2019	14:42	14:00-15:59	OF	40,2775	40,93666	20190912
13	2019	02:35	02:00-03:59	TRABZON-MERKEZ	39,69757	40,98824	20190726
14	2019	01:30	00:00-01:59	TRABZON-MERKEZ	39,76321	40,99607	20190303
15	2019	14:30	14:00-15:59	TRABZON-MERKEZ	39,70587	40,99435	20190622
16	2019	19:45	18:00-19:59	TRABZON-MERKEZ	39,69229	40,99842	20190412
17	2019	15:50	14:00-15:59	TRABZON-MERKEZ	39,69471	41,00358	20190518
18	2019	19:40	18:00-19:59	TRABZON-MERKEZ	39,77918	40,99159	20190407
19	2019	09:30	08:00-09:59	SÜRMENE	40,16695	40,91399	20191031
20	2019	22:30	22:00-23:59	TRABZON-MERKEZ	39,74	41,00068	20190202
21	2019	13:45	12:00-13:59	MAÇKA	39,53256	40,73046	20190818
22	2019	12:50	12:00-13:59	SÜRMENE	40,13542	40,91215	20190309
23	2019	18:00	18:00-19:59	OF	40,29174	40,96012	20191229
24	2019	18:45	18:00-19:59	TRABZON-MERKEZ	39,77034	40,99944	20191028
25	2019	17:30	16:00-17:59	DERNEKPAZARI	40,2551	40,80679	20191009

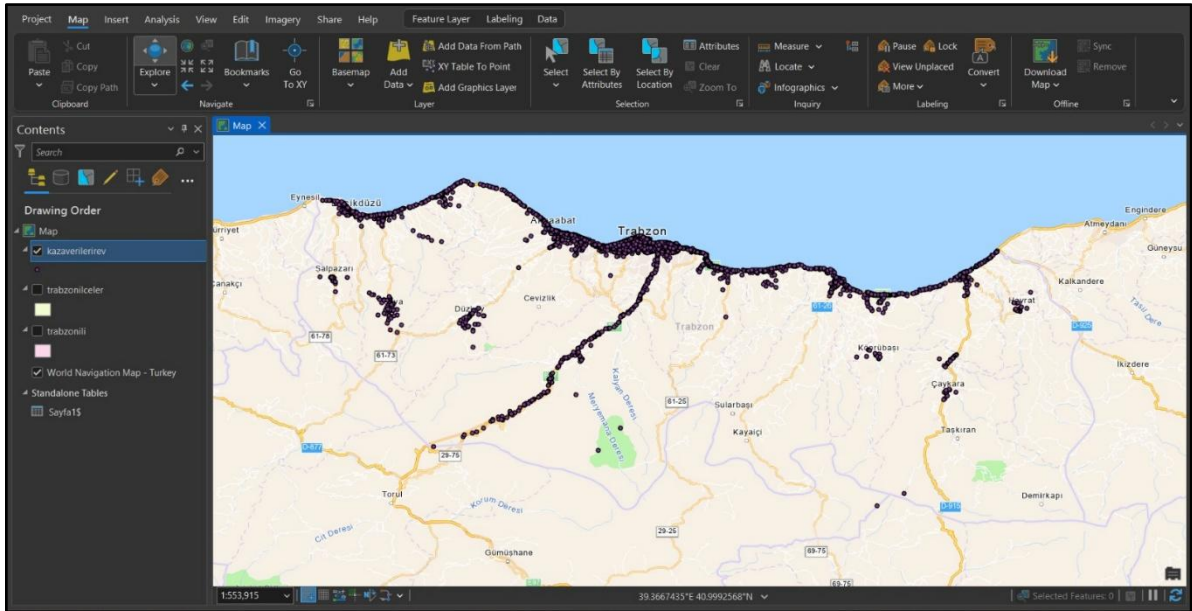
Şekil 4. Trafik kazası verilerinin Excel ortamındaki görünümü

Hazırlanan veri seti daha sonra ArcGIS Pro yazılımına aktarılmıştır. CBS ortamına veri aktarımı sürecinde, Excel formatındaki tablo verisi “XY Table to Point” aracı kullanılarak noktasal veri formatına dönüştürülmüştür. Bu işlem sırasında her bir kazaya ait X ve Y koordinatları kullanılarak kazalar harita üzerinde coğrafi olarak konumlandırılmıştır. Bu süreçte ait uygulama ekranı Şekil 5’te gösterilmektedir.



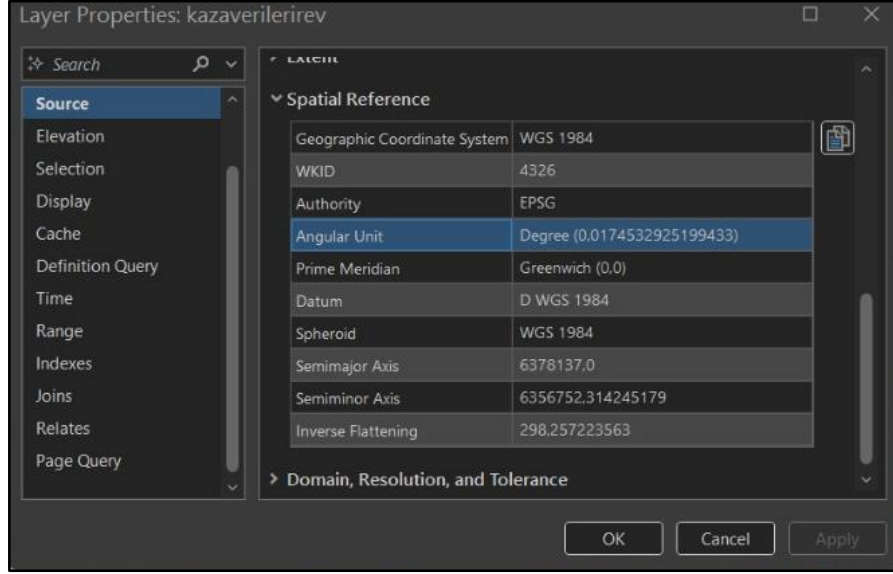
Şekil 5. Kaza verilerinin XY koordinatları kullanılarak CBS ortamına aktarılması

Koordinat bilgilerinin kullanılmasıyla oluşturulan noktasal kaza verilerinin harita üzerindeki dağılımı Şekil 6'da sunulmuştur. Bu aşamada tüm kazalar coğrafi konumlarına göre doğru şekilde yerleştirilmiş ve mekânsal analizlere uygun hale getirilmiştir.



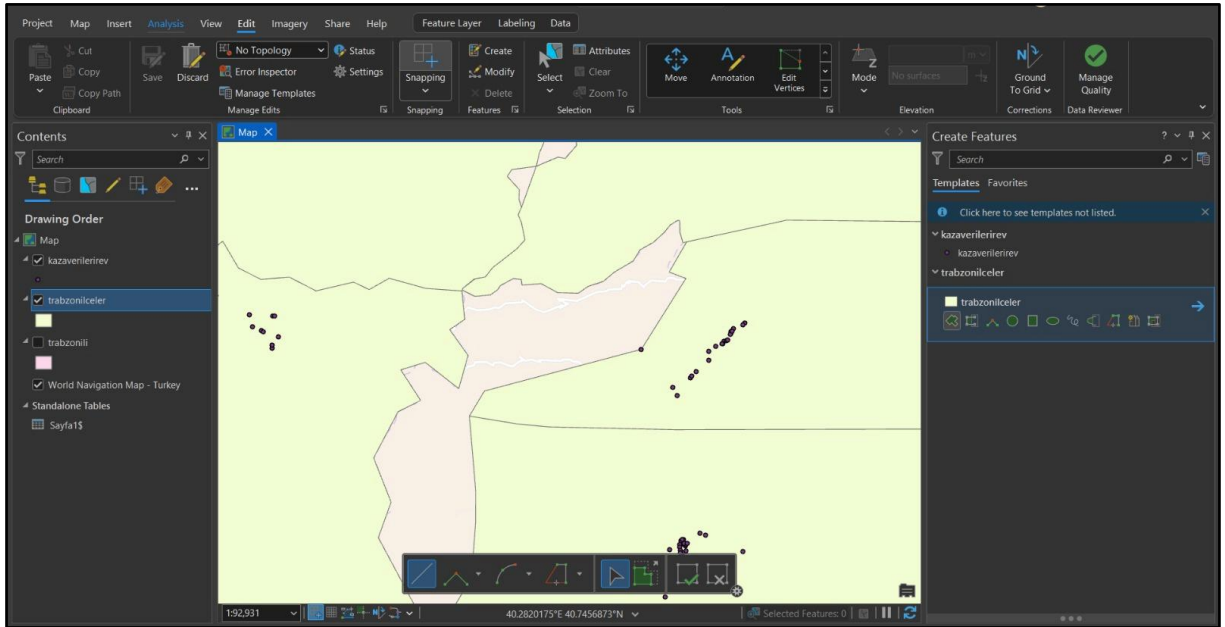
Şekil 6. Noktasal kaza verilerinin harita üzerindeki dağılımı

Oluşturulan veri katmanının konumsal doğruluğunu sağlamak amacıyla koordinat sistemi kontrol edilmiş ve WGS 84 (EPSG:4326) koordinat sistemi kullanılmıştır. Bu ayarlara ilişkin ekran görüntüsü Şekil 7'de verilmiştir.



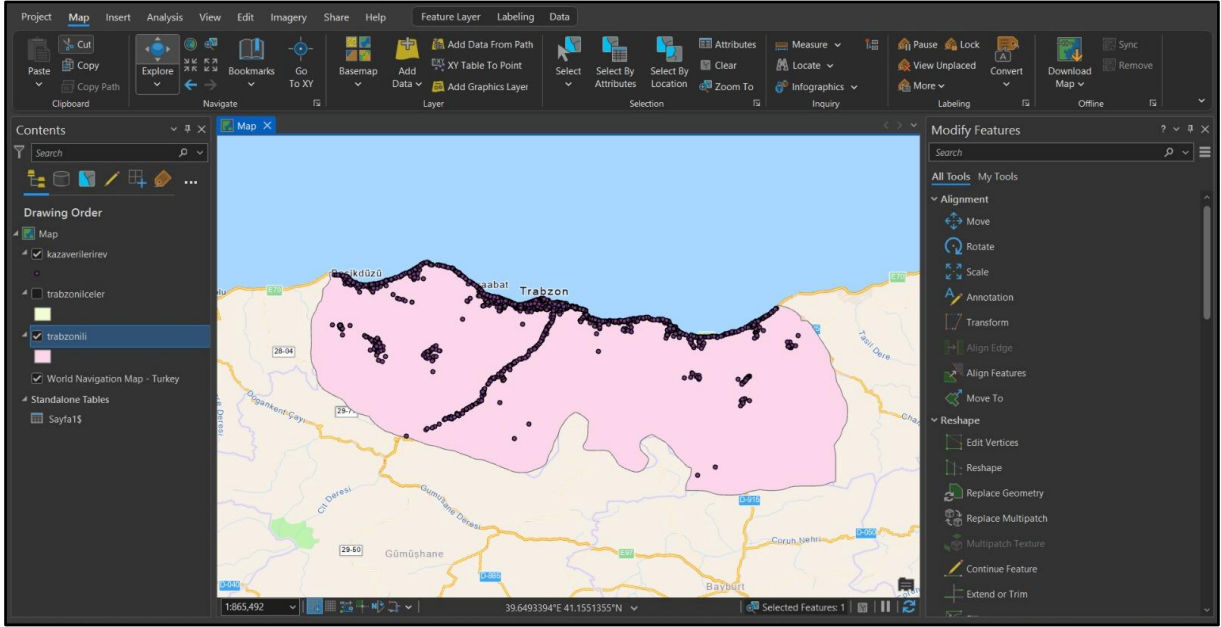
Şekil 7. Kaza verilerine ait koordinat sistemi (WGS 84) ayarları

Çalışma alanının idari sınırlarının belirlenmesi kapsamında Trabzon iline ait ilçe sınırları CBS ortamında polygon veri tipi kullanılarak oluşturulmuştur. İlçe sınırlarının çizilmesi sürecine ait görünüm Şekil 8'de gösterilmektedir. Bu işlem, kaza verilerinin ilçe bazında analiz edilebilmesine olanak sağlamaktadır.



Şekil 8. Trabzon ilçe sınırlarının CBS ortamında oluşturulması

Bununla birlikte, çalışma alanının genel sınırını belirlemek amacıyla Trabzon il sınırı da CBS ortamında vektör veri olarak oluşturulmuştur. Polygon oluşturma aracı kullanılarak çizilen il sınırı, analizlerin mekânsal olarak sınırlandırılmasını sağlamaktadır. Trabzon il sınırının oluşturulmasına ilişkin görünüm Şekil 9'da sunulmuştur.



Şekil 9. Trabzon il sınırının CBS ortamında vektör olarak oluşturulması

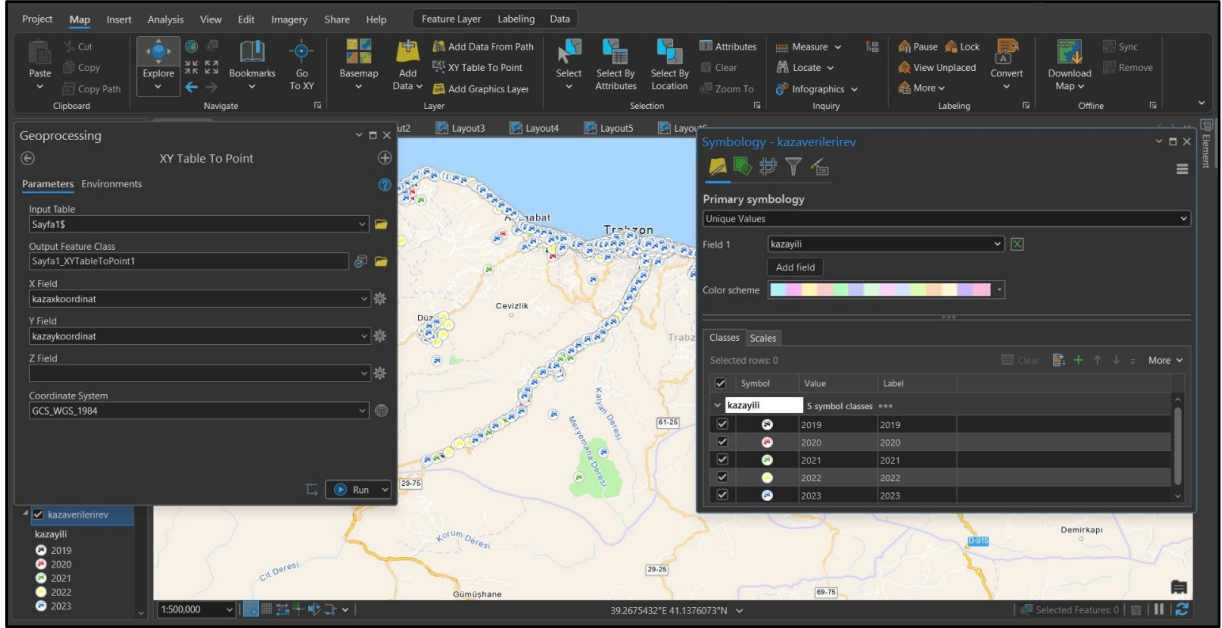
Veri temizleme, düzenleme ve CBS ortamına aktarım süreçlerinin tamamlanmasıyla birlikte trafik kazası verileri mekânsal analizler için hazır hale getirilmiştir. Bu aşama, çalışmanın devamında gerçekleştirilecek haritalandırma ve mekânsal analiz uygulamalarının temelini oluşturmaktadır.

### 3.7.2. Kaza verilerinin haritalandırılması

CBS ortamına aktarılan ve mekânsal olarak hazırlanan kaza verileri, bu aşamada haritalandırma işlemleri ile görselleştirilmiştir. Bu kapsamda, kaza verilerinin mekânsal olarak temsil edilmesi ve analizlere hazır hale getirilmesi amacıyla noktasal gösterim ve yoğunluk analizi uygulamaları gerçekleştirilmiştir.

İlk olarak, Excel ortamından aktarılan ve koordinat bilgileri kullanılarak oluşturulan kaza verileri, CBS ortamında nokta veri tipi olarak temsil edilmiştir. Her bir kaza olayı, X ve Y koordinat değerlerine bağlı olarak harita üzerinde bir nokta şeklinde konumlandırılmıştır. Bu

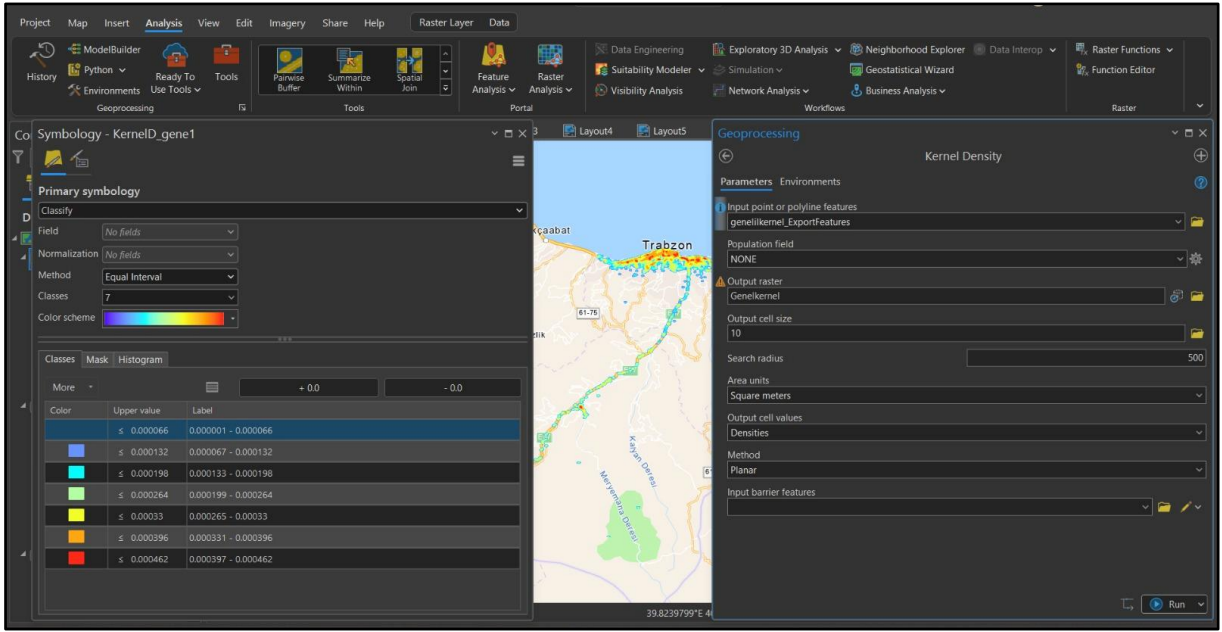
işlem sonucunda elde edilen noktasal veri katmanının CBS ortamındaki gösterimi Şekil 10'da sunulmaktadır.



Şekil 10. Trafik kazası verilerinin CBS ortamında noktasal olarak gösterimi

Noktasal veriler, mekânsal analizlerin temel girdisini oluşturmakta olup, kazaların konumsal olarak temsil edilmesini sağlamaktadır. Bu gösterim, verilerin kontrol edilmesi, doğrulanması ve analiz öncesi görsel olarak incelenmesi açısından önemli bir aşamadır.

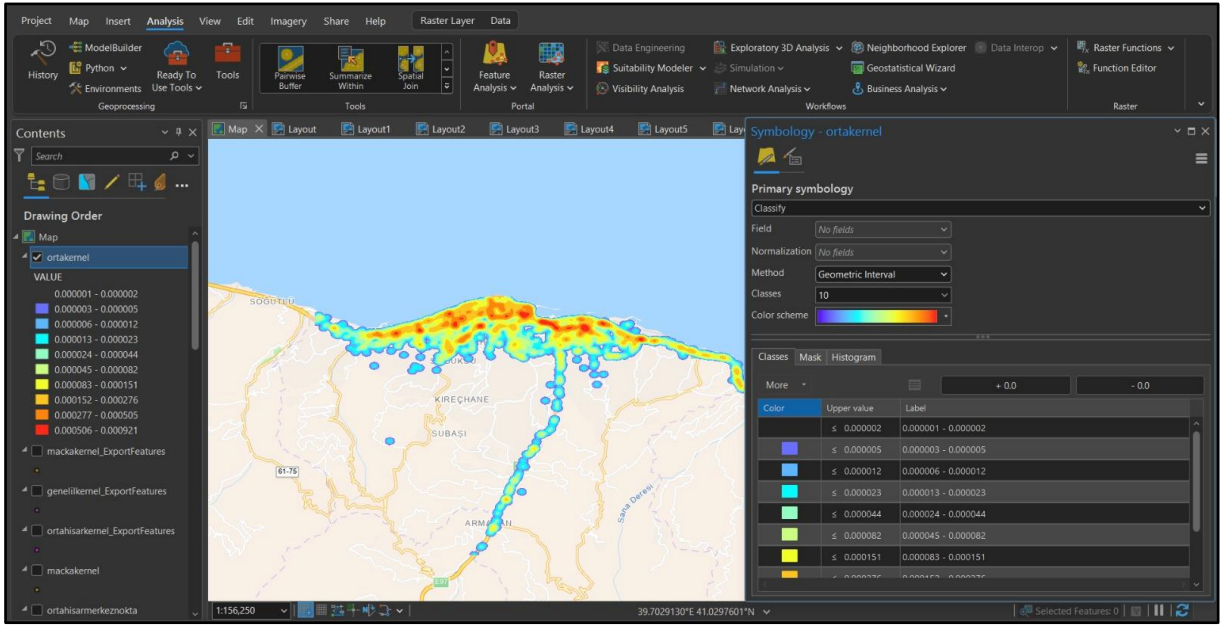
Kaza yoğunluklarının belirlenmesi amacıyla ArcGIS Pro yazılımında bulunan Spatial Analysis Tools araçları kullanılarak kernel density yöntemi uygulanmıştır. Bu analizde, noktasal kaza verileri temel alınarak her bir piksel için çevresindeki kaza yoğunluğu hesaplanmıştır. Kernel yoğunluk analizine ilişkin parametre ayarları ve uygulama arayüzü Şekil 11'de gösterilmektedir.



Şekil 11. Kernel yoğunluk analizi parametre ayarlarının uygulanması

Yoğunluk analizinde hücre (piksel) boyutu 10 birim olarak belirlenmiş ve analiz yarıçapı çalışma ölçeğine göre farklılaştırılmıştır. İl geneli analizlerde 500 metre yarıçap kullanılırken, ilçe bazlı analizlerde daha hassas sonuçlar elde edebilmek amacıyla 250 metre yarıçap tercih edilmiştir. Bu yaklaşım ile hem genel mekânsal eğilimlerin hem de yerel yoğunlaşmaların ortaya konulması amaçlanmıştır.

Kernel yoğunluk analizi sonucunda elde edilen raster veri, kazaların mekânsal yoğunluğunu temsil etmektedir. Bu yoğunluk yüzeyinin CBS ortamındaki gösterimi Şekil 12’de sunulmuştur.



Şekil 12. Kernel yoğunluk analizi sonucunda elde edilen yoğunluk yüzeyinin CBS ortamında gösterimi

Elde edilen yoğunluk verisinin daha anlaşılabilir hale getirilmesi amacıyla symbology ayarları kullanılarak sınıflandırma işlemi gerçekleştirilmiştir. Bu kapsamda yoğunluk değerleri geometrik dağılım yöntemi kullanılarak 6 sınıfa ayrılmış ve farklı yoğunluk seviyeleri renk geçişleri ile görselleştirilmiştir. Bu işlem, mekânsal yoğunluk farklılıklarının daha net bir şekilde ifade edilmesini sağlamaktadır.

Sonuç olarak, gerçekleştirilen haritalandırma işlemleri ile kaza verileri CBS ortamında mekânsal olarak temsil edilmiş ve analizlere uygun hale getirilmiştir. Bu aşama, çalışmanın ilerleyen bölümlerinde gerçekleştirilecek mekânsal değerlendirmeler için temel veri altyapısını oluşturmaktadır.

### 3.8. Eğitim ve Test Veri Setlerinin Oluşturulması

Makine öğrenmesi çalışmalarında oluşturulan modellerin performansını doğru bir şekilde değerlendirebilmek için veri setinin eğitim ve test olmak üzere iki ayrı veri grubuna ayrılması gerekmektedir. Bu işlem sayesinde model, veri setinin bir kısmı kullanılarak eğitilmekte ve daha sonra modelin performansı daha önce görmediği test verileri üzerinde değerlendirilmektedir.

Bu çalışmada veri seti modelleme sürecine geçmeden önce bağımlı değişken ve bağımsız değişkenler olarak iki ayrı veri yapısına ayrılmıştır. Hedef değişken olan “KAZA\_SIDDETI” bağımlı değişken olarak belirlenmiş, veri setinde yer alan diğer tüm değişkenler ise bağımsız değişkenler olarak modele dahil edilmiştir. Daha sonra veri seti eğitim ve test veri setleri olarak ikiye ayrılmıştır.

Veri setinin bölünmesi işlemi Python programlama dili kullanılarak “Scikit-Learn” kütüphanesinde yer alan “train\_test\_split()” fonksiyonu yardımıyla gerçekleştirilmiştir. Bu işlem sırasında veri setinin %80’i modelin eğitilmesi için eğitim veri seti olarak kullanılmış, %20’si ise model performansının değerlendirilmesi amacıyla test veri seti olarak ayrılmıştır.

### **3.9. Kullanılan Makine Öğrenmesi Yöntemleri**

Bu çalışmada trafik kazalarının şiddetini tahmin etmek amacıyla farklı makine öğrenmesi sınıflandırma algoritmaları kullanılmıştır. Modelleme sürecinde veri seti öncelikle veri temizleme, değişken dönüştürme ve kategorik değişkenlerin sayısallaştırılması gibi işlemlerden geçirilmiş, ardından modelleme aşamasına geçilmiştir. Analiz sürecinde veri seti bağımlı ve bağımsız değişkenler olarak iki ayrı yapıya ayrılmıştır. Hedef değişken olarak “KAZA\_SIDDETI” kullanılmış, veri setinde yer alan diğer değişkenler ise bağımsız değişkenler olarak modele dahil edilmiştir.

Modelleme aşamasında sınıf dengesizliği problemi bulunduğu için veri setine SMOTE yöntemi uygulanmış ve azınlık sınıfa ait sentetik gözlemler üretilerek veri seti dengelenmiştir. Daha sonra elde edilen veri seti eğitim ve test veri kümelerine ayrılmıştır. Veri setinin yüzde 80’i eğitim veri seti olarak kullanılmış, yüzde 20’si ise test veri seti olarak ayrılmıştır.

Bazı makine öğrenmesi algoritmaları değişken ölçeklerinden etkilenebildiği için modelleme sürecinde veri standartlaştırma işlemi de uygulanmıştır. Bu amaçla “Scikit-Learn” kütüphanesinde yer alan “StandardScaler” yöntemi kullanılmıştır. Standartlaştırma işlemi eğitim verisi üzerinde gerçekleştirilmiş ve aynı dönüşüm test veri setine de uygulanmıştır. Böylece değişkenlerin farklı ölçeklerde olmasından kaynaklanabilecek etkiler azaltılmıştır.

Bu çalışmada kullanılan makine öğrenmesi yöntemleri lojistik regresyon, karar ağacı, rastgele orman, AdaBoost, Gradient Boosting ve XGBoost algoritmalarından oluşmaktadır. Tablo 3’te makine öğrenmesi modellerinde kullanılacak parametreler verilmiştir.

Tablo 3. Makine öğrenmesi modellerinde kullanılacak parametreler.

Model	Temel Parametreler
Lojistik Regresyon	penalty=L1, C=0.01, solver=saga, max_iter=70
Karar Ağacı	criterion=gini, max_depth=11, min_samples_split=19
Random Forest	n_estimators=100, max_depth=100, max_features=sqrt
AdaBoost	n_estimators=1000, learning_rate=0.00001
Gradient Boosting	n_estimators=150, learning_rate=0.1, max_depth=4
XGBoost	"alpha=0.1" "lambda=10.0" "max_depth=None" "min_child_weight=1" "subsample=0.1" "colsample_bytree=1.0" "learning_rate=0.00001" "n_estimators=800"

### 3.9.1. Lojistik regresyon

Lojistik regresyon, bağımlı değişkenin iki sınıfa sahip olduğu sınıflandırma problemlerinde kullanılan istatistiksel temelli bir yöntemdir. Bu modelde bağımsız değişkenler kullanılarak bir gözlemin belirli bir sınıfa ait olma olasılığı hesaplanmaktadır. Bu çalışmada kullanılan lojistik regresyon modeli “penalty = L1, C = 0.01, solver = saga, max\_iter = 70 ve random\_state = 42” parametreleri ile kurulmuştur.

Modelde kullanılan “penalty” parametresi, modelde uygulanacak düzenleme türünü belirlemektedir. Bu çalışmada penalty parametresi L1 olarak belirlenmiştir. L1 düzenleme, model katsayılarının mutlak değerlerini ceza terimi olarak modele ekleyerek bazı katsayıların sifira yaklaşmasını sağlar. Bu sayede model gereksiz değişkenleri elimine ederek daha sade bir yapı oluşturur.

Modelde kullanılan C parametresi, düzenleme katsayısının tersini ifade etmektedir. C değerinin küçük olması daha güçlü bir düzenleme uygulanması anlamına gelmektedir. Bu çalışmada C değeri 0.01 olarak belirlenmiştir. Bu tercih modelin eğitim verisine aşırı uyum göstermesini engelleyerek daha genellebilir bir model elde edilmesine katkı sağlamaktadır.

Modelin optimizasyon sürecinde kullanılan “solver” parametresi, model katsayılarının hangi optimizasyon yöntemi ile hesaplanacağını belirlemektedir. Bu çalışmada saga çözücüsü

kullanılmıştır. Saga çözücüsü özellikle L1 düzenlileştirmeyi destekleyen ve yüksek boyutlu veri setlerinde etkin şekilde çalışan bir optimizasyon algoritmasıdır.

Modelde kullanılan “max\_iter” parametresi, modelin maksimum iterasyon sayısını ifade etmektedir. Bu çalışmada bu değer 70 olarak belirlenmiştir. Ayrıca modelin tekrarlanabilirliğini sağlamak amacıyla “random\_state” parametresi 42 olarak belirlenmiştir. Bu çalışmada lojistik regresyon modeli, değişkenlerin kaza şiddeti üzerindeki etkisini yorumlayabilmek ve temel bir sınıflandırma modeli oluşturabilmek amacıyla kullanılmıştır.

### 3.9.2. Karar ağacı

Karar ağacı, veri setini belirli karar kurallarına göre dallara ayırarak sınıflandırma yapan, yapısı itibarıyla yorumlanabilir bir makine öğrenmesi yöntemidir. Bu yöntemde her düğüm bir karar noktasını, dallar karar yollarını ve yaprak düğümler ise nihai sınıfı temsil etmektedir. Trafik kazası gibi çok sayıda değişkenin etkili olduğu çalışmalarda karar ağaçları, hangi değişkenin hangi sırayla ayırım yaptığını göstermesi bakımından önemli bir yöntemdir.

Bu çalışmada karar ağacı modeli “criterion=“gini”, max\_depth=11, min\_impurity\_decrease=0.000663, min\_samples\_split=19, min\_samples\_leaf=14, max\_features=None, random\_state=42” parametreleriyle kurulmuştur. “criterion=“gini”” parametresi, düğümlerin hangi safsızlık ölçütüne göre bölüneceğini belirlemektedir. Gini ölçütü, sınıflar arasındaki düzensizliği en aza indirecek bölünmeyi seçtiği için sınıflandırma problemlerinde yaygın ve hızlı bir yöntemdir. Dosyada da bu tercihin kazaya dayalı sınıflandırmalarda standart bir yaklaşım olduğu belirtilmiştir.

max\_depth=11 parametresi, ağacın en fazla 11 seviye derinliğe ulaşmasına izin vermektedir. Karar ağaçlarında derinliğin fazla olması, modelin eğitim verisindeki ayrıntıları ezberlemesine ve test verisinde performans kaybına yol açmasına neden olabilir. Bu nedenle derinliğin sınırlandırılması, özellikle gürültülü ve yüksek boyutlu veri yapılarında aşırı öğrenmeyi kontrol etmek için önemlidir. “min\_samples\_split=19” parametresi, bir düğümün yeni alt düğümlere ayrılabilmesi için en az 19 gözlem içermesi gerektiğini göstermektedir. Bu sayede çok az gözlem içeren düğümler üzerinden rastlantısal bölünmeler yapılmasının önüne geçilmektedir. “min\_samples\_leaf=14” ise bir yaprak düğümde en az 14 gözlem bulunmasını zorunlu kılar. Bu da modelin çok küçük alt gruplar üzerinden karar vermesini engelleyerek daha kararlı bir yapı oluşturur.

“min\_impurity\_decrease=0.000663” parametresi, bir düğümün bölünebilmesi için belirli bir minimum bilgi kazancı sağlanmasını zorunlu kılmaktadır. Bu eşik sayesinde modele anlamlı katkı sağlamayan bölünmeler engellenmekte, dolayısıyla daha sade ve daha genellenebilir bir ağaç yapısı elde edilmektedir. “max\_features=None” parametresi her bölünmede tüm değişkenlerin değerlendirilmesi anlamına gelmektedir. Bu tercih, tek ağaçlı modelde değişkenlerin etkisini mümkün olduğunca tam değerlendirmek amacıyla yapılmıştır. “random\_state=42” ise model sonuçlarının tekrarlanabilirliğini sağlamaktadır. Bu parametre yapısı genel olarak değerlendirildiğinde, karar ağacının özellikle aşırı öğrenme riski azaltılmıştır.

### 3.9.3. Rastgele orman

Rastgele orman, çok sayıda karar ağacının bir araya getirilmesiyle oluşturulan topluluk öğrenme yöntemlerinden biridir. Bu yöntemde her ağaç veri setinin farklı bir alt örneği üzerinde eğitilmekte ve nihai sınıflandırma, tüm ağaçların ortak kararı ile elde edilmektedir. Bu yapı tek bir karar ağacına göre daha dengeli ve daha kararlı sonuçlar üretmektedir. Trafik kazası verileri gibi çok boyutlu veri yapılarında rastgele orman hem tahmin başarısı hem de değişken önemlerini ortaya koyabilmesi açısından güçlü bir yöntemdir.

Çalışmada rastgele orman modeli “n\_estimators=100, max\_depth=100, min\_samples\_split=100, min\_samples\_leaf=100, max\_features='sqrt', bootstrap=True, random\_state=42” parametreleri ile kurulmuştur. “n\_estimators=100”, modelde 100 adet karar ağacı oluşturulduğunu göstermektedir. Çok sayıda ağacın kullanılması, tek tek ağaçlardan kaynaklanabilecek rastlantısal hataların ortalamalanmasını sağlar ve model kararlılığını artırır. Bu sayı, hesaplama maliyeti ile model performansı arasında dengeli bir tercih olarak değerlendirilebilir.

“max\_depth=100” ilk bakışta çok yüksek görünse de modelde buna karşılık “min\_samples\_split=100” ve “min\_samples\_leaf=100” gibi oldukça sınırlayıcı parametreler kullanılmıştır. “min\_samples\_split=100”, bir düğümün bölünebilmesi için en az 100 gözlem bulunmasını zorunlu kılmaktadır. “min\_samples\_leaf=100” ise yaprak düğümlerin en az 100 gözlem içermesini sağlamaktadır. Bu iki parametre, ağaçların çok küçük ve aşırı özelleşmiş yapılar oluşturmasını engelleyerek overfitting riskini azaltmaktadır. Dolayısıyla modelde teorik

olarak yüksek derinliğe izin verilmiş olsa da pratikte yaprak büyüklüğü kısıtları nedeniyle ağaçlar kontrollü şekilde gelişmektedir.

“max\_features='sqrt'” parametresi her bölünmede toplam değişken sayısının karekökü kadar değişkenin rastgele seçilerek değerlendirilmesini sağlamaktadır. Bu tercih, ağaçlar arasında çeşitliliği artırır ve ensemble yapının gücünü yükseltir. bootstrap=True parametresi ise her ağacın veri setinin yeniden örneklenmiş farklı alt kümeleri üzerinde eğitilmesini sağlamaktadır. Bu yaklaşım da yine model çeşitliliğini artırarak genellenebilirliği güçlendirmektedir. “random\_state=42” parametresi modelin tekrarlanabilir olmasını sağlamaktadır. Dosyadaki yorumlara göre bu yapı, tek ağaçlı modele göre daha kararlı, daha genelleyici ve nadir kaza türlerini de belirli ölçüde öğrenebilen bir model kurma amacı taşımaktadır.

#### **3.9.4. Adaboost**

AdaBoost algoritması, zayıf öğrencilerin ardışık biçimde bir araya getirilmesiyle güçlü bir sınıflandırıcı oluşturan boosting temelli bir makine öğrenmesi yöntemidir. Bu yöntemin temel prensibi, her yeni modelin önceki aşamalarda yanlış sınıflandırılan gözlemlere daha fazla ağırlık vererek öğrenme sürecini sürdürmesidir. Böylece modelleme süreci ilerledikçe zor örnekler üzerinde öğrenme kapasitesi artmakta ve modelin sınıflandırma performansı kademeli olarak iyileştirilmektedir. Bu özellik, özellikle sınıf dengesizliğinin bulunduğu veri setlerinde ve nadir ancak önemli olayların tahmin edilmesinde AdaBoost algoritmasını önemli bir yöntem haline getirmektedir.

Bu çalışmada kullanılan AdaBoost modeli “n\_estimators = 1000 ve learning\_rate = 0.00001” parametreleri ile kurulmuştur. Modelde kullanılan “n\_estimators” parametresi, modelde kaç adet zayıf öğrencinin kullanılacağını ifade etmektedir. Bu çalışmada bu değer 1000 olarak belirlenmesi, modelin çok sayıda ardışık öğrenci kullanarak hataları kademeli biçimde azaltmasını amaçlamaktadır. Yüksek sayıda tahminci kullanılması modelin öğrenme sürecinin daha ayrıntılı ve kontrollü bir şekilde gerçekleşmesine olanak sağlamaktadır.

Modelde kullanılan “learning\_rate” parametresi ise her iterasyonda modelin ne kadar öğrenme gerçekleştireceğini belirlemektedir. Bu çalışmada “learning\_rate” değeri 0.00001 olarak belirlenmiştir. Öğrenme hızının oldukça düşük seçilmesi, modelin her iterasyonda çok küçük güncellemeler yapmasını sağlamaktadır. Bu durum modelin daha temkinli bir öğrenme süreci izlemesine yardımcı olmakta ve aşırı öğrenme riskini azaltabilmektedir.

Bununla birlikte öğrenme hızının çok düşük olması modelin öğrenme sürecinin daha uzun sürmesine ve veri setindeki örüntüleri daha yavaş öğrenmesine neden olabilmektedir. Bu nedenle yüksek tahminci sayısı ile düşük öğrenme hızının birlikte kullanılması, modelin daha stabil ve kontrollü bir öğrenme süreci gerçekleştirmesine olanak sağlamaktadır.

AdaBoost algoritması, özellikle sınıf dengesizliği içeren trafik kazası verilerinde nadir fakat kritik kaza türlerini daha etkili bir şekilde yakalayabilmek amacıyla kullanılmıştır. Bu yaklaşım, modelin zor sınıflandırılan gözlemler üzerinde daha fazla öğrenme gerçekleştirmesini sağlayarak kaza şiddeti tahmininde daha dengeli bir model oluşturmayı amaçlamaktadır.

### 3.9.5. Gradient Boosting

Gradient Boosting algoritması, zayıf öğrencilerin ardışık olarak bir araya getirilmesiyle güçlü bir tahmin modeli oluşturmayı amaçlayan boosting temelli makine öğrenmesi yöntemlerinden biridir. Bu yöntemde modelleme süreci aşamalı olarak ilerlemekte ve her yeni model, bir önceki modelin tahmin hatalarını azaltacak şekilde oluşturulmaktadır. Gradient Boosting yaklaşımında model hataları bir kayıp fonksiyonu üzerinden değerlendirilmekte ve bu hataların azaltılması amacıyla gradyan iniş yöntemi kullanılmaktadır. Bu sayede modelleme süreci iteratif olarak geliştirilmekte ve her adımda model performansının artırılması hedeflenmektedir.

Gradient Boosting algoritması özellikle karmaşık veri yapılarında ve değişkenler arasındaki doğrusal olmayan ilişkilerin bulunduğu durumlarda başarılı sonuçlar verebilmektedir. Trafik kazaları gibi çok sayıda değişkenin birlikte etkili olduğu veri setlerinde bu algoritma, değişkenler arasındaki karmaşık örüntüleri öğrenebilme yeteneği sayesinde etkili bir sınıflandırma yöntemi olarak değerlendirilmektedir.

Bu çalışmada kullanılan Gradient Boosting modeli “ $n\_estimators = 150$ ,  $learning\_rate = 0.1$ ,  $max\_depth = 4$ ,  $subsample = 0.9$ ,  $min\_samples\_split = 3$  ve  $min\_samples\_leaf = 1$ ” parametreleri ile kurulmuştur. Modelde kullanılan “ $n\_estimators$ ” parametresi, modelde oluşturulan karar ağacı sayısını ifade etmektedir. Bu çalışmada bu değerin 150 olarak belirlenmesi, modelin yeterli sayıda zayıf öğrenci kullanarak tahmin hatalarını kademeli olarak azaltmasını sağlamayı amaçlamaktadır.

Modelde kullanılan “ $learning\_rate$ ” parametresi, modelin öğrenme hızını belirlemektedir. Bu parametre her iterasyonda yapılan güncellemenin büyüklüğünü ifade etmektedir. Bu çalışmada

“learning\_rate” değeri 0,1 olarak belirlenmiştir. Orta düzeyde seçilen bu değer, modelin öğrenme sürecinin dengeli bir şekilde ilerlemesini sağlamaktadır. Çok yüksek öğrenme hızları modelin aşırı öğrenmesine neden olabilirken, çok düşük öğrenme hızları modelin öğrenme sürecini gereksiz şekilde yavaşlatabilmektedir.

Modelde kullanılan “max\_depth” parametresi, oluşturulan karar ağaçlarının maksimum derinliğini ifade etmektedir. Bu çalışmada “max\_depth” değeri 4 olarak belirlenmiştir. Ağaç derinliğinin sınırlı tutulması, modelin aşırı öğrenmesini önlemek ve daha genellenebilir bir yapı elde etmek amacıyla tercih edilmiştir.

Modelde kullanılan “subsample” parametresi, her iterasyonda veri setinin belirli bir kısmının kullanılmasını sağlamaktadır. Bu çalışmada “subsample” değeri 0,9 olarak belirlenmiştir. Bu yaklaşım modelin varyansını azaltarak daha stabil sonuçlar elde edilmesine katkı sağlamaktadır.

Modelde kullanılan “min\_samples\_split” parametresi bir düğümün alt düğümlere ayrılabilmesi için gerekli minimum gözlem sayısını belirlemektedir. Bu değer 3 olarak belirlenmiştir. Ayrıca “min\_samples\_leaf” parametresi bir yaprak düğümde bulunması gereken minimum gözlem sayısını ifade etmektedir ve bu değer 1 olarak belirlenmiştir. Bu parametreler modelin veri setindeki daha küçük örüntüleri öğrenebilmesine olanak sağlamaktadır.

Gradient Boosting algoritması, trafik kazası verilerinde değişkenler arasındaki karmaşık ilişkileri öğrenebilmek ve kaza şiddeti tahmininde daha güçlü bir sınıflandırma modeli oluşturabilmek amacıyla kullanılmıştır.

### **3.9.6. XGBoost**

Extreme Gradient Boosting (XGBoost) algoritması, Gradient Boosting yaklaşımının geliştirilmiş ve optimize edilmiş bir versiyonu olup, daha yüksek performans ve daha hızlı hesaplama yeteneği sunan bir makine öğrenmesi yöntemidir. XGBoost, ardışık olarak oluşturulan karar ağaçları aracılığıyla model hatalarını minimize etmeyi amaçlamakta ve bu süreçte düzenleme tekniklerini etkin bir şekilde kullanarak aşırı öğrenmeyi kontrol altına almaktadır. Bu algoritma, kayıp fonksiyonunun minimize edilmesi sürecinde gradyan ve ikinci türev bilgilerini kullanarak daha hassas ve güçlü tahminler üretebilmektedir.

XGBoost algoritması, özellikle yüksek boyutlu veri setlerinde, eksik veri içeren yapılarda ve değişkenler arasındaki karmaşık, doğrusal olmayan ilişkilerin bulunduğu problemlerde başarılı sonuçlar vermektedir. Trafik kazaları gibi çok sayıda değişkenin birlikte etkili olduğu veri setlerinde, değişkenler arasındaki karmaşık örüntüleri öğrenebilme yeteneği sayesinde güçlü bir sınıflandırma yöntemi olarak değerlendirilmektedir.

Bu çalışmada kullanılan XGBoost modeli, “alpha = 0.1, lambda = 10.0, max\_depth = None, min\_child\_weight = 1, subsample = 0.1, colsample\_bytree = 1.0, learning\_rate = 0.00001 ve n\_estimators = 800” parametreleri ile kurulmuştur. Modelde kullanılan “alpha” ve “lambda” parametreleri sırasıyla L1 ve L2 düzenleme katsayılarını ifade etmektedir. Bu çalışmada L1 düzenleme katsayısı düşük (0.1) tutulurken, L2 düzenleme katsayısı yüksek (10.0) belirlenmiştir. Bu yaklaşım, modelin aşırı öğrenmesini engellemek ve ağırlıkların kontrol altında tutulmasını sağlamak amacıyla tercih edilmiştir

Modelde kullanılan “max\_depth” parametresi, karar ağaçlarının maksimum derinliğini ifade etmektedir. Bu çalışmada bu parametre sınırlanmamış (None) olup ağaçların derinleşmesine izin verilmiştir. Ayrıca “min\_child\_weight” parametresi 1 olarak belirlenmiş olup, her yaprakta minimum gözlem sayısının düşük tutulması modelin veri setindeki küçük örüntüleri öğrenebilmesine olanak sağlamaktadır. Ancak bu durum, modelin aşırı öğrenme riskini artırabilecek bir yapı oluşturmaktadır

Modelde kullanılan “subsample” parametresi, her iterasyonda veri setinin belirli bir kısmının kullanılmasını sağlamaktadır ve bu çalışmada 0.1 olarak belirlenmiştir. Bu değer, modelin her ağaçta farklı veri alt kümeleri kullanarak çeşitlilik kazanmasını sağlamaktadır. Bununla birlikte “colsample\_bytree” parametresi 1.0 olarak belirlenmiş olup, her ağaçta tüm özelliklerin kullanılması sağlanmıştır

Modelde kullanılan “learning\_rate” parametresi, modelin öğrenme hızını belirlemekte olup bu çalışmada oldukça düşük bir değer olan 0.00001 olarak seçilmiştir. Bu yaklaşım modelin çok yavaş ancak kontrollü bir şekilde öğrenmesini sağlamak ve aşırı öğrenme riskini azaltmaktadır. Buna karşılık “n\_estimators” parametresi 800 olarak belirlenmiş olup, modelin yeterli sayıda ağaç ile öğrenme sürecini sürdürmesi amaçlanmıştır. Ancak düşük öğrenme hızı nedeniyle daha fazla iterasyona ihtiyaç duyulabileceği değerlendirilmektedir.

### 3.10. Model Performans Değerlendirme Ölçütleri

Makine öğrenmesi modellerinin başarısını değerlendirmek amacıyla farklı performans ölçütlerinden yararlanılmaktadır. Bir modelin yalnızca doğru tahmin üretmesi yeterli olmayıp, bu tahminlerin güvenilirliğinin ve genellenebilirliğinin çeşitli ölçütler kullanılarak analiz edilmesi gerekmektedir. Özellikle trafik kazaları gibi sınıf dağılımının dengesiz olabileceği veri setlerinde model performansının farklı açılardan değerlendirilmesi önem taşımaktadır. Bu nedenle tek bir ölçüt yerine birden fazla performans metriği kullanılarak modelin doğruluğu, hata oranı ve sınıflandırma başarısı birlikte değerlendirilmektedir.

Bu çalışmada geliştirilen makine öğrenmesi modellerinin performansını değerlendirmek amacıyla accuracy, precision, recall, F1 skoru ve Cohen's Kappa katsayısı kullanılmıştır. Ayrıca model tahmin hatalarının büyüklüğünü incelemek amacıyla MAE (Mean Absolute Error) ve MSE (Mean Squared Error) hata ölçütleri de hesaplanmıştır. Bu ölçütler, eğitim ve test veri kümeleri üzerinde hesaplanarak modelin genellenebilirliği ve aşırı öğrenme (overfitting) durumu analiz edilmiştir.

Model performansının değerlendirilmesi sürecinde eğitim veri seti üzerinde elde edilen performans değerleri ile test veri seti üzerinde elde edilen değerler karşılaştırılmıştır. Eğitim ve test sonuçları arasındaki farkların düşük olması modelin genellenebilirliğinin yüksek olduğunu göstermektedir. Bu yaklaşım, makine öğrenmesi literatüründe modelin gerçek dünya verileri üzerindeki performansını değerlendirmek amacıyla yaygın olarak kullanılmaktadır.

#### 3.10.1. Accuracy (Doğruluk)

Accuracy, sınıflandırma modellerinin performansını değerlendirmek için kullanılan en temel ölçütlerden biridir. Accuracy değeri, model tarafından doğru şekilde sınıflandırılan gözlemlerin toplam gözlem sayısına oranını ifade etmektedir. Bu ölçüt modelin genel sınıflandırma doğruluğunu göstermektedir.

Accuracy değeri, Denklem (3.1)'de verilen formül kullanılarak hesaplanmaktadır.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

(3.1)

Burada;

TP (True Positive) gerçek pozitif ve doğru tahmin edilen gözlemleri,  
TN (True Negative) gerçek negatif ve doğru tahmin edilen gözlemleri,  
FP (False Positive) yanlış pozitif tahminleri,  
FN (False Negative) ise yanlış negatif tahminleri ifade etmektedir.

Bu çalışmada accuracy değeri, oluşturulan modellerin genel sınıflandırma başarısını değerlendirmek amacıyla kullanılmıştır. Ancak trafik kazası verilerinde sınıf dağılımı çoğu zaman dengesiz olabildiği için accuracy değeri tek başına yeterli görülmemiş ve diğer performans ölçütleri ile değerlendirilmiştir.

### 3.10.2. Precision (Kesinlik)

Precision ölçütü, model tarafından pozitif olarak tahmin edilen gözlemler içerisinde gerçekten pozitif olanların oranını ifade etmektedir. Başka bir ifadeyle precision değeri modelin pozitif tahminlerinin doğruluk oranını göstermektedir.

Precision değeri, Denklem (3.2)'de verilen formül kullanılarak hesaplanmaktadır.

$$Precision = \frac{TP}{TP + FP} \quad (3.2)$$

Precision ölçütü özellikle yanlış pozitif tahminlerin önemli olduğu durumlarda model performansını değerlendirmek için kullanılmaktadır. Trafik kazası analizlerinde model tarafından yüksek riskli olarak tahmin edilen olayların gerçekten riskli olup olmadığının doğru şekilde belirlenmesi önem taşımaktadır. Bu nedenle precision değeri modelin tahmin güvenilirliğini değerlendirmek açısından önemli bir ölçüt olarak kullanılmıştır.

### 3.10.3. Recall (Duyarlılık)

Recall ölçütü, veri setinde bulunan gerçek pozitif gözlemlerin model tarafından doğru şekilde tahmin edilme oranını ifade etmektedir. Bu ölçüt modelin pozitif sınıfa ait gözlemleri yakalama başarısını göstermektedir.

Recall değeri, Denklem (3.3)'te verilen formül kullanılarak hesaplanmaktadır.

$$Recall = \frac{TP}{TP+FN} \quad (3.3)$$

Recall değeri özellikle kritik olayların kaçırılmaması gereken durumlarda önemli bir performans ölçütü olarak kabul edilmektedir. Trafik kazası verilerinde özellikle ağır sonuçlar doğuran kazaların doğru şekilde tespit edilmesi önemli olduğundan recall değeri model performansının değerlendirilmesinde dikkate alınmıştır.

#### 3.10.4. F1 Skoru

F1 skoru, precision ve recall değerlerinin birlikte değerlendirilmesini sağlayan bir performans ölçütüdür. Precision ve recall değerlerinin harmonik ortalaması alınarak hesaplanmaktadır.

F1 skoru, Denklem (3.4)'te verilen formül kullanılarak hesaplanmaktadır.

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (3.4)$$

F1 skoru özellikle sınıf dağılımının dengesiz olduğu veri setlerinde model performansını daha dengeli bir şekilde değerlendirmek amacıyla kullanılmaktadır. Bu çalışmada F1 skoru, modelin hem pozitif sınıfı yakalama başarısını hem de yanlış pozitif tahmin oranını birlikte değerlendirmek amacıyla kullanılmıştır.

#### 3.10.5. Cohen's Kappa

Cohen's Kappa katsayısı, model tarafından yapılan sınıflandırma ile gerçek sınıflar arasındaki uyum düzeyini ölçen istatistiksel bir performans ölçütüdür. Bu ölçüt model tarafından elde edilen doğruluk oranını rastlantısal doğruluk ile karşılaştırarak daha güvenilir bir değerlendirme yapılmasını sağlamaktadır.

Cohen's Kappa katsayısı, Denklem (3.5)'te verilen formül kullanılarak hesaplanmaktadır.

$$\kappa = \frac{P_o - P_e}{1 - P_e} \quad (3.5)$$

Burada;

$P_o$ : Gözlenen doğruluk oranını,

$P_e$ : Rastlantısal olarak doğru sınıflandırma olasılığını ifade etmektedir.

Kappa katsayısı  $-1$  ile  $1$  arasında değer almaktadır. Değerin  $1$ 'e yakın olması model ile gerçek sınıflar arasında güçlü bir uyum olduğunu göstermektedir. Bu çalışmada Cohen's Kappa katsayısı, model başarısının rastlantısal doğruluktan bağımsız olarak değerlendirilmesi amacıyla kullanılmıştır.

### 3.10.6. MAE ve MSE

MAE (Mean Absolute Error) ve MSE (Mean Squared Error) ölçütleri model tahmin hatalarının büyüklüğünü değerlendirmek amacıyla kullanılan hata ölçütleridir. Bu ölçütler model tarafından yapılan tahminler ile gerçek değerler arasındaki farkın büyüklüğünü ortaya koymaktadır.

MAE değeri, Denklem (3.6)'da verilen formül kullanılarak hesaplanmaktadır.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (3.6)$$

Burada  $y_i$  gerçek değeri,  $\hat{y}_i$  ise model tarafından tahmin edilen değeri ifade etmektedir.

MSE değeri, Denklem (3.7)'de verilen formül kullanılarak hesaplanmaktadır.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3.7)$$

MAE ölçütü modelin ortalama hata büyüklüğünü gösterirken, MSE ölçütü büyük hatalara daha fazla ağırlık vermektedir. Bu çalışmada MAE ve MSE değerleri kullanılarak model tahminlerinin gerçek değerlere ne kadar yakın olduğu analiz edilmiştir. Ayrıca eğitim ve test veri setleri arasındaki hata farkları incelenerek modelin genellenebilirliği değerlendirilmiştir.

### 3.11. Confusion Matrix Analizi

Sınıflandırma modellerinin performansını daha ayrıntılı bir şekilde değerlendirebilmek amacıyla karışıklık matrisi (Confusion Matrix) kullanılmaktadır. Confusion Matrix, model tarafından yapılan tahminlerin gerçek sınıflar ile karşılaştırılmasını sağlayan ve sınıflandırma hatalarının türlerini ortaya koyan bir değerlendirme yöntemidir. Bu yöntem sayesinde modelin hangi sınıfları doğru veya yanlış tahmin ettiği açık bir şekilde analiz edilebilmektedir.

Confusion Matrix, dört temel bileşenden oluşmaktadır. Bunlar doğru pozitif (True Positive), doğru negatif (True Negative), yanlış pozitif (False Positive) ve yanlış negatif (False Negative) değerleridir. Doğru pozitif değerleri, gerçekte pozitif olan ve model tarafından doğru şekilde pozitif olarak tahmin edilen gözlemleri ifade etmektedir. Doğru negatif değerleri ise gerçekte negatif olan ve model tarafından doğru şekilde negatif olarak tahmin edilen gözlemleri göstermektedir. Yanlış pozitif değerleri gerçekte negatif olan ancak model tarafından pozitif olarak tahmin edilen gözlemleri, yanlış negatif değerleri ise gerçekte pozitif olan ancak model tarafından negatif olarak tahmin edilen gözlemleri ifade etmektedir.

Karışıklık matrisi, sınıflandırma performansını ayrıntılı biçimde değerlendirebilmek açısından önemli bir araçtır. Accuracy, precision, recall ve F1 skoru gibi performans ölçütleri de bu matris üzerinden hesaplanmaktadır. Bu nedenle Confusion Matrix analizi, model performansını yorumlayabilmek için temel bir değerlendirme yöntemi olarak kabul edilmektedir.

Bu çalışmada geliştirilen makine öğrenmesi modellerinin performansını değerlendirmek amacıyla Confusion Matrix analizi uygulanmıştır. Bu analiz sayesinde modelin ölümlü ve yaralanmalı kazaları ne ölçüde doğru sınıflandırdığı incelenmiş ve yanlış sınıflandırma türleri belirlenmiştir. Ayrıca Confusion Matrix sonuçları, model performansının güvenilirliğini değerlendirmek ve sınıflandırma hatalarının hangi durumlarda ortaya çıktığını analiz etmek amacıyla kullanılmıştır.

### 3.12. ROC Eğrisi ve AUC Analizi

Sınıflandırma modellerinin performansını farklı eşik değerleri altında değerlendirebilmek amacıyla ROC (Receiver Operating Characteristic) eğrisi ve AUC (Area Under Curve) ölçütü kullanılmaktadır. ROC eğrisi, modelin doğru pozitif oranı (True Positive Rate) ile yanlış pozitif oranı (False Positive Rate) arasındaki ilişkiyi göstermektedir. Bu eğri sayesinde modelin farklı karar eşikleri altında sınıflandırma performansı analiz edilebilmektedir.

ROC eğrisinde yatay eksen yanlış pozitif oranını, dikey eksen ise doğru pozitif oranını temsil etmektedir. Doğru pozitif oranı, Denklem (3.8)'de verilen formül kullanılarak hesaplanmaktadır.

$$TPR = \frac{TP}{TP+FN} \quad (3.8)$$

Yanlış pozitif oranı ise Denklem (3.9)'da verilen formül ile hesaplanmaktadır:

$$FPR = \frac{FP}{FP+TN} \quad (3.9)$$

ROC eğrisi modelin sınıfları ne ölçüde ayırt edebildiğini görsel olarak göstermektedir. Eğrinin üst sol köşeye yakın olması modelin sınıflandırma performansının yüksek olduğunu ifade etmektedir. Eğrinin köşegene yakın olması ise modelin rastlantısal tahminlere yakın performans gösterdiğini göstermektedir.

ROC eğrisinin altında kalan alan AUC (Area Under Curve) olarak adlandırılmaktadır. AUC değeri 0 ile 1 arasında değişmektedir. AUC değerinin 1'e yakın olması modelin sınıfları ayırt etme gücünün yüksek olduğunu göstermektedir. AUC değerinin 0.5'e yakın olması ise modelin rastlantısal tahmin seviyesinde performans gösterdiğini ifade etmektedir.

Bu çalışmada ROC eğrisi analizi kullanılarak geliştirilen makine öğrenmesi modellerinin farklı eşik değerleri altında sınıflandırma performansı değerlendirilmiştir. ROC analizi sayesinde modelin ölümlü ve yaralanmalı kazaları ayırt etme kapasitesi incelenmiş ve modelin sınıflandırma gücü görsel olarak analiz edilmiştir. ROC eğrisi ve AUC değeri, model performansının kapsamlı şekilde değerlendirilmesini sağlayan önemli bir yöntem olarak kullanılmıştır.

### 3.13. Çapraz Doğrulama (Cross Validation)

Makine öğrenmesi modellerinin performansını daha güvenilir bir şekilde değerlendirebilmek amacıyla çapraz doğrulama (cross validation) yöntemi kullanılmaktadır. Çapraz doğrulama, veri setinin farklı alt parçalara ayrılarak modelin birden fazla kez eğitilmesi ve test edilmesi prensibine dayanmaktadır. Bu yöntem sayesinde model performansının yalnızca tek bir eğitim-

test ayırımına bağılı kalması engellenmekte ve modelin genellenebilirliğı daha sağılıklı bir şekilde deęerlendirilebilmektedir.

Çapraz doęrulama yaklaşımda veri seti belirli sayıda alt kümeye ayrılmakta ve her iterasyonda bu alt kümelerden biri test verisi olarak kullanılırken kalan veri eğitim verisi olarak kullanılmaktadır. Bu işlem tüm alt kümeler için tekrar edilmekte ve elde edilen performans deęerlerinin ortalaması alınarak modelin genel performansı belirlenmektedir. Bu yaklaşım özellikle veri setinin sınırlı olduęu durumlarda model performansını daha güvenilir şekilde deęerlendirmeye olanak sağlamaktadır.

Bu çalışmada model performansının genellenebilirliğini incelemek amacıyla eğitim ve test veri setleri üzerinde elde edilen performans metrikleri karşılaştırılmıştır. Eğitim ve test sonuçları arasındaki farkların oldukça düşük olması modelin veri setine aşırı uyum göstermediğini ve genellenebilir bir yapı sergilediğini göstermektedir. Eğitim ve test performanslarının birbirine oldukça yakın olması, modelin yalnızca eğitim verisini ezberlemediğini ve yeni veri üzerinde de benzer performans gösterebildiğini ortaya koymaktadır. Bu yaklaşım makine öğrenmesi literatüründe modelin genellenebilirliğini deęerlendirmek için yaygın olarak kullanılan bir yöntemdir.

Ayrıca model karmaşıklığı ile eğitim ve doęrulama performansı arasındaki ilişki de analiz edilmiştir. Model karmaşıklığı arttıkça eğitim ve doęrulama performanslarının birlikte artması, modelin veri yapısını giderek daha iyi öğrendiğini göstermektedir. Eğitim ve doęrulama performans eğrilerinin birbirine yakın seyretmesi ise modelin aşırı öğrenme problemi yaşamadığını ve uygun bir bias–variance dengesine ulaşıldığını göstermektedir.

Bu nedenle çapraz doęrulama yaklaşımı, geliştirilen makine öğrenmesi modellerinin genellenebilirliğini deęerlendirmek ve model performansının güvenilirliğini analiz etmek amacıyla kullanılmıştır.

### **3.14. Özellik Önem Analizi (Feature Importance)**

Makine öğrenmesi modellerinde deęişkenlerin hedef deęişken üzerindeki etkisini belirlemek amacıyla özellik önem analizi (feature importance) kullanılmaktadır. Bu analiz, modelin tahmin üretirken hangi deęişkenleri daha fazla kullandığını ve hangi deęişkenlerin model üzerinde daha yüksek etkiye sahip olduğunu belirlemeye yardımcı olmaktadır. Özellik önem analizi

sayesinde veri setinde yer alan deęişkenlerin görelî etkileri incelenebilmekte ve model kararlarının hangi faktörlere dayandığı ortaya konulabilmektedir.

Bu çalışmada özellik önem analizi, özellikle ağaç tabanlı makine öğrenmesi modelleri kullanılarak gerçekleştirilmiştir. Ağaç tabanlı modellerde her bir deęişkenin önemi, model içerisinde yapılan bölünmelerde sağladığı bilgi kazancına göre hesaplanmaktadır. Bir deęişken model içerisinde ne kadar fazla ve ne kadar etkili bölünmelere yol açıyorsa o deęişkenin önem değeri de o kadar yüksek olmaktadır.

Özellik önem analizi sonucunda yol özellikleri, demografik faktörler ve araç özellikleri gibi deęişkenlerin kaza şiddetinin tahmin edilmesinde önemli rol oynadığı görülmektedir. Özellikle yol sınıfı, kavşak durumu, yol tipi ve yol geometrisi gibi altyapı deęişkenleri modelde önemli belirleyiciler arasında yer almaktadır. Bunun yanında kazaya karışan kişilerin yaş grupları, cinsiyetleri ve kazadan etkilenen kişi tipi gibi demografik faktörler de model tahminlerinde etkili deęişkenler arasında bulunmaktadır. Ayrıca araç model yılı, araç hareketi ve kaza araç sayısı gibi deęişkenler de kaza şiddetinin tahmin edilmesinde önemli rol oynamaktadır.

Özellik önem analizi sayesinde modelin yalnızca tahmin performansı değil, aynı zamanda hangi faktörlere dayanarak karar verdiği de analiz edilebilmektedir. Bu durum özellikle trafik güvenliği çalışmalarında önemli bir avantaj sağlamaktadır. Çünkü model tarafından belirlenen önemli deęişkenler, trafik kazalarının oluşumunda etkili olan faktörlerin anlaşılmasına ve önleyici güvenlik politikalarının geliştirilmesine katkı sağlayabilmektedir.

Bu nedenle özellik önem analizi, geliştirilen makine öğrenmesi modellerinin karar mekanizmasını incelemek ve trafik kazalarının oluşumunda etkili olan temel faktörleri belirlemek amacıyla kullanılmıştır.

## 4. BULGULAR

### 4.1. Mekânsal Analiz Bulguları

Bu bölümde, Trabzon iline ait 2018–2023 yılları arasındaki trafik kazası verileri Coğrafi Bilgi Sistemleri (CBS) kullanılarak mekânsal açıdan analiz edilmiştir. Analiz kapsamında kazaların noktasal dağılımı incelenmiş, kernel yoğunluk yöntemi ile kaza yoğunlukları belirlenmiş ve ilçe bazında mekânsal farklılıklar değerlendirilmiştir. Ayrıca seçili ilçeler üzerinde detaylı analizler gerçekleştirilerek kazaların mekânsal örüntüleri ortaya konulmuştur.

#### 4.1.1. Trafik kazalarının mekânsal noktasal dağılımının analizi

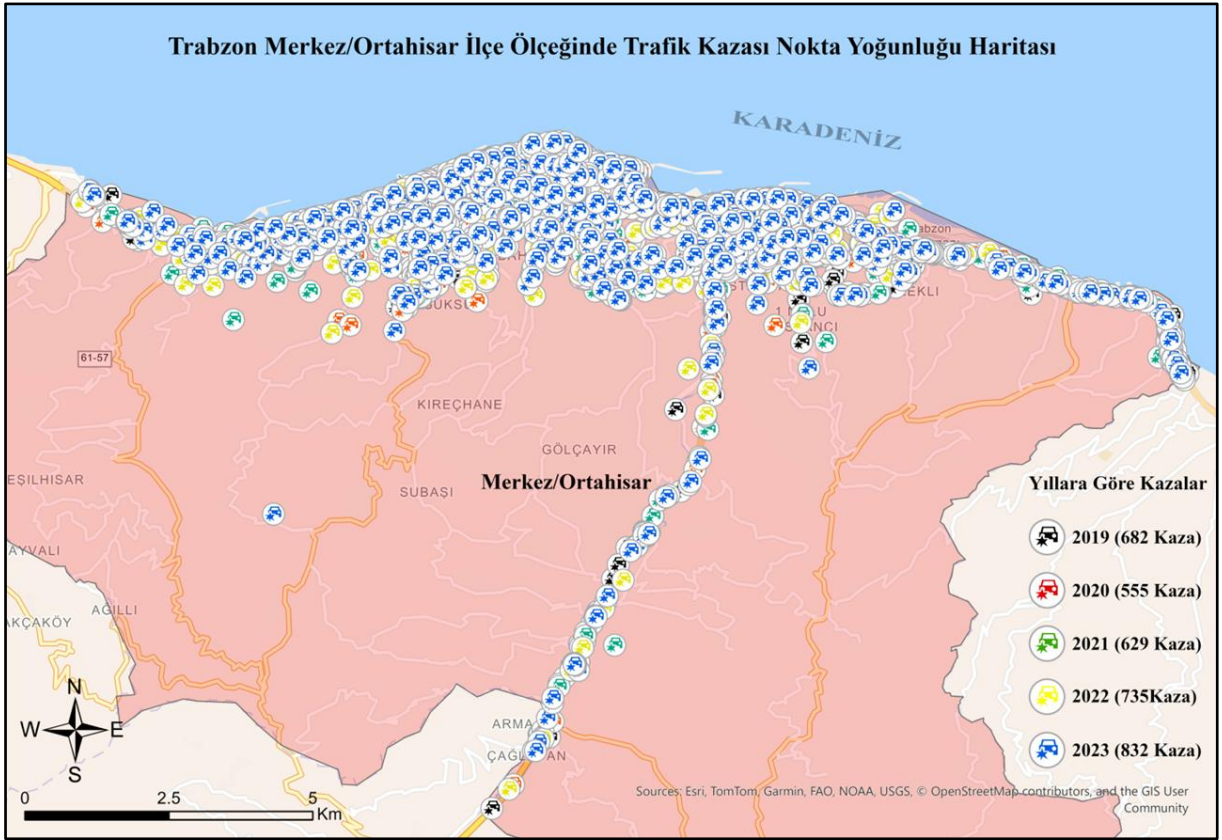
Trabzon iline ait 2018–2023 yılları arasındaki trafik kazalarının mekânsal dağılımı, noktasal kaza haritaları kullanılarak detaylı bir şekilde analiz edilmiştir. Noktasal gösterim yöntemi, her bir kaza olayının coğrafi koordinatları kullanılarak harita üzerinde ayrı bir nokta şeklinde temsil edilmesine dayanmakta olup, kazaların mekânsal örüntülerinin doğrudan gözlemlenmesine imkân sağlamaktadır. Bu yöntem, özellikle kazaların belirli bölgelerde kümelenme eğilimi gösterip göstermediğinin ortaya konulması açısından önemli bir analiz aracıdır.

Trabzon il geneline ait noktasal kaza dağılımı Şekil 13’te sunulmaktadır. Harita incelendiğinde, kazaların büyük ölçüde Karadeniz kıyı hattı boyunca uzanan ana ulaşım aksları üzerinde yoğunlaştığı açıkça görülmektedir. Özellikle sahil yolu boyunca kesintisiz bir şekilde devam eden yoğunluk, bu güzergâhın hem trafik hacmi hem de kullanım sıklığı açısından kritik bir öneme sahip olduğunu göstermektedir. Bununla birlikte, şehir merkezine bağlanan ana arterler üzerinde de belirgin bir kaza kümelenmesi gözlemlenmektedir. Bu durum hem şehir içi hem de şehirler arası ulaşımın kesiştiği bölgelerde trafik yoğunluğunun artmasına bağlı olarak kaza riskinin yükseldiğini göstermektedir.

İl genelinde sahil şeridinden iç kesimlere doğru gidildikçe kaza yoğunluğunun belirgin şekilde azaldığı ve dağılımın daha seyrek hale geldiği dikkat çekmektedir. Bu durum, kırsal alanlarda trafik hacminin daha düşük olması ve yol kullanım yoğunluğunun sınırlı kalması ile açıklanabilir. Ayrıca topografik yapı ve yerleşim dağılımı da kazaların mekânsal dağılımını etkileyen önemli faktörler arasında değerlendirilmektedir.







Şekil 15. Ortahisar ilçesine ait trafik kazalarının noktasal dağılımı

Genel olarak değerlendirildiğinde, trafik kazalarının mekânsal dağılımının homojen bir yapı göstermediği, aksine belirli bölgelerde yoğunlaşarak kümesel ve doğrusal örüntüler oluşturduğu belirlenmiştir. İl geneli analizinde kazaların ana ulaşım aksları boyunca yoğunlaştığı görülürken, ilçe bazında bu dağılımın yerel özelliklere bağlı olarak değişiklik gösterdiği ortaya konulmuştur. Özellikle kıyı hattı, şehir merkezi ve ana ulaşım koridorları yüksek riskli bölgeler olarak öne çıkmaktadır.

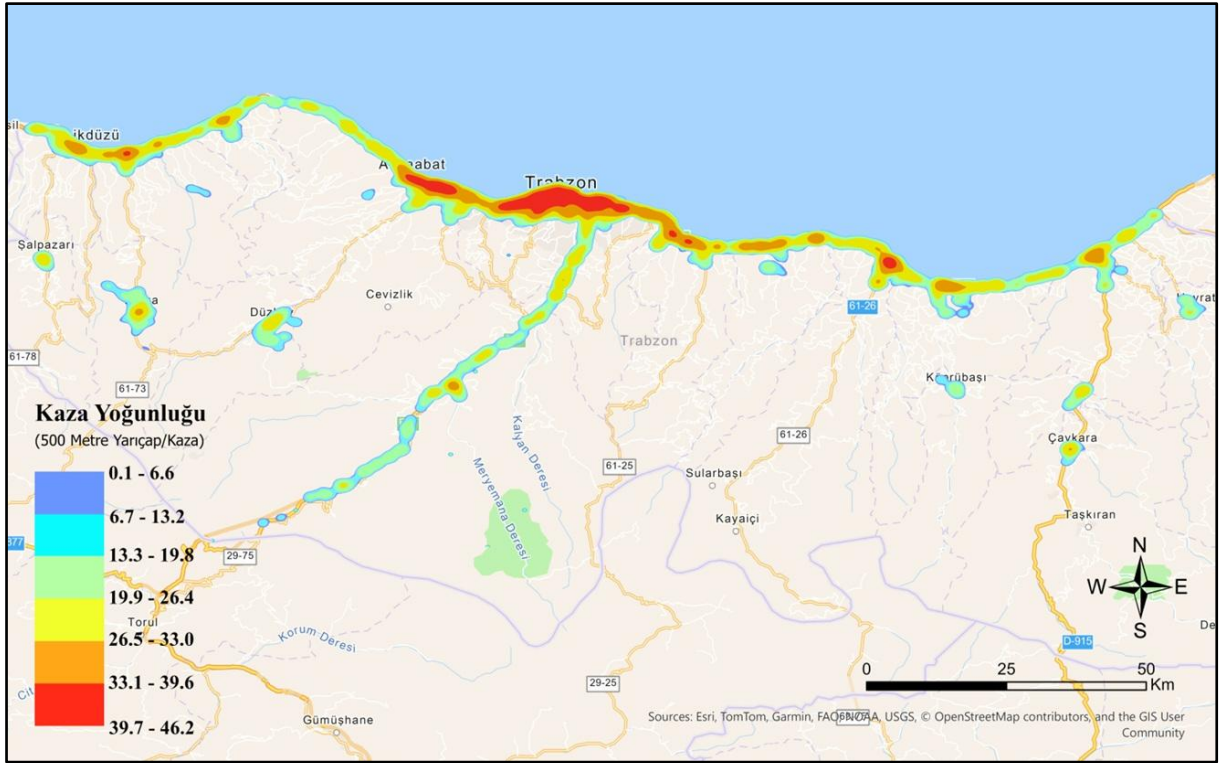
Bu bulgular, trafik kazalarının rastgele değil, belirli mekânsal faktörlere bağlı olarak oluştuğunu göstermekte olup, riskli bölgelerin belirlenmesi ve önleyici tedbirlerin geliştirilmesi açısından önemli bir temel oluşturmaktadır.

#### 4.1.2. Trafik kazalarının yoğunluk analizi bulguları

Trafik kazalarının mekânsal yoğunluklarının belirlenmesi amacıyla Kernel yoğunluk analizi uygulanmış ve elde edilen sonuçlar haritalar üzerinde görselleştirilmiştir. Bu yöntem, belirli bir yarıçap içerisinde bulunan kaza noktalarının yoğunluğunu hesaplayarak sürekli bir yüzey oluşturmakta ve kazaların mekânsal olarak hangi bölgelerde yoğunlaştığını daha net bir şekilde

ortaya koymaktadır. Böylece noktasal dağılımda doğrudan görülemeyen yoğunluk eğilimleri analiz edilebilmektedir.

Trabzon il geneline ait Kernel yoğunluk haritası Şekil 16'da sunulmuştur. Harita incelendiğinde, kaza yoğunluğunun özellikle kıyı hattı boyunca çok yüksek seviyelere ulaştığı görülmektedir. Kırmızı ve turuncu renklerle temsil edilen yüksek yoğunluk bölgeleri, sahil yolu ve şehir merkezi çevresinde belirgin bir şekilde öne çıkmaktadır. Bu durum, Trabzon ilinde trafik kazalarının büyük ölçüde kıyı boyunca uzanan ana ulaşım akslarında yoğunlaştığını göstermektedir. Ayrıca şehir merkezinden iç kesimlere doğru uzanan ana arterler üzerinde de orta düzeyde yoğunluk alanları dikkat çekmektedir. İç kesimlerde ise yoğunluğun belirgin şekilde azaldığı ve daha çok düşük yoğunluklu (mavi-yeşil) alanların hâkim olduğu görülmektedir.

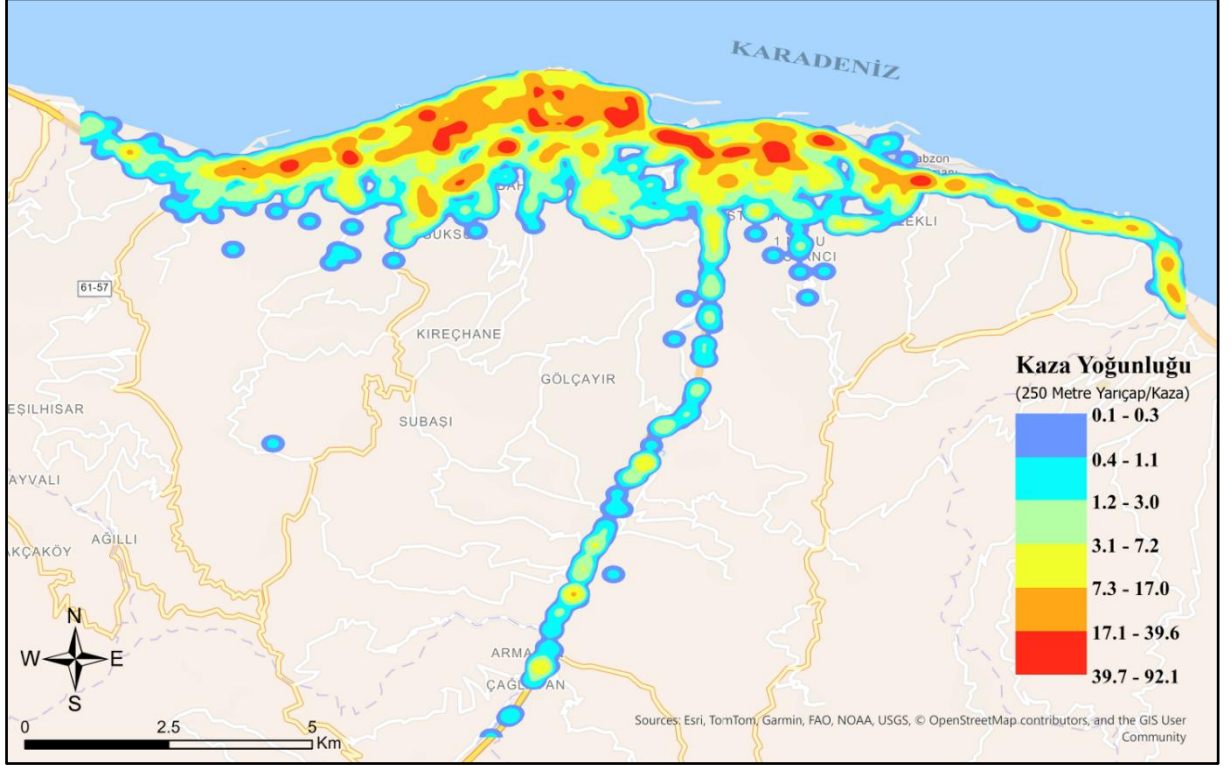


Şekil 16. Trabzon il geneline ait trafik kazası Kernel yoğunluk haritası

İlçe ölçeğinde yapılan Kernel analizleri, kazaların mekânsal dağılımının daha detaylı incelenmesine imkân sağlamıştır. Bu kapsamda Maçka ilçesine ait Kernel yoğunluk haritası Şekil 17'de verilmiştir. Harita incelendiğinde, kazaların belirli bir ulaşım koridoru boyunca yoğunlaştığı ve doğrusal bir yoğunluk yapısı oluşturduğu görülmektedir. Özellikle ana yol



Ortahisar ilçesinde ayrıca noktasal analizde görülen kümelenmelerin Kernel analizi ile daha belirgin hale geldiği ve bu bölgelerin açık bir şekilde “yüksek riskli alanlar” olarak tanımlanabildiği görülmektedir. Bu da Kernel yoğunluk analizinin, kazaların mekânsal risk bölgelerini belirlemede güçlü bir yöntem olduğunu ortaya koymaktadır.



Şekil 18. Ortahisar ilçesine ait trafik kazası Kernel yoğunluk haritası

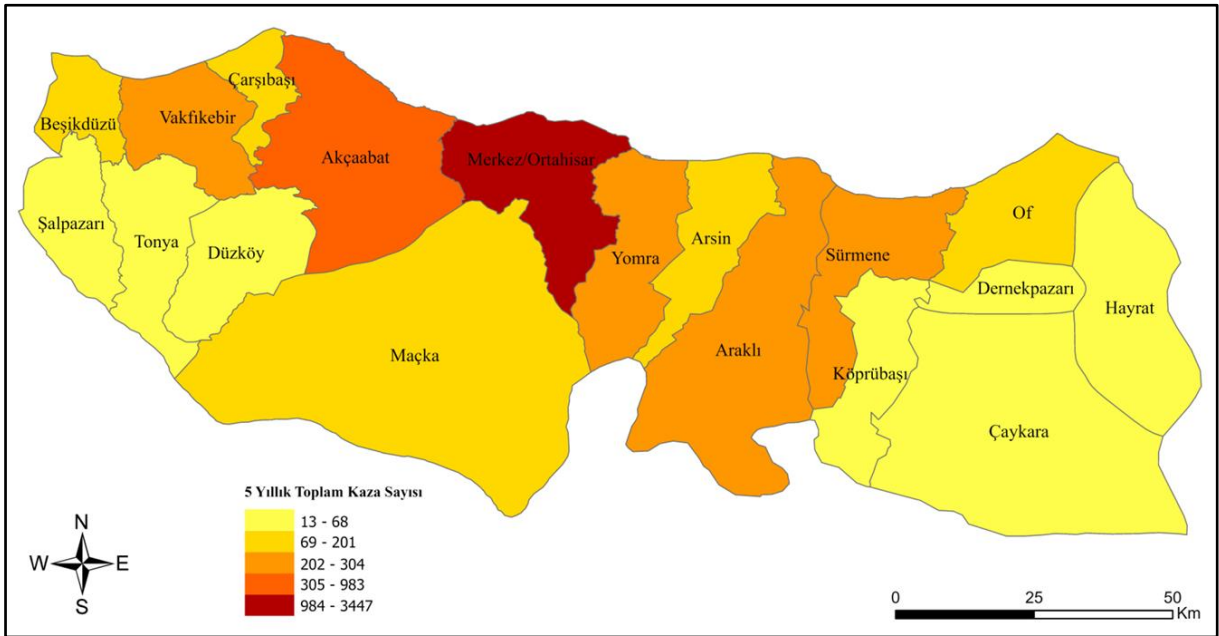
Genel olarak değerlendirildiğinde, Kernel yoğunluk analizi sonuçları trafik kazalarının rastgele dağılmadığını, aksine belirli mekânsal bölgelerde yoğunlaşarak sıcak noktalar oluşturduğunu göstermektedir. İl geneli ve ilçe ölçeğinde elde edilen bulgular birlikte değerlendirildiğinde, özellikle kıyı hattı, şehir merkezi ve ana ulaşım koridorlarının yüksek riskli bölgeler olduğu açıkça ortaya konulmuştur. Bu sonuçlar, trafik güvenliği açısından öncelikli müdahale alanlarının belirlenmesi ve planlama çalışmalarının yönlendirilmesi açısından önemli bir temel oluşturmaktadır.

#### 4.1.3. İlçe düzeyinde trafik kazası yoğunluklarının değerlendirilmesi

Trabzon ilinde meydana gelen trafik kazalarının ilçe bazında dağılımını incelemek amacıyla, toplam kaza sayıları kullanılarak yoğunluk haritası oluşturulmuştur. Bu harita, kazaların idari

sınırlar içerisinde nasıl dağıldığını ortaya koyarak mekânsal farklılıkların değerlendirilmesine imkân sağlamaktadır.

Trabzon iline ait ilçe bazlı kaza yoğunluk haritası Şekil 19’da sunulmaktadır. Harita incelendiğinde, ilçeler arasında belirgin farklılıklar olduğu ve kazaların homojen bir dağılım göstermediği açıkça görülmektedir. Özellikle Merkez/Ortahisar ilçesinin en yüksek kaza yoğunluğuna sahip olduğu ve en koyu renk ile temsil edildiği dikkat çekmektedir. Bu durum, Ortahisar’ın hem nüfus yoğunluğu hem de trafik hacmi açısından ilin en yoğun bölgesi olmasıyla doğrudan ilişkilidir. Ayrıca şehir merkezi olması nedeniyle ticari faaliyetler, ulaşım ağlarının kesişimi ve günlük hareketlilik bu bölgede kaza riskini artırmaktadır.



Şekil 19. Trabzon iline ait ilçe bazlı trafik kazası yoğunluk haritası

Ortahisar ilçesini takiben Akçaabat, Yomra, Araklı ve Sürmene ilçelerinde de kaza yoğunluğunun yüksek olduğu görülmektedir. Bu ilçelerin ortak özelliği, kıyı hattı boyunca uzanan ana ulaşım aksı üzerinde yer almalarıdır. Özellikle sahil yolu üzerindeki yoğun trafik akışı, bu bölgelerde kaza sayılarının artmasına neden olmaktadır. Bu durum, önceki noktasal ve Kernel yoğunluk analizleri ile de uyum göstermekte olup, kıyı hattının trafik kazaları açısından kritik bir bölge olduğunu doğrulamaktadır.

Buna karşılık, Çaykara, Hayrat, Dernekpazarı, Tonya, Düzköy ve Şalpazarı gibi iç kesimlerde yer alan ilçelerde kaza sayılarının daha düşük olduğu ve açık renk tonları ile temsil edildiği

görülmektedir. Bu ilçelerde nüfusun seyrek olması, trafik yoğunluğunun görece düşük olması, yerleşimlerin daha dağınık yapıda bulunması ve ulaşım ağlarının sınırlı olması kaza sayılarının azalmasında etkili faktörler olarak değerlendirilmektedir.

Maçka ilçesi ise orta düzeyde bir yoğunluk göstermekte olup, özellikle belirli ulaşım koridorları üzerinde yoğunlaşan kazalar bu sonucu açıklamaktadır. İlçede genel yayılım düşük olsa da ana yol güzergâhları üzerinde kaza sayılarının artış gösterdiği anlaşılmaktadır.

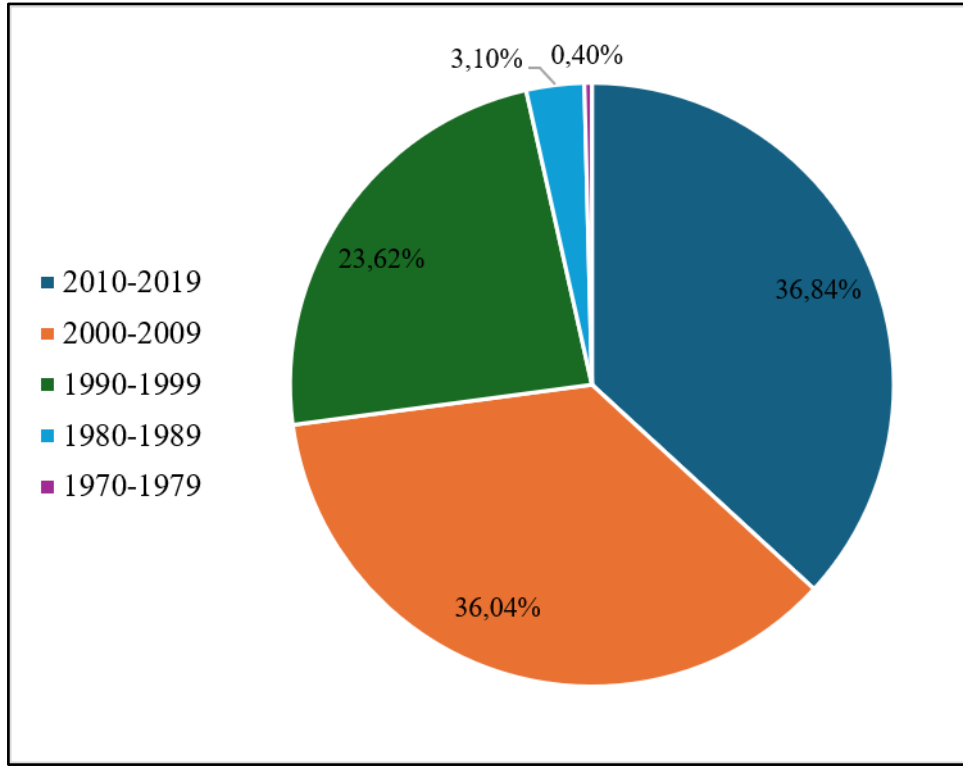
Genel olarak değerlendirildiğinde, ilçe bazlı analiz sonuçları trafik kazalarının mekânsal olarak belirli bölgelerde yoğunlaştığını ve özellikle kıyı hattı ile şehir merkezinin yüksek riskli alanlar olduğunu ortaya koymaktadır. Bu bulgular, noktasal ve Kernel yoğunluk analizleri ile birlikte değerlendirildiğinde, Trabzon ilinde trafik kazalarının mekânsal dağılımının büyük ölçüde ulaşım ağı, nüfus yoğunluğu ve kentsel yapı ile ilişkili olduğu sonucunu desteklemektedir.

## **4.2. Tanımlayıcı İstatistik Bulguları**

Bu bölümde veri setinde yer alan temel değişkenlerin dağılımları incelenmiş ve kazaların genel yapısına ilişkin bulgular ortaya konulmuştur. Tanımlayıcı istatistikler, veri setinin yapısını anlamak ve sonraki analizlerin yorumlanmasını desteklemek açısından önemli bir aşamadır. Bu kapsamda araç model yılı, yaya ve sürücü yaş grupları ile kaza şiddeti değişkenlerinin dağılımları grafikler aracılığıyla değerlendirilmiştir.

### **4.2.1. Araç model yılı gruplarının dağılımı**

Kazaya karışan araçların model yılına göre dağılımı, araçların teknolojik gelişim düzeyi ve güvenlik donanımları açısından önemli bilgiler sunmaktadır. Araç model yılı, dolaylı olarak araçların sahip olduğu aktif ve pasif güvenlik sistemlerini yansıttığı için kaza şiddeti üzerinde etkili olabilecek bir değişken olarak değerlendirilmektedir. Bu kapsamda Şekil 20'de araç model yılı gruplarının dağılımı verilmiştir.



Şekil 20. Kazaya karışan araçların model yılı gruplarına göre dağılımı.

Grafik incelendiğinde, kazaya karışan araçların büyük bir kısmının 2010–2019 (%36,8) ve 2000–2009 (%36,0) model yıllarına ait olduğu görülmektedir. Bu iki grubun toplamda yaklaşık %73'lük bir paya sahip olması, kazaların büyük ölçüde orta ve yeni sayılabilecek araç gruplarında yoğunlaştığını göstermektedir.

Buna karşılık 1990–1999 (%23,6) model yılına ait araçların da önemli bir orana sahip olduğu görülmektedir. Bu durum, eski model araçların halen trafikte aktif olarak kullanıldığını ve kazalara karışmaya devam ettiğini ortaya koymaktadır.

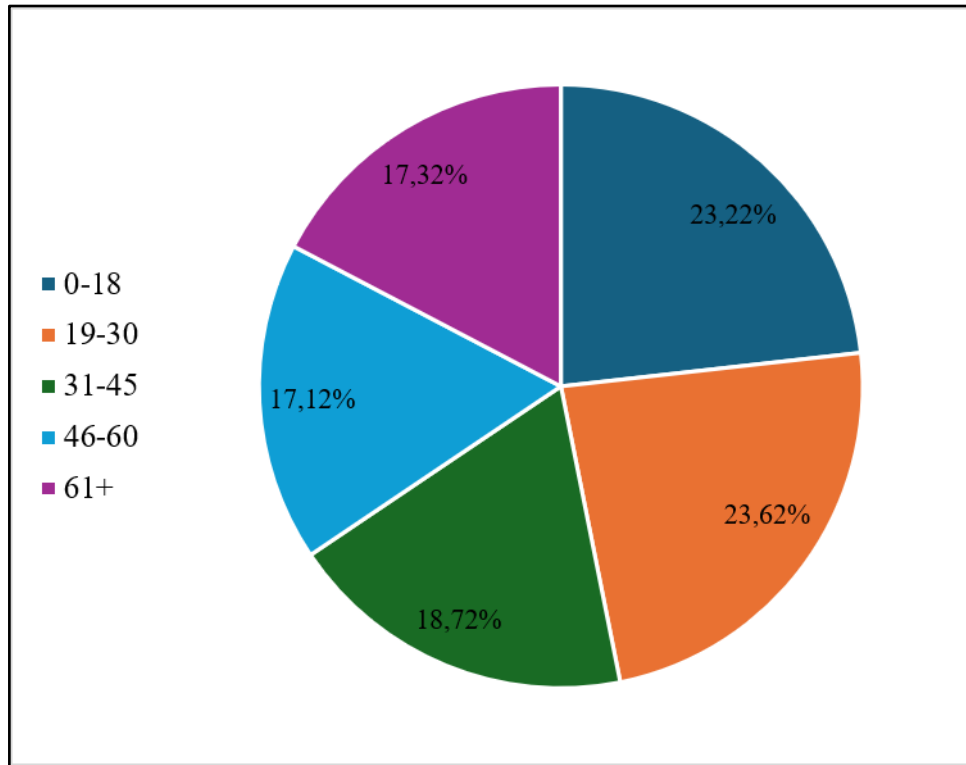
Daha eski araç grupları olan 1980–1989 (%3,1) ve 1970–1979 (%0,4) model yıllarına ait araçların ise oldukça düşük oranlarda temsil edildiği görülmektedir. Bu durum, bu araçların trafikteki sayısının sınırlı olması ile açıklanabilir.

Elde edilen bulgular, kazaların yalnızca eski model araçlarla sınırlı olmadığını, aksine daha yeni ve orta yaş grubundaki araçların da önemli ölçüde kazalara karıştığını göstermektedir. Bu durum, kaza oluşumunun yalnızca araç teknolojisi ile açıklanamayacağını, sürücü davranışları, trafik yoğunluğu ve çevresel faktörlerin de önemli rol oynadığını ortaya koymaktadır.

Bununla birlikte, eski model araçların güvenlik donanımlarının daha sınırlı olması nedeniyle bu araçların karıştığı kazalarda sonuçların daha ağır olabileceği değerlendirilmektedir. Bu nedenle araç model yılı değişkeni, kaza şiddetini açıklamada önemli bir faktör olarak ilerleyen modelleme aşamalarında dikkate alınmıştır.

#### 4.2.2. Yaya yaş gruplarının dağılımı

Kazaya karışan yayaların yaş gruplarına göre dağılımı, trafik kazalarında hangi yaş gruplarının daha fazla yer aldığını ortaya koyması açısından önemli bir göstergedir. Yaya yaşı hem trafik davranışları hem de fiziksel kırılganlık düzeyi ile doğrudan ilişkili bir değişken olarak değerlendirilmektedir. Bu kapsamda Şekil 21’de kazaya karışan yayaların yaş gruplarına göre dağılımı sunulmaktadır.



Şekil 21. Kazaya karışan yayaların yaş gruplarına göre dağılımı.

Grafik incelendiğinde, kazaya karışan yayaların en yüksek oranda 19–30 yaş (%23,6) ve 0–18 yaş (%23,2) gruplarında yoğunlaştığı görülmektedir. Bu iki grubun toplamda yaklaşık %47’lik bir paya sahip olması, genç ve çocuk yaştaki bireylerin kazalarda önemli bir yer tuttuğunu göstermektedir. Bu durum, özellikle genç bireylerin trafikte daha aktif olması ve çocukların trafik kurallarına uyum konusunda daha hassas bir grup olması ile açıklanabilir.

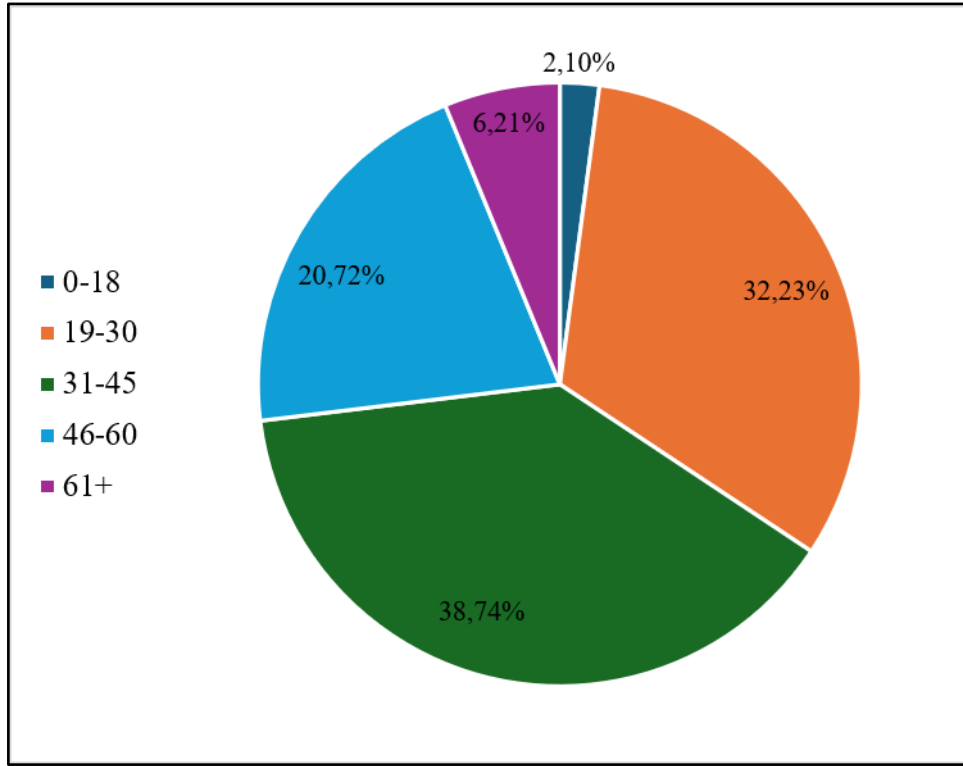
Bununla birlikte 31–45 yaş grubu (%18,7) ve 46–60 yaş grubu (%17,1) da önemli oranlara sahiptir. Bu yaş gruplarının kazalarda yer alması, günlük yaşamda trafikte aktif olarak bulunmaları ile ilişkilendirilebilir.

Dikkat çeken bir diğer bulgu ise 61 yaş ve üzeri bireylerin (%17,3) kazalardaki payıdır. Bu oran, ileri yaş grubunun kazalarda azımsanmayacak bir şekilde temsil edildiğini göstermektedir. Yaşlı bireylerin fiziksel hareket kabiliyetinin azalması, reaksiyon sürelerinin uzaması ve algısal sınırlılıkları, bu grubun trafik kazalarına karşı daha kırılgan olmasına neden olabilmektedir.

Genel olarak değerlendirildiğinde, kazaya karışan yayaların yaş dağılımının oldukça dengeli olduğu, ancak genç ve çocuk yaş grubunun biraz daha öne çıktığı görülmektedir. Bu durum, trafik güvenliği açısından özellikle gençler ve çocuklara yönelik bilinçlendirme çalışmalarının önemini ortaya koymaktadır. Ayrıca ileri yaş grubunun kırılgan yapısı da dikkate alınarak, yaya güvenliğine yönelik düzenlemelerin bu grupları koruyacak şekilde planlanması gerektiği değerlendirilmektedir.

#### **4.2.3. Sürücü yaş gruplarının dağılımı**

Kazaya karışan sürücülerin yaş gruplarına göre dağılımı, trafik kazalarının hangi yaş grubundaki sürücülerde daha yoğun gerçekleştiğini ortaya koymak açısından önemli bir göstergedir. Sürücü yaşı, deneyim düzeyi, risk alma eğilimi ve trafik davranışları ile doğrudan ilişkili bir değişken olarak değerlendirilmektedir. Bu kapsamda Şekil 22’de kazaya karışan sürücülerin yaş gruplarına göre dağılımı sunulmaktadır.



Şekil 22. Kazaya karışan sürücülerin yaş gruplarına göre dağılımı.

Grafik incelendiğinde, kazaya karışan sürücülerin büyük bir kısmının 31–45 yaş grubu (%38,7) ve 19–30 yaş grubu (%32,2) içerisinde yer aldığı görülmektedir. Bu iki grubun toplamda yaklaşık %71’lik bir paya sahip olması, kazaların büyük ölçüde aktif sürüş yaş aralığında gerçekleştiğini göstermektedir.

Özellikle 31–45 yaş grubunun en yüksek orana sahip olması, bu yaş grubunun hem trafikte daha fazla yer alması hem de araç kullanım süresinin yüksek olması ile açıklanabilir. Bu durum, kaza oluşumunun yalnızca deneyim eksikliği ile değil, aynı zamanda trafik maruziyeti ile de ilişkili olduğunu göstermektedir.

19–30 yaş grubunun yüksek oranı (%32,2) ise genç sürücülerde görülen risk alma eğilimi, hız yapma davranışı ve deneyim eksikliği gibi faktörlerle ilişkilendirilebilir. Bu yaş grubunun kazalarda yüksek oranda yer alması, trafik güvenliği açısından önemli bir risk grubu olduğunu ortaya koymaktadır.

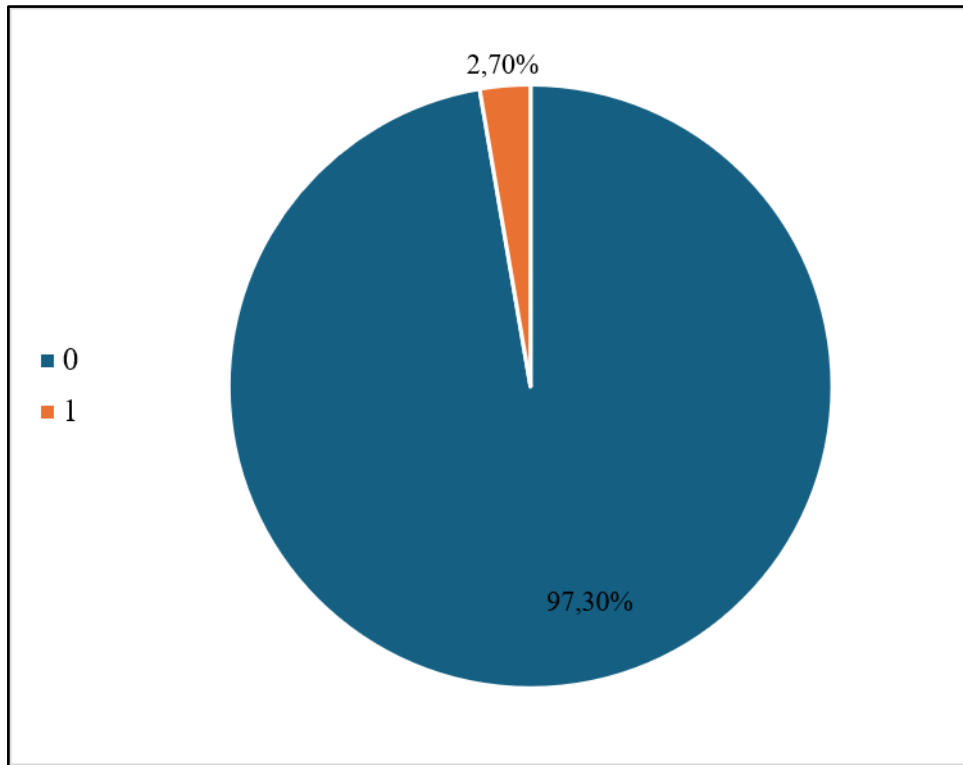
Buna karşılık 46–60 yaş grubu (%20,7) daha düşük bir oranla temsil edilirken, 61 yaş ve üzeri sürücüler (%6,2) oldukça sınırlı bir paya sahiptir. Bu durum, ileri yaş grubunun trafikte daha az

aktif olması ile açıklanabilir. Ayrıca 0–18 yaş grubunun (%2,1) çok düşük bir orana sahip olması, bu grubun yasal olarak sürücülük faaliyetinde bulunamaması ile uyumludur.

Genel olarak değerlendirildiğinde, kazaya karışan sürücülerin büyük ölçüde aktif çalışma çağındaki bireylerden oluştuğu görülmektedir. Bu durum, trafik güvenliği politikalarının özellikle genç ve orta yaş grubundaki sürücülere yönelik olarak geliştirilmesinin önemini ortaya koymaktadır.

#### 4.2.4. Kaza şiddeti dağılımı

Kaza şiddeti değişkeni, çalışmanın temel hedef değişkeni olup kazaların ölümlü ve ölümlü olmayan şeklinde sınıflandırılmasını ifade etmektedir. Bu çalışmada kaza şiddeti değişkeni ikili (binary) olarak tanımlanmış olup, 0 değeri ölümlü kazaları, 1 değeri ise ölümlü olmayan kazaları temsil etmektedir. Bu kapsamda Şekil 23'te kaza şiddetinin dağılımı gösterilmektedir.



Şekil 23. Kaza şiddeti sınıflarının dağılımı.

Grafik incelendiğinde, kazaların büyük çoğunluğunun ölümlü olmayan kazalardan (%97,3) oluştuğu, ölümlü kazaların ise oldukça düşük bir orana (%2,7) sahip olduğu görülmektedir.

Bu bulgu, trafik kazalarının doğası gereği ölümlü kazaların nadir gerçekleştiğini ortaya koymaktadır. Ancak her ne kadar oran olarak düşük olsa da ölümlü kazaların sonuçlarının çok daha ağır olması nedeniyle bu sınıfın analiz edilmesi kritik öneme sahiptir.

Ayrıca bu dağılım, veri setinde belirgin bir sınıf dengesizliği bulunduğunu göstermektedir. Bu durum, makine öğrenmesi modellerinin değerlendirilmesinde yalnızca doğruluk ölçütünün yeterli olmayacağını ortaya koymaktadır. Model, çoğunluk sınıfını doğru tahmin ederek yüksek doğruluk elde edebilirken, azınlık sınıfı olan ölümlü kazaları yeterince iyi tahmin edemeyebilir.

Bu nedenle çalışmada model performansının değerlendirilmesinde precision, recall, F1 skoru ve Cohen's Kappa gibi ölçütler de dikkate alınmıştır. Özellikle ölümlü kazaların doğru tespit edilmesi açısından recall metriği önemli bir gösterge olarak değerlendirilmiştir.

Genel olarak değerlendirildiğinde, kaza şiddeti dağılımı hem veri setinin yapısını ortaya koymakta hem de modelleme sürecinde kullanılacak yöntemlerin belirlenmesine doğrudan katkı sağlamaktadır.

#### **4.2.5. Kaza özelliklerinin il geneli ve ilçe bazında karşılaştırmalı analizi**

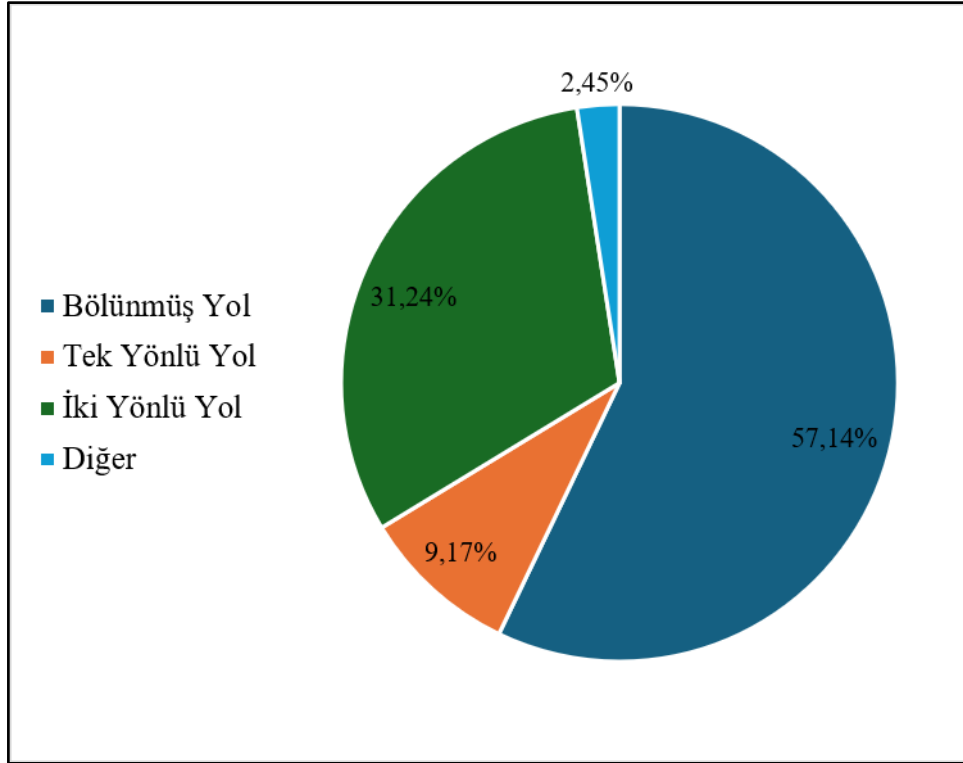
Trabzon ilinde meydana gelen trafik kazalarının yalnızca genel dağılımlarının incelenmesi, kazaların yapısını anlamak açısından yeterli olmamaktadır. Bu nedenle kazaların yol tipi, kavşak durumu ve gerçekleştiği zaman dilimi gibi temel özellikleri hem il geneli hem de ilçe bazında karşılaştırmalı olarak değerlendirilmiştir.

Bu yaklaşım, kazaların oluşumunda etkili olan faktörlerin mekânsal düzeyde nasıl farklılaştığını ortaya koymayı amaçlamaktadır. İl genelinde elde edilen bulgular, Trabzon'daki trafik kazalarına ilişkin genel eğilimleri yansıtırken, ilçe bazlı analizler bu eğilimlerin yerel koşullara bağlı olarak değişkenlik gösterebildiğini ortaya koymaktadır.

Özellikle trafik yoğunluğu, yol altyapısı, coğrafi yapı ve aydınlatma koşulları gibi faktörlerin ilçeler arasında farklılık göstermesi, kaza özelliklerinin de bu doğrultuda çeşitlenmesine neden olmaktadır. Bu bağlamda gerçekleştirilen karşılaştırmalı analiz, kazaların yalnızca sayısal dağılımlarını değil, aynı zamanda oluşum koşullarını da daha kapsamlı bir şekilde değerlendirme imkânı sunmaktadır.

#### 4.2.5.1. Yol tipine göre dağılım

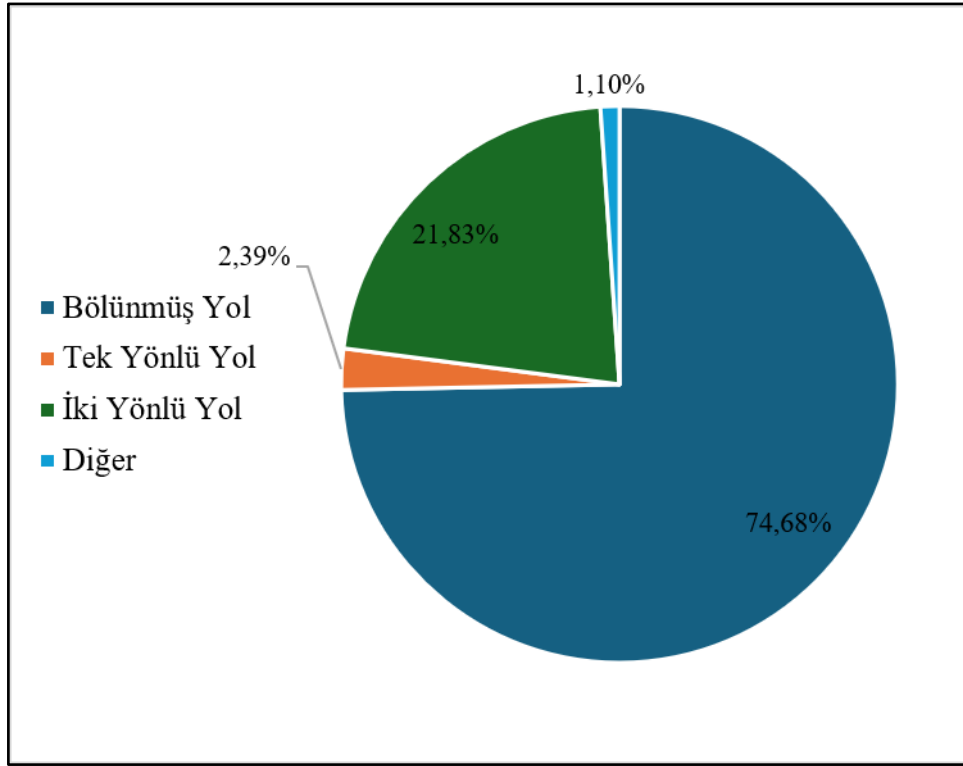
Trabzon il genelinde trafik kazalarının yol tipine göre dağılımı incelendiğinde, kazaların büyük bir kısmının bölünmüş yollarda (%57,14) meydana geldiği görülmektedir (Şekil 24). Bunu iki yönlü yollar (%31,24) ve tek yönlü yollar (%9,17) takip etmektedir. Diğer yol tiplerinin oranı ise oldukça düşük düzeyde (%2,45) kalmaktadır.



Şekil 24. Trabzon il genelinde yol tipine göre kaza dağılımı.

Bu durum, kazaların ağırlıklı olarak ana ulaşım arterlerinde yoğunlaştığını ve trafik yoğunluğu ile hızın kaza oluşumu üzerinde belirleyici bir rol oynadığını göstermektedir.

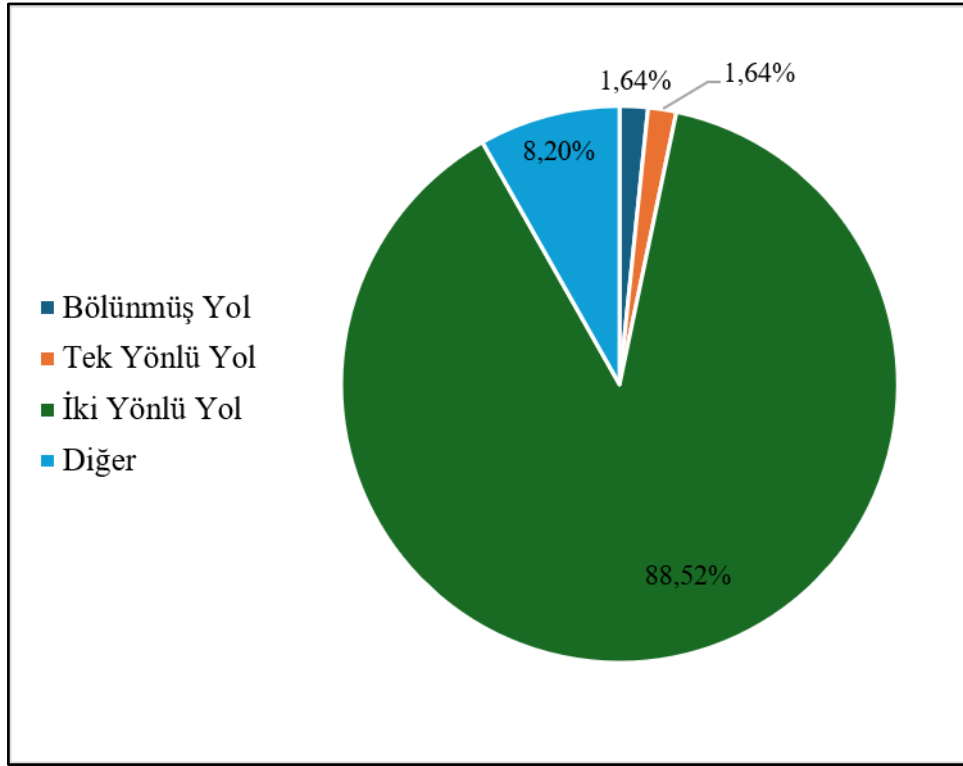
İlçe bazlı değerlendirmelerde ise bu genel dağılımın bazı ilçelerde belirgin farklılıklar gösterdiği görülmektedir. Örneğin, Arsin ilçesinde kazaların büyük ölçüde bölünmüş yollarda (%74,68) meydana geldiği dikkat çekmektedir (Şekil 25). Bu oran, il geneline kıyasla daha yüksek olup, ilçede ana ulaşım akslarının ve yüksek kapasiteli yolların etkisinin daha belirgin olduğunu göstermektedir.



Şekil 25. Arsin ilçesinde yol tipine göre kaza dağılımı.

Buna karşılık Şalpazarı ilçesinde yol tipi dağılımının farklı bir yapı sergilediği görülmektedir (Şekil 26). Bu ilçede kazaların büyük bir kısmı iki yönlü yollarda (%88,52) meydana gelirken, bölünmüş yolların oranı oldukça düşük seviyede (%1,64) kalmaktadır.

Bu durum, Şalpazarı ilçesinin kırsal yol yapısına sahip olması ve bölünmüş yol altyapısının sınırlı olması ile açıklanabilir. Aynı zamanda bu tür yollarda karşılıklı trafik akışının bulunması, çarpışma riskini artıran önemli bir faktör olarak değerlendirilmektedir.



Şekil 26. Şalpazarı ilçesinde yol tipine göre kaza dağılımı.

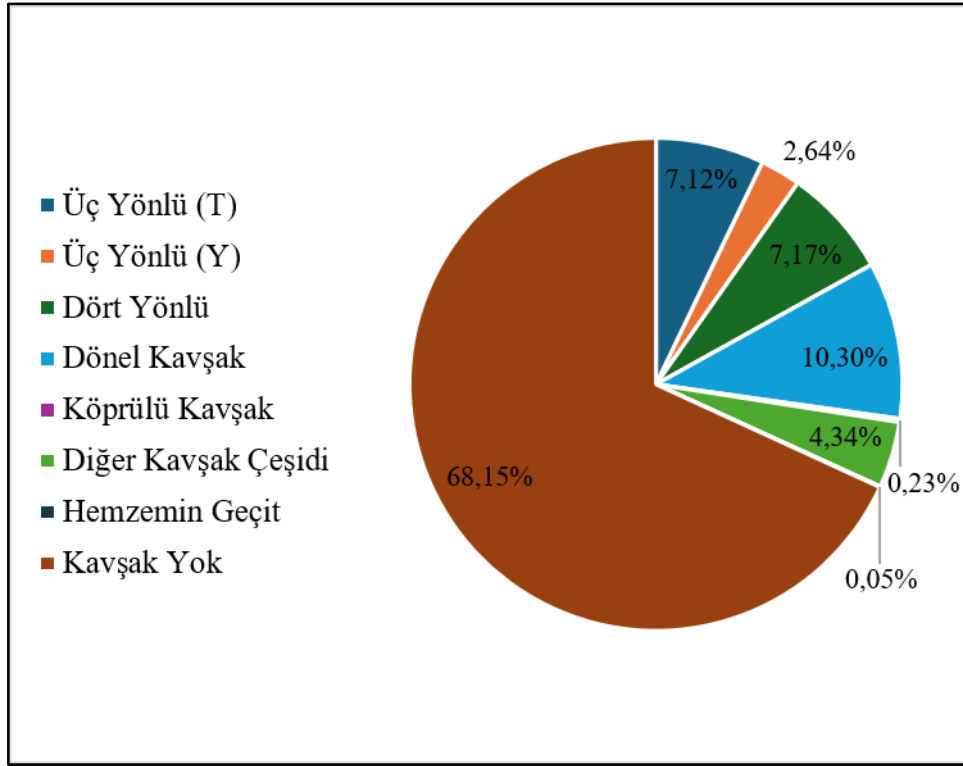
Genel olarak değerlendirildiğinde, yol tipinin trafik kazalarının oluşumunda önemli bir belirleyici olduğu ve bu etkinin ilçe bazında yol altyapısı ve trafik yapısına bağlı olarak önemli ölçüde değişkenlik gösterdiği anlaşılmaktadır.

#### 4.2.5.2. Kavşak durumuna göre dağılım

Trabzon il genelinde trafik kazalarının kavşak durumuna göre dağılımı incelendiğinde, kazaların büyük çoğunluğunun kavşak bulunmayan yol kesimlerinde (%68,14) meydana geldiği görülmektedir (Şekil 27).

Bunu sırasıyla dönel kavşaklar (%10,30), dört yönlü kavşaklar (%7,17) ve üç yönlü kavşaklar (%2,64) takip etmektedir. Diğer kavşak türlerinin oranı ise oldukça düşük düzeyde kalmaktadır.

Bu bulgu, trafik kazalarının yalnızca kavşak gibi karmaşık trafik noktalarında değil, aynı zamanda düz yol kesimlerinde de yoğunlaştığını göstermektedir. Kavşak bulunmayan kesimlerde sürücülerin hızlarını artırma eğiliminde olmaları ve dikkat düzeylerinin azalması, bu durumu açıklayan temel faktörler arasında yer almaktadır.



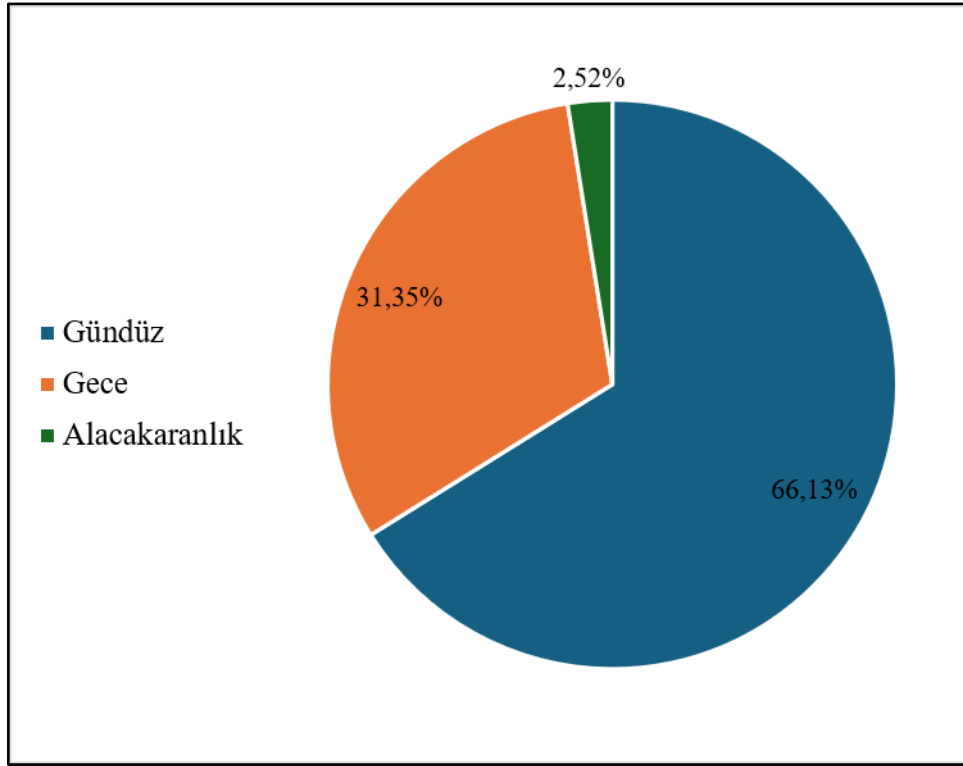
Şekil 27. Trabzon il genelinde kavşak durumuna göre kaza dağılımı.

İlçe bazlı değerlendirmelerde ise kavşak türlerine göre kaza dağılımının genel olarak benzer bir yapı sergilediği görülmektedir. Bu durum, kavşak etkisinin il genelinde homojen bir dağılım gösterdiğini ve belirli bir ilçeye özgü farklılaşmanın sınırlı olduğunu ortaya koymaktadır. Genel olarak değerlendirildiğinde, kavşak varlığından ziyade yol kesiminin yapısı ve sürücü davranışlarının trafik kazalarının oluşumunda daha belirleyici olduğu anlaşılmaktadır.

#### 4.2.5.3. Gündüz/gece durumuna göre dağılım

Trabzon il genelinde trafik kazalarının gerçekleştiği zaman dilimlerine göre dağılımı incelendiğinde, kazaların büyük çoğunluğunun gündüz saatlerinde (%66,13) meydana geldiği görülmektedir (Şekil 28). Gece saatlerinde gerçekleşen kazalar %31,35 oranına sahipken, alacakaranlık döneminin oranı ise %2,52 düzeyinde kalmaktadır.

Bu bulgu, trafik kazalarının büyük ölçüde trafik yoğunluğunun en yüksek olduğu zaman dilimlerinde gerçekleştiğini göstermektedir. Gündüz saatlerinde araç ve yaya hareketliliğinin artması, kaza sayısının bu zaman diliminde yoğunlaşmasına neden olmaktadır.



Şekil 28. Trabzon il genelinde gündüz/gece durumuna göre kaza dağılımı.

Buna karşılık gece saatlerinde trafik yoğunluğu daha düşük olmasına rağmen kazaların önemli bir kısmının bu zaman diliminde gerçekleşmesi dikkat çekicidir. Bu durum, gece koşullarında görüş mesafesinin azalması, sürücü yorgunluğu ve dikkat dağınıklığı gibi faktörlerin kaza oluşumunda etkili olduğunu göstermektedir.

Alacakaranlık döneminin düşük oranda temsil edilmesi ise bu zaman diliminin gün içerisindeki süresinin sınırlı olması ile ilişkilendirilebilir.

Genel olarak değerlendirildiğinde, trafik yoğunluğu kazaların sayısını artırırken, gece koşulları kaza riskini farklı bir boyutta etkileyen önemli bir faktör olarak öne çıkmaktadır.

### 4.3. Makine Öğrenmesi Modellerinin Sonuçları

Bu bölümde, lojistik regresyon modeli dışında kullanılan makine öğrenmesi yöntemlerinin performansları değerlendirilmiş ve elde edilen bulgular karşılaştırmalı olarak sunulmuştur. Karar ağacı, rastgele orman, AdaBoost ve Gradient Boosting modelleri kullanılarak kaza şiddeti sınıflandırılmış ve her bir modelin doğruluk, genellenebilirlik ve yorumlanabilirlik açısından performansı incelenmiştir. Elde edilen sonuçlar, farklı algoritmaların kaza verisi üzerindeki

öğrenme davranışlarını ortaya koymakta ve model seçiminde önemli bir karşılaştırma zemini sunmaktadır.

#### 4.3.1. Karar ağacı sonuçları

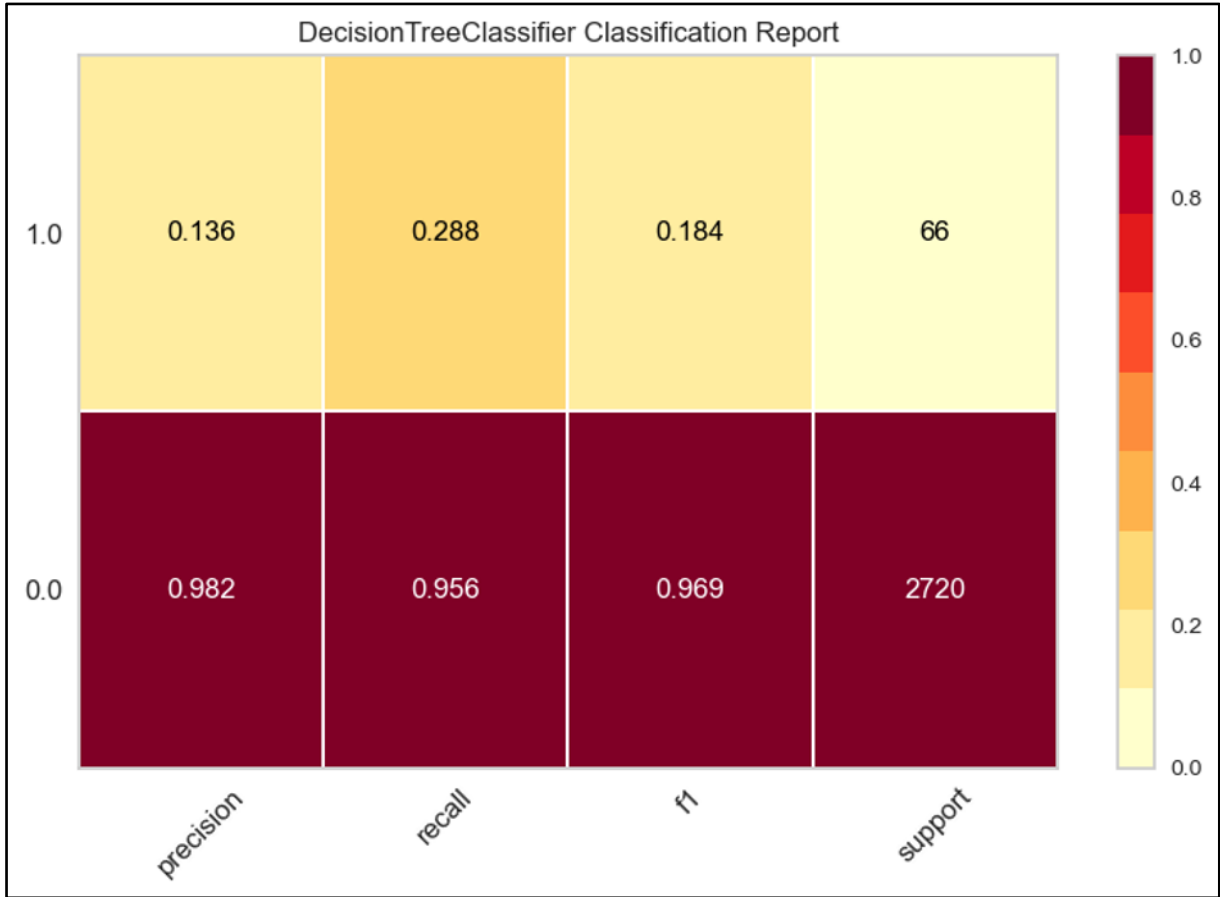
Bu çalışmada trafik kazalarının şiddet düzeyinin tahmin edilmesinde Karar Ağacı (Decision Tree) algoritması kullanılmıştır. Karar ağaçları, verileri belirli kurallar doğrultusunda ardışık olarak bölerek sınıflandırma gerçekleştiren ve elde edilen sonuçların yorumlanabilirliğinin yüksek olması nedeniyle trafik güvenliği çalışmalarında yaygın olarak tercih edilen makine öğrenmesi yöntemlerinden biridir. Model oluşturulurken aşırı öğrenmenin (overfitting) önüne geçebilmek amacıyla ağacın maksimum derinliği 11 olarak belirlenmiş, düğüm bölünmeleri için minimum örnek sayısı 19 ve yaprak düğümlerde bulunması gereken minimum örnek sayısı ise 14 olarak tanımlanmıştır. Ayrıca yalnızca anlamlı bilgi kazancı sağlayan bölünmelerin gerçekleştirilmesi amacıyla minimum safsızlık azalımı parametresi kullanılmıştır.

Model performansının değerlendirilmesi sonucunda doğruluk (Accuracy) değeri %93,97, Precision değeri %13,57, Recall değeri %28,79, F1 skoru %18,45 ve Cohen's Kappa katsayısı 0,1573 olarak hesaplanmıştır. Elde edilen doğruluk oranı oldukça yüksek olmakla birlikte veri setinde bulunan sınıf dengesizliği nedeniyle bu değer tek başına değerlendirilmesi yeterli değildir. Çünkü veri setindeki gözlemlerin büyük çoğunluğunu ölümlü olmayan kazalar oluşturmaktadır. Bu nedenle model performansının yorumlanmasında Precision, Recall ve F1 skorları daha anlamlı sonuçlar sunmaktadır.

Modelin genellenebilirliğini değerlendirmek amacıyla eğitim ve test performansları karşılaştırılmıştır. Eğitim ve test F1 skorları arasındaki farkın 0,1226 olduğu belirlenmiştir. Bu sonuç modelin eğitim verisini test verisine göre daha başarılı öğrendiğini ve belirli düzeyde aşırı öğrenme eğilimi gösterdiğini ortaya koymaktadır. Karar ağaçlarının yapısal olarak aşırı öğrenmeye yatkın modeller olması nedeniyle bu durum beklenen bir sonuç olarak değerlendirilebilir. Bununla birlikte model üzerinde budama işlemleri uygulanmasına rağmen aşırı öğrenme eğiliminin tamamen ortadan kaldırılamadığı görülmüştür.

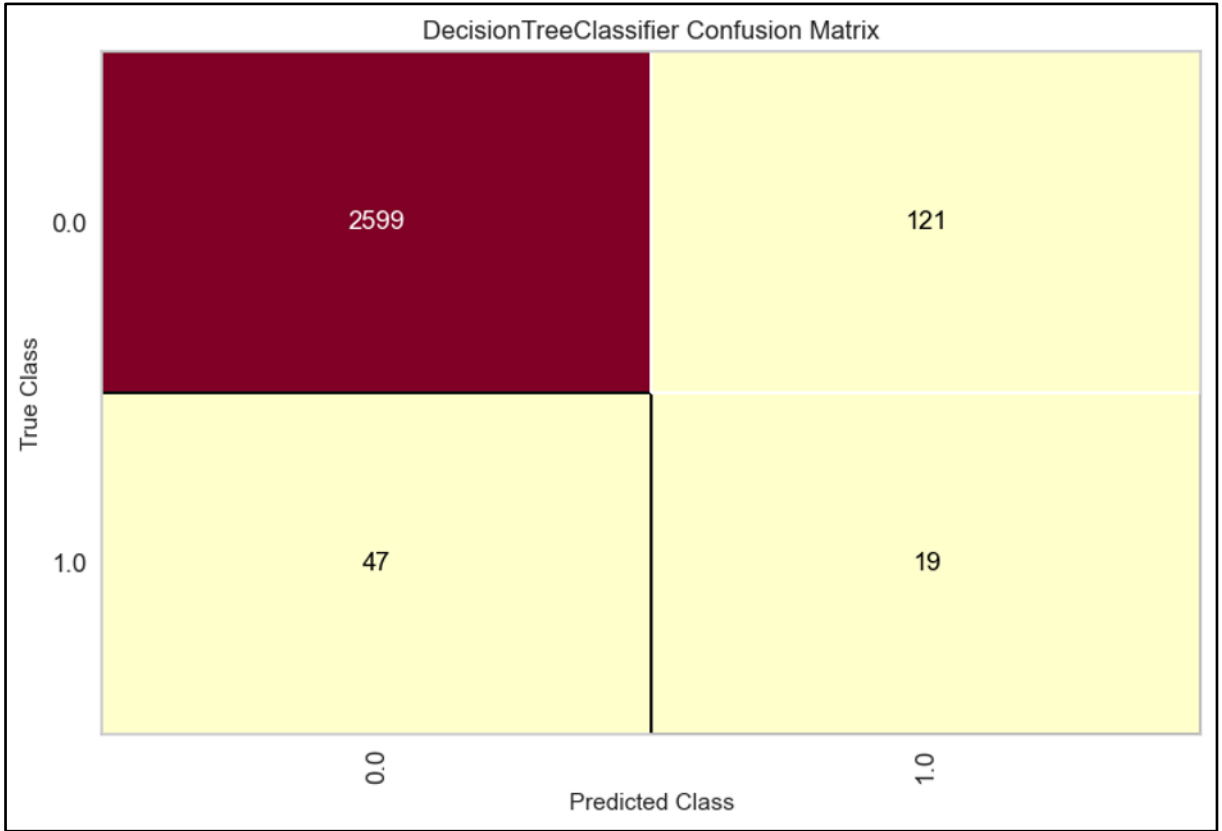
Beş katlı Stratified Cross Validation analizi sonucunda sırasıyla 0,973089, 0,973613, 0,973613, 0,973613 ve 0,973075 doğruluk skorları elde edilmiştir. Ortalama çapraz doğrulama başarısı 0,973401 olarak hesaplanırken standart sapma değeri 0,000260 olarak bulunmuştur. Standart

sapma deęerinin oldukça düşük olması model performansının farklı veri alt kümelerinde kararlı sonuçlar verdiğini göstermektedir. Ancak çapraz doğrulama doğruluğunun yüksek olmasına rağmen Recall ve F1 skorlarının düşük seviyelerde kalması, veri setindeki sınıf dengesizliğinin doğruluk metriğini yükselttiğini göstermektedir.



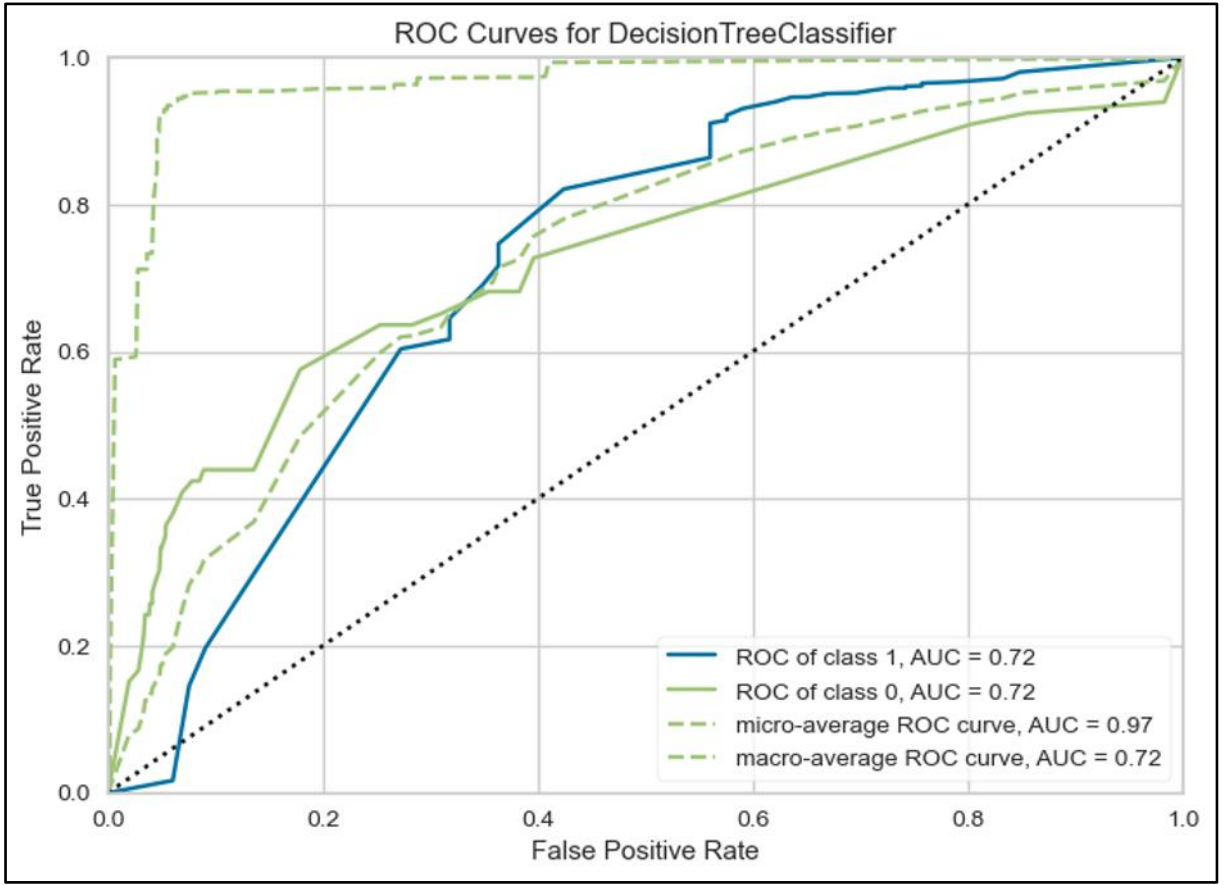
Şekil 29. Karar ağacı modeli sınıflandırma raporu

Şekil 29’da Karar Ağacı modeline ait sınıflandırma raporu sunulmaktadır. Sınıf bazında elde edilen sonuçlar incelendiğinde yaralanmalı kazalar için Precision, Recall ve F1 skorlarının sırasıyla 0,982, 0,956 ve 0,969 olduğu görülmektedir. Buna karşılık ölümlü kazalar için Precision değeri 0,136, Recall değeri 0,288 ve F1 skoru 0,184 olarak hesaplanmıştır. Bu sonuçlar modelin çoğunluk sınıfını oluşturan yaralanmalı kazaları oldukça başarılı şekilde sınıflandırabildiğini göstermektedir. Ancak ölümlü kazaların belirlenmesinde performansın daha düşük olduğu görülmektedir



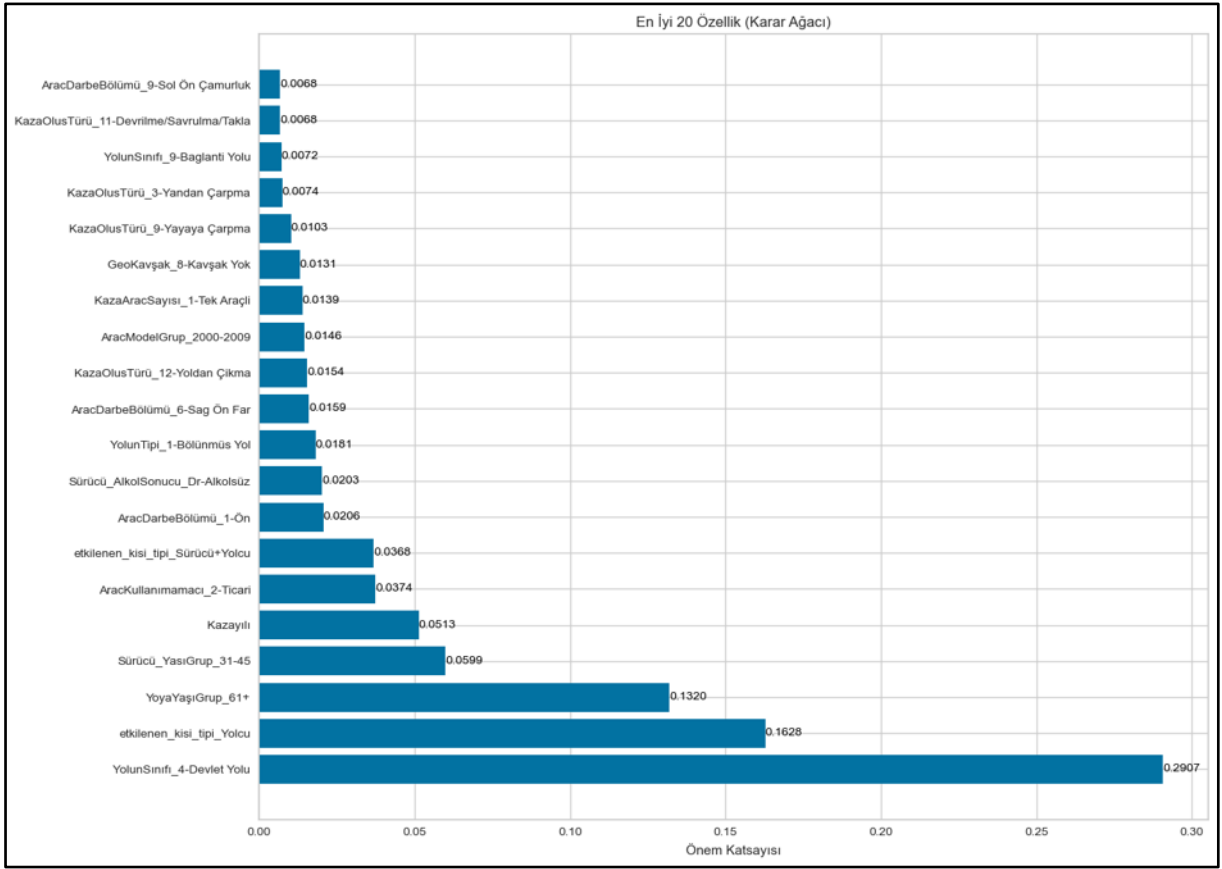
Şekil 30. Karar ağacı modeli karışıklık matrisi

Şekil 30'da Karar Ağacı modeline ait karışıklık matrisi gösterilmektedir. Karışıklık matrisi incelendiğinde toplam 2599 yaralanmalı kazanın doğru şekilde sınıflandırıldığı, 121 yaralanmalı kazanın yanlışlıkla ölümlü olarak tahmin edildiği görülmektedir. Ölümlü kazalar açısından değerlendirildiğinde ise toplam 66 ölümlü kazanın yalnızca 19'unun doğru şekilde tahmin edildiği, 47 ölümlü kazanın ise yaralanmalı olarak sınıflandırıldığı belirlenmiştir. Sonuçlar modelin yanlış pozitif üretimini önemli ölçüde azalttığını göstermektedir. Ancak bunun karşılığında gerçek ölümlü kazaların önemli bir kısmını gözden kaçırdığı görülmektedir. Bu durum modelin daha muhafazakâr bir sınıflandırma yaklaşımı sergilediğini ortaya koymaktadır.



Şekil 31. Karar ağacı modeli ROC eğrisi ve AUC değeri

Şekil 31’de Karar Ağacı modeline ait ROC eğrileri ve AUC değerleri sunulmaktadır. ROC eğrisi incelendiğinde modelin düşük yanlış pozitif oranlarında ölümlü kazaları ayırt etmekte zorlandığı görülmektedir. Sınıf 0 ve sınıf 1 için elde edilen AUC değerleri 0,72 olarak hesaplanmıştır. Ayrıca Macro Average AUC değeri 0,72 ve Micro Average AUC değeri 0,97 olarak bulunmuştur. Macro Average AUC değerinin 0,72 olması modelin sınıfları ayırt etme performansının orta düzeyde olduğunu göstermektedir. Buna karşılık Micro Average AUC değerinin oldukça yüksek olması veri setindeki çoğunluk sınıfının baskın etkisinden kaynaklanmaktadır. Bu nedenle model performansının değerlendirilmesinde Macro Average AUC değerinin daha anlamlı olduğu söylenebilir.



Şekil 32. Karar ağacı modeli için özellik önem düzeyleri

Şekil 32’de Karar Ağacı modelinde kullanılan değişkenlere ait önem düzeyleri gösterilmektedir. Sonuçlar incelendiğinde en yüksek önem değerine sahip değişkenin 0,2907 değeri ile “Yolun Sınıfı – Devlet Yolu” olduğu görülmektedir. Bu değişken modeldeki toplam açıklayıcılığın yaklaşık %29’unu oluşturmaktadır. İkinci sırada 0,1628 önem değeri ile “Etkilenen Kişi Tipi – Yolcu”, üçüncü sırada ise 0,1320 önem değeri ile “Yaya Yaş Grubu – 61+” değişkeni yer almaktadır. Bunun yanında sürücü yaş grubu, araç kullanım amacı, araç darbe bölümü ve sürücü alkol durumu gibi değişkenlerin de model performansı üzerinde etkili olduğu belirlenmiştir. Özellikle devlet yollarında meydana gelen kazalar ile ileri yaş grubundaki yayaların ölüm riski açısından daha kritik faktörler olduğu görülmektedir.

Genel olarak değerlendirildiğinde Karar Ağacı modeli %93,97 doğruluk oranı ile çalışmada kullanılan modeller arasında en yüksek doğruluk performansına ulaşmıştır. Ayrıca Precision, F1 skoru ve Cohen’s Kappa değerleri bakımından diğer temel sınıflandırma modellerine göre daha başarılı sonuçlar vermiştir. Buna karşın Recall değerinin düşük seviyede kalması modelin ölümlü kazaların önemli bir bölümünü tespit edemediğini göstermektedir. Sonuç olarak Karar Ağacı modeli yüksek yorumlanabilirliği ve değişken önemlerini ortaya koyabilme yeteneği

sayesinde trafik kazalarının şiddetini etkileyen faktörlerin belirlenmesinde önemli avantajlar sunmaktadır. Bununla birlikte ölümlü kazaların tahmin performansının artırılması amacıyla daha gelişmiş topluluk öğrenmesi yöntemleri ile desteklenmesi önerilmektedir.

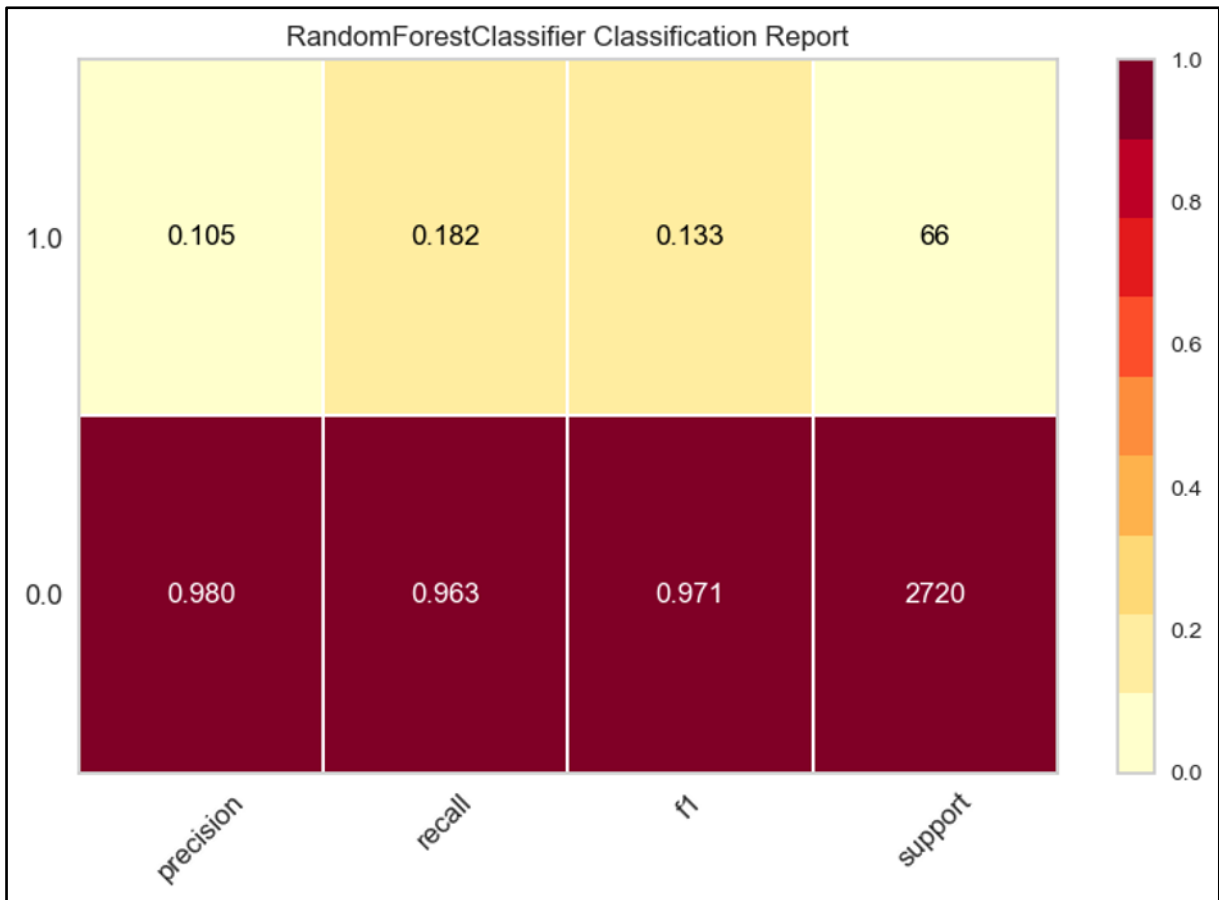
#### 4.3.2. Rastgele orman sonuçları

Bu çalışmada trafik kazalarının şiddet düzeyinin tahmin edilmesi amacıyla Rastgele Orman (Random Forest) algoritması kullanılmıştır. Rastgele Orman algoritması, birden fazla karar ağacının bir araya getirilmesiyle oluşturulan topluluk öğrenmesi (ensemble learning) yöntemlerinden biridir. Algoritma, farklı veri örnekleri üzerinde eğitilen karar ağaçlarının sonuçlarını birleştirerek nihai sınıflandırmayı gerçekleştirmektedir. Bu yaklaşım sayesinde tek bir karar ağacında görülebilen yüksek varyans azaltılmakta ve modelin genellenebilirliğinin artırılması amaçlanmaktadır. Model oluşturulurken maksimum ağaç derinliği 35 olarak belirlenmiş, minimum bölünme örnek sayısı ve yaprak düğümlerde bulunması gereken minimum örnek sayısı 35 olarak tanımlanmıştır. Ayrıca model içerisinde 5 adet karar ağacı kullanılmıştır.

Model performansının değerlendirilmesi sonucunda doğruluk (Accuracy) değeri %94,40, Precision değeri %10,53, Recall değeri %18,18, F1 skoru %13,33 ve Cohen's Kappa katsayısı 0,1065 olarak hesaplanmıştır. Doğruluk oranının yüksek olması modelin genel sınıflandırma başarısının yüksek olduğunu göstermektedir. Ancak veri setinde bulunan sınıf dengesizliği nedeniyle doğruluk metriğinin tek başına değerlendirilmesi yanıltıcı olabilmektedir. Bu nedenle model performansının yorumlanmasında Precision, Recall ve F1 skorları daha anlamlı ölçütler olarak dikkate alınmıştır.

Modelin genellenebilirliğini değerlendirmek amacıyla eğitim ve test performansları karşılaştırılmıştır. Eğitim ve test F1 skorları arasındaki farkın 0,1451 olduğu belirlenmiştir. Bu değer, kabul edilen 0,10 eşik değerinin üzerinde olduğundan modelin belirli düzeyde aşırı öğrenme eğilimi gösterdiğini ortaya koymaktadır. Rastgele Orman algoritmaları genellikle tek karar ağaçlarına göre aşırı öğrenmeye daha dayanıklı modeller olmasına rağmen, elde edilen sonuçlar modelin eğitim verisindeki örüntüleri test verisine kıyasla daha başarılı öğrendiğini göstermektedir. Bu durum modelin yeni veriler üzerinde aynı başarıyı sergileyemeyeceğini düşündürmektedir.

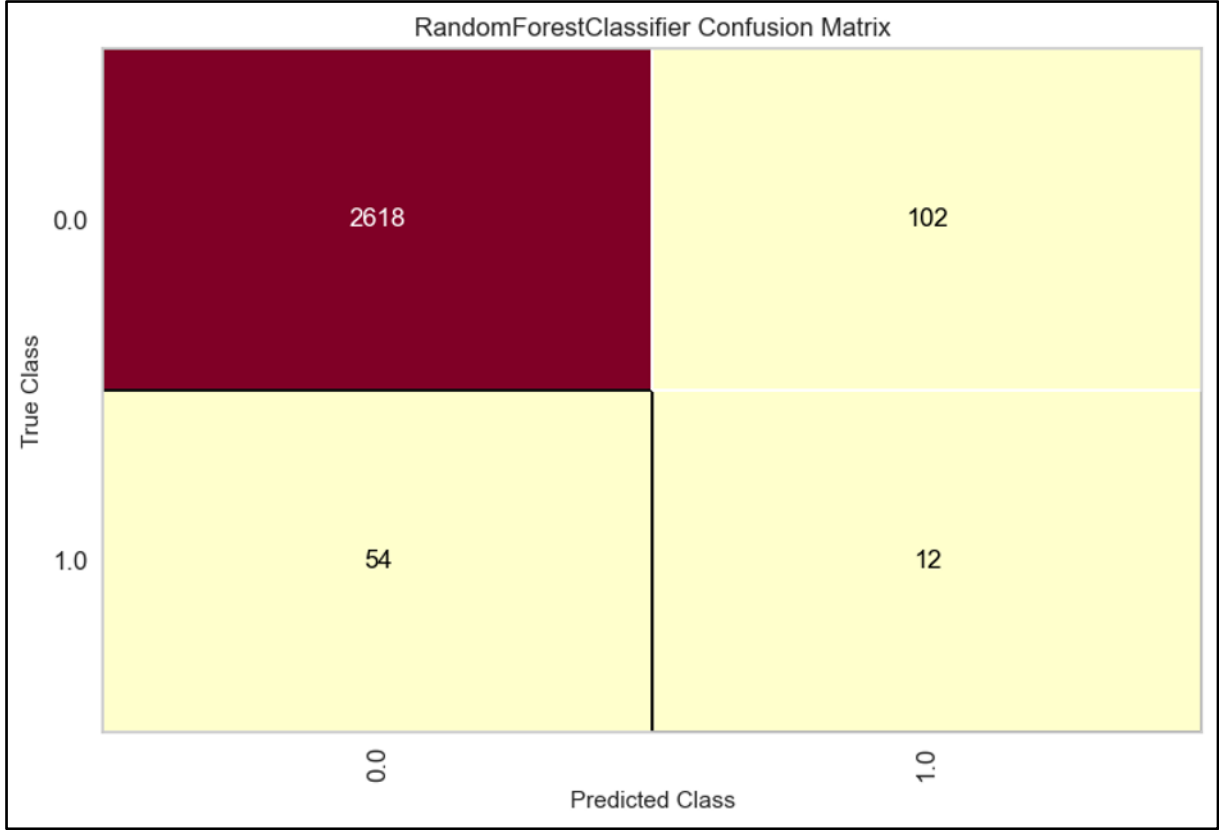
Modelin farklı veri alt kümelerindeki performans kararlılığını değerlendirmek amacıyla beş katlı Stratified Cross Validation uygulanmıştır. Analiz sonucunda sırasıyla 0,973089, 0,973613, 0,973613, 0,973613 ve 0,973075 doğruluk skorları elde edilmiştir. Ortalama çapraz doğrulama başarısı 0,973401 olarak hesaplanırken standart sapma değeri 0,000260 olarak bulunmuştur. Katlar arasında elde edilen sonuçların birbirine oldukça yakın olması model performansının kararlı olduğunu göstermektedir. Bununla birlikte çapraz doğrulama doğruluğunun yüksek olmasına rağmen Recall ve F1 skorlarının düşük seviyelerde kalması, veri setindeki sınıf dengesizliğinin doğruluk metriğini yükselttiğini göstermektedir.



Şekil 33. Rastgele orman modeli sınıflandırma raporu

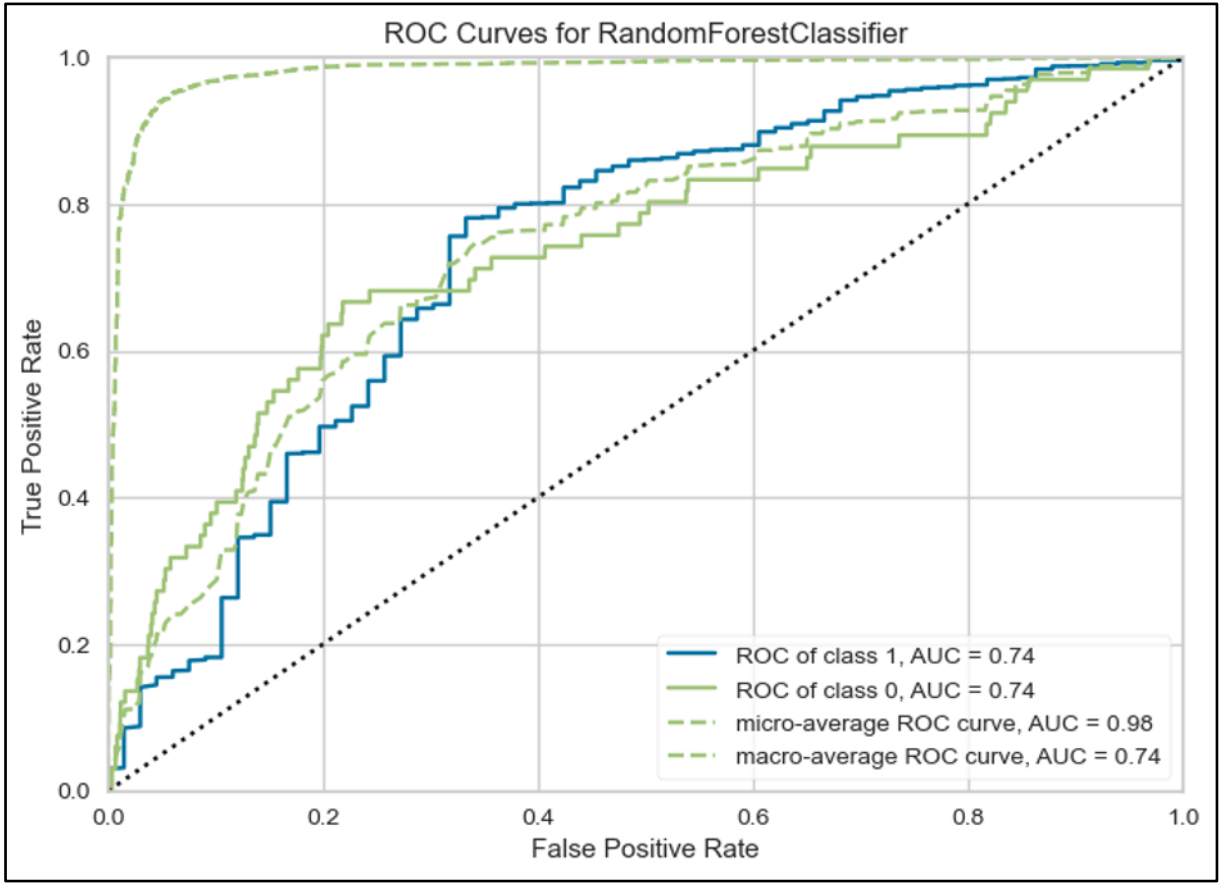
Şekil 33'te Rastgele Orman modeline ait sınıflandırma raporu sunulmaktadır. Sınıf bazında elde edilen sonuçlar incelendiğinde yaralanmalı kazalar için Precision, Recall ve F1 skorlarının sırasıyla 0,980, 0,963 ve 0,971 olduğu görülmektedir. Buna karşılık ölümlü kazalar için Precision değeri 0,105, Recall değeri 0,182 ve F1 skoru 0,133 olarak hesaplanmıştır. Sonuçlar modelin yaralanmalı kazaları oldukça başarılı şekilde sınıflandırabildiğini göstermektedir. Ancak ölümlü kazaların belirlenmesinde performansın belirgin şekilde düştüğü görülmektedir.

Özellikle Recall değerinin düşük olması, modelin gerçek ölümlü kazaların büyük bölümünü tespit edemediğini göstermektedir.



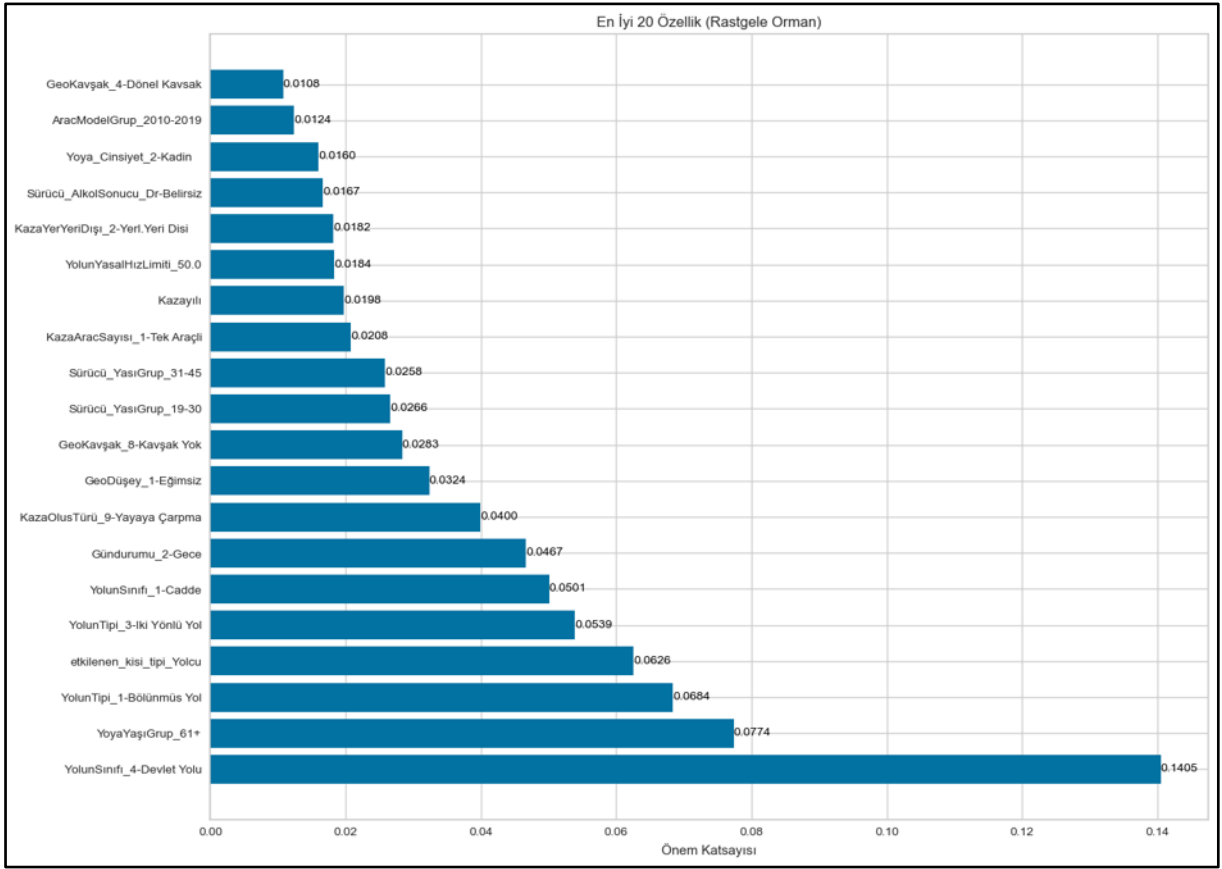
Şekil 34. Rastgele orman modeli karışıklık matrisi

Şekil 34'te Rastgele Orman modeline ait karışıklık matrisi gösterilmektedir. Karışıklık matrisi incelendiğinde toplam 2618 yaralanmalı kazanın doğru şekilde sınıflandırıldığı, 102 yaralanmalı kazanın yanlışlıkla ölümlü olarak tahmin edildiği görülmektedir. Ölümlü kazalar açısından değerlendirildiğinde ise toplam 66 ölümlü kazanın yalnızca 12'sinin doğru şekilde tahmin edildiği, 54 ölümlü kazanın ise yaralanmalı olarak sınıflandırıldığı belirlenmiştir. Bu sonuçlar modelin yanlış pozitif üretiminde düşük seviyelerde kaldığını göstermektedir. Ancak bunun karşılığında ölümcül kazaların büyük bölümünün gözden kaçırıldığı görülmektedir. Karar Ağacı modeli ile karşılaştırıldığında, Rastgele Orman modelinin ölümlü kazaları tespit etme başarısının daha düşük olduğu anlaşılmaktadır.



Şekil 35. Rastgele orman modeli ROC eğrisi ve AUC değeri

Şekil 35'te Rastgele Orman modeline ait ROC eğrileri ve AUC değerleri sunulmaktadır. ROC eğrisi incelendiğinde eğrinin genel olarak referans doğrusunun üzerinde yer aldığı ve dışbükey bir yapı oluşturduğu görülmektedir. Bu durum modelin rastgele tahminden daha başarılı olduğunu göstermektedir. Sınıf 0 ve sınıf 1 için elde edilen AUC değerleri 0,74 olarak hesaplanmıştır. Ayrıca Macro Average AUC değeri 0,74 ve Micro Average AUC değeri 0,98 olarak bulunmuştur. Macro Average AUC değerinin 0,74 olması modelin sınıfları ayırt etme performansının kabul edilebilir düzeyde olduğunu göstermektedir. Buna karşılık oldukça yüksek olan Micro Average AUC değeri veri setindeki çoğunluk sınıfının etkisinden kaynaklanmaktadır. Bu nedenle model performansının değerlendirilmesinde Macro Average AUC değerinin daha anlamlı olduğu söylenebilir.



Şekil 36. Rastgele orman modeli için özellik önem düzeyleri

Şekil 36’da Rastgele Orman modelinde kullanılan değişkenlere ait önem düzeyleri sunulmaktadır. Sonuçlar incelendiğinde en yüksek önem değerine sahip değişkenin 0,1405 değeri ile “Yolun Sınıfı – Devlet Yolu” olduğu görülmektedir. Bunun yanında “Yaya Yaş Grubu – 61+”, “Yolun Tipi – Bölünmüş Yol”, “Etkilenen Kişi Tipi – Yolcu” ve “Yolun Tipi – İki Yönlü Yol” değişkenleri de model üzerinde önemli etkiye sahip faktörler arasında yer almaktadır. Ayrıca gece meydana gelen kazalar, yayaya çarpma türündeki kazalar ve yerleşim yeri dışında gerçekleşen kazaların da ölüm riski üzerinde etkili olduğu belirlenmiştir. Elde edilen bulgular, diğer makine öğrenmesi modellerinden elde edilen sonuçlarla büyük ölçüde benzerlik göstermektedir.

Genel olarak değerlendirildiğinde Rastgele Orman modeli %94,40 doğruluk oranı ile yüksek bir genel başarı göstermiştir. Bununla birlikte Recall (%18,18) ve F1 (%13,33) değerlerinin düşük olması modelin ölümcül kazaların önemli bir bölümünü tespit edemediğini ortaya koymaktadır. AUC değerinin 0,74 olması ise modelin sınıfları ayırt etme kapasitesinin tamamen başarısız olmadığını göstermektedir. Sonuç olarak Rastgele Orman modeli, ölümcül

kazaların tahmini açısından beklenen performansı sağlayamamış olsa da trafik kazalarının şiddetini etkileyen risk faktörlerinin belirlenmesi ve yorumlanması açısından önemli bilgiler sunmaktadır.

### 4.3.3. Adaboost sonuçları

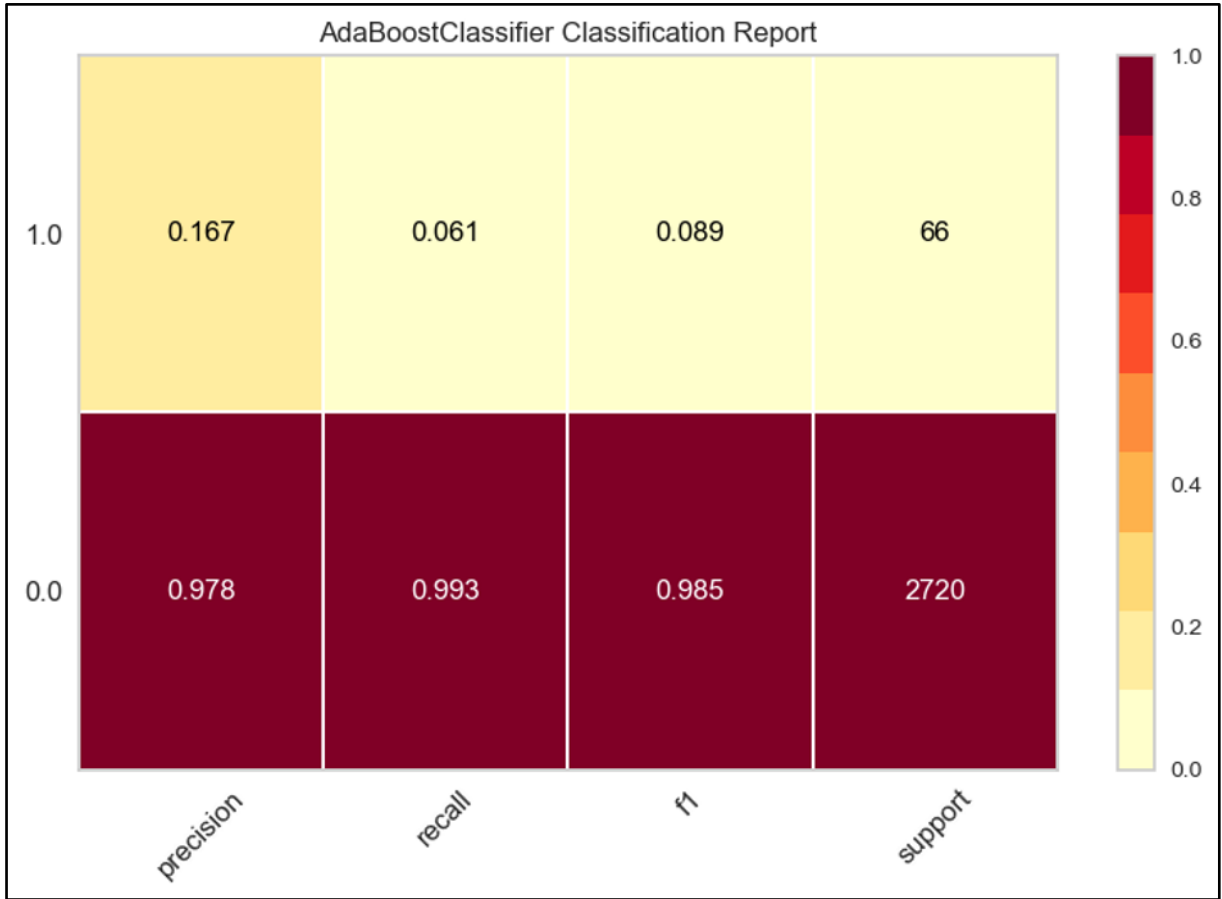
Bu çalışmada trafik kazalarının şiddet düzeyinin tahmin edilmesi amacıyla AdaBoost (Adaptive Boosting) algoritması kullanılmıştır. AdaBoost, zayıf öğrenicilerin ardışık olarak eğitilmesi prensibine dayanan topluluk öğrenmesi yöntemlerinden biridir. Algoritmanın temel amacı, önceki aşamalarda yanlış sınıflandırılan gözlemlere daha yüksek ağırlık vererek sonraki modellerin bu örnekleri daha iyi öğrenmesini sağlamaktır. Bu çalışmada temel öğrenici olarak maksimum derinliği 2 olan karar ağaçları kullanılmış, model 120 adet karar ağacı ile oluşturulmuş ve öğrenme oranı 0,25 olarak belirlenmiştir. Ayrıca daha kararlı sonuçlar elde edebilmek amacıyla SAMME algoritması tercih edilmiştir.

Model performansının değerlendirilmesi sonucunda doğruluk (Accuracy) değeri %97,06, Precision değeri %16,67, Recall değeri %6,06, F1 skoru %8,89 ve Cohen's Kappa katsayısı 0,0772 olarak hesaplanmıştır. Elde edilen doğruluk oranı çalışmada kullanılan modeller arasında en yüksek değerlerden biri olmasına rağmen veri setindeki sınıf dengesizliği nedeniyle tek başına yeterli bir performans göstergesi değildir. Precision değerinin %16,67 olması, model tarafından ölümlü olarak tahmin edilen kazaların yaklaşık altıda birinin gerçekten ölümlü olduğunu göstermektedir. Buna karşılık Recall değerinin yalnızca %6,06 olması, modelin gerçek ölümlü kazaların büyük bölümünü tespit edemediğini ortaya koymaktadır. Bu nedenle modelin ölümlü kazaları belirleme performansının sınırlı olduğu söylenebilir.

Modelin genellenebilirliğini değerlendirmek amacıyla eğitim ve test performansları karşılaştırılmıştır. Eğitim ve test F1 skorları arasındaki farkın 0,0718 olduğu belirlenmiştir. Bu değer 0,10 eşik değerinin altında olması modelde ciddi bir aşırı öğrenme problemi bulunmadığını göstermektedir. AdaBoost modelinin eğitim ve test verileri üzerinde benzer performans göstermesi, modelin genellenebilirliğinin kabul edilebilir düzeyde olduğunu ortaya koymaktadır. Elde edilen sonuçlar Karar Ağacı ve Rastgele Orman modellerine kıyasla daha düşük aşırı öğrenme eğilimine işaret etmektedir.

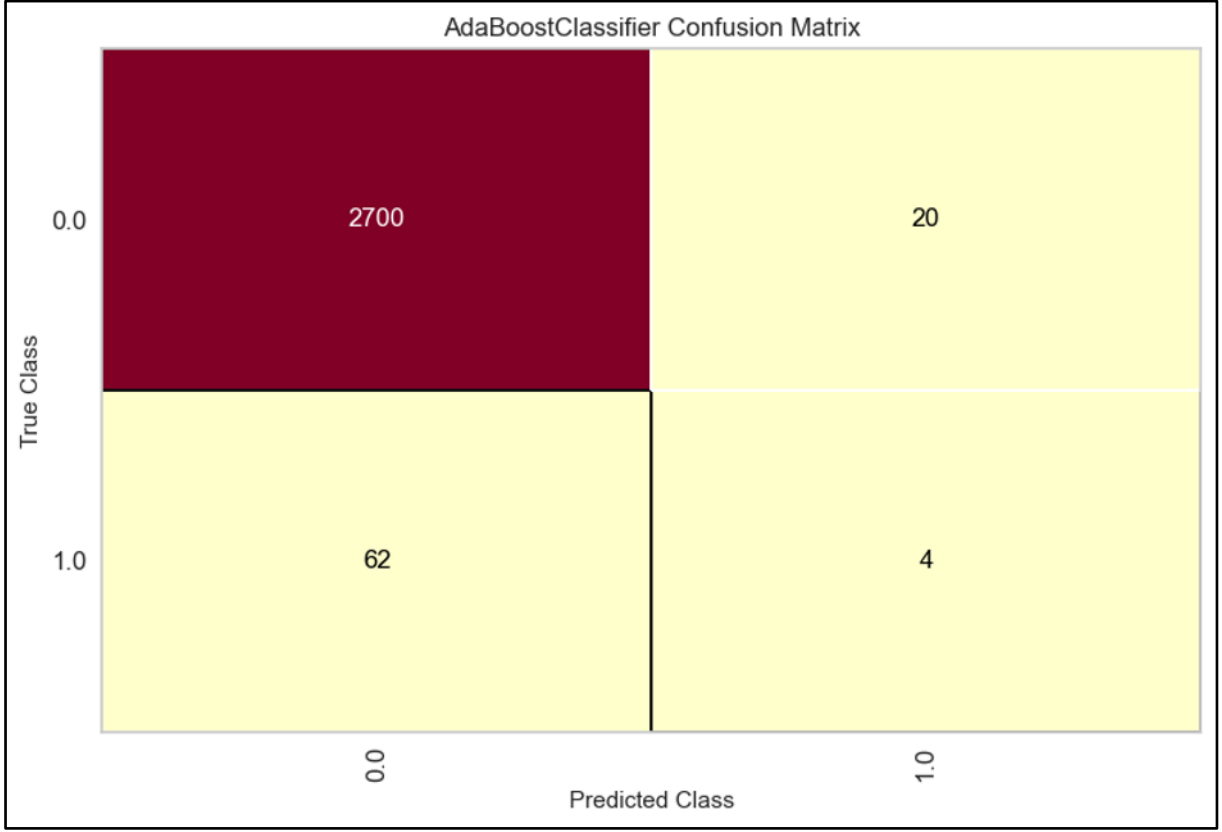
Modelin kararlılığını değerlendirmek amacıyla beş katlı Stratified Cross Validation uygulanmıştır. Analiz sonucunda sırasıyla 0,973089, 0,973613, 0,973613, 0,973613 ve

0,973075 doğruluk skorları elde edilmiştir. Ortalama çapraz doğrulama başarısı 0,973401 olarak hesaplanırken standart sapma değeri 0,000260 olarak bulunmuştur. Katlar arasındaki varyasyonun oldukça düşük olması model performansının farklı veri alt kümelerinde kararlı sonuçlar verdiğini göstermektedir. Bununla birlikte çapraz doğrulama doğruluğunun yüksek olmasına rağmen Recall ve F1 skorlarının düşük olması, veri setindeki sınıf dengesizliğinin doğruluk metriğini önemli ölçüde etkilediğini göstermektedir.



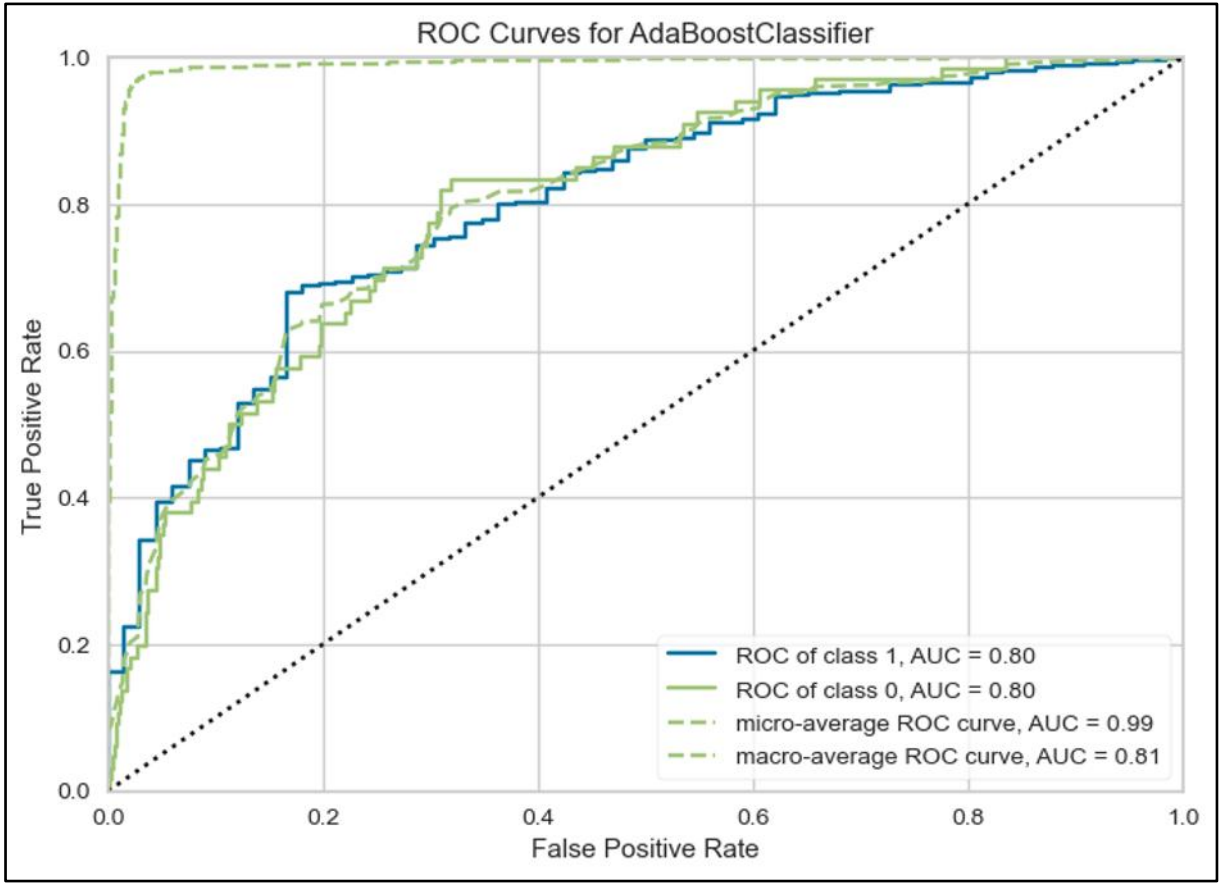
Şekil 37. AdaBoost modeli sınıflandırma raporu

Şekil 37’de AdaBoost modeline ait sınıflandırma raporu sunulmaktadır. Sınıf bazında elde edilen sonuçlar incelendiğinde yaralanmalı kazalar için Precision, Recall ve F1 skorlarının sırasıyla 0,978, 0,993 ve 0,985 olduğu görülmektedir. Buna karşılık ölümlü kazalar için Precision değeri 0,167, Recall değeri 0,061 ve F1 skoru 0,089 olarak hesaplanmıştır. Sonuçlar modelin yaralanmalı kazaları oldukça başarılı şekilde sınıflandırabildiğini göstermektedir. Ancak ölümlü kazalar açısından değerlendirildiğinde Recall ve F1 skorlarının oldukça düşük olduğu görülmektedir. Özellikle Recall değerinin %6,1 seviyesinde kalması, modelin gerçek ölümlü kazaların büyük bölümünü gözden kaçırdığını ortaya koymaktadır.



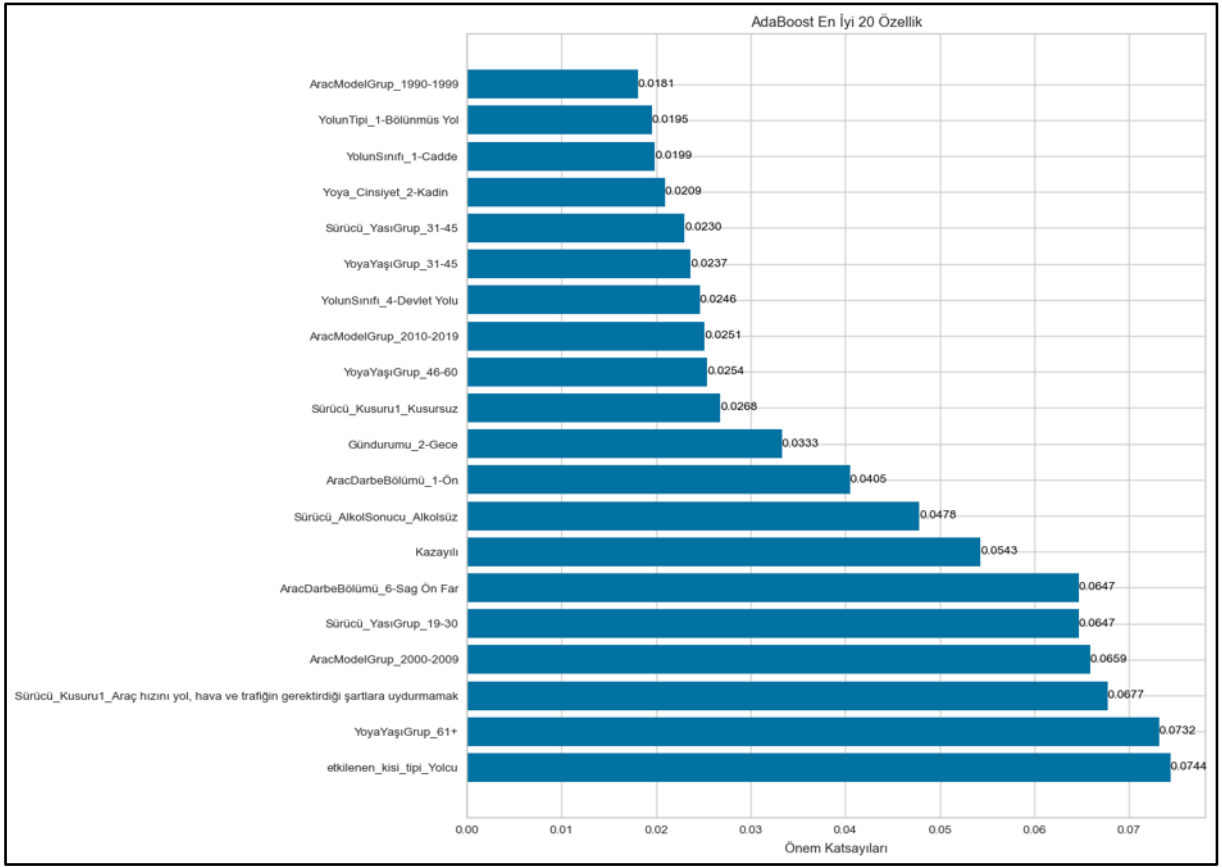
Şekil 38. AdaBoost modeli karışıklık matrisi

Şekil 38’de AdaBoost modeline ait karışıklık matrisi gösterilmektedir. Karışıklık matrisi incelendiğinde toplam 2720 yaralanmalı kazanın 2700 tanesinin doğru şekilde sınıflandırıldığı, yalnızca 20 yaralanmalı kazanın yanlışlıkla ölümlü olarak tahmin edildiği görülmektedir. Buna karşılık toplam 66 ölümlü kazanın yalnızca 4 tanesi doğru şekilde tespit edilirken, 62 ölümlü kaza yaralanmalı olarak sınıflandırılmıştır. Bu sonuçlar modelin yanlış pozitif üretimini oldukça düşük seviyede tuttuğunu göstermektedir. Ancak bunun karşılığında ölümcül kazaların büyük bölümünü gözden kaçırdığı görülmektedir. Dolayısıyla model son derece muhafazakâr bir sınıflandırma yaklaşımı sergilemektedir.



Şekil 39. AdaBoost modeli ROC eğrisi ve AUC değeri

Şekil 39’da AdaBoost modeline ait ROC eğrileri ve AUC değerleri sunulmaktadır. ROC eğrisi incelendiğinde eğrinin referans doğrusunun belirgin şekilde üzerinde yer aldığı ve dışbükey bir yapı oluşturduğu görülmektedir. Sınıf 0 ve sınıf 1 için elde edilen AUC değerleri 0,80 olarak hesaplanmıştır. Ayrıca Macro Average AUC değeri 0,81 ve Micro Average AUC değeri 0,99 olarak bulunmuştur. Macro Average AUC değerinin 0,81 olması modelin sınıfları ayırt etme performansının iyi düzeyde olduğunu göstermektedir. Buna rağmen Recall değerinin oldukça düşük olması, modelin olasılık bazında ayırt edicilik kapasitesinin yüksek olduğunu ancak kullanılan sınıflandırma eşik değerinin ölümlü kazaların belirlenmesi açısından yeterince duyarlı olmadığını düşündürmektedir.



Şekil 40. AdaBoost modeli için özellik önem düzeyleri

Şekil 40’ta AdaBoost modelinde kullanılan değişkenlere ait önem düzeyleri gösterilmektedir. Özellik önem grafiği incelendiğinde “Etkilenen Kişi Tipi – Yolcu”, “Yaya Yaş Grubu – 61+”, “Sürücü Kusuru”, “Araç Model Grubu”, “Sürücü Yaş Grubu”, “Kazayılı”, “Sürücü Alkol Sonucu” ve “Araç Darbe Bölümü” gibi değişkenlerin model üzerinde etkili olduğu görülmektedir. Ayrıca devlet yolu, gece koşulları ve ileri yaş gruplarına ilişkin değişkenlerin de önem düzeylerinin yüksek olduğu belirlenmiştir. Bu sonuçlar diğer makine öğrenmesi modellerinden elde edilen bulgularla büyük ölçüde uyumludur ve ölümcül kazaların oluşumunda belirli risk faktörlerinin öne çıktığını göstermektedir.

Genel olarak değerlendirildiğinde AdaBoost modeli %97,06 doğruluk oranı ile çalışmadaki en yüksek doğruluk değerlerinden birine ulaşmıştır. Ayrıca %16,67 Precision değeri ile ölümlü kaza tahminlerinde en yüksek kesinlik oranlarından birini üretmiştir. Buna karşın Recall değerinin yalnızca %6,06 olması modelin ölümcül kazaların büyük bölümünü tespit edemediğini göstermektedir. Bununla birlikte 0,81 Macro AUC değeri modelin sınıfları ayırt etme kapasitesinin aslında yüksek olduğunu ortaya koymaktadır. Sonuç olarak AdaBoost modeli yüksek doğruluk ve yüksek AUC değerlerine rağmen düşük Recall nedeniyle ölümcül

kazaların belirlenmesinde beklenen performansı gösterememiştir. Ancak karar eşiklerinin optimize edilmesi ve maliyet duyarlı öğrenme yaklaşımlarının uygulanması durumunda model performansının önemli ölçüde artırılabilceği değerlendirilmektedir.

#### 4.3.4. Gradient Boosting sonuçları

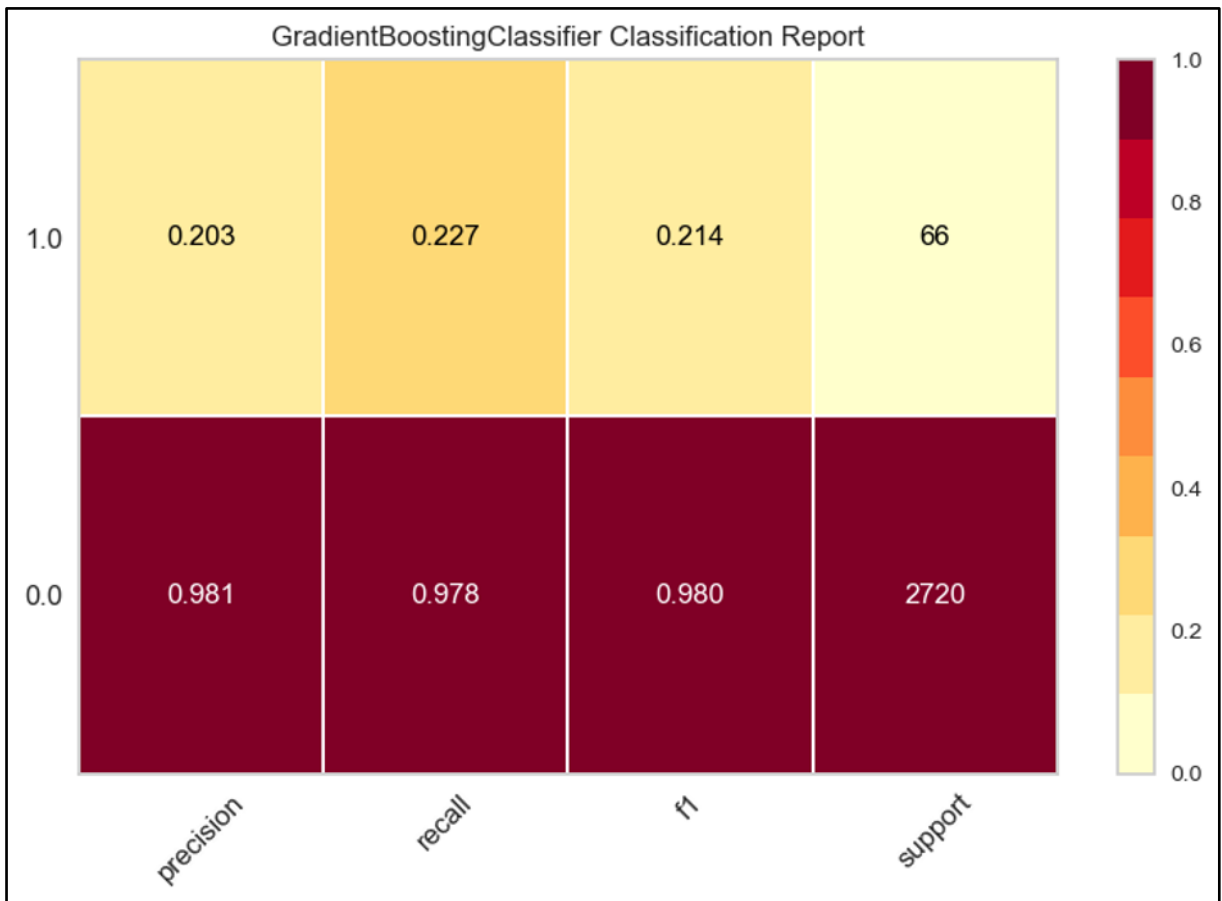
Bu çalışmada trafik kazalarının şiddet düzeyinin tahmin edilmesi amacıyla Gradient Boosting algoritması kullanılmıştır. Gradient Boosting, zayıf öğrencilerin ardışık olarak eğitildiği ve her yeni modelin bir önceki modelin yaptığı hataları düzeltmeye çalıştığı güçlü bir topluluk öğrenmesi yöntemidir. Karar ağaçlarının sıralı olarak bir araya getirilmesiyle oluşturulan bu yaklaşım, özellikle doğrusal olmayan ilişkilerin modellenmesinde ve karmaşık veri yapılarının öğrenilmesinde başarılı sonuçlar vermektedir. Model oluşturulurken düşük öğrenme oranı (learning rate=0,015), yüksek ağaç sayısı (n\_estimators=800), düşük derinlikli ağaçlar (max\_depth=2) ve rastgele alt örnekleme (subsample=0,55) kullanılmıştır. Ayrıca modelin gereksiz öğrenmeye devam etmesini önlemek amacıyla erken durdurma (early stopping) mekanizmasından yararlanılmıştır.

Model performansının değerlendirilmesi sonucunda doğruluk (Accuracy) değeri %96,05, Precision değeri %20,27, Recall değeri %22,73, F1 skoru %21,43 ve Cohen's Kappa katsayısı 0,1941 olarak hesaplanmıştır. Elde edilen doğruluk oranı oldukça yüksek olmakla birlikte veri setindeki sınıf dengesizliği nedeniyle tek başına yeterli bir performans göstergesi değildir. Precision değerinin %20,27 olması model tarafından ölümlü olarak sınıflandırılan kazaların yaklaşık beşte birinin gerçekten ölümlü olduğunu göstermektedir. Recall değerinin %22,73 olması ise modelin gerçek ölümlü kazaların yaklaşık dörtte birini tespit edebildiğini ortaya koymaktadır. F1 skorunun %21,43 olması, modelin Precision ve Recall değerleri arasında diğer modellere kıyasla daha dengeli bir performans sergilediğini göstermektedir. Ayrıca Cohen's Kappa katsayısının 0,1941 olarak hesaplanması, model ile gerçek sınıflandırmalar arasındaki uyumun çalışmada kullanılan diğer modellere göre daha yüksek olduğunu ortaya koymaktadır.

Modelin genellenebilirliğini değerlendirmek amacıyla eğitim ve test performansları karşılaştırılmıştır. Eğitim ve test F1 skorları arasındaki farkın yalnızca 0,0125 olduğu belirlenmiştir. Bu farkın oldukça düşük olması modelin aşırı öğrenme göstermediğini ve yüksek genellenebilirliğe sahip olduğunu göstermektedir. Çalışmada değerlendirilen modeller arasında en düşük aşırı öğrenme düzeylerinden birinin Gradient Boosting modelinde elde edilmesi,

kullanılan düşük öğrenme oranı, erken durdurma mekanizması ve alt örnekleme yaklaşımının başarılı şekilde çalıştığını göstermektedir.

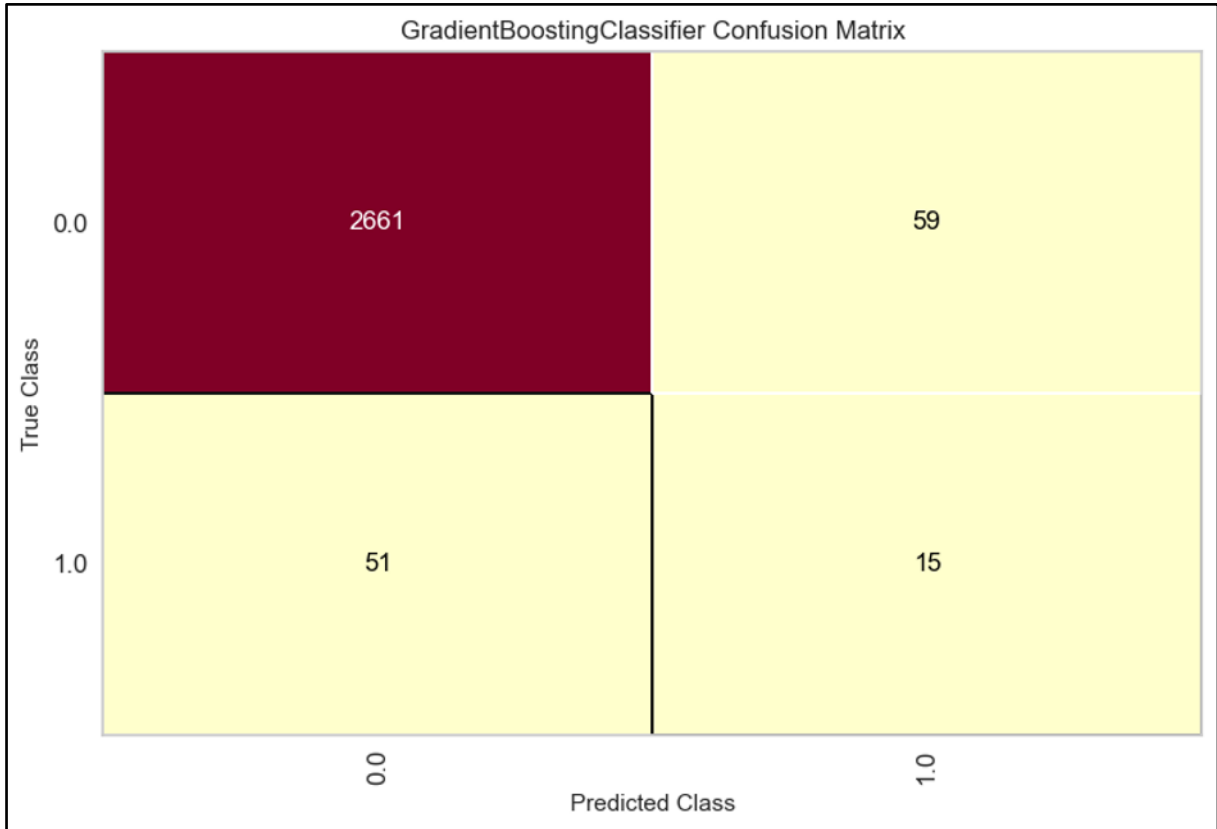
Model performansının farklı veri alt kümelerinde ne derece değiştiğini değerlendirmek amacıyla beş katlı Stratified Cross Validation uygulanmıştır. Analiz sonucunda sırasıyla 0,973089, 0,973613, 0,973613, 0,973075 ve 0,973075 doğruluk skorları elde edilmiştir. Ortalama çapraz doğrulama başarısı 0,973293 olarak hesaplanırken standart sapma değeri 0,000261 olarak bulunmuştur. Katlar arasında elde edilen sonuçların birbirine oldukça yakın olması model performansının kararlı olduğunu göstermektedir. Bununla birlikte diğer modellerde olduğu gibi doğruluk temelli çapraz doğrulama sonuçlarının sınıf dengesizliğinden etkilenebileceği göz önünde bulundurulmalıdır.



Şekil 41. Gradient Boosting modeli sınıflandırma raporu

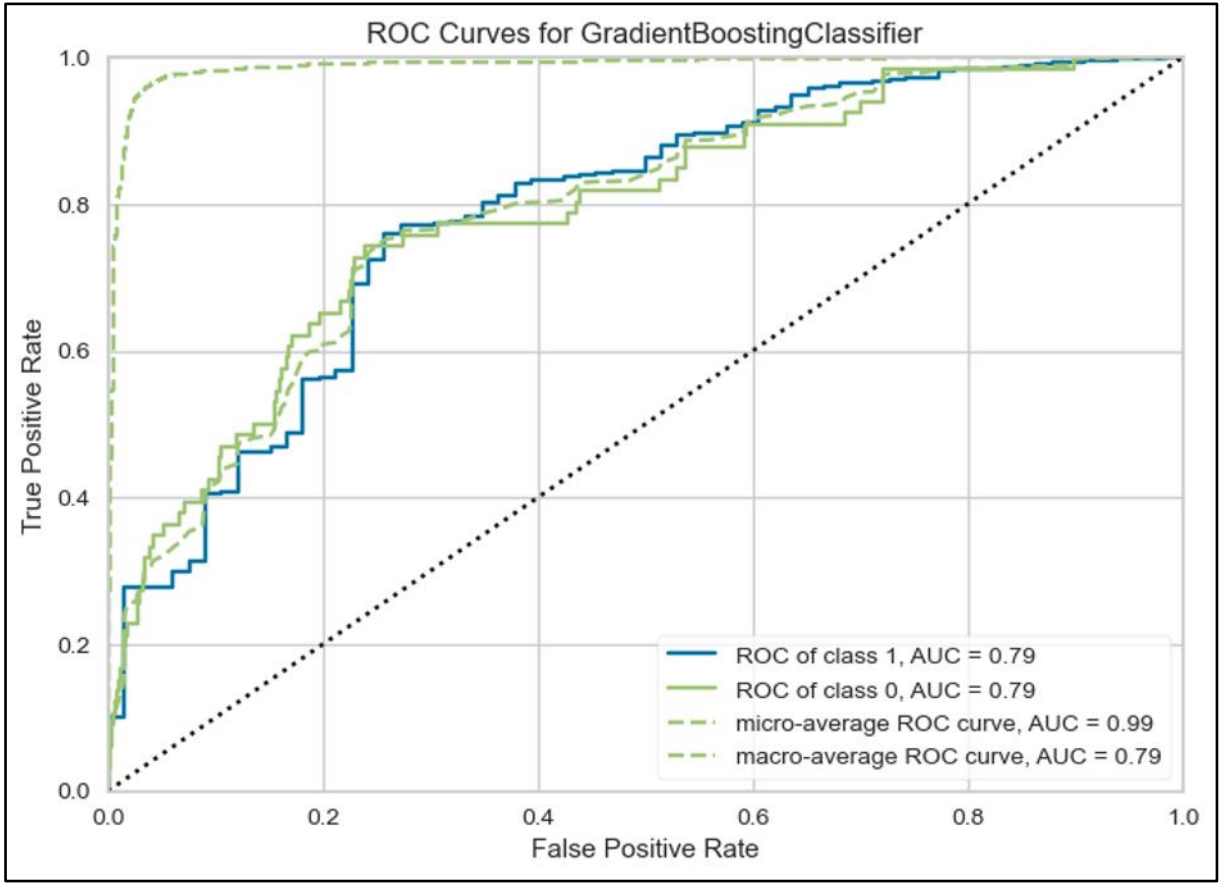
Şekil 41’de Gradient Boosting modeline ait sınıflandırma raporu sunulmaktadır. Sınıf bazında elde edilen sonuçlar incelendiğinde yaralanmalı kazalar için Precision, Recall ve F1 skorlarının sırasıyla 0,981, 0,978 ve 0,980 olduğu görülmektedir. Buna karşılık ölümlü kazalar için

Precision değeri 0,203, Recall değeri 0,227 ve F1 skoru 0,214 olarak hesaplanmıştır. Sonuçlar modelin yaralanmalı kazaları oldukça başarılı şekilde sınıflandırabildiğini göstermektedir. Ölümlü kazalar açısından değerlendirildiğinde ise elde edilen Precision, Recall ve F1 değerleri çalışmada kullanılan diğer modellerle karşılaştırıldığında daha dengeli bir performansa işaret etmektedir. Özellikle F1 skorunun Karar Ağacı, Rastgele Orman, AdaBoost ve Lojistik Regresyon modellerinden daha yüksek olması dikkat çekicidir.



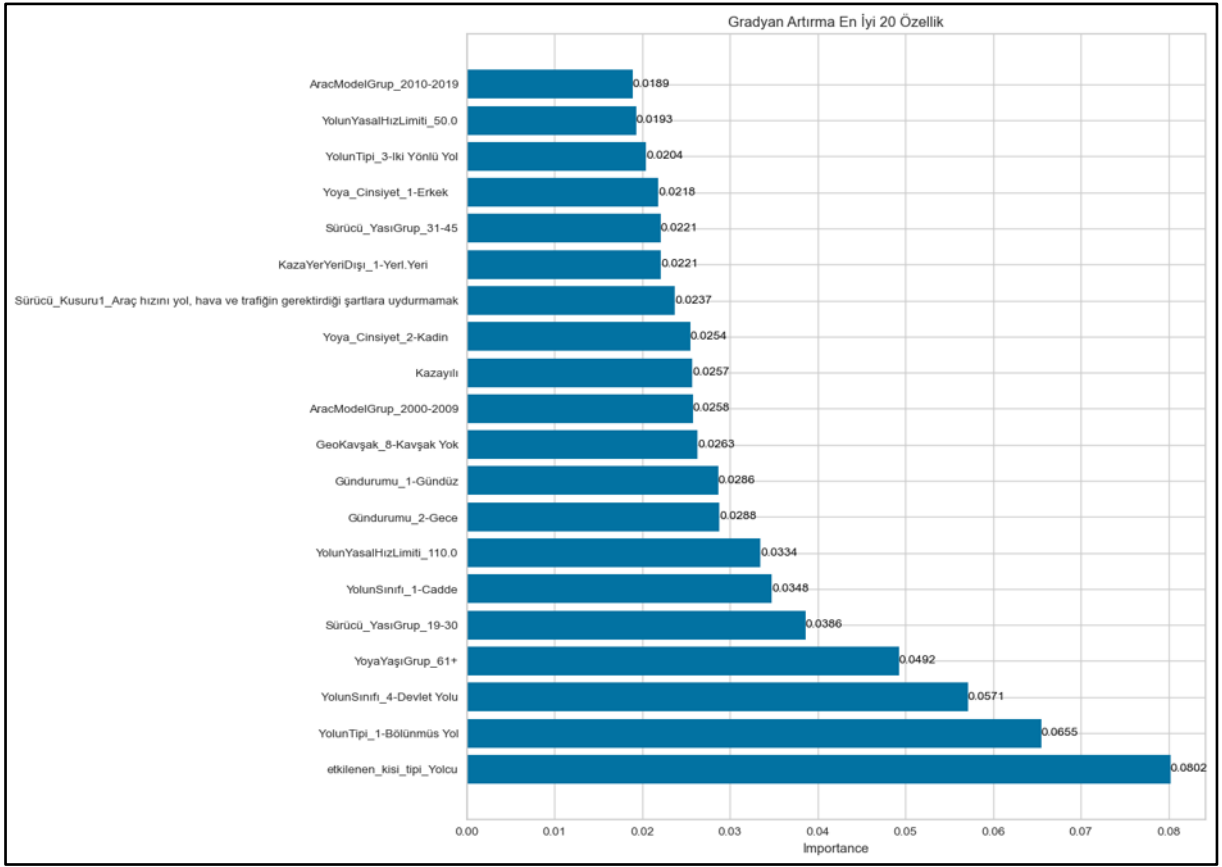
Şekil 42. Gradient Boosting modeli karışıklık matrisi

Şekil 42’de Gradient Boosting modeline ait karışıklık matrisi gösterilmektedir. Karışıklık matrisi incelendiğinde toplam 2661 yaralanmalı kazanın doğru şekilde sınıflandırıldığı, 59 yaralanmalı kazanın yanlışlıkla ölümlü olarak tahmin edildiği görülmektedir. Ölümlü kazalar açısından değerlendirildiğinde ise toplam 66 ölümlü kazanın 15 tanesinin doğru şekilde tespit edildiği, 51 ölümlü kazanın ise yaralanmalı olarak sınıflandırıldığı belirlenmiştir. Sonuçlar modelin yanlış pozitif üretme eğiliminin düşük olduğunu göstermektedir. Karar Ağacı modeli 19 ölümlü kazayı doğru tahmin etmiş olsa da Gradient Boosting modeli daha az yanlış pozitif üreterek daha dengeli bir sınıflandırma performansı sergilemiştir.



Şekil 43. Gradient Boosting modeli ROC eğrisi ve AUC değeri

Şekil 43’de Gradient Boosting modeline ait ROC eğrileri ve AUC değerleri sunulmaktadır. ROC eğrisi incelendiğinde eğrinin referans çizgisinin üzerinde ve belirgin şekilde dışbükey bir yapı oluşturduğu görülmektedir. Sınıf 0 ve sınıf 1 için elde edilen AUC değerleri 0,79 olarak hesaplanmıştır. Ayrıca Macro Average AUC değeri 0,79 ve Micro Average AUC değeri 0,99 olarak bulunmuştur. Macro Average AUC değerinin 0,79 olması modelin ölümcül ve yaralanmalı kazaları ayırt etme kapasitesinin oldukça iyi düzeyde olduğunu göstermektedir. Elde edilen AUC değeri 0,80 sınırına oldukça yakın olup çalışmada kullanılan modeller arasında en yüksek ayırt edicilik performanslarından birini temsil etmektedir. Micro Average AUC değerinin yüksek olması ise veri setindeki çoğunluk sınıfının baskın etkisinden kaynaklanmaktadır.



Şekil 44. Gradient Boosting modeli için özellik önem düzeyleri

Şekil 44’te Gradient Boosting modelinde kullanılan değişkenlere ait önem düzeyleri gösterilmektedir. Sonuçlar incelendiğinde en yüksek önem değerine sahip değişkenin “Etkilenen Kişi Tipi – Yolcu” olduğu görülmektedir. Bunun yanında “Yolun Tipi – Bölünmüş Yol”, “Yolun Sınıfı – Devlet Yolu”, “Yaya Yaş Grubu – 61+”, “Sürücü Yaş Grubu – 19-30”, “Gece Koşulları” ve “Hız Limiti – 110 km/s” değişkenleri de model üzerinde önemli etkiye sahip faktörler arasında yer almaktadır. Ayrıca sürücünün hızını yol ve trafik şartlarına göre ayarlamaması değişkeninin ilk yirmi değişken içerisinde yer alması trafik güvenliği açısından dikkat çekici bir bulgu olarak değerlendirilmektedir. Önceki modellerde de öne çıkan devlet yolu, ileri yaş grubundaki yayalar, gece koşulları ve yolcu durumu gibi değişkenlerin bu modelde de üst sıralarda yer alması, bu değişkenlerin ölümcül kazalarla tutarlı biçimde ilişkili olduğunu göstermektedir.

Genel olarak değerlendirildiğinde Gradient Boosting modeli %96,05 doğruluk, %20,27 Precision, %22,73 Recall, %21,43 F1 skoru ve 0,1941 Cohen’s Kappa değeri ile çalışmada değerlendirilen modeller arasında en dengeli performansı göstermiştir. Model, en yüksek F1 skorlarından birine sahip olması, en düşük aşırı öğrenme düzeyini göstermesi ve yüksek ROC-

AUC performansı sergilemesi bakımından dikkat çekmektedir. Ayrıca yanlış pozitif ve yanlış negatif tahminler arasında daha dengeli bir dağılım oluşturmuştur. Bu bulgular doğrultusunda Gradient Boosting modeli, çalışmada değerlendirilen modeller arasında ölümcül trafik kazalarının tahmini açısından en başarılı ve en genellenebilir modellerden biri olarak değerlendirilebilir.

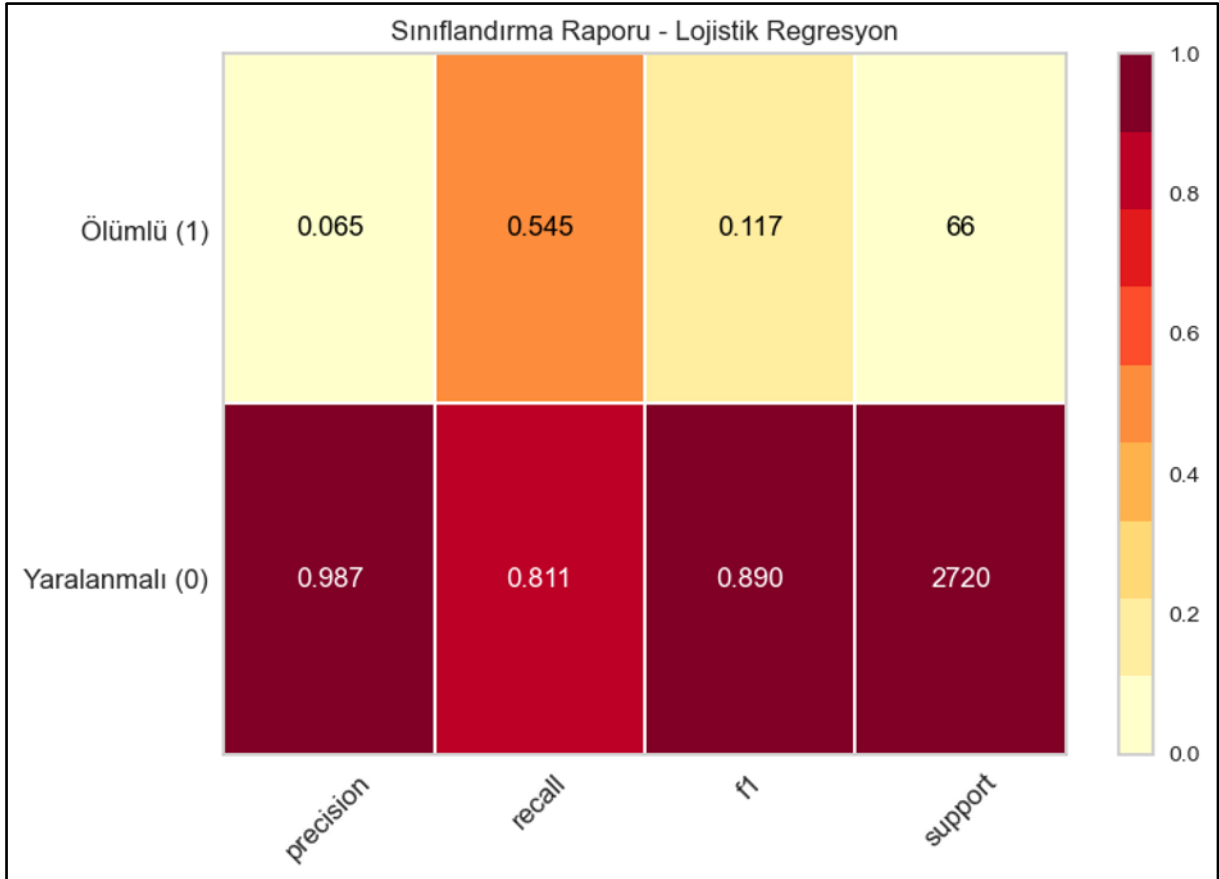
#### 4.3.5. Lojistik regresyon sonuçları

Bu çalışmada trafik kazalarının şiddet düzeyinin tahmin edilmesi amacıyla Lojistik Regresyon algoritması kullanılmıştır. Lojistik Regresyon, bağımlı değişkenin belirli bir sınıfa ait olma olasılığını tahmin eden ve özellikle ikili sınıflandırma problemlerinde yaygın olarak kullanılan istatistiksel bir makine öğrenmesi yöntemidir. Model oluşturulurken aşırı öğrenmenin önüne geçebilmek amacıyla L2 düzenleme uygulanmış, çözümleyici olarak ikili sınıflandırma problemlerinde yaygın şekilde kullanılan liblinear algoritması tercih edilmiştir. Ayrıca veri setinde ölümlü ve yaralanmalı kazalar arasında önemli düzeyde sınıf dengesizliği bulunması nedeniyle sınıf ağırlıkları dengelenmiş (class\_weight=balanced) ve azınlık sınıfının öğrenme sürecindeki etkisi artırılmıştır.

Model performansının değerlendirilmesi sonucunda doğruluk (Accuracy) değeri %80,44, Precision değeri %6,53, Recall değeri %54,55, F1 skoru %11,67 ve Cohen's Kappa katsayısı 0,0777 olarak hesaplanmıştır. Elde edilen doğruluk oranı ilk bakışta yüksek görünmekle birlikte veri setindeki sınıf dengesizliği nedeniyle tek başına yeterli bir performans göstergesi değildir. Precision değerinin %6,53 olması modelin ölümlü olarak sınıflandırdığı kazaların yalnızca küçük bir kısmının gerçekten ölümlü olduğunu göstermektedir. Buna karşılık Recall değerinin %54,55 olması modelin gerçek ölümlü kazaların yaklaşık yarısını tespit edebildiğini ortaya koymaktadır. Bu durum modelin ölümcül kazaları kaçırmama konusunda görece başarılı olduğunu ancak çok sayıda yanlış alarm ürettiğini göstermektedir.

Modelin genellenebilirliğini değerlendirmek amacıyla eğitim ve test performansları karşılaştırılmıştır. Eğitim ve test veri setleri arasındaki F1 skoru farkının 0,087 olduğu belirlenmiştir. Genel olarak bu farkın 0,10 değerinin altında olması modelde ciddi düzeyde aşırı öğrenme bulunmadığını göstermektedir. Sonuçlar, lojistik regresyon modelinin eğitim verisini ezberlemek yerine genel örüntüleri öğrenebildiğini ve yeni veriler üzerinde benzer performans sergileyebildiğini ortaya koymaktadır.

Modelin kararlılığını değerlendirmek amacıyla beş katlı Stratified Cross Validation uygulanmıştır. Analiz sonucunda sırasıyla 0,750, 0,741, 0,770, 0,794 ve 0,801 doğruluk skorları elde edilmiştir. Ortalama doğruluk değeri 0,771 olarak hesaplanırken standart sapma değeri 0,0235 olarak bulunmuştur. Standart sapma değerinin düşük olması model performansının farklı veri alt kümelerinde önemli değişiklik göstermediğini ve kararlı sonuçlar ürettiğini göstermektedir.



Şekil 45. Lojistik regresyon modeli sınıflandırma raporu

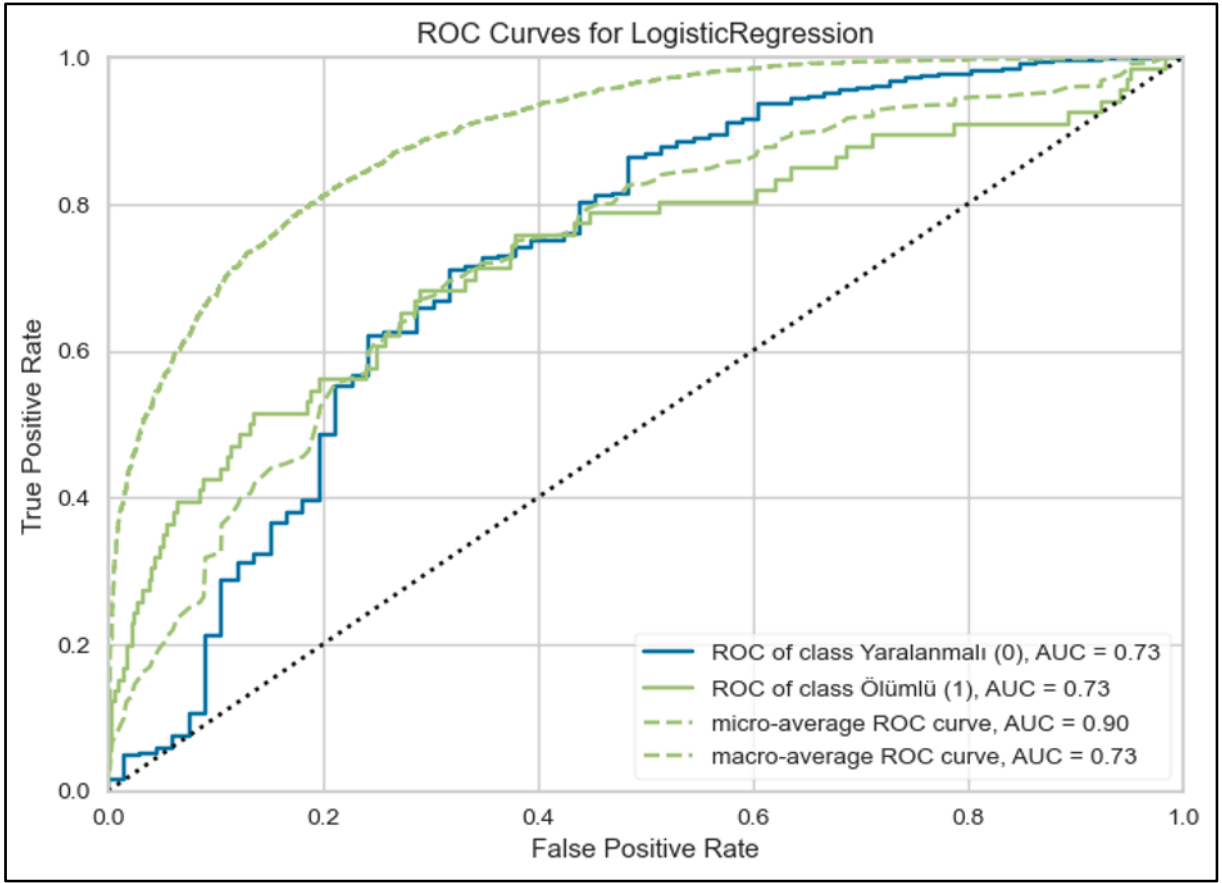
Şekil 45'te Lojistik Regresyon modeline ait sınıflandırma raporu sunulmaktadır. Sınıf bazında elde edilen sonuçlar incelendiğinde yaralanmalı kazalar için Precision, Recall ve F1 skorlarının sırasıyla 0,987, 0,811 ve 0,890 olduğu görülmektedir. Buna karşılık ölümlü kazalar için Precision değeri 0,065, Recall değeri 0,545 ve F1 skoru 0,117 olarak hesaplanmıştır. Sonuçlar modelin yaralanmalı kazaları oldukça başarılı şekilde sınıflandırabildiğini göstermektedir. Ancak ölümlü kazalar açısından değerlendirildiğinde Precision değerinin oldukça düşük olması nedeniyle modelin çok sayıda yanlış pozitif ürettiği anlaşılmaktadır. Buna rağmen Recall

değerinin diğer birçok modele göre yüksek olması, modelin ölümcül kazaların önemli bir bölümünü tespit edebildiğini göstermektedir.

True Class	Yaralanmalı (0)	2205	515
	Ölümlü (1)	30	36
		Yaralanmalı (0)	Ölümlü (1)
		Predicted Class	

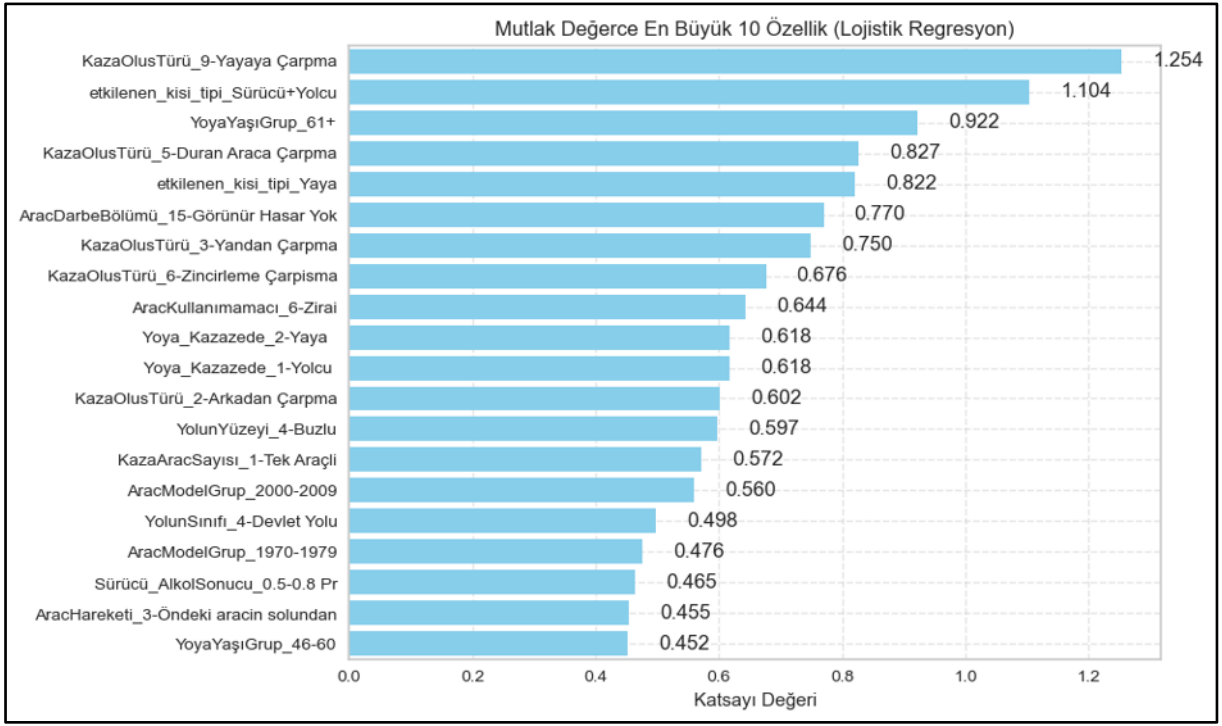
Şekil 46. Lojistik regresyon modeli karışıklık matrisi

Şekil 46’da Lojistik Regresyon modeline ait karışıklık matrisi gösterilmektedir. Karışıklık matrisi incelendiğinde toplam 2205 yaralanmalı kazanın doğru şekilde sınıflandırıldığı, 515 yaralanmalı kazanın yanlışlıkla ölümlü olarak tahmin edildiği görülmektedir. Ölümlü kazalar açısından değerlendirildiğinde ise toplam 66 ölümlü kazanın 36 tanesinin doğru şekilde tahmin edildiği, 30 ölümlü kazanın ise yaralanmalı olarak sınıflandırıldığı belirlenmiştir. Sonuçlar modelin ölümlü kazaları belirleme konusunda duyarlı olduğunu göstermektedir. Ancak 515 adet yanlış pozitif tahmin üretilmesi modelin ölüm riskini olduğundan daha yüksek tahmin etme eğiliminde olduğunu ortaya koymaktadır.



Şekil 47. Lojistik regresyon modeli ROC eğrisi ve AUC değeri

Şekil 47’de Lojistik Regresyon modeline ait ROC eğrileri ve AUC değerleri sunulmaktadır. ROC eğrisi incelendiğinde eğrinin rastgele sınıflandırmayı temsil eden referans doğrusunun üzerinde ve dışbükey bir yapıda olduğu görülmektedir. Sınıf 0 ve sınıf 1 için elde edilen AUC değerleri 0,73 olarak hesaplanmıştır. Ayrıca Macro Average AUC değeri 0,73 ve Micro Average AUC değeri 0,90 olarak bulunmuştur. Macro Average AUC değerinin 0,73 olması modelin ölümlü ve yaralanmalı kazaları ayırt etme kapasitesinin orta düzeyde olduğunu göstermektedir. Literatürde 0,70–0,80 aralığındaki AUC değerleri kabul edilebilir ayırt edicilik düzeyi olarak değerlendirilmektedir. Micro Average AUC değerinin daha yüksek olması ise veri setindeki çoğunluk sınıfının etkisinden kaynaklanmaktadır.



Şekil 48. Lojistik regresyon modeli için özellik önem düzeyleri

Şekil 48’de Lojistik Regresyon modelinde elde edilen katsayıların mutlak değerlerine göre önem düzeyleri gösterilmektedir. Sonuçlar incelendiğinde ölüm olasılığını artıran en güçlü değişkenin “Kaza Oluş Türü – Yayaya Çarpma” olduğu görülmektedir ( $\beta=1,254$ ). Bunun yanında “Etkilenen Kişi Tipi – Sürücü ve Yolcu” ( $\beta=1,104$ ), “Yaya Yaş Grubu – 61+” ( $\beta=0,922$ ), “Yaya/Kazazede – Yaya” ( $\beta=0,618$ ), “Yolun Sınıfı – Devlet Yolu” ( $\beta=0,498$ ) ve “Yaya Yaş Grubu – 46–60” ( $\beta=0,452$ ) değişkenlerinin de ölüm riskini artıran önemli faktörler olduğu belirlenmiştir. Buna karşılık “Duran Araca Çarpma”, “Etkilenen Kişi Tipi – Yaya”, “Görünür Hasar Yok”, “Yandan Çarpma”, “Zincirleme Çarpışma” ve “Zirai Araç” değişkenlerinin ölüm olasılığını azaltan faktörler arasında yer aldığı görülmüştür. Özellikle yayaya çarpma kazalarının ve ileri yaş grubundaki yayaların ölümcül sonuçlarla güçlü biçimde ilişkili olması dikkat çekici bir bulgudur.

Genel olarak değerlendirildiğinde Lojistik Regresyon modeli veri setindeki genel örüntüleri başarılı şekilde öğrenmiş ve kararlı sonuçlar üretmiştir. Ancak ölümlü kaza sınıfı için elde edilen düşük Precision (%6,53), düşük F1 skoru (%11,67) ve düşük Cohen’s Kappa (0,0777) değerleri modelin ölümcül kazaları ayırt etme gücünün sınırlı olduğunu göstermektedir. Buna karşın %54,55’lik Recall değeri, modelin ölümcül kazaları tespit etme konusunda diğer bazı modellere göre daha duyarlı olduğunu ortaya koymaktadır. Sonuç olarak Lojistik Regresyon

modeli yüksek yorumlanabilirlik sağlması ve değişkenlerin ölüm riski üzerindeki etkilerini ortaya koyabilmesi açısından önemli avantajlar sunmakla birlikte, tahmin performansı bakımından daha gelişmiş topluluk öğrenmesi yöntemlerinin gerisinde kalmıştır.

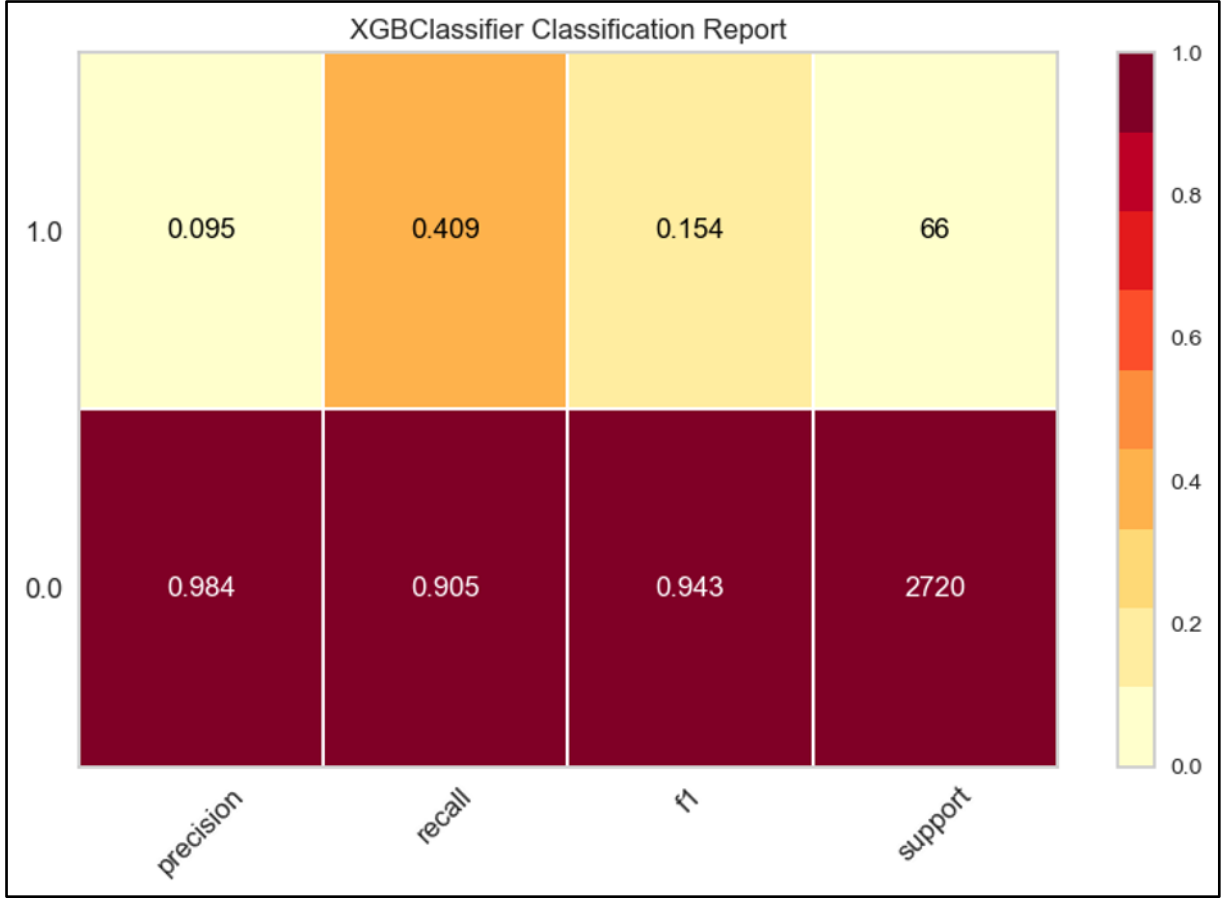
#### 4.3.6. XGBoost sonuçları

Bu çalışmada trafik kazalarının şiddet düzeyinin tahmin edilmesi amacıyla Extreme Gradient Boosting (XGBoost) algoritması kullanılmıştır. XGBoost, gradyan artırma prensibine dayanan ve karar ağaçlarını ardışık biçimde eğiterek tahmin performansını artıran gelişmiş bir topluluk öğrenmesi yöntemidir. Özellikle doğrusal olmayan ilişkilerin modellenmesinde, yüksek boyutlu veri setlerinde ve sınıflandırma problemlerinde literatürde en başarılı algoritmalar arasında gösterilmektedir. Model oluşturulurken L1 ve L2 düzenleme uygulanmış, maksimum ağaç derinliği sınırlandırılmamış, 1000 adet karar ağacı kullanılmış ve oldukça düşük bir öğrenme oranı ( $learning\_rate=0,00001$ ) tercih edilmiştir. Bu yapı modelin daha kontrollü öğrenmesini sağlamak amacıyla oluşturulmuştur.

Model performansının değerlendirilmesi sonucunda doğruluk (Accuracy) değeri %89,34, Precision değeri %9,47, Recall değeri %40,91, F1 skoru %15,38 ve Cohen's Kappa katsayısı 0,1200 olarak hesaplanmıştır. Doğruluk oranı Gradient Boosting, AdaBoost, Rastgele Orman ve Karar Ağacı modellerine göre daha düşük olmasına rağmen veri setindeki sınıf dengesizliği dikkate alındığında bu durum modelin başarısız olduğu anlamına gelmemektedir. Precision değerinin %9,47 olması model tarafından ölümlü olarak sınıflandırılan kazaların yalnızca yaklaşık onda birinin gerçekten ölümlü olduğunu göstermektedir. Buna karşılık Recall değerinin %40,91 olması modelin gerçek ölümlü kazaların yaklaşık %41'ini tespit edebildiğini ortaya koymaktadır. Bu sonuç XGBoost modelinin ölümcül kazaların belirlenmesinde oldukça duyarlı olduğunu göstermektedir.

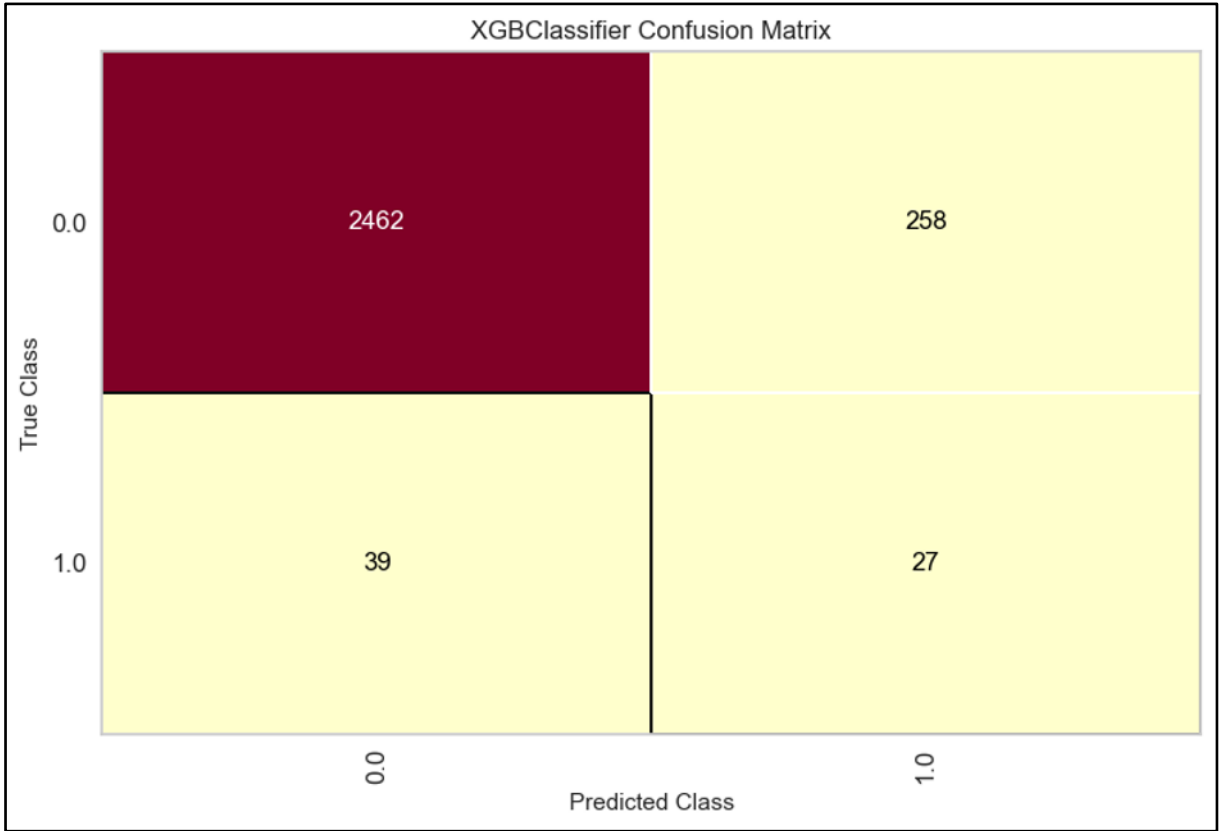
Modelin farklı veri alt kümelerindeki performans kararlılığını değerlendirmek amacıyla beş katlı Stratified Cross Validation uygulanmıştır. Analiz sonucunda sırasıyla 0,973089, 0,973613, 0,973613, 0,973613 ve 0,973075 doğruluk skorları elde edilmiştir. Ortalama çapraz doğrulama başarısı 0,973401 olarak hesaplanırken standart sapma değeri 0,000260 olarak bulunmuştur. Standart sapmanın oldukça düşük olması model performansının farklı örneklemeler üzerinde kararlı olduğunu ve benzer sonuçlar üretebildiğini göstermektedir. Bununla birlikte doğruluk

temelli çapraz doğrulama sonuçlarının veri setindeki sınıf dengesizliğinden etkilenebileceği göz önünde bulundurulmalıdır.



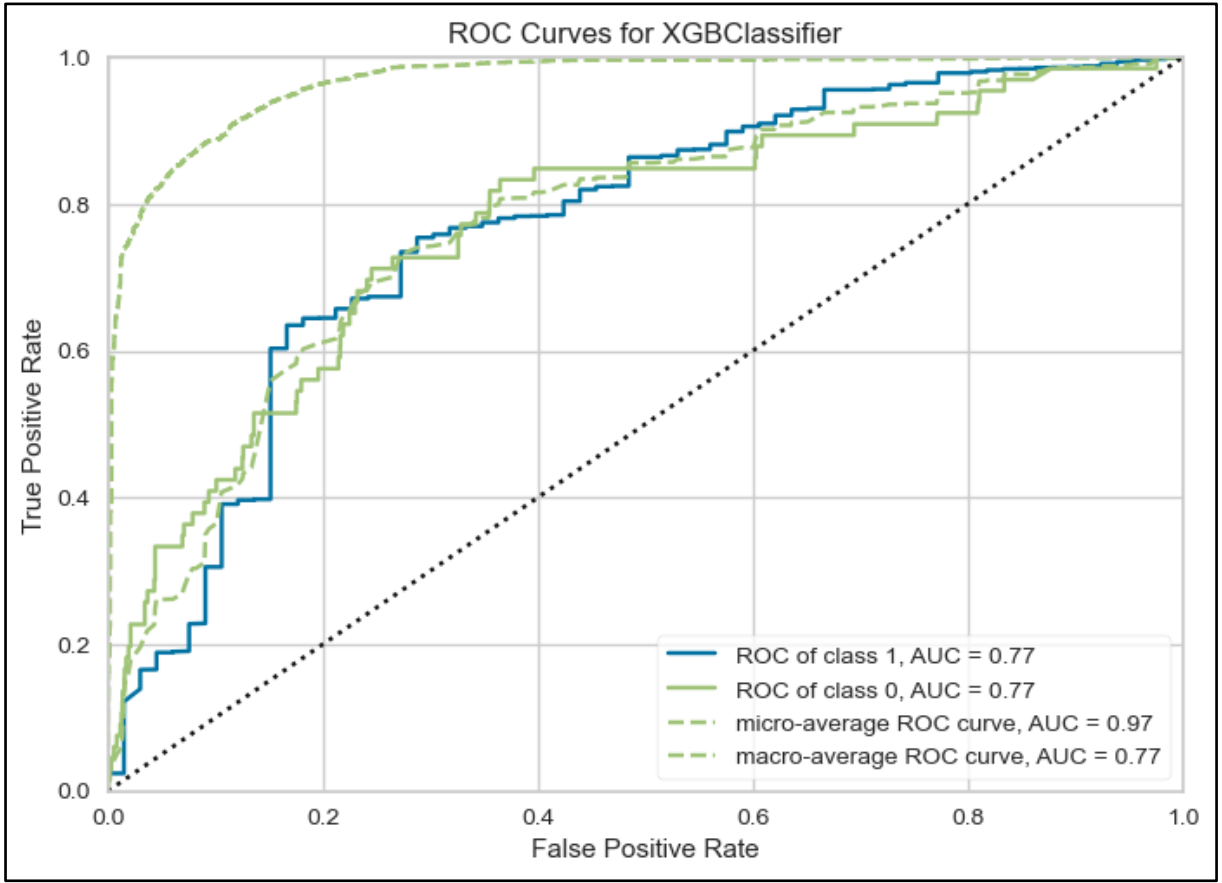
Şekil 49. XGBoost modeli sınıflandırma raporu

Şekil 49’da XGBoost modeline ait sınıflandırma raporu sunulmaktadır. Sınıf bazında elde edilen sonuçlar incelendiğinde yaralanmalı kazalar için Precision, Recall ve F1 skorlarının sırasıyla 0,984, 0,905 ve 0,943 olduğu görülmektedir. Buna karşılık ölümlü kazalar için Precision değeri 0,095, Recall değeri 0,409 ve F1 skoru 0,154 olarak hesaplanmıştır. Sonuçlar modelin yaralanmalı kazaları başarılı şekilde sınıflandırdığını göstermektedir. Ölümlü kazalar açısından değerlendirildiğinde ise Recall değerinin %40,9 olması dikkat çekicidir. Bu sonuç çalışmada kullanılan birçok modelden daha yüksek olup modelin ölümcül kazaları belirleme konusunda güçlü bir performans sergilediğini göstermektedir. Ancak Precision değerinin düşük olması modelin çok sayıda yanlış alarm ürettiğini ortaya koymaktadır.



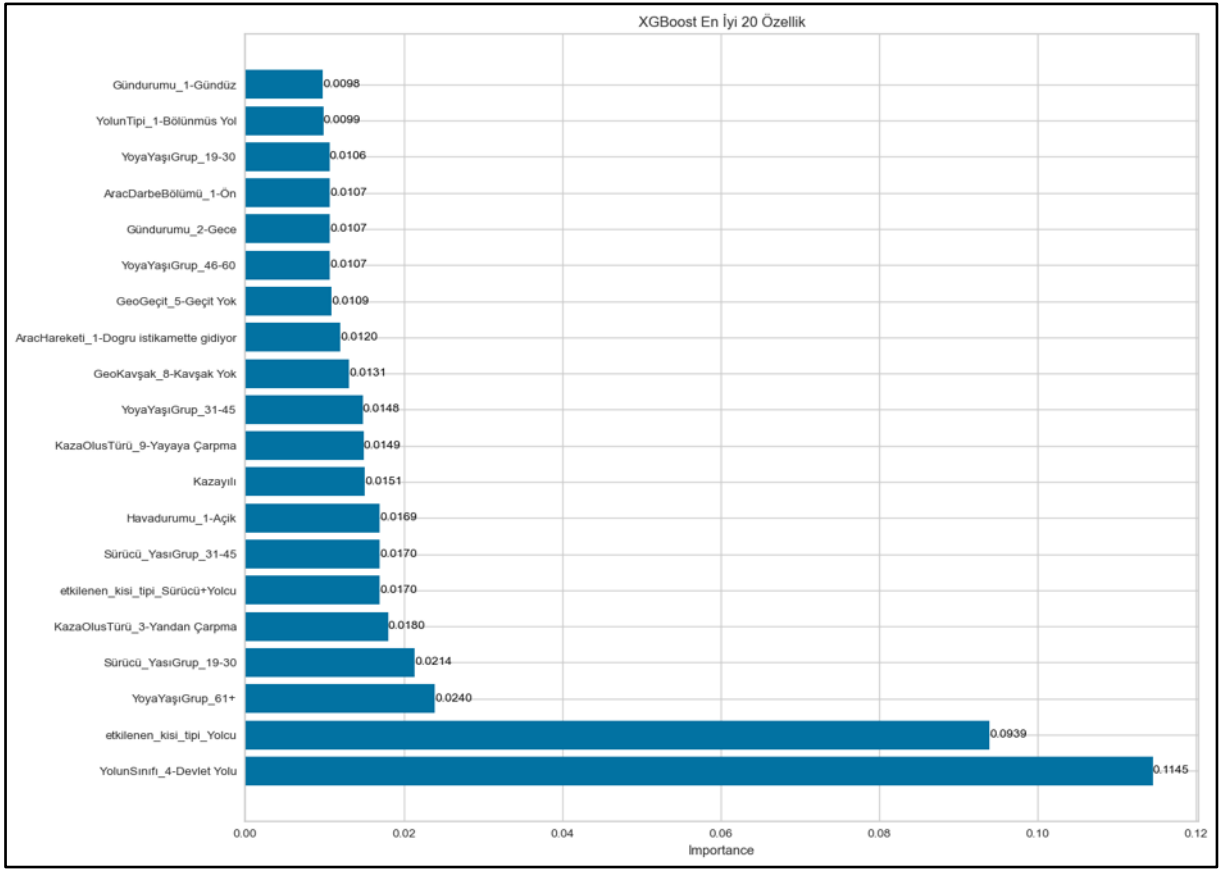
Şekil 50. XGBoost modeli karışıklık matrisi

Şekil 50’da XGBoost modeline ait karışıklık matrisi gösterilmektedir. Karışıklık matrisi incelendiğinde toplam 2462 yaralanmalı kazanın doğru şekilde sınıflandırıldığı, 258 yaralanmalı kazanın yanlışlıkla ölümlü olarak tahmin edildiği görülmektedir. Ölümlü kazalar açısından değerlendirildiğinde ise toplam 66 ölümlü kazanın 27 tanesinin doğru şekilde tahmin edildiği, 39 ölümlü kazanın ise yaralanmalı olarak sınıflandırıldığı belirlenmiştir. Bu sonuçlar modelin ölümcül kazaların belirlenmesinde diğer birçok modele göre daha başarılı olduğunu göstermektedir. Karar Ağacı modeli 19, Gradient Boosting modeli 15, Rastgele Orman modeli 12 ve AdaBoost modeli yalnızca 4 ölümlü kazayı doğru tahmin edebilirken, XGBoost modeli 27 ölümlü kazayı doğru şekilde belirleyebilmiştir. Bununla birlikte bu başarı, yanlış pozitif tahmin sayısındaki artış ile elde edilmiştir.



Şekil 51. XGBoost modeli ROC eğrisi ve AUC değeri

Şekil 51’de XGBoost modeline ait ROC eğrileri ve AUC değerleri sunulmaktadır. ROC eğrisi incelendiğinde eğrinin referans çizgisinin üzerinde ve dışbükey bir yapı oluşturduğu görülmektedir. Sınıf 0 ve sınıf 1 için elde edilen AUC değerleri 0,77 olarak hesaplanmıştır. Ayrıca Macro Average AUC değeri 0,77 ve Micro Average AUC değeri 0,97 olarak bulunmuştur. Macro Average AUC değerinin 0,77 olması modelin ölümcül ve yaralanmalı kazaları ayırt etme kapasitesinin kabul edilebilir düzeyin üzerinde olduğunu göstermektedir. Bu değer Gradient Boosting modelinin (0,79) biraz altında olmakla birlikte Karar Ağacı, Rastgele Orman ve Lojistik Regresyon modellerinden daha yüksektir. Bu nedenle modelin ayırt edicilik performansının başarılı olduğu söylenebilir.



Şekil 52. XGBoost modeli için özellik önem düzeyleri

Şekil 52’de XGBoost modelinde kullanılan değişkenlere ait önem düzeyleri gösterilmektedir. Sonuçlar incelendiğinde en yüksek önem değerine sahip değişkenin “Yolun Sınıfı – Devlet Yolu” olduğu görülmektedir. Bunu sırasıyla “Etkilenen Kişi Tipi – Yolcu”, “Yaya Yaş Grubu – 61+”, “Sürücü Yaş Grubu – 19-30”, “Kaza Oluş Türü – Yandan Çarpma” ve “Etkilenen Kişi Tipi – Sürücü+Yolcu” değişkenleri takip etmektedir. Ayrıca gece koşulları, yayaya çarpma kazaları, kavşak bulunmayan yollar ve bölünmüş yol değişkenlerinin de model üzerinde önemli etkileri olduğu belirlenmiştir. Daha önce Karar Ağacı, Rastgele Orman ve Gradient Boosting modellerinde de öne çıkan devlet yolu, ileri yaş grubundaki yayalar, yolcu durumu ve gece koşulları gibi değişkenlerin bu modelde de üst sıralarda yer alması, bu faktörlerin ölümcül trafik kazalarıyla güçlü biçimde ilişkili olduğunu göstermektedir.

Genel olarak değerlendirildiğinde XGBoost modeli %89,34 doğruluk, %40,91 Recall, %15,38 F1 skoru ve 0,77 Macro AUC değeri ile ölümcül kazaların belirlenmesinde oldukça başarılı sonuçlar üretmiştir. Modelin en önemli avantajı çalışmada kullanılan modeller arasında en yüksek Recall değerlerinden birine ulaşmasıdır. Bu durum ölümcül kazaların gözden kaçırılmasını azaltmaktadır. Ancak Precision değerinin düşük olması modelin çok sayıda yanlış

alarm üretmesine neden olmaktadır. Sonuç olarak XGBoost modeli ölümcül kazaların tespit edilmesinin öncelikli olduğu senaryolarda önemli avantajlar sunmakta, ancak genel performans dengesi açısından değerlendirildiğinde Gradient Boosting modelinin daha yüksek F1 skoru ve daha dengeli yapısı nedeniyle üstünlük sağladığı görülmektedir.

#### 4.4. Modellerin Karşılaştırılması

Bu çalışmada trafik kazalarının şiddet düzeyinin tahmin edilmesi amacıyla Lojistik Regresyon, Karar Ağacı, Rastgele Orman, AdaBoost, Gradient Boosting ve XGBoost algoritmaları kullanılmıştır. Modellerin performansları doğruluk (Accuracy), kesinlik (Precision), duyarlılık (Recall), F1 skoru, Cohen's Kappa katsayısı, hata ölçütleri ve aşırı öğrenme düzeyleri bakımından karşılaştırılmıştır.

Tablo 4. Makine öğrenmesi modellerinin performans karşılaştırması

<b>Model</b>	<b>Accuracy</b>	<b>Precision</b>	<b>Recall</b>	<b>F1 Skoru</b>	<b>Cohen's Kappa</b>
Lojistik Regresyon	0,8044	0,0653	0,5455	0,1167	0,0777
Karar Ağacı	0,9397	0,1357	0,2879	0,1845	0,1573
Rastgele Orman	0,9440	0,1053	0,1818	0,1333	0,1065
AdaBoost	0,9706	0,1667	0,0606	0,0889	0,0772
Gradient Boosting	0,9605	0,2027	0,2273	0,2143	0,1941
XGBoost	0,8934	0,0947	0,4091	0,1538	0,1200

Tablo 4. incelendiğinde doğruluk oranı açısından en başarılı modelin %97,06 ile AdaBoost modeli olduğu görülmektedir. Ancak veri setindeki sınıf dengesizliği nedeniyle doğruluk oranı tek başına yeterli bir performans ölçütü değildir. Bu nedenle diğer performans kriterlerinin de değerlendirilmesi gerekmektedir.

Tablo 5. Precision değerlerine göre model sıralaması

Sıralama	Model	Precision
1	Gradient Boosting	0,2027
2	AdaBoost	0,1667
3	Karar Ağacı	0,1357
4	Rastgele Orman	0,1053
5	XGBoost	0,0947
6	Lojistik Regresyon	0,0653

Tablo 5. incelendiğinde en yüksek Precision değerinin 0,2027 ile Gradient Boosting modelinde elde edildiği görülmektedir. Bu sonuç model tarafından ölümlü olarak sınıflandırılan kazaların diğer modellere kıyasla daha yüksek doğrulukla tahmin edildiğini göstermektedir.

Tablo 6. Recall değerlerine göre model sıralaması

Sıralama	Model	Precision
1	Lojistik Regresyon	0,5455
2	XGBoost	0,4091
3	Karar Ağacı	0,2879
4	Gradient Boosting	0,2273
5	Rastgele Orman	0,1818
6	AdaBoost	0,0606

Tablo 6’da modellerin Recall sonuçları verilmiştir. En yüksek Recall değeri 0,5455 ile Lojistik Regresyon modelinde elde edilmiştir. Bu sonuç modelin ölümcül kazaların yaklaşık yarısından fazlasını doğru şekilde belirleyebildiğini göstermektedir. XGBoost modeli ise ikinci sırada yer almıştır.

Tablo 7. F1 skorlarına göre model sıralaması

Sıralama	Model	Precision
1	Gradient Boosting	0,2143
2	Karar Ağacı	0,1845
3	XGBoost	0,1538
4	Rastgele Orman	0,1333
5	Lojistik Regresyon	0,1167
6	AdaBoost	0,0889

Tablo 7. incelendiğinde en yüksek F1 skorunun 0,2143 ile Gradient Boosting modelinde elde edildiği görülmektedir. F1 skoru Precision ve Recall değerlerinin birlikte değerlendirilmesini sağladığından, dengesiz veri setlerinde model performansının değerlendirilmesinde önemli bir ölçüt olarak kabul edilmektedir. Sonuçlar Gradient Boosting modelinin bu iki performans ölçütü arasında en dengeli yapıya sahip olduğunu göstermektedir.

Tablo 8. Cohen's Kappa katsayılarına göre model sıralaması

Sıralama	Model	Precision
1	Gradient Boosting	0,1941
2	Karar Ağacı	0,1573
3	XGBoost	0,1200
4	Rastgele Orman	0,1065
5	Lojistik Regresyon	0,0777
6	AdaBoost	0,0772

Tablo 8. incelendiğinde en yüksek Cohen's Kappa katsayısının 0,1941 ile Gradient Boosting modelinde elde edildiği görülmektedir. Bu durum model ile gerçek sınıflandırmalar arasındaki uyumun diğer modellere göre daha yüksek olduğunu göstermektedir.

Tablo 9. Modellerin hata ölçütlerinin karşılaştırılması

Model	MAE	MSE
AdaBoost	0,0294	0,0294
Gradient Boosting	0,0395	0,0395
Rastgele Orman	0,0560	0,0560
Karar Ağacı	0,0603	0,0603
XGBoost	0,1066	0,1066
Lojistik Regresyon	0,1956	0,1956

Tablo 9. incelendiğinde en düşük hata değerlerinin AdaBoost modelinde elde edildiği görülmektedir. Bununla birlikte Recall ve F1 skorlarının oldukça düşük olması nedeniyle bu sonuç tek başına modelin en başarılı model olduğunu göstermemektedir.

Tablo 10. Modellerin aşırı öğrenme düzeylerinin karşılaştırılması

Model	Train-Test F1 Farkı
Gradient Boosting	0,0125
AdaBoost	0,0718
XGBoost	0,0764
Lojistik Regresyon	0,0872
Karar Ağacı	0,1226
Rastgele Orman	0,1451

Tablo 10'da modellerin aşırı öğrenme düzeyleri gösterilmektedir. Eğitim ve test F1 skorları arasındaki fark dikkate alındığında en düşük fark değerinin 0,0125 ile Gradient Boosting modelinde elde edildiği görülmektedir. Bu sonuç modelin yüksek genellenebilirliğe sahip olduğunu göstermektedir. Karar Ağacı ve Rastgele Orman modellerinde ise belirgin düzeyde aşırı öğrenme eğilimi gözlemlenmiştir.

Tablo 11. Performans ölçütlerine göre en başarılı modeller

Performans Ölçütü	En Başarılı Model	Değer
Accuracy	AdaBoost	0,9706
Precision	Gradient Boosting	0,2027
Recall	Lojistik Regresyon	0,5455
F1 Skoru	Gradient Boosting	0,2143
Cohen's Kappa	Gradient Boosting	0,1941
En Düşük Fark	Gradient Boosting	0,0125

Tablo 11 'de modellerin performans ölçütlerine göre elde ettikleri en iyi sonuçlar özetlenmiştir. Sonuçlar incelendiğinde farklı performans ölçütlerinde farklı modellerin öne çıktığı görülmektedir. AdaBoost modeli doğruluk oranı bakımından en başarılı model olurken, Lojistik Regresyon modeli ölümcül kazaları yakalama açısından en yüksek Recall değerine ulaşmıştır. Buna karşın Precision, F1 skoru, Cohen's Kappa katsayısı ve aşırı öğrenme düzeyi bakımından en başarılı sonuçlar Gradient Boosting modelinde elde edilmiştir.

Genel değerlendirme yapıldığında, trafik kazalarında ölümcül olayların tahmin edilmesinde yalnızca doğruluk oranının yeterli olmadığı görülmektedir. Veri setinin dengesiz yapısı nedeniyle Precision, Recall, F1 skoru, Cohen's Kappa katsayısı ve genellenebilirlik düzeyi birlikte değerlendirilmelidir. Bu kapsamda Gradient Boosting modeli en yüksek Precision, en yüksek F1 skoru, en yüksek Cohen's Kappa katsayısı ve en düşük aşırı öğrenme düzeyine sahip olması nedeniyle çalışmada kullanılan modeller arasında en dengeli ve en güvenilir model olarak belirlenmiştir.

## 5. TARTIŞMA VE SONUÇ

Bu çalışmada, Trabzon ilinde meydana gelen trafik kazaları hem mekânsal analiz yöntemleri hem de makine öğrenmesi algoritmaları kullanılarak çok boyutlu bir yaklaşımla incelenmiştir. Elde edilen bulgular, trafik kazalarının mekânsal olarak rastgele bir dağılım göstermediğini, aksine belirli bölgelerde yoğunlaşma eğiliminde olduğunu ortaya koymuştur. Bu durum, trafik kazalarının oluşumunda coğrafi, çevresel ve altyapısal faktörlerin önemli rol oynadığını göstermektedir.

Gerçekleştirilen mekânsal analizler sonucunda, özellikle kent merkezlerinde ve ana ulaşım arterleri boyunca kaza yoğunluğunun daha yüksek olduğu belirlenmiştir. Kernel yoğunluk analizi ile elde edilen bulgular, kazaların belirli “sıcak noktalar” etrafında kümелendiğini açık biçimde ortaya koymuştur. Bu sonuç, trafik güvenliği çalışmalarında kaynakların daha etkin kullanılabilmesi için riskli bölgelerin öncelikli olarak ele alınması gerektiğini göstermektedir. Özellikle yol geometrisinin karmaşık olduğu, trafik hacminin yüksek olduğu ve çevresel koşulların olumsuz etkiler oluşturduğu alanlarda kaza riskinin arttığı görülmektedir.

Makine öğrenmesi modellerinden elde edilen sonuçlar incelendiğinde, farklı algoritmaların kaza şiddetini tahmin etme konusunda değişen performanslar sergilediği tespit edilmiştir. Lojistik regresyon modeli daha yorumlanabilir bir yapı sunarken, karar ağacı ve özellikle topluluk öğrenme yöntemleri olan Random Forest, Gradient Boosting ve AdaBoost algoritmalarının daha yüksek tahmin başarısı sağladığı belirlenmiştir. Bu durum, trafik kazalarının oluşumunda etkili olan değişkenler arasındaki ilişkilerin doğrusal olmaktan ziyade karmaşık ve çok boyutlu bir yapıya sahip olduğunu göstermektedir. Ağaç tabanlı modellerin bu karmaşık ilişkileri daha iyi öğrenebilmesi, onların daha başarılı sonuçlar üretmesini sağlamıştır.

Veri setinde bulunan sınıf dengesizliği problemi, model performansını etkileyen önemli bir unsur olarak öne çıkmıştır. Ölümlü kazaların sayıca daha az olması nedeniyle modelin çoğunluk sınıfına yönelme eğilimi gösterdiği görülmüş, bu durum SMOTE yöntemi uygulanarak giderilmeye çalışılmıştır. SMOTE yöntemi ile veri setinin dengelenmesi sonucunda, özellikle azınlık sınıfa ait gözlemlerin tahmin edilmesinde belirgin bir iyileşme sağlanmıştır. Bu sonuç, trafik kazaları gibi nadir ancak kritik olayların analizinde veri dengeleme tekniklerinin kullanılmasının gerekliliğini ortaya koymaktadır.

Model performans ölçütleri birlikte değerlendirildiğinde, kullanılan algoritmaların genel olarak başarılı ve tutarlı sonuçlar ürettiği görülmüştür. Eğitim ve test veri setleri arasında elde edilen

performans değerlerinin birbirine yakın olması, modellerin aşırı öğrenme problemi yaşamadığını ve genellenebilir bir yapıya sahip olduğunu göstermektedir. Bu durum, geliştirilen modellerin gerçek dünya verileri üzerinde de benzer başarı düzeyini koruyabileceğini göstermesi açısından önemlidir.

Elde edilen bulgular literatürde yer alan çalışmalarla büyük ölçüde örtüşmektedir. Trafik kazalarının çok sayıda faktörün etkileşimi sonucunda ortaya çıktığı ve bu nedenle geleneksel istatistiksel yöntemlerin tek başına yeterli olmadığı, makine öğrenmesi ve mekânsal analiz yaklaşımlarının birlikte kullanılmasının daha güçlü ve kapsamlı sonuçlar sağladığı bu çalışma ile de doğrulanmıştır. Özellikle veri temelli yaklaşımların trafik güvenliği çalışmalarında giderek daha fazla önem kazandığı anlaşılmaktadır.

Bu çalışma sonucunda trafik kazalarının belirli bölgelerde yoğunlaştığı, yol ve çevre koşulları ile demografik faktörlerin kaza şiddeti üzerinde önemli etkiler oluşturduğu ve makine öğrenmesi algoritmalarının bu tür karmaşık yapıları modellemede etkili bir araç olduğu açıkça ortaya konulmuştur. Ayrıca ağaç tabanlı topluluk yöntemlerinin, diğer modellere kıyasla daha yüksek performans sergilediği ve sınıf dengesizliği probleminin giderilmesinin model başarısını önemli ölçüde artırdığı belirlenmiştir.

Çalışmanın uygulamaya yönelik önemli çıktıları da bulunmaktadır. Elde edilen sonuçlar doğrultusunda, trafik kazalarının yoğunlaştığı bölgelerde altyapı iyileştirmelerinin yapılması, trafik denetimlerinin artırılması ve veri temelli karar destek sistemlerinin geliştirilmesi önerilmektedir. Ayrıca yerel yönetimlerin ve ilgili kurumların, trafik güvenliği politikalarını oluştururken mekânsal analiz sonuçlarını dikkate almasının, daha etkili ve hedef odaklı çözümler geliştirilmesine katkı sağlayacağı değerlendirilmektedir.

Bununla birlikte çalışmanın bazı sınırlılıkları bulunmaktadır. Araştırmanın yalnızca Trabzon ili ile sınırlı olması ve belirli bir zaman aralığını kapsamaması, elde edilen sonuçların genellenebilirliğini kısmen sınırlandırmaktadır. Ayrıca veri setinde yer almayan sürücü davranışları, psikolojik faktörler ve anlık karar süreçleri gibi değişkenlerin modele dahil edilememiş olması da çalışmanın önemli kısıtları arasında yer almaktadır. Bunun yanında SMOTE yöntemi ile üretilen sentetik verilerin gerçek veri dağılımını tam olarak temsil edememe ihtimali de göz önünde bulundurulmalıdır.

Gelecek alıřmalarda daha geniř kapsamlı veri setlerinin kullanılması, farklı Őehirlerin karřılařtırmalı olarak incelenmesi ve derin ğrenme tabanlı modellerin uygulanması ile daha kapsamlı sonuçlar elde edilebilir. Ayrıca gerek zamanlı veri kaynaklarının entegrasyonu ve mikro lekli mekânsal analizlerin gerekleřtirilmesi, trafik kazalarının daha detaylı bir Őekilde incelenmesine olanak saėlayacaktır.

Sonuç olarak, bu alıřma makine ğrenmesi ve Coėrafi Bilgi Sistemleri tabanlı yaklařımların birlikte kullanımının trafik kazalarının analizi ve tahmininde etkili ve gçlü bir yöntem olduėunu ortaya koymaktadır. Elde edilen bulgular hem akademik literatüre katkı saėlamakta hem de trafik gvenliėinin artırılmasına ynelik uygulamalara ışık tutmaktadır.

## KAYNAKÇA

- Akanksha Mishra. (2024). A comprehensive review of artificial intelligence and machine learning: Concepts, trends, and applications. *International Journal of Scientific Research in Science and Technology*, 11(5), 126–142. doi:10.32628/IJSRST2411587
- Alpan, M. P. (2025). Kırsal kesimde gerçekleşen trafik kazaları ve karayolu geometrisi ilişkisi: Zonguldak örneği (Doktora tezi).
- Amiri, M. A., Afshari, S., & Soltani, A. (2025). Machine learning approaches to traffic accident severity prediction: Addressing class imbalance. *Machine Learning with Applications*, 22, 100792. doi:10.1016/J.MLWA.2025.100792
- Anselin, L. (1995). Local indicators of spatial association—LISA. *Geographical Analysis*, 27(2), 93–115. doi:10.1111/J.1538-4632.1995.TB00338.X
- Anselin, L. (2007). *Spatial regression analysis in R: A workbook*. Center for Spatially Integrated Social Science. Retrieved from <http://www.csiss.org/>
- Aydoğdu, M., & Bakırcı, M. (2021). Süleymanpaşa ilçesinde (Tekirdağ) nüfusun zamansal gelişimi ve mahallelere göre yoğunluk analizi. *Doğu Coğrafya Dergisi*, 26(46), 273–294. doi:10.17295/ATAUNIDCD.961068
- Behboudi, N., Moosavi, S., & Ramnath, R. (2024). Recent advances in traffic accident analysis and prediction: A comprehensive review of machine learning techniques. Retrieved from <https://arxiv.org/pdf/2406.13968>
- Berhanu, Y., Schröder, D., Wodajo, B. T., & Alemayehu, E. (2024). Machine learning for predictions of road traffic accidents and spatial network analysis for safe routing. *Results in Engineering*, 23, 102737. doi:10.1016/J.RINENG.2024.102737
- Bommasani, R., et al. (2021). On the opportunities and risks of foundation models.
- Borhan, M. N., Ibrahim, A. N. H., Aziz, A., & Yazid, M. R. M. (2018). The relationship between demographic, personal, and social factors of motorcyclists and risk taking behavior. *Accident Analysis & Prevention*, 121, 94–100. doi:10.1016/J.AAP.2018.09.004
- Brown, T. B., et al. (2020). Language models are few-shot learners. In *Advances in Neural Information Processing Systems* (Vol. 33, pp. 1877–1901).
- Burrough, P. A., McDonnell, R. A., & Lloyd, C. D. (n.d.). *Principles of geographical information systems* (3rd ed.).

- Cunha, J., Silva, J. S., Ribeiro, R., & Gomes, P. (2025). Sustainable road safety: Predicting traffic accident severity in Portugal using machine learning. *Sustainability*, 17(24). doi:10.3390/SU172411199
- Emniyet Genel Müdürlüğü Trafik Başkanlığı. (n.d.). İstatistikler. Retrieved April 12, 2026, from <https://www.trafik.gov.tr/istatistikler37>
- GLOBAL STATUS REPORT ON ROAD SAFETY 2018. (n.d.). World Health Organization.
- Güneş, P., Nihan, F., & Sönmez, Ö. (2026). Mekan okuma ve analiz yöntemlerinin kentsel gelişim alanlarında incelenmesi. *Ankara Üniversitesi Gayrimenkul Çalışmaları Dergisi*, 1(1), 64–88. doi:10.1155/2010/974917
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE CVPR* (pp. 770–778). doi:10.1109/CVPR.2016.90
- IBM. (n.d.). Supervised learning. Retrieved from <https://www.ibm.com/think/topics/supervised-learning>
- Jima, D., & Sipos, T. (2022). The impact of road geometric formation on traffic crash severity. *Sustainability*, 14(14), 8475. doi:10.3390/SU14148475
- Kotsyubynska, Y., et al. (2026). Machine learning and deep learning for predicting traffic crash injury severity: A systematic review. *Journal of Road Safety*, 37(1), 46–60. doi:10.33492/JRS-D-26-1-2721386
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems* (Vol. 25, pp. 1097–1105).
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521, 436–444. doi:10.1038/nature14539
- Levine, S., et al. (2020). Offline reinforcement learning: Tutorial, review, and perspectives. Retrieved from <https://arxiv.org/pdf/2005.01643>
- Li, H., & Chen, L. (2025). Traffic accident risk prediction based on deep learning. *PLOS ONE*, 20(5), e0320656. doi:10.1371/JOURNAL.PONE.0320656
- Makalesi, A., et al. (2025). Makine öğrenmesi ve derin öğrenmeye dayalı duygu analizi. *The Black Sea Journal of Sciences*, 15(2), 923–959. doi:10.31466/KFBD.1536270
- Mccarty, D., et al. (2024). Risky behaviors and road safety. *PLOS ONE*, 19(1), e0296663. doi:10.1371/JOURNAL.PONE.0296663
- Mnih, V., et al. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540), 529–533. doi:10.1038/nature14236

- Mostafa, A. M., et al. (2025). AI-based prediction of traffic crash severity. *Scientific Reports*, 15, 27468. doi:10.1038/s41598-025-10970-7
- National Highway Traffic Safety Administration. (2015). Traffic safety facts: Electronic stability control (ESC).
- O'Sullivan, D., & Unwin, D. J. (2010). *Geographic information analysis* (2nd ed.). doi:10.1002/9780470549094
- Prudencio, R. F., et al. (2023). A survey on offline reinforcement learning. doi:10.1109/TNNLS.2023.3250269
- Razzaq, K., & Shah, M. (2025). Machine learning and deep learning paradigms. *Computers*, 14(3), 93. doi:10.3390/COMPUTERS14030093
- Schulman, J., et al. (2017). Proximal policy optimization algorithms. Retrieved from <https://arxiv.org/pdf/1707.06347>
- Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (2nd ed.). MIT Press.
- Tobler, W. R. (1970). A computer movie simulating urban growth. *Economic Geography*, 46, 234. doi:10.2307/143141
- Türkiye İstatistik Kurumu. (2025). Motorlu kara taşıtları istatistikleri: Trabzon ili. <https://data.tuik.gov.tr>
- Türkiye İstatistik Kurumu. (2026). Adrese dayalı nüfus kayıt sistemi sonuçları: Trabzon ili. <https://data.tuik.gov.tr>
- Wang, F. (2006). Quantitative methods and applications in GIS. doi:10.1201/9781420004281
- World Health Organization. (2018). Global status report on road safety 2018.
- Xie, Z., & Yan, J. (2013). Detecting traffic accident clusters. *Journal of Transport Geography*, 31, 64–71. doi:10.1016/J.JTRANGEO.2013.05.009
- Yomralıoğlu, T. (2000). Coğrafi bilgi sistemleri: Temel kavramlar ve uygulamalar.
- Yumak, A., et al. (2025). Machine learning approach to accident severity patterns. *Applied Sciences*, 15(23), 12824. doi:10.3390/APP152312824
- Zhang, J., et al. (2025). A survey on self-supervised learning. *Neurocomputing*, 655, 131409. doi:10.1016/J.NEUCOM.2025.131409