T.C.
BURSA TEKNİK ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

RASTGELE ORMAN VE KARAR AĞACI YÖNTEMLERİNİN KARŞILAŞTIRILMASI: BİSİKLETLİ KAZA ANALİZİ ÖRNEK ÇALIŞMASI

YÜKSEK LİSANS TEZİ

Ebru SADIK

İnşaat Mühendisliği Anabilim Dalı

İnşaat Mühendisliği Tezli Yüksek Lisans Programı

TEMMUZ 2025

T.C.
BURSA TEKNİK ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

RASTGELE ORMAN VE KARAR AĞACI YÖNTEMLERİNİN KARŞILAŞTIRILMASI: BİSİKLETLİ KAZA ANALİZİ ÖRNEK ÇALIŞMASI

YÜKSEK LİSANS TEZİ

Ebru SADIK

(21434979021)

ORCID: 0009-0005-7272-4344

İnşaat Mühendisliği Anabilim Dalı

İnşaat Mühendisliği Tezli Yüksek Lisans Programı

Danışman: Dr. Öğr. Üyesi Nurten AKGÜN

ORCID: 0000-0003-3888-3913

TEMMUZ 2025

BTÜ, Lisansüstü Eğitim Enstitüsü'nün 21434979021 numaralı Yüksek Lisans Öğrencisi Ebru SADIK, ilgili yönetmeliklerin belirlediği gerekli tüm şartları yerine getirdikten sonra hazırladığı "RASTGELE ORMAN VE KARAR AĞACI YÖNTEMLERİNİN KARŞILAŞTIRILMASI: BİSİKLETLİ KAZA ANALİZİ ÖRNEK ÇALIŞMASI" başlıklı tezini aşağıda imzaları olan jüri önünde başarı ile sunmuştur.

Tez Danışmanı : **Dr. Öğr. Üyesi Nurten AKGÜN** .....
Bursa Teknik Üniversitesi

Jüri Üyeleri : **Dr. Öğr. Üyesi Ziya ÇAKICI** .....
İzmir Demokrasi Üniversitesi

Dr. Öğr. Üyesi Ahmet Münir ÖZDEMİR.....
Bursa Teknik Üniversitesi

Savunma Tarihi : 28.07.2025
Teslim Tarihi :

20.04.2016 tarihli Resmi Gazete'de yayımlanan Lisansüstü Eğitim ve Öğretim Yönetmeliğinin 9/2 ve 22/2 maddeleri gereğince; Bu Lisansüstü teze, Bursa Teknik Üniversitesi'nin abonesi olduğu intihal yazılım programı kullanılarak Lisansüstü Eğitim Enstitüsü'nün belirlemiş olduğu ölçütlere uygun rapor alınmıştır.

İNTİHAL BEYANI

Bu tezde görsel, işitsel ve yazılı biçimde sunulan tüm bilgi ve sonuçların akademik ve etik kurallara uyularak tarafımdan elde edildiğini, tez içinde yer alan ancak bu çalışmaya özgü olmayan tüm sonuç ve bilgileri tezde kaynak göstererek belgelediğimi, aksinin ortaya çıkması durumunda her türlü yasal sonucu kabul ettiğimi beyan ederim.

Öğrencinin Adı SOYADI: Ebru SADIK

İmzası:



Aileme...

ÖNSÖZ

Akademik yolculuğum sürecinde desteğini, bilgisini, vaktini esirgemeyen, süreç boyunca gösterdiği anlayışlı ve sabırlı yaklaşımı için çok değerli danışmanım Dr. Öğr. Üyesi Nurten AKGÜN'e ve tezimin yazılma sürecinde bana hep destek olan aileme sevgilerimi ve teşekkürlerimi sunarım.

Temmuz 2025

Ebru SADIK
(İnşaat Mühendisi)

İÇİNDEKİLER

Sayfa

İNTİHAL BEYANI ..... **v**
ÖNSÖZ ..... **vii**
İÇİNDEKİLER ..... **viii**
KISALTMALAR ..... **x**
SEMBOLLER ..... **xi**
ÇİZELGE LİSTESİ ..... **xii**
ŞEKİL LİSTESİ ..... **xiii**
ÖZET ..... **xiv**
SUMMARY ..... **xvi**
1. GİRİŞ **1**
1.1 Literatür Araştırması 1
1.2 Tezin Amacı, Kapsamı ve Hedefleri 7
2. DÖNEL KAVŞAKLAR **8**
2.1 Mini Dönel Kavşak 9
2.2 Kentsel Kompakt Dönel Kavşaklar 10
2.3 Kentsel Tek Şeritli Dönel Kavşaklar 10
2.4 Kentsel Çift Şeritli Dönel Kavşaklar 11
2.5 Kırsal Tek Şeritli Dönel Kavşaklar 11
2.6 Kırsal Çift Şeritli Dönel Kavşaklar 12
2.7 Dönel Kavşak Güvenliği 13
2.8 Dönel Kavşak Geometrik Elemanları 15
2.8.1 Yaklaşma genişliği 16
2.8.2 Giriş genişliği 16
2.8.3 İç ada çapı 17
2.8.4 Görüş mesafesi 17
2.8.5 Sapma açısı ve yarıçapı 18
3. BETİMSEL İSTATİSTİK **20**
3.1 Merkezi Eğilim ve Dağılım Ölçüleri 20
3.1.1 Ortanca (medyan) 20
3.1.2 Varyans ve standart sapma 20
3.1.3 Ortalama 20
3.1.4 Değer aralığı 21
3.1.5 En az ve en çok 21
4. Kİ KARE TESTİ **22**
5. MAKİNE ÖĞRENMESİ YÖNTEMLERİ **24**
5.1 Makine Öğrenimi Türleri 24
5.1.1 Denetimli öğrenme 24
5.1.2 Denetimsiz öğrenme 25
5.1.3 Yarı denetimli öğrenme 25
5.2 Karar Ağaçları 25

5.2.1 C4.5 ve C5.0 algoritmaları ... 28
5.2.2 CART algoritması ... 28
5.2.3 CHAID algoritması ... 29
5.2.4 QUEST algoritması ... 29
5.3 Rastgele Orman ... 29
5.3.1 Ağacın yapısı ... 31
5.3.2 Rastgele orman algoritmasının aşamaları ... 32
5.4 Rastgele Orman ve Karar Ağacı Karşılaştırması ... 33
6. MATERYAL VE YÖNTEM ... 35
7. SONUÇLAR ... 39
8. TARTIŞMA ... 48
KAYNAKLAR ... 50
ÖZGEÇMİŞ ... 55

KISALTMALAR

ABD : Amerika Birleşik Devletleri
AID : Automatic Interaction Detection
ANN : Artificial Neural Network
CART : Classification and Regression Trees
CHAID : Chi-squared Automatic Interaction Detector
DTs : Decision Trees
KNN : K-Nearest Neighbors
OHSP : Office of Highway Safety Planning
OOB : Out- Of- Bag
PART : Partial Decision Trees
RF : Random Forest
SVM : Support Vector Machine

SEMBOLLER

$\boldsymbol{\sigma^2}$ **:** Varyans

$\boldsymbol{\sigma}$ **:** Standart sapma

$\boldsymbol{B}$ **:** Ağaç sayısı

$\boldsymbol{m}$ **:** Tahminci sayısı

$\boldsymbol{n}$ **:** Gözlem sayısı

$\boldsymbol{p}$ **:** Tahminci değişken

ÇİZELGE LİSTESİ

Sayfa

Çizelge 6.1 : Parametreler..... **35**
Çizelge 6.1 (devam) : Parametreler..... **36**
Çizelge 6.2 : Parametrelerin kodlanma şekli. **36**
Çizelge 6.2 (devam) : Parametrelerin kodlanma şekli **37**
Çizelge 6.3 : Ki kare tesindeki temsil. **37**
Çizelge 7.1 : Betimsel istatistik sonuçları..... **39**
Çizelge 7.2 : Betimsel istatistik sonuçları..... **40**
Çizelge 7.3 : Betimsel istatistik sonuçları..... **40**
Çizelge 7.4 : Ki kare testi sonuçları. **43**
Çizelge 7.5 : Rastgele orman sınıflandırma modeli karışıklık matrisi sonuçları..... **44**
Çizelge 7.6 : Rastgele orman tahmin sonuçları **45**
Çizelge 7.7 : Rastgele orman özellik önem seviyesi **45**
Çizelge 7.8 : Karar ağacı tahmin sonuçları..... **47**

ŞEKİL LİSTESİ

Sayfa

Şekil 2.1 : Üç kollu dönel kavşak **8**
Şekil 2.2 : Dört ve beş kollu dönel kavşak **9**
Şekil 2.3 : Mini dönel kavşak **9**
Şekil 2.4 : Kentsel kompakt dönel kavşak **10**
Şekil 2.5 : Kentsel tek şeritli dönel kavşak **10**
Şekil 2.6 : Kentsel çift şeritli dönel kavşak **11**
Şekil 2.7 : Kırsal tek şeritli dönel kavşak **12**
Şekil 2.8 : Kırsal çift şeritli dönel kavşak **12**
Şekil 2.9 : Geleneksel kavşak ve dönel kavşak çakışma noktası karşılaştırması **13**
Şekil 2.10 : Geleneksel kavşaklarda bisikletlilere özgü çakışmalar **14**
Şekil 2.11 : Dönel kavşaklarda bisiklet çakışmaları **15**
Şekil 2.12 : Modern dönel kavşak geometrik elemanları **16**
Şekil 2.13 : İç ada çapı (merkez ada çapı) **17**
Şekil 2.14 : Kavşak görüş mesafesi **18**
Şekil 2.15 : Sapma açısı ve yarıçapı **18**
Şekil 2.16 : Trafik doğrultusunda sapma sağlanması **19**
Şekil 5.1 : Makine öğreniminin çalışma prensibi **24**
Şekil 5.2 : Karar ağacı kök, düğüm ve dal yapısı **26**
Şekil 5.3 : Rastgele orman modeli **30**
Şekil 5.4 : Rastgele ormanı oluşturan ağaçların yapısı **31**
Şekil 5.5 : Orijinal veri setinin ayrılma şeması **32**
Şekil 6.1 : Karışıklık matrisi **38**
Şekil 7.1 : Rastgele orman oob hata oranı **44**
Şekil 7.2 : Rastgele orman özellik önem seviyesi **46**
Şekil 7.3 : Karar ağacı model özeti **46**
Şekil 7.4 : Karar ağacı şeması **47**

RASTGELE ORMAN VE KARAR AĞACI YÖNTEMLERİNİN KARŞILAŞTIRILMASI: BİSİKLETLİ KAZA ANALİZİ ÖRNEK ÇALIŞMASI

ÖZET

Kaza analizlerinde geleneksel regresyon yerine makine öğrenimi algoritmaları tercih edilmeye başlanmıştır. Makine öğrenimi algoritmaları bir öğrenme sürecine sahiptir bunun yanında sınıflandırma ve tahmin modelleri kurabilir, büyük ve küçük veri setleriyle çalışabilirler. Bu çalışmada dönel kavşaklarda meydana gelmiş bisikletli kazaları analiz edilmiştir. Veri setindeki bisikletli yaralanma derecesini bağımsız değişkenlere göre sınıflandırmak için makine öğrenmesi yöntemlerinden ikisi olan Karar Ağacı ve Rastgele Orman algoritmaları kullanılmıştır. Literatürde yapılan çalışmalar sınıflandırma açısından Karar Ağacı ve Rastgele Orman algoritmalarının yüksek doğruluk oranı verdiğini göstermektedir. Çalışmada Karar Ağacı ve Rastgele Orman algoritmalarının güçlü bir sınıflandırıcı olmasından yararlanarak hangi bağımsız değişkenin yaralanma derecesi üzerinde ne kadar etkisi olduğunu göstermek amaçlanmıştır. Ayrıca ortaya çıkan sonuçlarda iki makine öğrenmesi yönteminin karşılaştırması da yapılmıştır. Motorlu araçlar açısından kaza yaralanma derecesini büyük ölçüde azaltan dönel kavşaklar aynı etkiyi bisikletliler için göstermemektedir. Bisiklet sürücülerinin dönel kavşaklarda dezavantajlı durumda bulunmasından dolayı bisikletli kazaları analiz edilmiştir. Analizi gerçekleştirmek için 439 adet örnek kullanılmıştır. Veri seti İngiltere'nin Newcastle kentinde farklı dönel kavşaklarda meydana gelen bisikletli-motorlu kaza bilgilerini içermektedir. Veride yaralanma derecesi hafif ve ağır olmak üzere 2 kategoriye ayrılmıştır. Analizde yaralanma derecesi bağımlı değişkendir; Bağımsız değişkenler ise dönel kavşak geometrik elemanları, sürücülere ait sosyodemografik bilgiler, sürücü davranışı, hız limiti, hava koşullarıdır. Veri setinde bağımlı değişken olan yaralanma derecesi kategoriktir ve sınıflandırma işlemi yapılmıştır. Analizlerin gerçekleştirilmesi için R Programı tercih edilmiştir. Karar Ağacı ve Rastgele Orman birer denetimli öğrenme algoritmasıdır. Karar Ağacı ve Rastgele Orman algoritmalarında eğitim ve test için veriler %70-%30 şeklinde ayrılmıştır. Sınıflandırma doğruluğunu ölçmek için karışıklık matrisi kullanılmıştır. Karışıklık matrisinde, kullanılan yöntemin doğru ve yanlış tahmin ettiği değerler sayısal olarak ortaya konulur. Kullanılan algoritmanın doğruluk oranının yüksek olması iyi bir sınıflandırma algoritması seçimi için önemlidir. Analiz sonucunda Karar Ağacı %86,15 ve Rastgele Orman %83,33 sınıflandırma doğruluğu ortaya koymuştur. Literatürde genellikle Rastgele Ormanın Karar Ağacına göre daha yüksek doğruluk oranı verdiği görülmüştür. Bu çalışmada Rastgele Ormanın daha düşük sonuç vermesinin sebebinin veri seti kaynaklı olabileceği düşünülmektedir ve her seferinde seçilen %70'lik eğitim verisindeki değişim doğruluk oranını etkileyebilir. Kaza analizlerinde net bir makine öğrenimi algoritması önerilmesi doğru değildir. İleride yapılacak kaza analizi çalışmalarında farklı veri setleri üzerinde farklı makine öğrenimi algoritmaları da denenmelidir.

Anahtar kelimeler : Kaza analizi, Bisikletli, Makine öğrenmesi, Yaralanma derecesi, Rastgele orman, Karar ağacı.

COMPARISON OF RANDOM FOREST AND DECISION TREE METHODS: STUDY OF CYCLIST ACCIDENT ANALYSIS

SUMMARY

Machine learning algorithms have become increasingly popular in accident analysis, replacing traditional regression. Machine learning algorithms have a learning process and can build classification and prediction models, and can work with both large and small datasets. This study analyzed cyclist accidents that occurred at roundabouts. Two machine learning methods, Decision Tree and Random Forest algorithms, were used to classify cyclist injury severity in the dataset based on independent variables. Literature studies indicate that Decision Tree and Random Forest algorithms provide high accuracy in classification. This study aimed to demonstrate the impact of each independent variable on the injury severity by leveraging the robustness of Decision Tree and Random Forest algorithms. The results also compared the two machine learning methods. Roundabouts, which significantly reduce the injury severity for motor vehicles, do not have the same effect on cyclists. Because cyclists are disadvantaged at roundabouts, cyclist accidents were analyzed. 439 samples were used to perform the analysis. The dataset contains information on cyclist-motorist accidents occurring at various roundabouts in Newcastle, England. The data is divided into two categories: mild and severe. In the analysis, the degree of injury is the dependent variable; the independent variables are the roundabout geometric elements, sociodemographic information of the drivers, driver behavior, speed limit, and weather conditions. The dependent variable in the dataset, degree of injury, is categorical, and classification is performed. R software was used to perform the analyses. Decision Tree and Random Forest are both supervised learning algorithms. Data are split into 70% and 30% for training and testing in Decision Tree and Random Forest algorithms. A confusion matrix is used to measure classification accuracy. The confusion matrix numerically displays the values correctly and incorrectly predicted by the method used. A high algorithmic accuracy is crucial for selecting a good classification algorithm. The analysis revealed a classification accuracy of 86.15% for Decision Tree and 83.33% for Random Forest. In the literature, it is generally observed that Random Forest yields higher accuracy than Decision Tree. The lower results obtained with Random Forest in this study are thought to be due to the dataset, and the variation in the selected 70% training data each time may affect the accuracy rate. It is not appropriate to recommend a single machine learning algorithm for accident analysis. Future accident analysis studies should also test different machine learning algorithms on different datasets.

Keywords : Accident analysis, Cyclist, Machine learning, Casualty severity, Random forest, Decision tree.

1. GİRİŞ

Günümüzde trafik kazaları çok ciddi bir problem haline gelmiştir. Kazalar maddi ve manevi boyutta önemli kayıplar meydana getirir. Bu kayıplar kişi ve toplum için önemli bir sorun yarattığı gibi ülke ekonomisi için de önemli sorunlar meydana getirir. Trafik kazalarına çeşitli etkenler sebep olabilir. Bunlar insan, taşıt, yol ve çevre olarak sıralanabilir (Tuncuk, 2004).

Geleneksel kavşaklarla kıyaslandığında trafik kaza sayısının ve yaralanmalı kaza oranının büyük ölçüde azaldığı dönel kavşaklar motorlu araçlar için güvenli kabul edilmektedir. Ancak aynı etkiyi bisikletli kullanıcılar için sağlamamaktadır (Robinson ve diğ, 2000). Bisiklet kullanımının avantajları vardır örneğin ucuzdur, çevre dostudur, alandan tasarruf sağlar ve kentsel tıkanıklığın giderilmesi için ulaştırma sistemlerine fayda sağlar (Cornet ve Hook, 2025).

Trafik kaza yaralanma derecesini etkileyen pek çok parametre bulunmaktadır. Bu parametreleri belirlemek karar vericiler için de yol gösterici olacaktır. Trafik kaza yaralanma derecesi son zamanlarda geleneksel regresyon yöntemleri yerine makine öğrenmesi algoritmalarıyla tahmin edilmeye başlanmıştır.

Trafik kaza yaralanma derecesini makine öğrenmesi yöntemlerine göre sınıflandırmak için literatürde çeşitli çalışmalar yapılmıştır. Çeşitli makine öğrenimi algoritmaları ve farklı veri setleri kullanılan bu çalışmalarda farklı makine öğrenmesi yöntemleri en yüksek sınıflandırma doğruluğunu vermiştir.

Bu çalışmada dönel kavşaklarda dezavantajlı durumda bulunan bisikletlilerin kaza yaralanma derecesi ve diğer faktörler, makine öğrenmesi yöntemlerinden ikisi olan Karar Ağacı ve Rastgele Orman algoritmalarıyla sınıflandırılmış ve tahmin modeli oluşturulmuştur. Ardından iki algoritmanın doğruluk oranları karşılaştırılmıştır.

1.1 Literatür Araştırması

Bolat ve diğ. (2022), çalışmalarında Kahramanmaraş merkez ve ilçelerinde gerçekleşen 7929 adet trafik kazası kaydını makine öğrenmesi yöntemleriyle analiz

etmişlerdir. Veri seti 2017 ile 2020 yılları arasında gerçekleşen ve 28 adet öznitelikten oluşan kaza kayıtlarını içermektedir. Kaza analizinde görsel karşılaştırma, sınıflandırma ve birliktelik kuralı yöntemleri kullanılmıştır. Trafik kazaları sonuçlarını içeren veri seti Bayes Ağları, Destek Vektör Makinesi, K-En Yakın Komşuluk, PART (Kısmi Karar Ağaçları), AdaboostM1, J48 ve Rastgele Orman algoritmaları kullanılarak sınıflandırılmıştır. Analiz sonucunda en iyi performansı J48 algoritması vermiştir. Diğer algoritmalara göre en düşük performansı Bayes Ağları algoritması vermiştir.

Almamlook ve diğ. (2019) trafik kaza analizinde en iyi performans gösteren modeli belirlemeyi amaçlamıştır. AdaBoost, Lojistik Regresyon, Bayes Ağları ve Rastgele Orman ile trafik kaza verileri analiz edilmiştir. Veriler Office of Highway Safety Planning (OHSP)' den alınmıştır. Bu çalışmanın veri seti Michigan'da 2010-2016 yılları arasında meydana gelen trafik kazalarıyla ilişkilidir. Çalışmadaki veri seti toplam 271.563 adet araç trafik kazası kaydını içermektedir. Veri dengesizliği için SMOTE algoritması kullanılmıştır. Rastgele Orman algoritmasının kaza şiddeti tahmininde diğer modellerden daha iyi performans sunduğu ve daha güvenilir sonuç verdiği gözlemlenmiştir.

Krishnaven ve Hemalatha (2011) tarafından yürütülen çalışmada, trafik kazaları sırasında meydana gelen yaralanmaların seviyelerini belirleyen tahmin modeli üretmek için bazı sınıflandırma modellerini ele almıştır. Hong Kong hükümetinin Ulaştırma Bakanlığı tarafından sağlanan kaza verileri analiz edilmiştir. Naive Bayes, AdaBoostM1, PART, J48, Rastgele Orman sınıflandırıcıları ile sınıflandırma işlemi yapılmıştır. En iyi performansı Rastgele Orman yöntemi göstermiştir.

Yan ve Shen (2022) trafik kazası yaralanma derecesi modellerinin, trafik güvenliği yönetimi ve kontrolü için önemli olduğunu belirtmiştir. Bu kapsamda, yüksek tahmin doğruluğu ve model yorumlanabilirliği elde etmek için, Rastgele Orman ve Bayes Optimizasyonu'nu bütünleştiren hibrit bir model geliştirmişlerdir. Bu çalışma, ABD deki kaza veri setine dayanmaktadır ve hedef veri olarak Pennsylvania eyaletinin Montgomery ilçesindeki kaza verileri seçilmiştir. Her kaza kaydı; kaza yeri, hava durumu, zaman ve önemli noktalar dâhil olmak üzere çok çeşitli veri öznitelikleriyle tanımlanır. Çalışmanın sonucunda, Bayes Optimizasyonu ve Rastgele Orman yöntemlerinin hibrit versiyonunun geleneksel algoritmalardan daha yüksek doğruluk sağladığı belirlenmiştir.

Chen ve Chen (2020), 2015 ile 2019 yılları arasında Tayvan otoyollarında meydana gelen kazaları incelemeyi amaçlamışlardır. İstatistiksel ve parametrik olmayan veri madenciliği tekniği olan Lojistik Regresyon, CART (Classification and Regression Tree) ve Rastgele Orman kullanılarak, yöntemlerin tahmin yeteneklerini ve doğruluklarını karşılaştırmışlardır. Bu çalışmada, tahmin performansının doğruluk, duyarlılık ve özgüllük olmak üzere üç değerlendirme ölçütü bulunmaktadır. Rastgele Orman'ın üç model arasında en etkili tahmin modeli olduğu bulunmuştur.

Yavuz ve diğ. (2021) uyguladığı çalışmada, Antalya merkezinde ve ilçelerinde 2012'den 2016' ya kadar meydana gelen 3181 örnekten oluşan kaza veri setini analiz etmiştir. Kazalar ölümlü ve yaralanmalı olarak gerçekleşmiştir. Çalışmada makine öğrenmesi yöntemleri kullanılarak sınıflandırma analizi yapılmıştır. Çalışmada 10 adet bağımsız değişken kullanılmıştır. Bağımlı değişken ise kaza türüdür. Bu çalışmada Lojistik Regresyon Ağaçları, Rassal Orman, Karar Ağaçları, Rassal Ağaç, J48 Ağaçları, Basit CART ve Naive Bayes yöntemleri uygulanmıştır. Kazaların analiz edilmesi için WEKA kullanılmıştır. Veri seti için üç çeşit karşılaştırma kriteri kullanılmıştır. Makine öğrenmesi algoritmalarının performanslarının farklı kriterlere göre kıyaslanması sonucunda Naive Bayes kaza verilerini en yüksek doğrulukla sınıflandıran yöntem olmuştur. Ardından en etkili sonuçları lojistik regresyon ve karar ağaçlarının verdiği görülmüştür.

Iranitalab ve diğ. (2017), 2012 ve 2015 yılları arasında ABD'nin Nebraska kentinde meydana gelen 68.448 iki araçlı kaza (bisikletler hariç) bilgisine sahip veri setini makine öğrenmesi yöntemleri ile analiz etmiştir. İlk olarak trafik kazalarının hasar seviyelerini tahmin etmede Support Vector Machine, Multinomial Logit, Random Forest ve Nearest Neighbor Classification dâhil olmak üzere dört yöntemin performansı karşılaştırılmıştır. Nearest Neighbor Classification'ın genel olarak ve ağır hasarlı çarpışmalarda en iyi tahmin performansına sahip olduğu belirlenmiştir. Random Forest ve Support Vector Machine sonraki en iyi iki sonuca sahiptir. Ancak, Multinomial Logit en zayıf yöntemdir. Kümeleme yöntemleri, Support Vector Machine 'in tahmin sonuçlarını etkilememiştir. Diğer yandan kümeleme yönteminin, Multinomial Logit, Nearest Neighbor Classification ve Random Forest'ın tahmin performansını iyileştirdiği gözlenmiştir. Gizli sınıf kümeleme ise, Multinomial Logit ve Random Forest yöntemlerinin performansını yükseltmiş, ancak Nearest Neighbor Classification'ın performansını zayıflattığı görülmüştür.

Gan ve Weng (2020) çalışmalarında, trafik tehlike unsurlarının dâhil olduğu veya olmadığı otoyollardaki kaza yaralanmalarının seviyesini tahmin etmek için Multinomial Logit ve Random Forest yöntemlerini kullanmışlardır. Tahmin modelleri, Ocak 2017'den Nisan 2019'a kadar olan Çin otoyol kaza verileriyle oluşturulmuştur. Model performansını değerlendirmek için kesinlik, geri çağırma ve F1 puanı kullanılmıştır. Otoyollarda meydana gelen trafik tehlike unsurlarını içeren ve içermeyen kazalar için Random Forest yöntemi, sürücü yaralanmasının ciddiyetini Multinomial Logit yönteminden daha doğru bir şekilde tahmin etmiştir.

Çelik ve Selvi (2022) çalışmalarında, kazaların oluşmasına neden olan etkenlerin araştırılmasını ve kaza şiddeti tahmin modelinin doğru bir şekilde geliştirilmesini önemli görmüştür. Trafik kaza verileri Teksas Austin, San Antonio ve Dallas şehirlerinden alınmıştır. Kazalar 2011-2021 yılları arasında gerçekleşmiştir. Lojistik Regresyon, Deep Learning, XGBoost, Support Vector Machine, Random Forest ve K- Nearest Neighbor makine öğrenimi teknikleri uygulamış ve tahmin sonuçlarını karşılaştırmıştır. Sonuçta en iyi performans gösteren algoritma Lojistik Regresyon olmuştur.

Umer ve diğ. (2020) ABD'de 2016- 2020 yılları arasında 49 eyaleti kapsayan, ülke çapında yaklaşık 3,5 milyon araç kazasını dikkate alan kaza şiddeti tahmin modelleri geliştirdi. Bu çalışmada Random Forest, AdaBoost, Extra Tree ve Gradient Boosting tekniklerinin yanı sıra iki klasik istatistiksel model olan Logistic Regression ve Stochastic Gradient Descent yöntemlerini oylama sınıflandırıcıları olarak trafik kazası şiddetinin tahmini için karşılaştırılmıştır. Analiz, Random Forest'ın en yüksek performans gösteren model olduğunu kanıtlamıştır (0,974 doğruluk, 0,954 hassasiyet, 0,930 geri çağırma ve 0,942 F puanı).

Al-Moqri ve diğ. (2020) tarafından yürütülen çalışmada, Yemen'deki araç kazalarının yaralanma şiddetini sınıflandırmak ve tahmin etmek için makine öğrenimi algoritmaları kullanılmıştır. Birincil amaç, yaralanma şiddetinin önde gelen nedenlerini değerlendirmekti. Seçilen makine öğrenme algoritmaları geleneksel istatistiksel yöntemlerle karşılaştırıldı. İki ana hastaneden iki ay içinde (Ağustos-Ekim 2015) toplanan veriler, 156 adet yerden bildirilen araç kazalarının 128 yaralı hastasını içeriyordu. Veriler üç yaralanma şiddeti kategorisine ayrılmıştır: Şiddetli, ciddi ve hafif. Veriler, SMOTE kullanılarak dengelenmiştir. Yaralanmaların ciddiyetini modellemek için Multinomial Logit modelinin performansını beş makine

öğrenimi sınıflandırıcısı, Naive Bayes, J48, Decision Tree, Random Forest, Support Vector Machine ve Multilevel Perceptron ile karşılaştırdı. Sonuçlar, makine öğrenimi tabanlı algoritmaların çoğunun, trafik yaralanmasının ciddiyetini tahmin etmede ve sınıflandırmada klasik istatistiksel yönteme göre daha iyi performans gösterdiğini belgeledi. Beş sınıflandırıcıdan Random Forest, %94,84 doğrulukla en iyi sınıflandırıcıdır. Kazanın yaralanma ciddiyetindeki en kritik faktörler yol tipi, toplam yaralı sayısı, kaza tipi, yol kullanıcısı, acil servise ulaşım yolu ve çarpışma eylemiydi.

Panicker ve Ramadurai (2020) iki tekerlekli motorlu araç sürücülerinin yaralanma ciddiyetini makine öğrenimi modeli geliştirerek tahmin etmişler ve böylece ciddiyete katkıda bulunan faktörleri belirlemişlerdir. 2016'dan 2018'e kadar Chennai'de meydana gelen 7654 adet kaza vakası analiz edilmiştir. Random Forest ve Conditional Random Forest yöntemlerinin performans analizi yaralanma şiddeti tahmininde kullanılmış ve her biri Ordered Probit ile karşılaştırılmıştır. Conditional Random Forest yöntemi, yaralanma ciddiyetini tahmin etmede hem Random Forest hem de Ordered Probit modellerinden daha iyi performans göstermiştir.

Lee ve diğ. (2020) çalışmalarında yağışlı hava şartlarında meydana gelen kaza şiddetini değerlendirmek için Random Forest, Neural Network ve Decision Tree tekniklerini kullanmıştır. Seul, Kore'de bulunan Naebu Otoyolu'nda 2007 – 2015 yılları arasında gerçekleşen 518 adet kaza dikkate alınmıştır. Model değerlendirmesi hata oranı, ortalama karesel hata ve ortalama karesel hatanın kök tahmini ölçütleriyle değerlendirilmiştir. Uygulanan üç makine öğrenmesi yöntemlerinden en iyi performansı Random Forest vermiştir.

Zhang ve diğ. (2018) tarafından yapılan çalışmada, ABD'de ki Florida Eyaleti'nden kaza yaralanması şiddeti verileri kullanılmıştır. En yaygın kullanılan iki istatistiksel model olan Ordered Probit ve Multinomial Logit yöntemleri ve ayrıca K-Nearest Neighbors, Decision Tree, Support Vector Machine ve Random Forest dâhil olmak üzere dört popüler makine öğrenmesi yöntemi analizde uygulanmıştır. Sonuçlar, makine öğrenmesi yöntemlerinin klasik istatistiksel tahmin yöntemlerinden daha yüksek tahmin doğruluğuna sahip olduğunu göstermiştir. Random Forest yöntemi, genel ve ciddi kazalarda en iyi tahmine sahipken, Ordered Probit model en zayıf performansa sahiptir. Ancak, makine öğrenmesi yöntemlerinin farklı veri ve

değişken grupları için aşırı uyum sorunu olduğu tespit edilmiştir. Bu nedenle uygulanacak makine öğrenmesi yönteminin dikkatle seçilmesi gerekmektedir.

Abraham ve diğ. (2005) tarafından ABD'de 1995'ten 2000 yılına kadar gerçekleşen ve beş yaralanma şiddeti sınıfına ayırdıkları 417.670 kazanın yaralanma şiddetini sınıflandırmak için dört farklı makine öğrenimi tekniği kullanmışlardır. Bunlar, hibrit öğrenmeyi kullanan Artificial Neural Networks (ANN, yapay sinir ağları), Decision Tree (karar ağaçları), Support Vector Machines (Destek vektör makineler, SVM), Decision Tree-ANN (DTANN)'dir. İş göremez hale getirmeyen yaralanma, İş göremez hale getiren yaralanma ve ölümcül yaralanma için hibrit algoritma; sinir ağları, karar ağaçları ve destek vektör makinelerinden daha iyi sonuç vermiştir. Yaralanma olmayan ve olası yaralanma sınıfları için hibrit algoritma, sinir ağlarından daha iyi performans göstermiştir. Karar ağaçları yaralanma olmayan ve olası yaralanma sınıflarını en iyi şekilde modellemiştir. Sonuç olarak Hibrit karar ağacı-sinir ağı yaklaşımı diğer algoritmalardan daha iyi sonuç vermiştir.

Akgün (2023) tarafından 439 adet kaza verisi incelenmiştir. Veriler İngilterenin karma kullanılan dönel kavşaklarından alınmıştır. Veri seti bisikletli kazalarını içermektedir. Çalışmada Kaza yaralanma şiddeti seviyesini Random Forest ve Support Vector Machine Algoritmalarıyla tahmin etmiştir ve iki modelin performanslarını karşılaştırmıştır. Random Forest algoritması %88,6 sınıflandırma doğruluğu ile tahmin modeli oluşturmuştur. Bu oran Support Vector Machine algoritması içinse %84,73' tür. Yanlış tahmin edilen veri sayısı, RF yönteminde 18; SVM yönteminde 20'dir. İki model de yüksek doğruluk oranı vermiştir.

Makine öğrenmesi yöntemi ile kaza analizi alanında ki genel kanı, en yaygın olarak kullanılan ve en yüksek performans veren yöntemin uygulanması gerektiğidir. Literatürdeki çeşitli çalışmaların sonuçlarına göre en iyi performans veren dört yöntem, performans ve doğruluk değerleri göz önünde bulundurularak en çoktan en aza doğru sıralı bir şekilde, Random Forest, Support Vector Machine, Decision Tree, ve K-Nearest Neighbor metodlarıdır (Santos ve diğ, 2022). Bu nedenle, bu tez çalışmasında, en yüksek performans veren Rastgele Orman ve Karar Ağacı yöntemleriyle dönel kavşaklarda meydana gelen bisikletli kazalarının yaralanma derecesinin sınıflandırılması amaçlanmıştır.

1.2 Tezin Amacı, Kapsamı ve Hedefleri

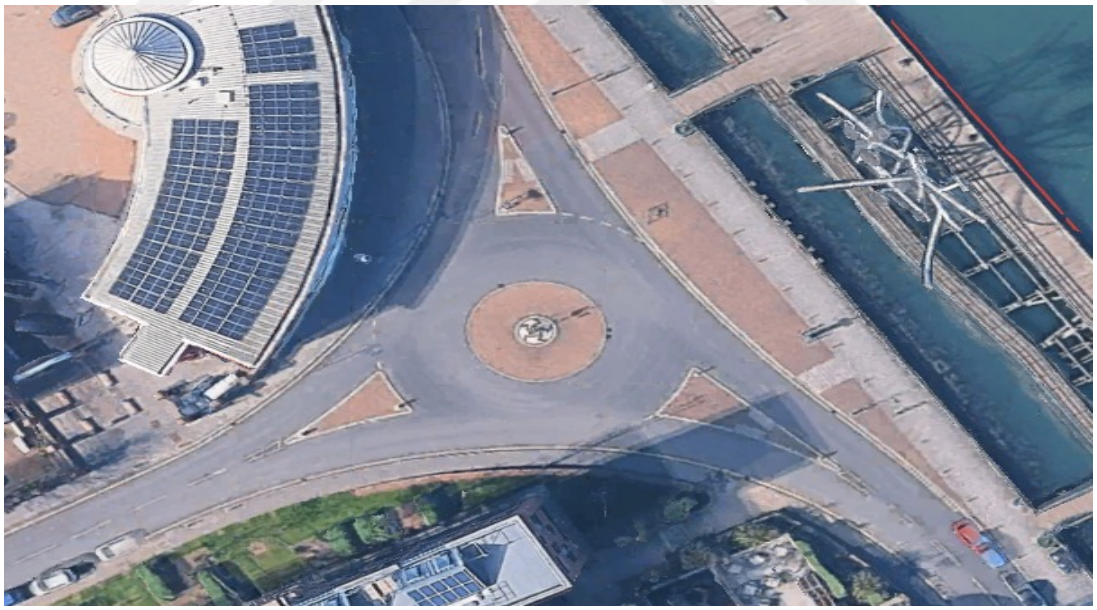
Dönel kavşaklar, motorlu taşıtlar için güvenli kabul edilmesine rağmen bisikletliler için aynı güvenlik etkisine sahip değildir. Bu nedenle çalışmada kullanılan veri seti dönel kavşaklarda meydana gelmiş bisiklet kazalarında kaydedilen faktörleri içermektedir. Bu veri setinin makine öğrenmesi yöntemleri ile analiz edilmesi amaçlanmıştır. Kazaların yaralanma derecesini bağımsız değişkenlere göre sınıflandırmak için makine öğrenmesi yöntemlerinden Rastgele Orman ve Karar Ağacı algoritmaları kullanılmış ve bu iki algoritmanın doğruluk sonuçları karşılaştırılmıştır. Çalışmanın hedefleri aşağıdaki gibi belirlenmiştir:

- Veri setinde bulunan faktörleri betimsel istatistik yöntemi ile incelemek,
- Her bir faktör arasındaki ilişkiyi Ki-Kare testi ile tespit etmek,
- Faktörleri ve hedef parametreyi makine öğrenmesi yöntemleri kullanarak sınıflandırmak ve tahmin modeli oluşturmak,
- Kullanılan iki algoritmanın doğruluk sonuçlarını karşılaştırmak.

2. DÖNEL KAVŞAKLAR

Dönel kavşak, kavşağın orta bölümünde trafiğin çevresinde seyrettiği bir ada bulunması ile oluşan kavşak tipidir (Şekil 2.1, Şekil 2.2). Tek yönlü ilerleyen bir trafik akımı sağlayan dönel kavşaklarda ayrılma ve birleşmeler bir örülme alanı yardımıyla oluşmaktadır. Dönel kavşak içerisinde kafa kafaya çarpışmalar meydana gelmemektedir. Tüm birleşme ve ayrılmalar teğetseldir. Dönel kavşaklar imalat yönünden eşdüzey kavşaklara göre daha ucuza yapılır (İnançlı, 2012).

Modern dönel kavşaklarda hareket yaklaşım kollarından gelen araçların dönel kavşak içindeki araca yol vermesi ile sağlanır. Modern dönel kavşaklarda kavşak içinde durmak veya park etmek yasaktır. Merkez ada üzerinden yaya geçişi uygun görülmez ve dönel kavşaklarda hızı azaltmak için trafik adalarına ihtiyaç vardır (İnançlı, 2012).



Şekil 2.1 : Üç kollu dönel kavşak.

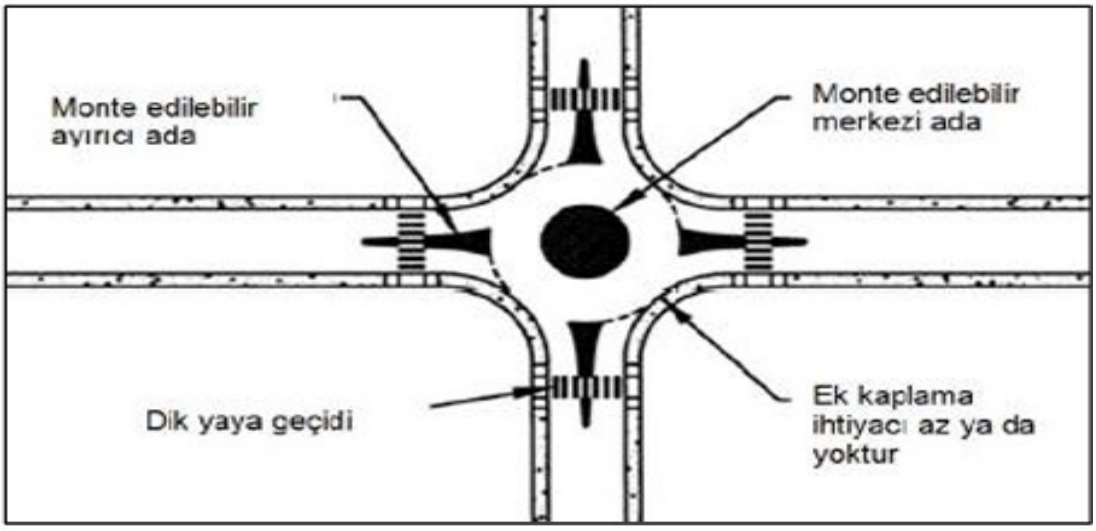


Şekil 2.2 : Dört ve beş kollu dönel kavşak.

Modern dönel kavşaklar kendi içinde gruplara ayrılır. Bu gruplara ayrılma işlemi dönel kavşağın konumu, şerit sayısı ve büyüklüğüne göre yapılır (Robinson ve diğ, 2000).

2.1 Mini Dönel Kavşak

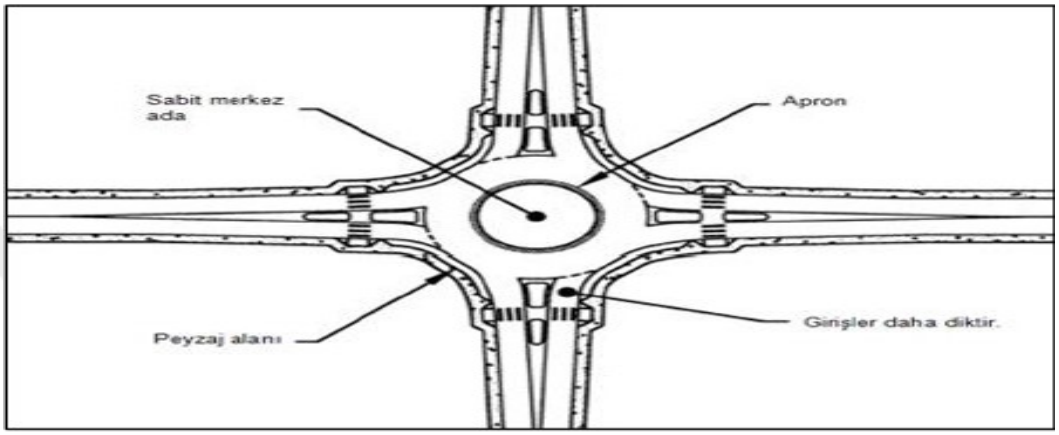
Bu tipteki dönel kavşaklar (Şekil 2.3), küçük kavşakladır ve şehir içinde hızın düşük olduğu alanlarda kullanılır. Çalışma hızları ortalama 60 km/sa (35 mil/sa) veya daha azdır. Küçük boyutlu olduğundan yayalar için daha güvenlidir. Yenileme çalışmaları daha ucuza gerçekleşir (Robinson ve diğ, 2000).



Şekil 2.3 : Mini dönel kavşak (Öğütveren, 2019).

2.2 Kentsel Kompakt Dönel Kavşaklar

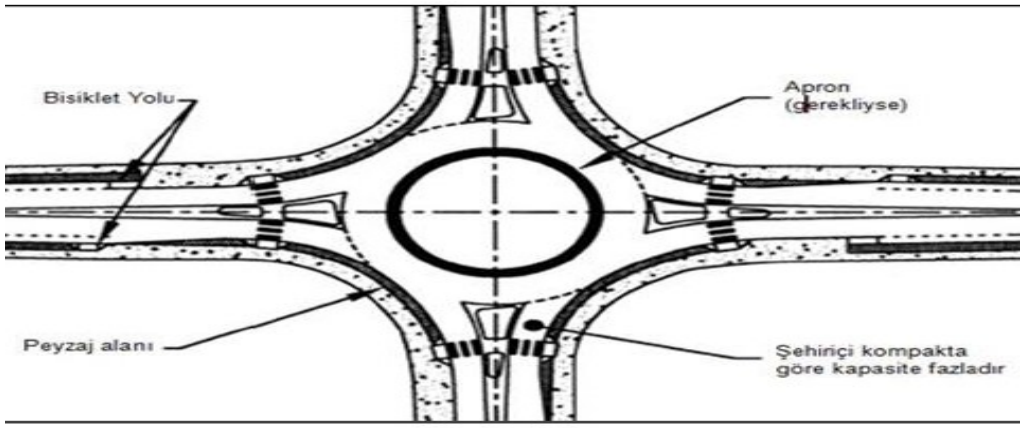
Bu dönel kavşak türü yayaların ve bisikletlilerin kullanımına uygun olarak tasarlanmıştır. Kentsel kompakt dönel kavşaklar etkili dönel kavşakların tüm tasarım gerekliliklerini karşılar. Kapasite kentsel kompakt dönel kavşaklar için kritik bir tasarım hedefi değildir (Robinson ve diğ, 2000). Şekil 2.4'de kentsel kompakt dönel kavşak gösterilmiştir.



Şekil 2.4 : Kentsel kompakt dönel kavşak (Öğütveren, 2019).

2.3 Kentsel Tek Şeritli Dönel Kavşaklar

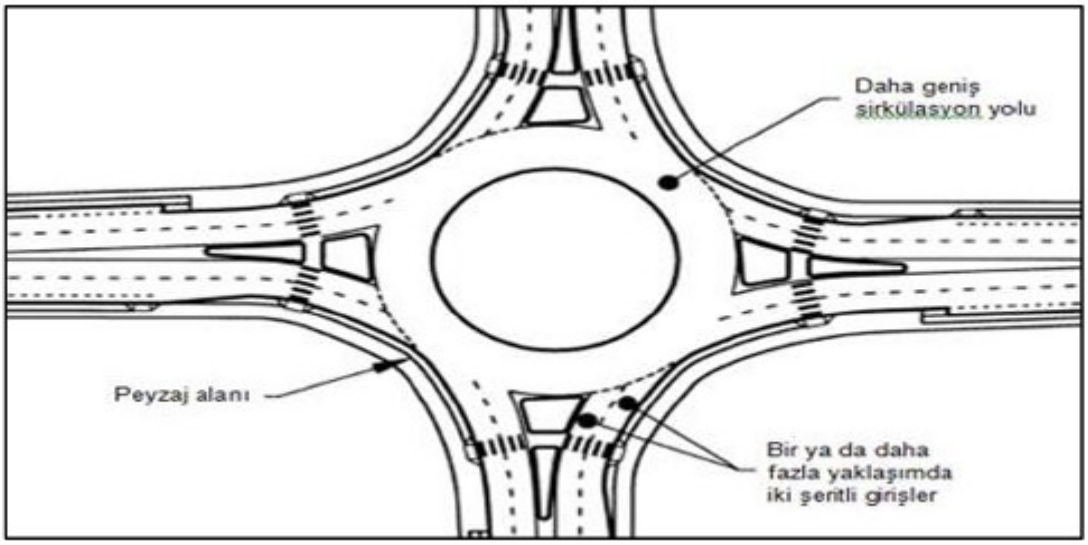
Tüm kollarda tek şeritli giriş ve tek şeritli bir dolaşım yolu vardır (Şekil 2.5). Kentsel kompakt dönel kavşaklara göre daha büyük dairsel dış çapa sahiptir. Tasarımları gereği; girişte dolaşım yolunda ve çıkışta daha yüksek hızlara olanak tanır (Robinson ve diğ, 2000).



Şekil 2.5 : Kentsel tek şeritli dönel kavşak (Öğütveren, 2019).

2.4 Kentsel Çift Şeritli Dönel Kavşaklar

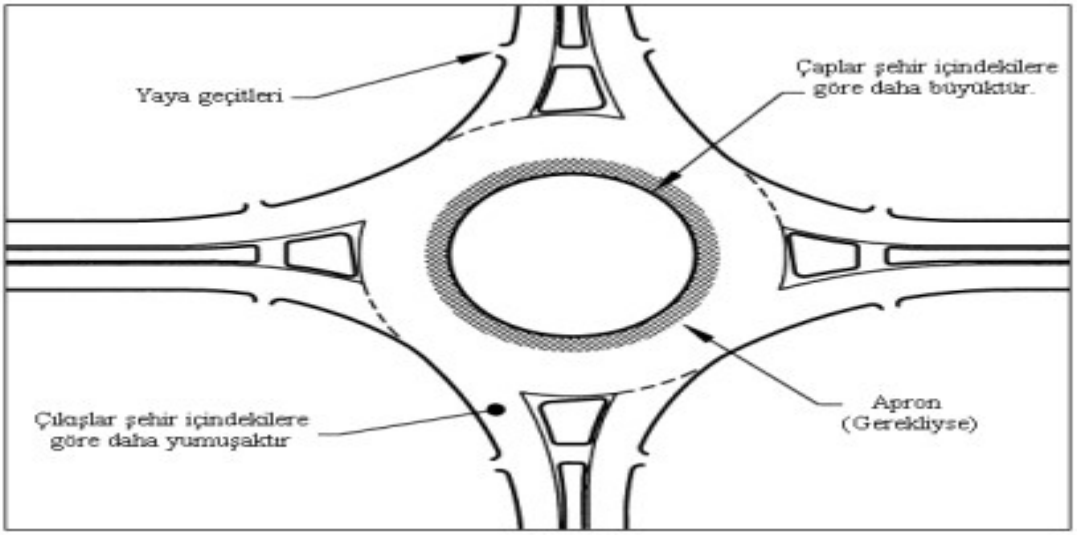
Bu tür kavşaklarda en az bir giriş çift şeritli olarak tasarlanır (Şekil 2.6). Yaklaşım kollarının giriş şeridi sayısı birden ikiye genişleyen kavşaklar da bu kategoridedir. Dolaşım yolları yan yana bir kaç araç hareket ettiği için daha geniş tasarlanır. Kentsel tek şeritli dönel kavşaklara girişte, dolaşım yolunda ve çıkıştaki hızları benzerlik gösterir. Bu tür kavşaklarda dönel kavşak boyunca araç hızlarının tutarlı olması önem arzeder. Kentsel çift şeritli dönel kavşaklar yaya ve bisikletli trafiğinde de kullanıldıkları için peyzaj alanına sahiptir (Robinson ve diğ, 2000).



Şekil 2.6 : Kentsel çift şeritli dönel kavşak (Öğütveren, 2019).

2.5 Kırsal Tek Şeritli Dönel Kavşaklar

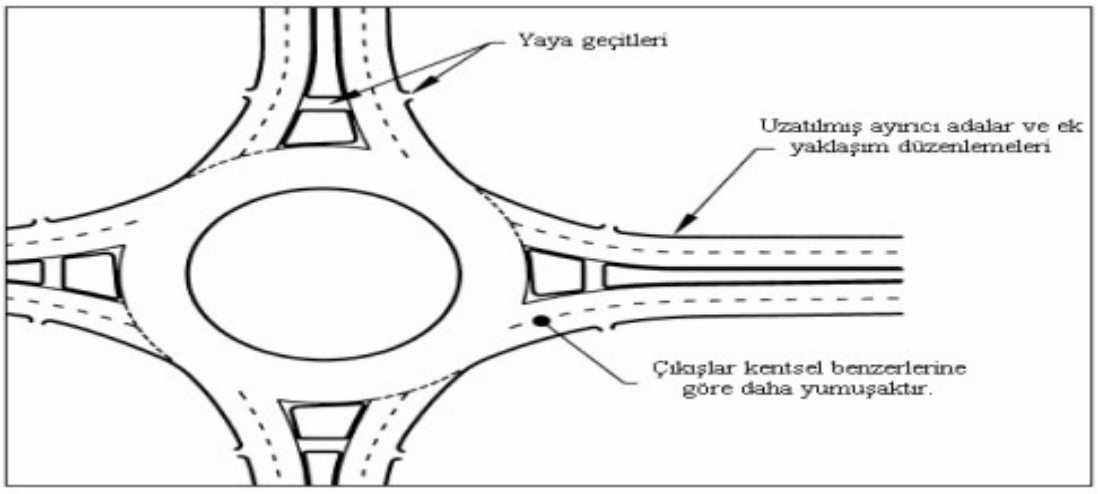
Bu kavşaklarda yaklaşma kollarından gelen aracın hızı 80-100 km/sa (50 - 60 mil/sa) arasındadır. Bu yüzden bu hız seviyelerine hizmet vermek için dolaşım yolunun çapı ve merkez ada büyük çaplarda tasarlanır. Sürücülerin kavşağa girmeden önce hızlarını düşürmek için yaklaşma kollarında geometrik elemanlara ve trafik kontrol cihazı uygulamalarına ihtiyaç vardır (Robinson ve diğ, 2000). Şekil 2.7'de kırsal tek şeritli dönel kavşak gösterilmiştir.



Şekil 2.7 : Kırsal tek şeritli dönel kavşak (İnançlı, 2012).

2.6 Kırsal Çift Şeritli Dönel Kavşaklar

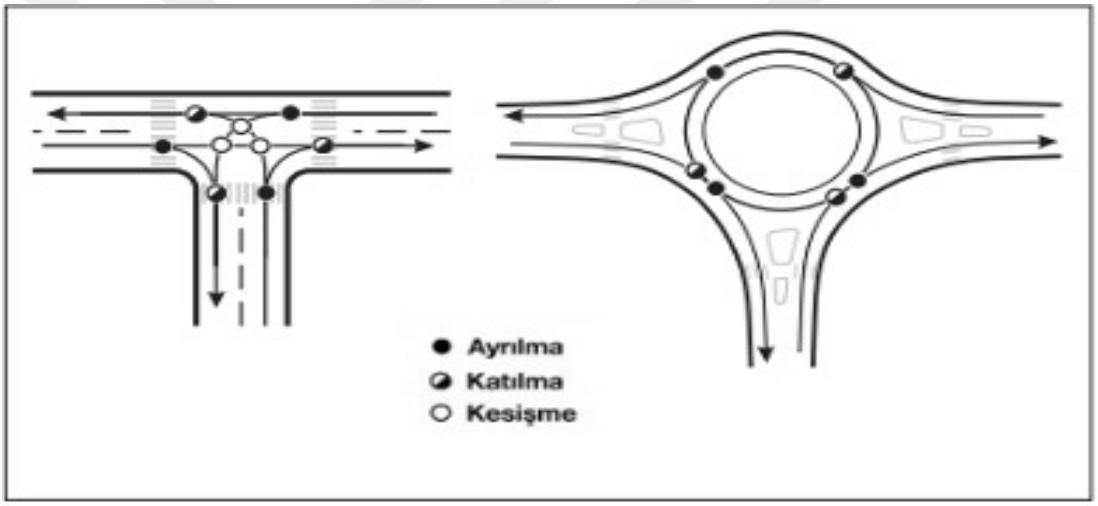
Kırsal tek şeritli kavşaklarda olduğu gibi işletme hızları 80-100 km/sa (50-60 mil/sa) aralığındadır. Bir veya daha fazla yaklaşım kolundan iki giriş şeridine veya bir şeritten iki şeride genişletilmiş girişe ihtiyaçları vardır. Bu özellik bakımından kırsal tek şeritli dönel kavşaklardan farklıdırlar. Bu kavşak türü çeşitli özellikleri ve tasarım parametreleriyle kentsel dönel kavşakların özelliklerini taşır (Şekil 2.8). Bu kavşakların kentsel kavşaklara göre farkı yüksek hızlardan dolayı daha büyük çaplarda tasarlanmalarıdır (Robinson ve diğ, 2000).



Şekil 2.8 : Kırsal çift şeritli dönel kavşaklar (İnançlı, 2012).

2.7 Dönel Kavşak Güvenliği

Denetimsiz standart kavşaklara oranla dönel kavşaklarda çakışma noktaları daha azdır. Dönel kavşaklarda kafa kafaya çarpışmalar ve kesişme çakışmaları görülmez. Dönel kavşaklar yol ver prensibi ile tasarlandığından daha çok açısal çarpışmalar ve örülme çakışmaları görülür bu da ölümlü, yaralanmalı ve maddi hasarlı kaza oranını azaltır. Dönel kavşaklar pek çok özelliğiyle denetimsiz eş düzey kavşaklara göre daha güvenlidir. Bu özellikler: dönel kavşaklarda çakışma oranının fazla olmayışı, merkezinde bir adanın olmasından dolayı adaya girerken hızı azaltmayı gerektirmesi, kafa kafaya çarpışmalar gibi riskli çarpışma türlerini ortadan kaldırmasıdır. (Yılmaz,2022). Şekil 2.9'da geleneksel üç kolu olan bir kavşak ve üç kolu olan dönel kavşak çakışma noktası karşılaştırması verilmiştir.

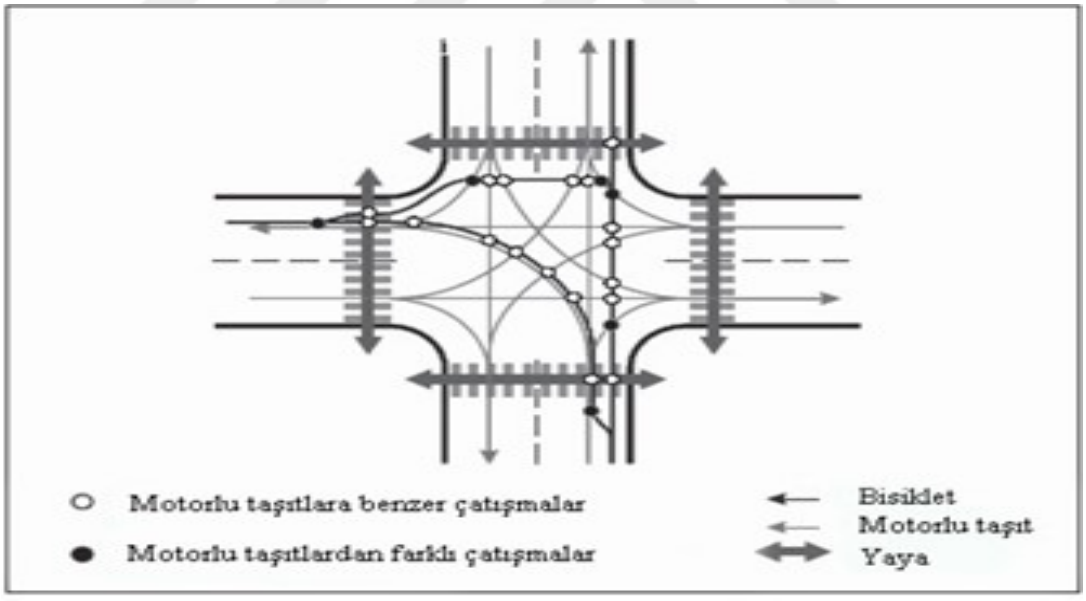


Şekil 2.9 : Geleneksel kavşak ve dönel kavşak çakışma noktası karşılaştırması (Kırbaş ve Ordu, 2021).

Dönel kavşaklarda motorlu araçlar açısından kafa kafaya çarpışmaların görülmediği bilinmektedir. Bu durum yaralanmalı kaza oranını büyük ölçüde azaltır. Dönel kavşaklar aynı zamanda bisikletli kullanıcıların da yolculuklarında kullandıkları alanlardır. Özellikle, tüm yolculuklar veya yolculukların bazı bölümleri için erişilebilir bir ulaşım aracı olarak bisiklet kullanımının teşvik edilmesi, öncelik verilmesi ve dünya çapında bisiklete yönelik altyapının iyileştirilmesi gerekmektedir. Hedefe ulaşılabilmesi için bisikletçilerin güvenliğinin küresel yol güvenliğinin temeli olması ve yol sisteminin bisikletçilerin ihtiyaçlarını dikkate alacak şekilde tasarlanıp yeniden düzenlenmesi gerekmektedir (Scarano ve diğ, 2023).

Araştırmalar, kavşakların dönel kavşağa dönüştürülmesinin motorlu araçlar için faydasını gösterirken bisikletliler için durumun farklı olduğunu göstermektedir. Bu durumun birden fazla açıklaması olabilir. İlk olarak, dönel kavşaklar, dairesel tasarımı nedeniyle sıradan bir kavşağa göre tüm yol kullanıcıları için daha zorlu bir yönlendirme görevi temsil eder. Bu, yönlendirme başarısızlığı olasılığını ve kaza riskini artırabilir. İkincisi, bir dönel kavşaktan geçmek, yolun dairesel sapması nedeniyle fiziksel olarak daha zordur. Bu durum tek bisikletlilerin karıştığı kazaların sayısını ve oranını artırabilir (Hels ve Orozova,2007).

Dönel kavşakların göreceli güvenlik avantajları yüksek trafik akışlarında bisikletli ve yayalar için azalır. Bisikletliler sinyalli ve dönel kavşaklarda motorlu taşıtlara benzer çatışmalar yaşamaktadır. Ancak, bisikletliler genellikle kavşaklar arasında yolun sağ tarafında sürdüğünden, motorlu taşıtlarla çakışan yollar nedeniyle ek çatışmalarla karşı karşıya kalırlar. Bisikletçilere özgü çatışmalar Şekil 2.10'da gösterildiği gibi (motorlu taşıtlar ve yayalara benzer sola dönüşü gösteren) geleneksel dört kollu kavşaklara her yaklaşımda meydana gelir (Robinson ve diğ, 2000).

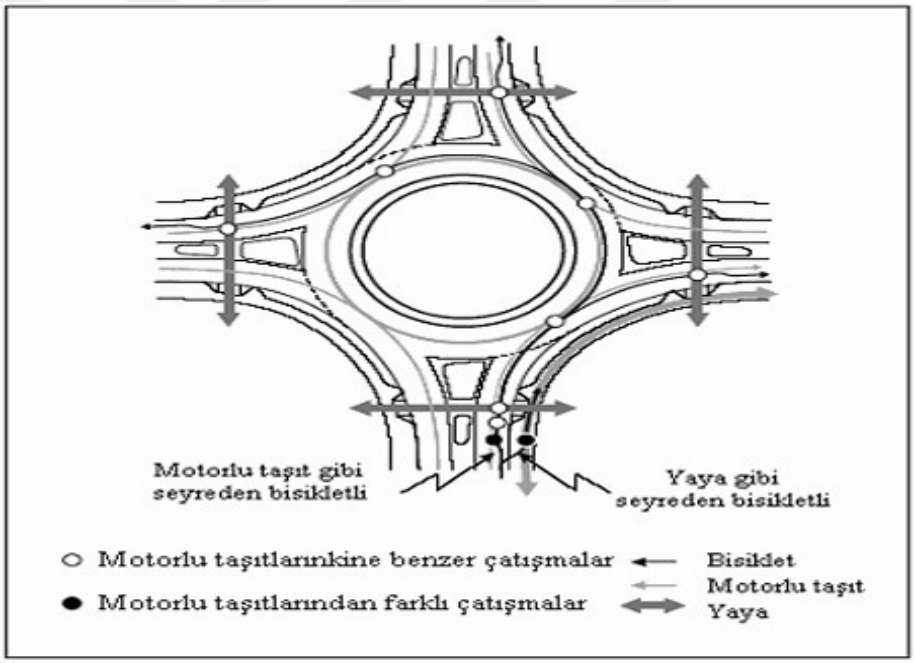


Şekil 2.10 : Geleneksel kavşaklarda bisikletlilere özgü çakışmalar (İnançlı,2012).

Dönel kavşaklarda bisikletlere araç veya yaya olarak seyahat etme seçeneği sunulabilir. Bu durumda bisikletçilerin yaşadığı çakışmalar, kavşağı nasıl geçmeyi seçtiklerine bağlıdır (Şekil 2.11). Tek şeritli bir dönel kavşakta araç olarak seyahat ederken, bisikletçinin trafik akışına karıştığı noktada ek bir çakışma meydana gelir; geri kalanlar motorlu araçlar için olanlara benzerdir. Bisikletlerin genellikle dolaşım

yolunun dış kısmında seyahat ettiği çift şeritli ve daha büyük dönel kavşaklarda, bisikletçiler, bisikletçinin kavşak etrafında dolaşmaya devam ettiği çıkış yapan araçlarla olası bir çatışmayla karşı karşıya kalır. Bisikletçiler, mümkün olduğunca çatışmalardan kaçınırken çemberi "geçmek" zorunda hissedebilirler (örneğin, sürücülere kollarıyla niyetlerini belirterek). Bisikletçilerin daha az görünür olması sebebiyle çift şeridi bulunan dönel kavşaklarda meydana gelen çatışmalara (birleşme ve çıkış) karşı daha savunmasızdır (Robinson ve diğ, 2000).

Yaya olarak seyahat eden bisikletliler için ek bir çakışma, bisikletlinin kaldırıma çıktığı noktada meydana gelir. Bu noktada bisikletli, bir yaya gibi dönel kavşağın etrafından dolaşır. Paylaşılan bisiklet-yaya yollarında veya kaldırımlarda, bisikletliler sürerse, bisikletli ve yayalar kesiştiği her noktada ilave yaya-bisiklet çakışmaları meydana gelir (Robinson ve diğ, 2000).

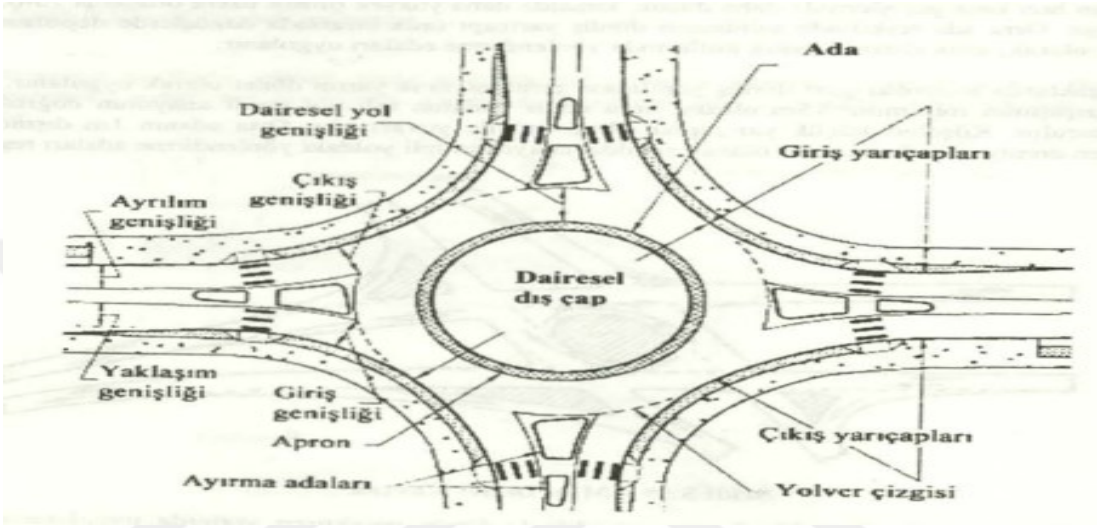


Şekil 2.11 : Dönel kavşaklarda bisiklet çakışmaları (İnançlı,2012).

2.8 Dönel Kavşak Geometrik Elemanları

Geleneksel eşdüzey (hemzemin) kavşaklarla kıyaslandığında dönel kavşaklar güvenlik bakımından daha üstündür çünkü dönel kavşaklarda çakışmalar azdır veya tamamen ortadan kaldırılır ayrıca kavşağa girerken ve çıkarken motorlu araçların hızlarını azaltmasını sağlamak gibi performans özellikleri de vardır (Rodegerdts,

2010). Bu çalışmada yaralanma derecesi üzerinde hız ve geometrik parametrelerin de ne kadar etkili olduğu ele alınmıştır. Dönel kavşaklarda hız limiti önemli bir etken olduğundan hızın azaltılması için geometrik etkenler düzgün tasarlanmalıdır. Hız limitinin kontrollü olması açısından yol geometrik tasarım etkenlerinin (Şekil 2.12) birbirine uyumlu olması ve birbirlerine uygun fiziksel özelliklerde bulunması sağlanmalıdır (Özinal ve Uz, 2021).



Şekil 2.12 : Modern dönel kavşak geometrik elemanları (KGM, 2005).

Analizde kullanılacak geometrik parametrelerin tanımları aşağıda verilmiştir.

2.8.1 Yaklaşma genişliği

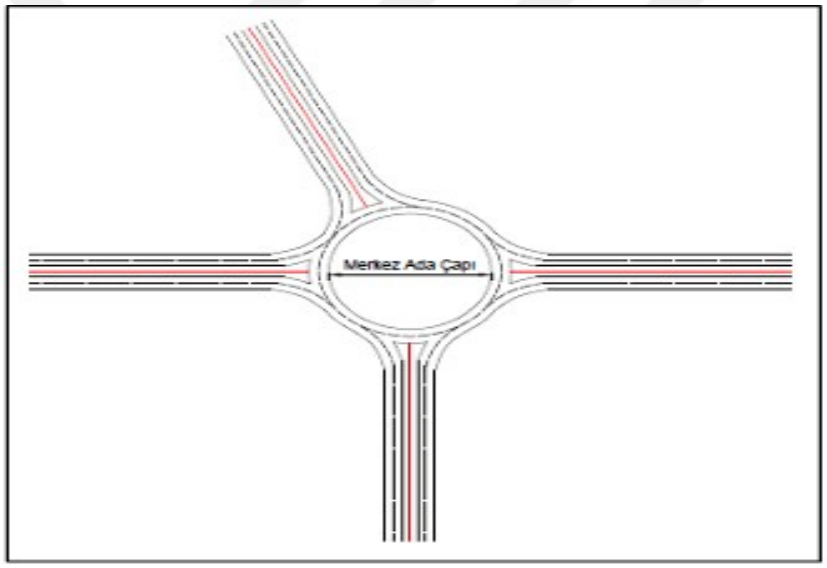
Yaklaşma Genişliği dönel kavşağa girmeden önce yaklaşan trafiğin kullandığı yolun genişliğidir. Yaklaşım genişliği genellikle yolun toplam genişliğinin yarısından fazla değildir (Robinson ve diğ, 2000). Dönel kavşaklardaki geometrik faktörlerden kavşak girişlerinin ve yaklaşım yollarının düzenlenmesi, kavşağa giren trafik hızının kontrol altına alınmasını sağlar. Kavşağın güvenlik seviyesinin yükselmesini sağlar. Bu yüzden kavşak tasarımı yapılırken göz önünde bulundurulması önemli olan faktörlerdendir (Özinal ve Uz, 2021).

2.8.2 Giriş genişliği

Giriş genişliği, Şekil 2.12'de görüldüğü gibi girişin merkezi adayla kavuştuğu yerdir (Robinson ve diğ, 2000). Giriş genişliği, trafik talebini karşılayacak ve giren araçların hızını düşürecek şekilde tasarlanmalıdır (Ersoy,2012).

2.8.3 İç ada çapı

İç ada dönel kavşağın merkezinde bulunur ve taşıtlar bu alanın etrafında dönüş hareketi yapar. İç adanın üzerinden geçiş uygun değildir (Öğütveren, 2019). Dönel kavşağın merkezinde yükseltilmiş olarak bulunur. İç ada yaklaşan sürücülerin kavşak algısının artırılmasını sağlar. İç adanın, dönel kavşak kullanıcılarının gerekli duruş görüş mesafesinde görebileceği biçimde dizayn edilmesi trafik güvenliği açısından önemlidir. Çünkü iç adanın algılanması kavşak girişindeki hızların kontrol edilmesi açısından önemlidir (Özinal ve Uz, 2021). İç ada çapı yaklaşım kollarında yeterli derecede sapma yapılmasına uygun tasarlanmalıdır ve kavşağı kullanan bütün taşıtların dönüş hareketi yapmalarına olanak sağlamalıdır (Öğütveren, 2019).Şekil 2.13'de iç ada çapı gösterilmektedir.

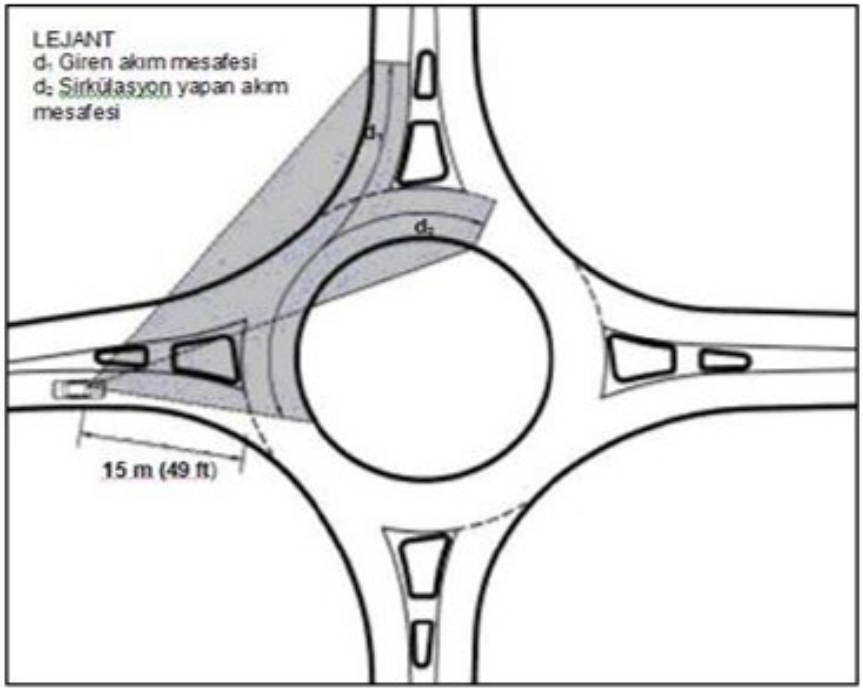


Şekil 2.13 : İç ada çapı (merkez ada çapı) (Öğütveren, 2019).

2.8.4 Görüş mesafesi

Gereken görüş mesafesinin sağlanması amacıyla kavşağa yaklaşım kollarından gelen dönel kavşak kullanıcılarının merkez ada, dönüş şeritleri ve ayırıcı adayı kolaylıkla görmeleri sağlanmalıdır. Ortamda bu geometrik faktörlerin görülmesini engelleyecek herhangi bir fiziksel engel bulunmamalıdır varsa da ortamdan kaldırılmalıdır (Öğütveren, 2019). Sürücü duruş çizgisinde beklerken kendinden önce kavşağa yaklaşım kolundan giren sürücülerin girişlerini görebildiği mesafe en az kritik aralık değeri kadar uzaklıktan olmalıdır. 70 m mesafe ve 50 km/sa seyir hızı, dönüş akımı düşük olan kavşaklar açısından yeterlidir (Taekratok,1998; Öğütveren, 2019). Yoğun

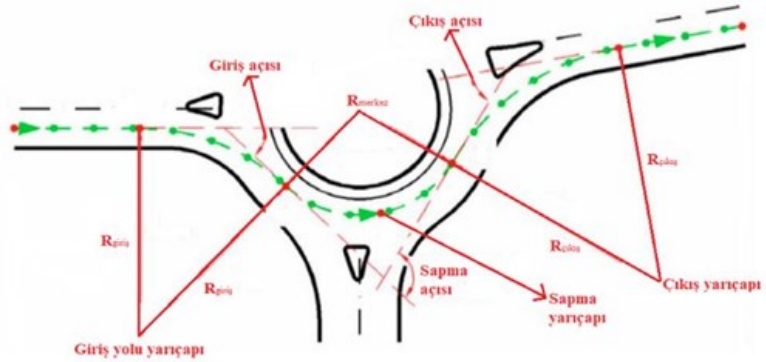
dönel kavşaklarda ise 4 saniye kritik aralık kabulü olarak yeterli bir süredir (Öğütveren, 2019). Bu çalışmada veri setinde görüş mesafesi bulunmamaktadır. Şekil 2.14' de kavşak görüş mesafesi gösterilmektedir.



Şekil 2.14 : Kavşak görüş mesafesi (Öğütveren, 2019).

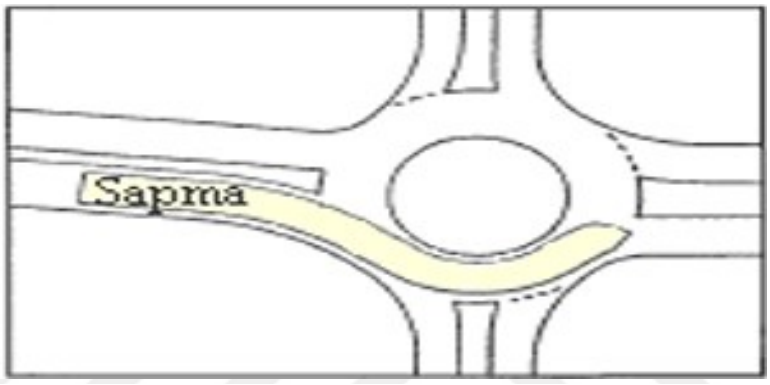
2.8.5 Sapma açısı ve yarıçapı

Sapma açısı bir dönel kavşakta karşı karşıya bulunan iki kolun arasındaki yörünge sapması ölçüsü olarak ifade edilebilir. Sapma açısı; dönel kavşak kullanıcılarının güzergâhlarından saptırılarak kavşağa girerken, çıkarken ve doğru geçişlerinde hızlarının düşürülmesini sağlar. Bu sebeple dönel kavşakların tercihinde büyük bir etkendir (Şekil 2.15). Dönel kavşaklarda yeterli derecede saptırma yaptırılması kavşak güvenliğini artırıp kaza oranını azaltır (Özinal ve Uz, 2021; Öğütveren, 2019).



Şekil 2.15 : Sapma açısı ve yarıçapı (Özinal ve Uz, 2021).

Dönel kavşak kullanıcılarının dolaşım güzergâhı üzerinde seyrettikleri yarıçap sapma yarıçapıdır (Şekil 2.15). Sapma yarıçapının hız ile bir ilişkisi vardır ve tasarımları dönel kavşakta hızları kontrol etmek için ayarlanmıştır. Dönel kavşaklarda trafik akışı, girişten itibaren sirkülasyon yoluna doğru iç adanın etkisiyle sağa doğru yönlendirilerek eğri bir güzergâh oluşturmaktadır (Şekil 2.16). Bu yönlendirmenin amacı araçların hızlarının düşürülmesine olanak sağlanmasıdır. Çünkü dönel kavşak içerisindeki araç hızları kavşak ve sürüş güvenliği açısından büyük etkiye sahiptir. Bu da uygun hızlara ulaşmayı en kritik tasarım hedefi yapmaktadır (Özinal ve Uz, 2021).



Şekil 2.16 : Trafik doğrultusunda sapma sağlanması (Öğütveren, 2019).

3. BETİMSEL İSTATİSTİK

Bu çalışmada betimsel (tanımlayıcı) istatistik, elimizdeki verilerin analiz ve organize edilmesini sağlamak için kullanılmıştır. Betimsel istatistik ile veriler özetlenmiş sayısal ve kategorik veriler için farklı ölçümler kullanılmıştır. Betimsel istatistik merkezi eğilim ve dağılım (yayılım) ölçüleri olarak ayrılmaktadır. Çalışmada kullanılmış olan merkezi eğilim ölçüleri ve dağılım (yayılım) ölçüleri aşağıda açıklanmıştır.

3.1 Merkezi Eğilim ve Dağılım Ölçüleri

Merkezi eğilim ölçüleri bir serinin merkezinin tespit edilmesinde kullanılır (Miran, 2021). Merkezi eğilim ölçülerinin grup içinde ne şekilde dağıldığını gösteren değerler ise merkezi dağılım ölçüleridir (Yüzer, 2009). Bazı merkezi eğilim ve dağılım ölçüleri diğerlerinden daha fazla kullanılır. Bu merkezi eğilim ölçüleri tepe değer, ortalama ve ortancadır. Dağılım ölçüleri ise, varyans, standart hata, değişim katsayısı ve standart sapmadır.

3.1.1 Ortanca (medyan)

Bir örnek topluluğunda bulunan değerler en büyük değerden en küçük değere doğru sıralandığında tam olarak orta noktaya gelen değer ortancadır.

3.1.2 Varyans ve standart sapma

Ortalama ile ilgili olan standart sapma bir veri setindeki gözlemlerin ortalamadan farkının karelerinin alınmış halinin ortalaması olarak tanımlanabilir. Sembolik olarak "σ" ile gösterilir. Standart sapma varyans'ın kareköküdür. Varyans "σ^2" ile sembolize edilir (Yüzer, 2009).

3.1.3 Ortalama

Ortalama, bir veri setindeki verilerin toplamının toplam veri sayısına bölünmesiyle bulunur.

3.1.4 Değer aralığı

Range olarak da adlandırılır. Değer aralığı mevcut veri setindeki en küçük değer ile en büyük değer arasındaki farkı ortaya koyar.

3.1.5 En az ve en çok

Min (minimum değer), mevcut veri setinde bulunan en küçük değeri temsil eder veri setinde bulunan en küçük ölçümdür. Max (maksimum değer), mevcut veri setinde bulunan en büyük değeri temsil eder. Veri setinde bulunan en büyük ölçümdür.

4. Kİ KARE TESTİ

Kategorik değişkenler çoğunlukla iki yönü bulunan tablolarla özetlenirler. Bu iki yönü bulunan tablolarda mevcut olan kategorik değişkenlerin birbirleriyle ilişkisini belirlerken ki-kare testi, Fisher'ın kesin olasılık testi, oran testi, log-doğrusal modeller gibi istatistikler kullanılabilir (Eyduran ve diğ, 2005).

Ki-kare istatistiği, bağımlı değişkenin nominal türde olduğu durumlarda gruplar arasındaki farklılıkları irdelemek amacıyla tasarlanıştır. Ki-kare istatistiği parametrik değildir. Diğer parametrik olmayan tüm istatistiklerde olduğu gibi Ki-kare de verilerin dağılımından etkilenmez ve sağlam bir analiz olasılığı sunar. Çalışma grupları arasında varyans eşitliği veya verilerde homoskedastisite gerektirmemektedir. İkili bağımsız türdeki değişkenlerin ve de çoklu grup çalışmalarının değerlendirilmesini sağlar. Bazı parametrik ve pek çok parametrik olmayan istatistiklerden farklı olarak Ki-kare istatistiği, grupların çalışmadaki performansını ortaya koyar. Bunun gibi detaylar araştırmacının sonuçları anlamasına katkı sağlar (McHugh, 2013).

Ki-kare testinde G ile sembolize edilen gözlenen frekanslar ve B ile sembolize edilen beklenen frekanslar vardır. Bu iki frekansın arasındaki fark istatistik açısından anlamlı mı yoksa değil mi kökenine dayanır (Güngör ve Bulut, 2008). Bu test, nitel veriler için kullanılmasının yanında ölçümle belirtilen sürekli değişkenler için de uygulanabilir. Sürekli değişkenler bu uygulama için belli bir dereceden az ya da çok şeklinde belirtilmelidir. Verilerin bu testte yüzdelik ve oran şeklinde gösterilmesi istenmez böyle bir durumda ki-kare testi uygulanmaz. (Güngör ve Bulut, 2008). Bu test, serbestlik derecesi (sd)'yle nitelendirilir. Dağılımın ortalaması ve varyansı serbestlik derecesine bağlıdır. Ortalama serbestlik derecesine; varyans ise serbestlik derecesinin iki katına eşittir. Ki-kare değerleri, sıfırdan artı sonsuza gider. Küçük serbestlik derecelerinde dağılım basık olmasına rağmen serbestlik derecesinin artmasıyla normal dağılıma yaklaşır. Ki-kare dağılımı, sürekli bir dağılımdır, çoğunlukla iki bağımsız niteliksel kriteri test eder. Ki-kare dağılımında sıfır hipotezi ve araştırma hipotezi kurulur. Sıfır hipotezi (H_0), kriterlerin birbirlerinden bağımsız

olduğunu; araştırma hipotezi (H_A) ise kriterler arasında ilişki bulunduğunu belirtir. Ki-kare testi genellikle (Güngör ve Bulut, 2008);

- İki veya daha çok grup arasında fark bulunup bulunmadığı testinde,
- İki değişken arasında bağ bulunup bulunmadığı testinde,
- Gruplar arası homojenlik testinde,
- Örneklemden elde edilen dağılımın istenen bir teorik dağılıma uyup uymadığının testinde (Uyum iyiliği testinde),
- Varyans için ki-kare testinde,
- Varyansla ilgili aralık tahmininde,
- Kontenjans katsayısının hesabında kullanılır.

Ki-kare bağımsızlık testinde veri setindeki iki değişken arasındaki ilişki araştırılır. Beklenen ve gözlenen frekans dağılımları arasındaki farkların rastgele hatalardan meydana gelip gelmediğini araştırmak için kullanılır. Bu tarz bir test örneğinde gözlenen bir frekans dağılımının Binom, normal ya da Poison dağılımına uygunluğu değerlendirilir; frekansların iki özelliğe göre dağılım göstermesi durumunda, bu iki özelliğin birbirinden bağımsızlık durumuna açıklık getirmeyi sağlar (Kubat ve Ayaşlıgil, 2002).

Araştırma sorusu çoğunlukla iki değişkenin birbiri ile ilişkili olup olmadığıdır. Bu çalışmada hız limiti ile yol vermeme arasında bir ilişki olup olmadığını veya yaralanma derecesi ile giriş sapma yarıçapı arasında bir ilişki olup olmadığı gibi bir araştırma sorusu belirlenebilir. Veri setindeki iki değişkenin birbirinden bağımsız olması durumu, aralarında ilişki olmadığı anlamına gelir. Bu durumda birinin dağılımının diğerinin dağılımına bağımlı olmaması şartıyla gerçekleşir ve bu iki değişkenin birbirinden bağımsız olduğu belirlenir. Veri setindeki İki değişken arasında ilişki bulunmaması halinde: Bir değişkenin özelliğinin değerini bilmek, diğer değişkenin özelliğinin değerinin bilinmesine olanak sağlamaz.

5. MAKİNE ÖĞRENMESİ YÖNTEMLERİ

1950'li yıllarda yapay sinir ağı ve yapay zekâ araştırmalarında makine öğrenmesi adı geçmeye başlamıştır. Makine öğrenmesi, yapay zekânın alt dalıdır. Hesaplama gücündeki durmaksızın büyüme nedeniyle Makine öğrenimi yöntemleri, şu anda hızla büyümektedir. Makine öğrenmesinde bilgisayarlar programlanmaya gerek duymadan kendiliğinden öğrenir. Makine öğrenmesi yöntemleri, insan aklını taklit eder ve bunu etraftan öğrendiği bilgilerle gerçekleştirir. Son dönemde trafik kaza verilerini yorumlamak ve analiz etmek amacıyla makine öğrenmesi algoritmaları sık kullanılmaktadır. Makine öğrenmesi yöntemlerinden sınıflandırma algoritmaları kaza analizlerinde sıklıkla tercih edilmektedir (Santos ve diğ, 2022). Makine öğrenmesinin çalışma prensibi aşağıda (şekil 5.1) gösterildiği gibidir.



Şekil 5.1 : Makine öğreniminin çalışma prensibi (Gökalp,2022).

5.1 Makine Öğrenimi Türleri

Makine öğrenimini türlerine ayırırken iki maddeye dikkat edilmesi gerekir. Bunlardan ilki makinelerin karar verirken kullandığı kurallar dizinidir. Diğeri ise farklı makine öğrenmesi yöntemlerinin algoritmalarını incelemektir. Makine öğrenimi yöntemleri kendi içinde sınıflara ayrılır bu sınıflar denetimli ve denetimsiz öğrenmedir. Bunlara zaman içinde başka sınıflar da ilave edilmiştir bu sınıflar yarı denetimli ve pekiştirmeli öğrenmedir. Bu şekilde makine öğrenmesi yöntemleri 4 sınıfa ayrılabilir (Gökalp, 2022).

5.1.1 Denetimli öğrenme

Denetimli öğrenmede analize eğitim veri setiyle başlanır. Öğrenme algoritması sonuç değerleri ile ilgili tahminler üretmek amacıyla bir çıkarsama işlevi yaratır. Eğitim

sürecinden sonra sistem, yeni bir girdi için hedef sağlayabilir. Denetimli öğrenme veri setindeki çıktıyla ilgilidir denilebilir (Özşahin ve diğ, 2022). Makine öğrenmesinde tahminlerin yapılmasında regresyon veya sınıflandırma yöntemi kullanılmaktadır. Sınıflandırma yöntemine örnek olarak gelen mailin spam olup olmadığını saptama gösterilebilir (Gökalp,2022).

Literatürde sıklıkla kullanılan denetimli makine öğrenimi algoritmaları aşağıda sıralanmıştır (Özşahin ve diğ, 2022).

- En Yakın Komşu → K-Nearest Neighbors (KNN)
- Lojistik Regresyon → Logistic Regression
- Karar Ağaçları → Decision Trees (DTs)
- Destek Vektör Makinaları → Support Vector Machine (SVM)
- Yapay Sinir Ağları → Artificial Neural Network (ANN)
- Rastgele Orman → Random Forest (RF)
- Doğrusal Regresyon → Linear Regression

5.1.2 Denetimsiz öğrenme

Gizli olan bir yapıyı etiketlenmemiş verilerden açıklamak amacıyla bir işlevi nasıl çıkarabileceğini araştırmak denetimsiz öğrenmenin araştırma alanıdır. Verilerini inceleyen sistem etiketlenmemiş verilerden gizli yapıları izah etmek amacıyla veri setlerinden tespitler yapabilir (Özşahin ve diğ, 2022).

5.1.3 Yarı denetimli öğrenme

Eğitim aşamasında etiketlenmiş verileri kullanan, denetlenen öğrenme teknik ve görevlerini de kullanan sınıfa yarı denetimli öğrenme denir. Etiketlenmemiş veri sayısı etiketlenmiş veri sayısından daha fazladır. Yarı denetimli öğrenme denetimli öğrenme ile denetimsiz öğrenme arasında kalır. Araştırmacılar, öğrenme doğruluğunda iyileşmenin etiketlenmiş verilerle birlikte az sayıda etiketlenmemiş verilerin kullanılmasıyla arttığını bulmuştur (Sarıcaoğlu, 2019).

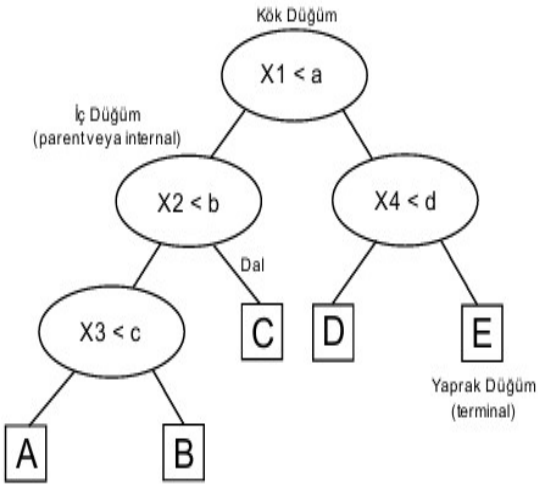
5.2 Karar Ağaçları

Çalışmada kullanılan makine öğrenimi algoritmalardan biri de Karar Ağaçlarıdır. Karar ağaçları, ağaç tabanlı bir algoritmadır. Sınıflandırma problemleri ve regresyon problemlerinde kullanılabilir. Karar ağaçları regresyon ve sınıflandırma modellerini;

ağaç yapısında bulunan kök düğüm, dallar ve yaprak düğümlerle üretir (Gacar ve Kocakoç, 2020). Bağımlı değişkenin kategorik veya sürekli olması durumuna göre sınıflandırma veya regresyon uygulanır. Sınıflandırma ağacını bağımlı değişkenin kategorik olduğu durumda, regresyon ağacını ise bağımlı değişkenin sürekli olduğu durumda oluşturur (Gürsel ve Akdur, 2024).

Veri madenciliği ile tahmin modelleri oluşturmada pek çok yöntem vardır. Karar ağaçları bu modellerin oluşturulmasında sık ve etkili kullanılan yöntemlerden biridir (Emel ve Taşkın, 2005). Diğer veri madenciliği yöntemlerinin zayıf yönü açıklama yapmadan iyi tahmin yapmalarıdır. Karar ağacında ise bu zayıflık söz konusu değildir (Zorman ve diğ, 2001). Karar ağaçlarının veri madenciliği sınıflama algoritmaları içinde bu kadar sık kullanılmasının sebeplerinden biri yorumlanabilirliklerinin kolay gerçekleşmesidir. Diğer sebepleri de veri tabanı sistemleri ile basit şekilde bütünleştirilebilmeleri ve kurulumlarının pahalı olmamasıdır (Emel ve Taşkın, 2005).

Karar ağaçları verileri daha küçük veri kümelerine ayırmak için belli başlı bölme standartları kullanır. Ağaç yapısının en tepesinde kök düğüm yer alır ortalarda iç düğüm ve sonuç bölümünde yaprak düğümlerden oluşur (Topçu ve Vupa, 2021). Şekil 5.2‘ de Karar Ağacı yapısı yer almaktadır. Karar ağaçlarında kök düğüm; bağımlı değişken üzerinde en çok etkiye sahip bağımsız değişkene, diğer bir ifadeyle özniteliklerden birine karşılık gelmektedir (Gürsel ve Akdur, 2024). Karar ağacında veri tabanının mümkün mertebe eşit parçalara bölünmesi için neredeyse bütün Karar Ağacı algoritmalarında entropi veya gini indeksi kullanılır. Bu şekilde yaprak düğüme ulaşılabilir (Topyıldız, 2019).



Şekil 5.2 : Karar ağacı kök, düğüm ve dal yapısı (Topçu ve Vupa, 2021).

Karar Ağacı avantajları:

- Parametrik olmayan bir yöntem olarak değerlendirilebilir. Bu sebepten parametrik yöntemlerin varsayımlarını gerektirmez (Gürsel ve Akdur, 2024).
- Karar ağaçlarının maliyeti azdır.
- Tek bir ağaçtan oluştuğu için anlaşılabilirliğinin ve yorumlanabilirliğinin basit olmasının yanında veri tabanlarıyla entegrasyonu da kolaydır.
- Güvenilirlikleri iyi düzeydedir (Çalış ve diğ, 2014).
- Sınıflandırma ve regresyon ağacı kurabilir hem sayısal hem de kategorik verilerle çalışabilir.

Karar Ağacı dezavantajları:

- Sürekli değişkenler için karar ağacı iyi bir yöntem değildir.
- Değişkenler arasında güçlü korelasyon varsa karar ağacının uygulanmasında dikkatli olunmalıdır (Korkmaz, 2025).
- Karar Ağacında budama yöntemi kullanılır. Karar ağacı çok büyürse aşırı öğrenme (over-fitting) gerçekleşebilir. Aşırı öğrenme ağacın çok büyümesiyle anlamsız sonuçlar üretmesidir. Bu sorundan budama yöntemi ile kaçınılabilir (Patel ve Upadhyay, 2012).

Veri madenciliğinde en bilinen yöntem sınıflandırmadır. Girdiler bir sınıflayıcı model tarafından çeşitli niteliklere göre sınıflara atanırlar. Eldeki nesnelerin bir sınıfa atanma veya atanmama durumunun veya hangi sınıflara atanacağının belirlenmesidir. Diğer bir deyişle nesneler veya durumların uygun sınıfının tahmin edilmesidir. Eğitim kümesi sınıflama girdileridir. Sınıflandırmada eğitim kümesinin oluştuğu gözlemlerden her biri bir sınıf etiketiyle etiketlenir. Modelin her bir gözleme niteliklere göre atadığı sınıf etiketi ise çıktıdır (Emel ve Taşkın, 2005).

Sınıflandırma makine öğrenmesinin önemli araçlarındandır. Tümevarımsal öğrenmenin esas noktaları eğitim ve test kümesini içerir. İki aşamada gerçekleşen sınıflandırma veri setinden seçilen bir miktar verinin eğitilmesi ve diğer aşamada modelin test edilmesidir. Eğitim aşamasında eğitim kümesinden öğrenilenlerle model oluşturulur; Test aşamasında ise test kümesiyle modelin performansı değerlendirilir. Model kesinliği karşılaştırma yöntemiyle bulunur. Karşılaştırma, test örneklerinin iyi bilinen sınıfı ile modelin tahmin ettiği sınıf arasındadır. Test kümesinin model tarafından doğru biçimde sınıflanma oranı kesinlik oranını verir. Model girdilerden

çıktıları üretir ardından sınıf etiketini tahmin etmek amacıyla kayıp veya sınıf etiketi belli olmayan yeni gözlemler için kullanılabilir (Emel ve Taşkın, 2005).

Karar ağaçları en önemli sınıflama araçlarından biridir ve basit bir öğrenme algoritması vardır. Ortaya konulmuş olan özbilginin gösterimi kolaylıkla anlaşılabilir. Karar Ağaçları sadece kararları göstermekle kalmaz açıklama getiren yapıları sebebiyle kararların açıklamasını da kapsarlar. Bu algoritmanın eğitim süreci tümevarımdan oluşur. Ağaç tümevarımı bir eğitim nesneleri kümesinden karar ağacı oluşturmaktır. Özbilgi keşfinin en sık tercih edilen yöntemlerinden bir tanesidir. Sınıflama veya tahmin için kullanılabilecek ağaç benzeri örüntüleri keşfetme için bir yöntemdir (Emel ve Taşkın, 2005).

Bilginin keşfi esnasında Karar Ağaçları pek çok test gerçekleştirir. Karar Ağaçları hedefi tahmin ederken en doğru sırayı bulmayı hedefler. Test olarak ifade edilen Karar Ağacındaki dallardır. Diğer testlerin gerçekleşmesini Karar Ağacı'ndaki dallar sağlar. Karar Ağacında test işlemi yaprak düğümün oluşmasıyla sona erer. Karar ağacında başlangıç kök düğümünden hedef yaprak düğümün oluşması "Kural" olarak isimlendirilir. Bu kurallar if-then (eğer-sonra) yapısındadır (Bounsaythip ve Rinta, 2001). Karar ağaçlarında kullanılan pek çok algoritma vardır. C4.5, C5.0, ID3, CART, CHAID ve QUEST bunlardan bazılarıdır (Çalış ve diğ, 2014).

5.2.1 C4.5 ve C5.0 algoritmaları

ID3 algoritmasının ilerletilmiş hali C4.5'tir. C4.5'in ilerletilmiş hali ise C5.0'dır. C5.0 özellikle büyük miktarlı veri setleri ile uygulanan bir algoritmadır. C5.0, algoritması boosting algoritmasını kullanır ve bunu doğruluğu yükseltmek için yapar. Bu seeple boosting ağaçları olarak da bilinir. C5.0 algoritması C.4.5 e göre hızlıdır. Aynı zamanda C5.0 algoritması hafızayı daha iyi kullanır. Sonuçlar C5.0 ve C4.5 için aynıdır. Aradaki fark C5.0'ın biçim açısından daha düzgün karar ağaçları göstermesidir (Çalış ve diğ, 2014).

5.2.2 CART algoritması

CART, Breiman tarafından 1984 yılında önerilmiş olan ve Freidman, Olsen ve Stone tarafından geliştirilen Karar Ağacı algoritmasıdır. CART, entropiden faydalanır ve bir düğümden iki dal oluşturur ve bu dallardan en iyisiyle devam eder. CART algoritması bölünmeler esnasında twoing ve gini algoritmalarını kullanır. CART

algoritmasında oluşan dallar homojen yapıdadır. CART algoritmasının temelinde ikili bölünme vardır. Nitelik değerini İkili bölünme ile böler ve bölünme sonucunda sınıfları oluşturur. Bu işlemler bölünme devam etmeyinceye kadar sürer (Arslan, 2022).

5.2.3 CHAID algoritması

Karar Ağacı algoritmalarının en çok kullanılanlarındandır. Ki kare istatistiğini en uygun bölünmeleri bulmak için kullanır. CHAID istatistiksel bir yöntemdir ve bölümlendirme için kullanılır. Kriter olarak CHAID, istatistiksel bir testin anlamlılığından yararlanarak bir olası ön kestirici değişkenin tüm değerlerini değerlendirir. Hedef değişkene göre homojen olarak değerlendirilen bütün değerleri birleştirir ve kalan tüm değerleri benzer olmayan şeklinde nitelendirir. Ardından, karar ağacındaki ilk dalın şekline göre en iyi ön kestirici değişkenin seçilmesiyle birlikte her bir düğümün seçilen değişkenin homojen değerlerinin bir grubunu oluşturmasını sağlar. Ağaç tamamen büyüyene kadar bu durum sürer. Hedef değişkenin ölçüm düzeyine göre kullanılan istatistiksel test belirlenir (Oğuzlar, 2004).

5.2.4 QUEST algoritması

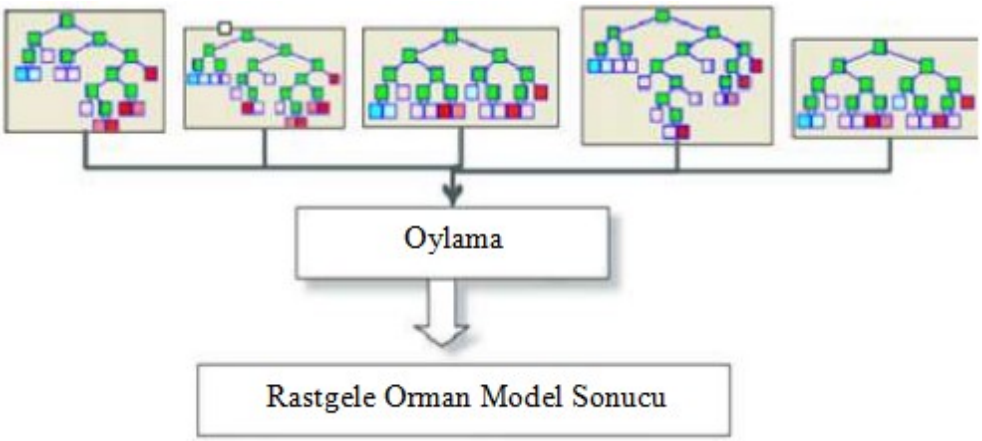
Loh ve Shih tarafından 1997'de geliştirilmiştir. Bu algoritma ikili ağaç yapısını kullanır. Bir sınıflandırma algoritmasıdır. İkili ağaç yapısında doğrudan durma ve budama kuralı teknikleri kullanılabilir. Bazı algoritmalar bölünme ve değişken seçimini aynı zamanda yaparken QUEST algoritması bu tarz durumlarla ayrı ayrı ilgilenir. QUEST'in tasarlanma amacı ağacın dallanması sırasında önyargılı seçimi daha genel hale getirmek ve bunun yanında hesaplama maliyetini azaltılmaktır (Çalış ve diğ, 2014).

5.3 Rastgele Orman

Leo Breiman 1996 senesinde "Bagging Sınıflama Yöntemi" ni geliştirmiştir. Bagging sınıflama Yönteminin varyansı düşürücü özelliği ve aşırı öğrenmeye karşı dirençli bir yapısı vardır. 2003 yılında ise Breiman ile Adele Cutler birlikte Bagging yöntemini ileriye taşımış ve Rastgele Orman Sınıflama Yöntemi'ni ortaya çıkartmışlardır (Korkem, 2013).

Ağaç tabanlı bir makine öğrenmesi olan Rastgele Orman yöntemi fazla sayıda karar ağacından meydana gelir ve bir orman sınıflayıcısıdır. Rastgele orman yöntemi ile sınıflama ağacı ve regresyon ağacı oluşturulabilir bunların yanında Rastgele Orman kümeleme için de kullanılabilir. Rastgele Orman yöntemi ile sınıflandırma mı yoksa regresyon mu yapılacağı sınıf değişkenine bağlıdır. Sınıflama ağaçları kurulması için söz konusu sınıf etiketinin kategorik olması gerekir. Eğer sınıf etiketi sürekliyse regresyon ağaçları kurulmaktadır. Rastgele Ormanı oluşturan karar ağacı, orijinal veri setinden bootstrap yöntemi ile örneklem seçilmesi ve her bir düğümde mevcut değişkenler içinden belirli bir sayıda değişkenin seçilmesiyle meydana gelir. Rastgele Orman metodunda ağaçlar CART algoritmasıyla oluşturulur ve oluşturulan ağaçlarda budama işlemi gerçekleştirilmez. CART algoritmasında veri setinin dallara ayırma işleminin hangi değişkenden başlayacağına bilgi kazancından faydalanarak karar verilir. Dallara ayırıcı değişkenin gini katsayısı ile uygun test kriteri bir diğer deyişle cutt-off değeri belirlenir. Rastgele Orman modelinin hatası Out-Of-Bag, OOB olarak da kısaltılabilir (Akman ve diğ, 2010).

Rastgele Orman modeli oylama yöntemine dayanır (Şekil 5.3). Ağaçlarda hata oranı en düşük olan ağaç en yüksek ağırlığı alırken hata oranı en yüksek olan ağaç en düşük değeri alır. Sınıf tahmini için ağaçlar kendisine değer biçilen ağırlığa göre oy verme işlemine sokulur (Akman ve diğ, 2010). Rastgele Orman algoritmasında regresyon için ağaçların tahminlerinin ortalamasını, sınıflandırma için tahminler içinden en çok oyu alan seçilir. Karar ağacı modelinin mevcut sorunlarından birisi veriyi ezberlemesi ve aşırı öğrenmedir. Rastgele Orman ise veri setinden yüzlerce farklı alt ağaçlar seçip bunları eğiterek bu sorunu çözer. (Akyiğit ve Taşçı, 2022).

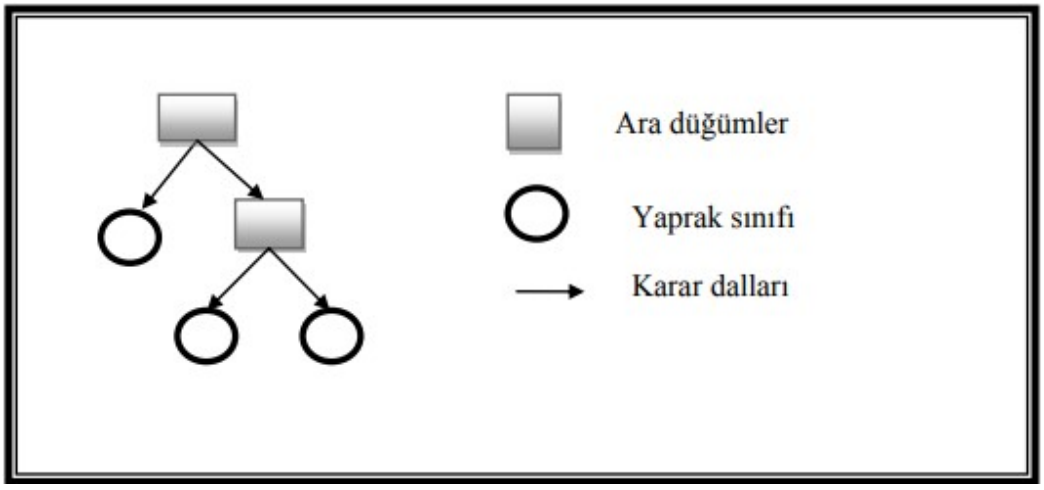


Şekil 5.3 : Rastgele orman modeli (Akyiğit ve Taşçı, 2022).

Rastgele Orman sınıflama yöntemi yakın dönemde çok tercih edilir haldedir. Bunun sebebi Rastgele Orman'ın yüksek sınıflama özelliğidir. Rastgele Orman pek çok ağaç oluşturarak oylama yöntemi ile sınıflandırıcıları birleştirir ve bu şekilde sınıflandırma performansını artırmış olur. Rastgele Orman algoritmasında oylama sonucunda en fazla oyu alan sınıf kazanan sınıftır. Sınıflama, veri tabanındaki gizli bulunan kalıpları meydana çıkarır. Sınıflama yönteminde veri tabanı daha küçük homojen şekilde gruplandırılır. Sınıflama tekniği her yeni gelen verinin hangi sınıfa atanacağını gösterir. Sınıflama tekniğinin temeli öğrenme algoritmasına dayanır. Sınıflama algoritmasıyla sınıflama modeli oluşturulur ve sınıfı bilinmeyen veri için sınıf tespiti yapılır. Burada önemli olan nokta iyi tespit edilmiş değişkenlerdir (Korkem, 2013).

5.3.1 Ağacın yapısı

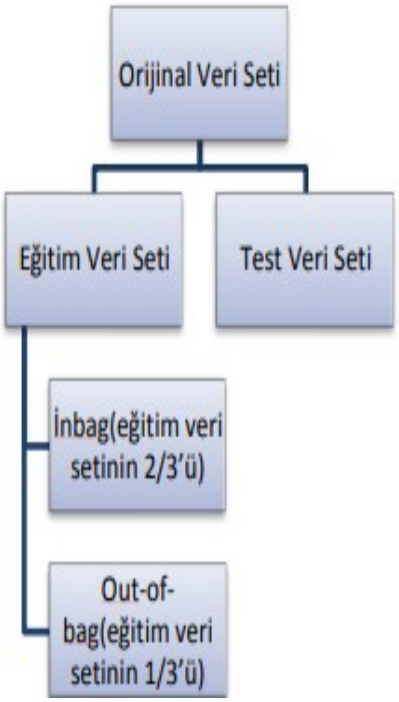
Rastgele Ormanı oluşturan ağacın yapısı yaprak ve dallardan oluşur. Niteliklerin her biri bir düğüm tarafından simgelenir. Ağaçta en son oluşan kısım "yaprak"tır. Terminal olarak da isimlendirilir. En üst bölüm ise köktür. Kök ve yapraklar arasında kalan kısım "dal" olarak adlandırılır (Şekil 5.4). Rastgele Orman algoritmasında, ağaç verinin tümünün meydana getirdiği tek bir düğümle başlar. Örneklerin hepsinin aynı sınıfta olduğu durumda düğüm yaprak olarak biter ve sınıf etiketi verilir. Örneklerin aynı sınıfta olmadığı durumda örnekleri sınıflara en iyi bölecek nitelik seçilir (Korkem, 2013).



Şekil 5.4 : Rastgele ormanı oluşturan ağaçların yapısı (Korkem, 2013).

5.3.2 Rastgele orman algoritmasının aşamaları

Rastgele Orman modelinin üzerine kurulu olduğu 2 parametre vardır. Bu parametreler; oluşturulacak olan ağaç sayısı "B" ve "m" ile gösterilen tahminci sayısıdır. "m" her düğüm ayrımında rastgele seçilir. Her bir Karar Ağacı oluşturma esnasında bootstrap yöntemiyle örneklem oluşturulur. Örneklem oluşturma işlemi orijinal veri setindeki gözlem sayısı yani n ile aynı ölçüde olacak biçimde oluşturulur. Örneklemin 2/3'ü ağacı oluşturmak için kullanılır ve bu inBag (Eğitim veri seti)'dir; Ardında kurulan modelin iç hata oranını test etmek amacıyla geriye kalan kısım yani örneklemin 1/3'ü test veri seti (out of bag veya OOB) olacak şekilde ayrılır (Yılmaz, 2014). Şekil 5.5'de Rastgele Orman algoritmasının eğitim ve test aşaması gösterilmektedir.



Şekil 5.5 : Orijinal veri setinin ayrılma şeması (Korkem, 2013).

Rastgele Orman yönteminin algoritması aşağıda verilen adımlar takip edilerek kurulmaktadır (Yılmaz, 2014):

1. n hacimli bir veri seti Bootstrap yöntemi ile seçilir. Bu n hacimli veri seti, ikiye ayrılır. Bu ayrılma işlemi inBag ve OOB şeklindedir. Burada inBag eğitim veri setini ve OOB ise test veri setini ifade eder.
2. Eğitim veri setleri (inBag) ile oluşturulabilecek en fazla genişliğe sahip olan bir CART Karar Ağacı meydana getirilir. Rastgele Orman modelinde

oluşturulan Karar Ağacı budanmaz. Ağacın meydana gelmesi sırasında düğümlerin bölünmesinde toplam p adet tahminci değişkenin m adedi tesadüfi seçilir. Bu noktada m değerinin p değerinden küçük olması gerekir. Bunun sebebi aşırı uyum sorununun gözlemlenmesinin ve ağacın aşırı büyümesinin istenmemesidir. Dallara ayırırken bilgi kazancından faydalanılır. Bu işlem "m" tane tahminciden bilgi kazancı en fazla olanla yapılır. Gini indeksi ile tespit edilen bu değişkenin hangi değerine göre ayrımın yapılacağı saptanır ve sonunda her bir düğüm için oluşturulacak yeni bir dal olmayıncaya kadar bu işlem devam eder.

3. Her yaprak düğüme bir sınıf atama işlemi yapılır ve ardından OOB (Test veri seti) ağacın en tepesinden bırakılır; Bu veri setinde bulunan her gözleme atanan sınıf kaydedilir.
4. 1. aşamadan 3. aşamaya kadar her bir aşama oluşturulacak ağaç sayısı yani "B" kez tekrarlanır.
5. Ağaç oluşturulurken inBag kullanılmıştır ve kullanılmamış olan OOB gözlemleri ile bir değerlendirmede bulunulur. İncelenen bir gözlemin kategorilerden hangisine kaç kere sınıflandırıldığı hesaplanır.
6. Sınıf ataması her gözleme, ağaç setleri üzerinden karar verilen bir oy fazlalığı ile yapılır. 2 kategorili olan bir sınıflandırma modelinin ele alındığı durumda, bir gözlem bütün ağaçlar üzerinden %51 ve daha fazla oy çoğunluğunu aldığı sınıfın etiketini taşır ve artık bu sınıf onun tahmin edilmiş sınıf değeridir.

5.4 Rastgele Orman ve Karar Ağacı Karşılaştırması

Yapı olarak Karar Ağacı tek bir ağaçtan oluşur ve veriyi dallara ayırarak karar verir, Rastgele Orman ise birden fazla karar ağacından oluşur. Her ağacın çıktısının ortalaması veya çoğunluk oyu alınarak karar verilir. Karar Ağacı algoritması tek bir ağaçtan oluştuğu için aşırı öğrenmeye daha yatkındır. Eğitim verisine fazla uyum sağlayarak yeni verilere genelleme yapmakta zorlanabilir. Rastgele Orman algoritması ise birden fazla ağaç kullandığı için aşırı öğrenmeyi azaltır ve daha iyi genelleme yapar. Karar Ağacı tek bir ağaç oluşturduğu için daha hızlıdır. Rastgele Orman ise birden fazla ağaç oluşturduğu için hesaplama maliyeti daha yüksektir. Karar ağacı daha küçük veri setlerinde iyi çalışabilir ama karmaşık problemlerde doğruluk düşebilir. Rastgele Orman daha yüksek doğruluk oranına sahiptir çünkü

farklı ağaçların sonuçlarını birleştirir. Karar Ağacının yorumu ve görselleştirmesi kolaydır; Rastgele Orman ise daha karmaşıktır çünkü birçok ağaçtan oluşur. Bu yüzden tek bir ağacı yorumlamak kadar kolay değildir. Özetlenecek olursa Karar Ağacı hız ve yorumlanabilirlik açısından, Rastgele Orman ise daha iyi doğruluk ve genelleme açısından kullanılabilir.

6. MATERYAL VE YÖNTEM

Bu tezde, dönel kavşaklarda meydana gelen bisikletli kazalarının sınıflandırılması amacı ile makine öğrenmesi yöntemlerinden olan Rastgele Orman ve Karar Ağacı algoritmaları kullanılmıştır. Çalışmada analizleri gerçekleştirmek için R programı kullanılmıştır.

Çalışmada kullanılacak kaza verisi, İngiltere'nin Newcastle kentinde bulunan dönel kavşaklarda 2014-2019 yılları arasında gerçekleşmiş olan bisikletli-motorlu taşıt kaza istatistiklerini içermektedir. Veri seti 439 adet örnekten oluşmaktadır. Tahmin modelindeki bağımlı değişken, kaza yaralanma derecesidir. Bu seviye iki kategoridedir: hafif yaralanma, ağır yaralanma. Verideki hafif yaralanma 370; ağır yaralanma 69 adettir. Analiz edilecek bağımsız değişkenler; yaralanan sürücünün sosyodemografik bilgileri, hava şartları, yol yüzeyi şartları, sürücü davranışları ve kavşak geometrik tasarım parametreleridir.

Bağımsız değişkenlerin bir kısmı kategorik verilerdir. Makine öğrenmesi yöntemlerine veri setindeki parametrelerin uyum sağlaması için "0", "1", ve "2" şeklinde kodlanmıştır. Bu bütün kategorik parametreler için gerçekleştirilmiştir. Çalışmada 19 adet bağımsız değişken kullanılmıştır. Çizelge 6.1'de parametrelerin tam ismi gösterilmektedir. Bu parametreler R programı tarafından anlaşılması için kısaltılmış ve İngilizce karakterlerle yazılmıştır. R programında modelin doğruluğunu etkilememesi için Kaza Kodu, Kaza Saati, Yıl, Kavşak Tipi parametreleri veriden çıkartılmıştır. Parametrelerimizin kodlanmış hali Çizelge 6.2 'de gösterilmiştir.

Çizelge 6.1: Parametreler.

Parametre Eski İsim	Parametre Yeni İsim
Kaza Kodu	Kod
Kaza Saati	Saat
Yıl	Yil
Kavşak Tipi	Kavsak

Çizelge 6.1(devam): Parametreler.

Parametre Eski İsim	Parametre Yeni İsim
Yaralanma Derecesi	Yaralanma
Cinsiyet	Cinsiyet
Yaş	Yas
Işık Durumu	Isik
Hava Şartları	Hava
Yol Yüzeyi	Yuzey
Bisikletliyi Görmeme	Bisikletli
Dikkatsizlik	Dikkatsizlik
Bisikletliye Çok Yakın Geçme	Yakin
Yol Vermeme	Yol
Manevra Yapamama	Manevra
Hız Limiti	Hiz
Yaklaşma Şerit Sayısı	Yaklasma
Yaklaşma Genişliği	Genislik
Giriş Sapma Yarıçapı	Sapma
Kol Sayısı	Kol
İç Ada Çapı	Ada
Giriş Genişliği	Giris
Giriş Şerit Sayısı	Serit
Dönüş Şerit Sayısı	Donus

Çizelge 6.2 : Parametrelerin kodlanma şekli.

Parametre	0	1	2
Yaralanma Derecesi	Hafif	Ağır	
Cinsiyet	Erkek	Kadın	
Işık Durumu	Karanlık	Aydınlık	
Hava Şartları	Karlı	Yağmurlu	Yağışsız
Yol Yüzeyi	Buzlu	Islak	Kuru
Bisikletliyi Görmeme	Evet	Hayır	

Çizelge 6.2 (devam) : Parametrelerin kodlanma şekli.

Parametre	0	1	2
Dikkatsizlik	Evet	Hayır	
Bisikletliye Çok Yakın Geçme	Evet	Hayır	
Yol Vermeme	Evet	Hayır	
Manevra Yapamama	Evet	Hayır	

Verileri yorumlayabilmek, kazalara karışanları tanımlamak ve kazaların nasıl ortamlarda gerçekleştiğini anlayabilmek için betimsel istatistik uygulanmıştır. Verilerin arasındaki ilişkiyi inceleyebilmek için Ki Kare testi uygulanmıştır. Çizelge 6.3'de Ki Kare testi hesabı yapılırken parametreleri gösteren temsili değerler verilmiştir. Bu çalışma için ağaç tabanlı denetimli makine öğrenimi algoritmalarından Rastgele Orman ve Karar Ağacı seçilmiştir. Bu iki yöntem ile sınıflandırma modeli oluşturulmuş ve test veri setleriyle sınıflandırma doğruluğu tahmin edilmiştir. Sınıflandırma doğruluğu karışıklık matrisi ile bulunmuştur.

Çizelge 6.3 : Ki kare testindeki temsil.

Ki Kare Testindeki Temsil	Parametre İsmi
1	Yaralanma Derecesi
2	Cinsiyet
3	Yaş
4	Işık Durumu
5	Hava Şartları
6	Yol Yüzeyi
7	Bisikletliyi Görmeme
8	Dikkatsizlik
9	Bisikletliye Çok Yakın Geçme
10	Yol Vermeme
11	Manevra Yapamama
12	Hız Limiti
13	Yaklaşma Şerit Sayısı
14	Yaklaşma Genişliği
15	Giriş Sapma Yarıçapı
16	Kol Sayısı
17	İç Ada Çapı
18	Giriş Genişliği
19	Giriş Şerit Sayısı
20	Dönüş Şerit Sayısı

Karışıklık matrisi makine öğrenimi yöntemlerinin ortaya koyduğu sonuçları değerlendirmek için kullanılır. Karışıklık matrisi modelin sonuçlarını hassasiyet, hata

oranı, duyarlılık, f-değeri, doğruluk gibi ölçümler üzerinden değerlendirir. Karışıklık matrisinde ölçümler Şekil 6.1 de gösterilen tablo üzerinden hesaplanır. Bu tabloda DP, doğru sınıflandırılan pozitif sınıf değerini; YP, yanlış sınıflandırılan pozitif sınıf değerini; YN, yanlış sınıflandırılan negatif sınıf değerini ve DN ise doğru sınıflandırılan negatif sınıf değerini verir (Güldal ve Çakıcı, 2017).

	Öngörülen Sınıf		
Gerçek Sınıf		**Sınıf 1 (Pozitif)**	**Sınıf 2 (Negatif)**
	Sınıf 1 (Pozitif)	Doğru Pozitif (DP)	Yanlış Negatif (YN)
	Sınıf 2 (Negatif)	Yanlış Pozitif (YP)	Doğru Negatif (DN)

Şekil 6.1 : Karışıklık matrisi (Güldal ve Çakıcı, 2017).

Çalışmada iki algoritmanın doğruluk performansı karşılaştırılacağı için Denklem 6.1'de karışıklık matrisinden doğruluk ölçümünün nasıl hesaplandığı gösterilmiştir (Güldal ve Çakıcı, 2017).

$$Doğruluk = \frac{DP+DN}{DP+YN+YP+DN} \tag{6.1}$$

7. SONUÇLAR

Betimsel istatistik sonuçlarına göre kazaya karışan erkeklerin sayısı kadınlardan daha fazla çıkmıştır. Bu durum kadınların bisiklet sürmekle erkeklere oranla daha az ilgilendiğini gösterebilir. Kazaya karışanların yaş ortalaması 39,22 olarak bulunmuştur. Bu da orta yaş grubundaki kullanıcıların bisiklete binmeyi tercih ettiğini gösterir. Kazaların büyük bir kısmı aydınlık ve güzel havada gerçekleşmiştir. Kaza sırasında yol yüzeyinin kuru olduğu durumda daha fazla kaza gerçekleşmiştir. Bu da bisikletli kullanıcıların kötü hava şartlarını tercih etmediğini gösterir. Kazalar çoğunlukla 30 mph hızda ve 4 kollu kavşaklarda meydana gelmiştir. Aşağıda uygulanan betimsel istatistiğin sonuçları verilmiştir. Yaklaşma şerit sayısı 1; giriş şerit sayısı 2 ve dönüş şerit sayısı 1 olan dönel kavşaklarda daha fazla kaza meydana gelmiştir. Yaklaşma genişliği ortalama 5,26; Giriş sapma yarıçapı ortalama 66,94; İç ada çapı ortalama 13,62 ve giriş genişliği ortalama 7,71 m olan dönel kavşaklarda daha fazla kaza meydana gelmiştir. Çizelge 7.1, 7.2 ve 7.3'te betimsel istatistik sonuçları verilmiştir.

Çizelge 7.1 : Betimsel İstatistik Sonuçları.

Kategorik Parametreler	0	1	2
Yaralanma derecesi	370	69	
Cinsiyet	384	55	
Işık Durumu	98	341	
Hava Şartları	12	54	373
Yol Yüzeyi	9	118	312
Bisikletliyi Görmeme	349	90	
Dikkatsizlik	99	340	
Bisikletliye Çok Yakın Geçme	58	381	
Yol Vermeme	101	338	
Manevra Yapmama	51	388	

Çizelge 7.2 : Betimsel İstatistik Sonuçları.

Kategorik Parametreler	1	2	3	4	5	6	20	30	40	50	60
Hız Limiti							5	323	41	13	42
Yaklaşma Şerit Sayısı	306	126	7								
Kol Sayısı			72	286	65	16					
Giriş Şerit Sayısı	189	221	28	1							
Dönüş Şerit Sayısı	270	162	7								

Çizelge 7.3 : Betimsel İstatistik Sonuçları.

Sayısal Parametreler	Mean	SD	Min	Max	Median	Range	Var
Yaş	39,22	14,17	5	80	39	5/80	200,8001
Yaklaşma Genişliği	5,26	1,79	3	11,37	4,77	3,00 /11,37	3,193841
Giriş Sapma Yarıçapı	66,94	21,38	19,23	99,98	68,12	19,23/ 99,98	457,0429
İç Ada Çapı	18,62	23,15	1	124,79	10	1,00/ 124,79	535,7227
Giriş Genişliği	7,71	2,55	3	20,27	7,41	3,00 / 20,27	6,499445

Ki-kare testinde p değeri 0.05'ten küçükse parametreler arasında anlamlı bir ilişkiden söz edilebilir. Çizelge 7.4' te ki kare testi sonuçları verilmiştir. Bu sonuçlara göre:

- Yaralanma derecesi ile hız Limiti, yaklaşma şerit sayısı, yaklaşma genişliği, giriş sapma yarıçapı, giriş genişliği, giriş şerit sayısı, dönüş şerit sayısı arasında bir ilişki tespit edilmiştir.
- Cinsiyet ile herhangi bir parametrenin ilişkisi çıkmamıştır.
- Yaş ile yol vermeme ve giriş sapma yarıçapı arasında;
- Işık durumu ile hava şartları ve yol yüzeyi arasında;
- Hava şartları ile ışık durumu, yol yüzeyi ve iç ada çapı arasında;
- Yol yüzeyi ile ışık durumu, hava şartları, bisikletliyi görmeme, bisikletliye çok yakın geçme, yaklaşma şerit sayısı ve yaklaşma genişliği arasında;
- Bisikletliyi görmeme ile yol yüzeyi, dikkatsizlik ve bisikletliye çok yakın geçme arasında;
- Dikkatsizlik ile bisikletliyi görmeme arasında;
- Bisikletliye çok yakın geçme ile yol yüzeyi, bisikletliyi görmeme, yol vermeme, manevra yapamama ve iç ada çapı arasında;
- Yol vermeme ile yaş ve bisikletliye çok yakın geçme arasında;
- Manevra yapamama ile bisikletliyi görmeme ve bisikletliye çok yakın geçme arasında;
- Hız limiti ile yaralanma derecesi, yaklaşma şerit sayısı, yaklaşma genişliği, giriş sapma yarıçapı, kol sayısı, iç ada çapı, giriş genişliği, giriş şerit sayısı, dönüş şerit sayısı arasında;
- Yaklaşma şerit sayısı ile yaralanma derecesi, yol yüzeyi, hız limiti, yaklaşma genişliği, giriş sapma yarıçapı, iç ada çapı, giriş genişliği, giriş şerit sayısı, dönüş şerit sayısı arasında;
- Yaklaşma genişliği ile yaralanma derecesi, yol yüzeyi, hız limiti, yaklaşma şerit sayısı, giriş sapma yarıçapı, kol sayısı, iç ada çapı, giriş genişliği, giriş şerit sayısı, dönüş şerit sayısı arasında;
- Giriş sapma yarıçapı ile yaralanma derecesi, yaş, hız limiti, yaklaşma şerit sayısı, yaklaşma genişliği, kol sayısı, iç ada çapı, giriş genişliği, giriş şerit sayısı, dönüş şerit sayısı arasında;

- Kol sayısı ile, hız limiti, yaklaşma genişliği, giriş sapma yarıçapı, iç ada çapı, giriş genişliği, giriş şerit sayısı, dönüş şerit sayısı arasında;
- İç ada çapı ile hava şartları, bisikletliye çok yakın geçme, hız limiti, yaklaşma şerit sayısı, yaklaşma genişliği, giriş sapma yarıçapı, kol sayısı, giriş genişliği, giriş şerit sayısı, dönüş şerit sayısı arasında;
- Giriş genişliği ile yaralanma derecesi, hız limiti, yaklaşma şerit sayısı, yaklaşma genişliği, giriş sapma yarıçapı, kol sayısı, iç ada çapı, giriş şerit sayısı, dönüş şerit sayısı arasında;
- Giriş şerit sayısı ile yaralanma derecesi, hız limiti, yaklaşma şerit sayısı, yaklaşma genişliği, giriş sapma yarıçapı, kol sayısı, iç ada çapı, giriş genişliği, dönüş şerit sayısı arasında;
- Dönüş şerit sayısı ile yaralanma derecesi, hız limiti, yaklaşma şerit sayısı, yaklaşma genişliği, giriş sapma yarıçapı, kol sayısı, iç ada çapı, giriş genişliği, giriş şerit sayısı arasında ilişkili bulunmuştur.

Çizelge 7.4 : Ki kare testi sonuçları.

Parametreler	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0.00	0.75	0.35	1	0.32	0.41	0.39	0.74	1	0.09	0.54	0.03	0,02	0.00	0.04	0.60	0.38	0.05	0.02	0.03
2	0.75	0.00	0.72	0.94	0.33	0.99	1	0.32	0.17	0.46	0.96	0.38	0.99	0.20	0.30	0.12	0.81	0.31	0.21	0.14
3	0.35	0.72	0.00	0.91	0.82	0.12	0.49	0.86	0.21	0.03	0.36	0.67	0.99	1	0.05	0.35	0.51	0.42	0.94	0.79
4	1	0.94	0.91	0.00	0.00	0,01	0.11	0.21	0.88	1	1	0.09	0.92	0.90	0.64	0.17	0.91	0.81	0.20	0.36
5	0.32	0.33	0.82	0.00	0.00	0.00	0.09	0.65	0.63	0.70	0.52	0.37	0.74	0.31	0.58	0.28	0.03	0.81	0.81	0.20
6	0.41	0.99	0.12	0,01	0.00	0.00	0.01	0.54	0.02	0.36	0.28	0.79	0.01	0.00	0.15	0.44	0.95	0.57	0.40	0.85
7	0.39	1	0.49	0.11	0.09	0.01	0.00	0.00	0.01	0.61	0.06	0.97	0.52	0.09	0.20	0.22	0.49	0.29	0.89	0.08
8	0.74	0.32	0.86	0.22	0.65	0.54	0.00	0.00	0.85	0.15	0.72	0.77	0.19	0.12	0.25	0.08	0.28	0.34	0.17	0.92
9	1	0.17	0.21	0.88	0.63	0.02	0.01	0.85	0.00	0.01	0.00	0.10	0.88	0.47	0.53	0.35	0.02	0.60	0.50	0.78
10	0.09	0.46	0.03	1	0.70	0.36	0.61	0.15	0.01	0.00	0.53	0.83	0.18	0.39	0.13	0.60	0.27	0.27	0.07	0.18
11	0.54	0.96	0.36	1	0.52	0.28	0.06	0.72	0.00	0.53	0.00	0.54	0.40	0.19	0.29	0.93	0.30	0.35	0.16	0.84
12	0.03	0.38	0.67	0.09	0.37	0.79	0.97	0.77	0.10	0.83	0.54	0.00	0	0	0	0.00	0	0.00	0	0
13	0,02	0.99	0.99	0.92	0.74	0.01	0.52	0.19	0.88	0.18	0.40	0	0.00	0.00	0.00	0,09	0	0	0.00	0.00
14	0.00	0.20	1	0.90	0.31	0.00	0.09	0.11	0.47	0.39	0.19	0	0.00	0.00	0.00	0	0.00	0.00	0	0
15	0.04	0.30	0.05	0.64	0.58	0.15	0.20	0.25	0.53	0.13	0.29	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0
16	0.60	0.12	0.35	0.17	0.28	0.44	0.22	0.08	0.35	0.60	0.93	0.00	0,09	0	0.00	0.00	0.00	0	0.03	0
17	0.38	0.81	0.51	0.91	0.03	0.95	0.49	0.28	0.03	0.27	0.30	0	0	0.00	0.00	0.00	0.00	0.00	0	0.00
18	0.05	0.31	0.42	0.81	0.81	0.57	0.29	0.34	0.60	0.27	0.35	0.00	0	0.00	0.00	0	0.00	0.00	0,03	0
19	0.02	0.21	0.94	0.20	0.81	0.40	0.89	0.17	0.50	0.07	0.16	0	0.00	0	0.00	0.03	0	0,03	0.00	0.00
20	0.03	0.14	0.79	0.36	0.20	0.85	0.08	0.92	0.78	0.19	0.84	0	0.00	0	0	0	0.00	0	0.00	0.00

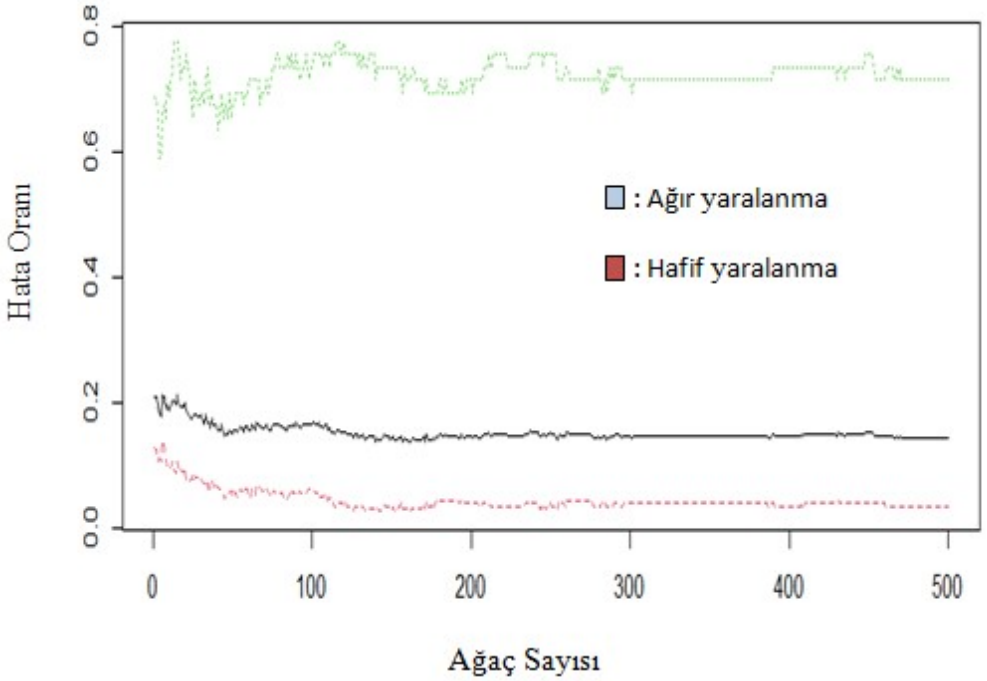
.

Rastgele Orman modelinde kullanılmak üzere %70 eğitim verisi ve %30 test verisi oluşturulmuştur. 307 adet eğitim verisi ve 132 adet test verisi oluşmuştur. Modeli oluşturduğumuz veri kümesi eğitim verileridir. Modelin doğruluğu ise test veri seti üzerinden yürütülecektir. Rastgele orman algoritması ile %70'lik eğitim veri seti kullanılarak bir sınıflandırma modeli oluşturulmuştur. Her bölünmede denenen değişken sayısı 4'tür. Modelin uygulanması sırasında 500 adet ağaç oluşturulmuştur. OOB verisinde %14.33' lük hata oranı meydana gelmiştir (Çizelge 7.5).

Çizelge 7.5 : Rastgele orman sınıflandırma modeli karışıklık matrisi sonuçları.

	0	1	Sınıflandırma Hatası
0	249	9	0.03488372
1	35	14	0.71428571

Şekil 7.1'de görüldüğü gibi ağaç sayısı bir noktadan sonra doğruluk oranını etkilememektedir. Yaklaşık 100 ağaçtan sonra doğruluk oranı sabit kalmıştır. Bu da daha fazla ağaç oluşturmanın bir anlamı olmadığını gösterir.



Şekil 7.1 : Rastgele orman oob hata oranı.

Model doğruluğu test veri kümesi üzerinde test edilmiştir. Doğruluğu değerlendirmek için tahmin işlevi kullanılmıştır. Sonuç olarak Rastgele Orman %83,33 doğruluk performansı ortaya koymuştur. Toplam 22 parametreyi yanlış tahmin etmiştir. Çizelge 7.6 da Karışıklık Matrisi sonuçları gösterilmektedir.

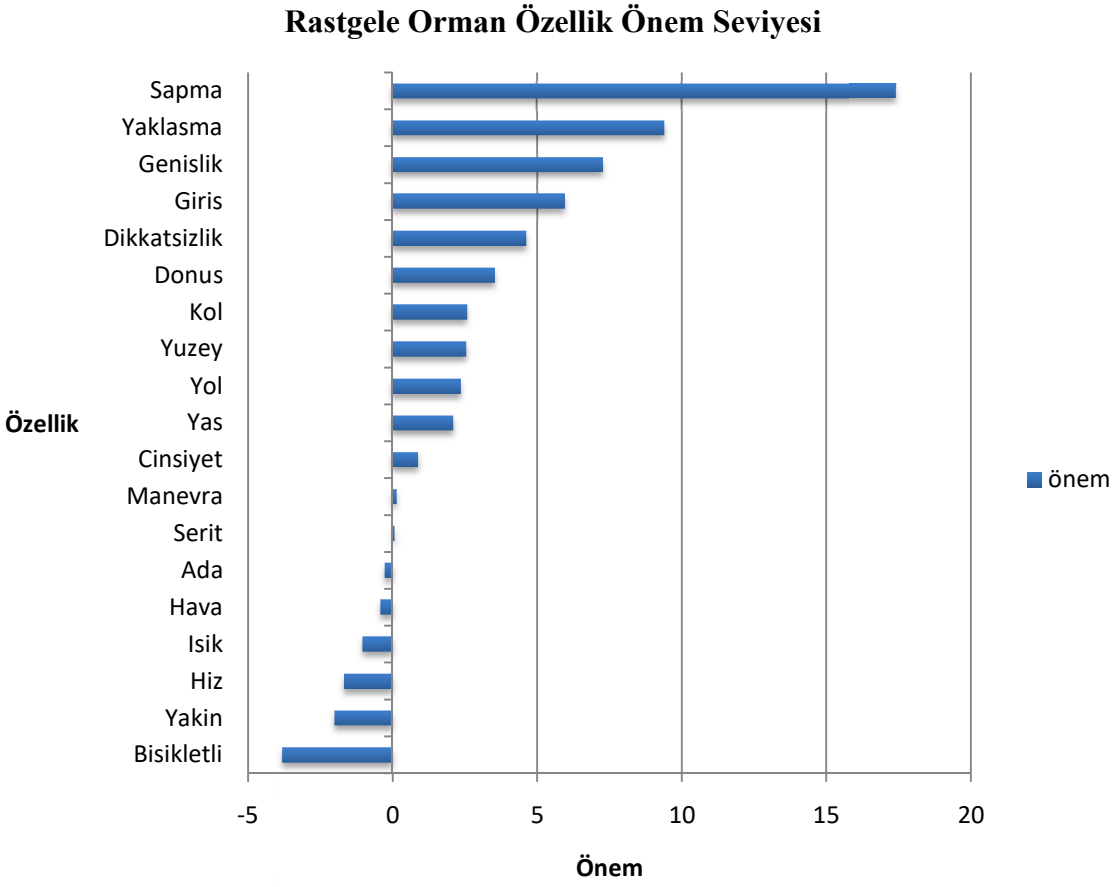
Çizelge 7.6 : Rastgele orman tahmin sonuçları.

	Hafif	Ciddi	Tahmin Doğruluğu
Hafif	108	18	%83,33
Ciddi	4	2	

Çizelge 7.7 ve Şekil 7.2' de Rastgele Orman modelinin özellik önem seviyesi verilmiştir. Model için en önemli parametre giriş sapma yarıçapıdır.

Çizelge 7.7 : Rastgele orman özellik önem seviyesi.

Özellik	Önem
Cinsiyet	0,86203613
Yas	2,07762732
Isik	-1,04626523
Hava	-0,43612238
Yuzey	2,52083928
Bisikletli	-3,80239537
Dikkatsizlik	4,58879173
Yakin	-1,99458789
Yol	2,35003131
Manevra	0,14793372
Hiz	-1,67860901
Yaklasma	9,36327936
Genislik	7,25453930
Sapma	17,39441318
Kol	2,56200532
Ada	-0,27099227
Giris	5,94645346
Serit	0,05396668
Donus	3,53079874



Şekil 7.2 : Rastgele orman özellik önem seviyesi.

Karar Ağacı Modelinde veri setinden seçilen %70 eğitim ve %30 test veri seti kullanılmıştır. Eğitim setlerinden bir sınıflandırma modeli oluşturulmuştur. Sınıflandırma modelinin özeti aşağıda (Şekil 7.3) görülmektedir.

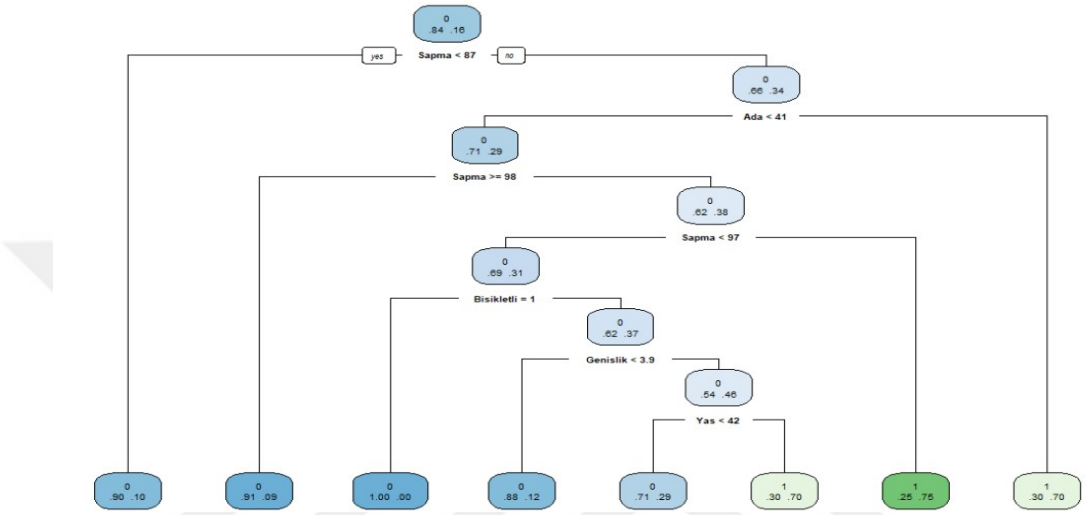
```
> classification_tree
n= 309

node), split, n, loss, yval, (yprob)
      * denotes terminal node

  1) root 309 49 0 (0.84142395 0.15857605)
    2) Sapma< 86.72 229 22 0 (0.90393013 0.09606987) *
    3) Sapma>=86.72 80 27 0 (0.66250000 0.33750000)
      6) Ada< 40.65 70 20 0 (0.71428571 0.28571429)
       12) Sapma>=98.115 23  2 0 (0.91304348 0.08695652) *
       13) Sapma< 98.115 47 18 0 (0.61702128 0.38297872)
         26) Sapma< 97.47 39 12 0 (0.69230769 0.30769231)
           52) Bisikletli>=0.5 7  0 0 (1.00000000 0.00000000) *
           53) Bisikletli< 0.5 32 12 0 (0.62500000 0.37500000)
            106) Genislik< 3.885 8  1 0 (0.87500000 0.12500000) *
            107) Genislik>=3.885 24 11 0 (0.54166667 0.45833333)
              214) Yas< 42 14  4 0 (0.71428571 0.28571429) *
              215) Yas>=42 10  3 1 (0.30000000 0.70000000) *
         27) Sapma>=97.47 8  2 1 (0.25000000 0.75000000) *
      7) Ada>=40.65 10  3 1 (0.30000000 0.70000000) *
```

Şekil 7.3 : Karar ağacı model özeti.

Şekil 7.4'te Karar Ağacı Modelinin oluşturduğu sınıflandırma ağacı görülmektedir. Ağaç giriş sapma yarıçapı ile başlamış olup evet/hayır sınıflandırması yapmıştır. Model en önemli parametreyi giriş sapma yarıçapı olarak belirlemiştir. Giriş sapma yarıçapının 87 den küçük olduğu durumda model %90 oranında hafif yaralanma sonucunu verir. Giriş sapma yarıçapının 87 den büyük olduğu durumda diğer parametrelere göre kaza yaralanma derecesi sınıflandırılmıştır.



Şekil 7.4 : Karar ağacı şeması.

Karar Ağacı Modelinde test veri setleriyle oluşturulan tahmin modeli sonrasında oluşan karışıklık matrisi Çizelge 7.8' de verilmiştir. Model 18 parametreyi yanlış tahmin etmiştir.

Çizelge 7.8 : Karar ağacı tahmin sonuçları.

	Hafif	Ciddi	TAHMİN DOĞRULUĞU
Hafif	106	14	%86.15
Ciddi	4	6	

Net sonuçlara bakıldığında Karar Ağacı Modeli Rastgele Ormana göre daha yüksek doğruluk oranı vermiştir. Karar Ağacı Modeli %86,15; Rastgele Orman Modeli %83,33 sınıflandırma doğruluğunda tahmin modeli oluşturmuştur. İki model içinde en önemli parametrenin giriş sapma yarıçapı olduğu görülmüştür.

8. TARTIŞMA

Bu çalışmada bisikletli kaza analizi için Rastgele Orman ve Karar Ağacı yöntemleri kulanılarak; veri setindeki bağımsız değişkenler ile kaza yaralanma derecesi sınıflandırılmış ve tahmin modelleri oluşturulmuştur. Ardından iki modelin doğruluk performansları karşılaştırılmıştır. Rastgele Orman ve Karar Ağacı yöntemlerinde tahmin doğruluk oranları sırasıyla %83,33 ve %86,15 olarak ortaya çıkmıştır. Rastgele Orman yöntemi için bu sonuç literatür ile kıyaslandığında düşük bir seviyede kalmıştır. Karar Ağacı yöntemi ise beklenen bir sonuç ortaya koymuştur. Rastgele Orman algoritması sınıflandırma modelinde ciddi yaralanmada %71'lik bir hata oranı vermiştir. Test verileriyle kurulan tahmin modelinde ise ciddi yaralanmada daha fazla yanlış tahmin yapmıştır. Bunun sebebi ciddi yaralanmalı kaza sayısının az olması olarak gösterilebilir. Bunun yanında Rastgele Orman yöntemi hafif yaralanmada daha az yanlış tahminde bulunmuştur. Bu durum, veri sayısının fazla olması durumunda modelin daha istikrarlı bir performans göstereceğini bize göstermektedir. Bu çalışmanın sınırlı yanı veri setindeki örnek sayısının az olmasıdır.

Karar Ağacı en önemli parametreyi giriş sapma yarıçapı olarak belirlemiştir. Model giriş sapma yarıçapının 87 den küçük olduğu durumda yaralanma derecesi %90 hafif yaralanma olarak sınıflandırmıştır. 87 den büyük olması durumunda %66 hafif yaralanma sonucu vermiş ve yeniden düğüm oluşturmuştur. İç ada çapının 41 den büyük olduğu durumda %30 ciddi yaralanma meydana gelmiştir; 41 den küçük olduğu durumda %71 hafif yaralanma sonucunu vermiştir. Ardından bir düğüm daha oluşmuş ve giriş sapma yarıçapının 98 den büyük ve 98' e eşit olduğu durumda %91 hafif yaralanma sonucu ortaya çıkmıştır. Karar Ağacında yaralanma derecesinin bisikletliyi görmeme, yaklaşma genişliği ve yaş gibi parametrelere bağlı olduğu da ortaya çıkmıştır. Karar Ağacı özellikle sapma yarıçapını ve iç ada çapını önemli parametreler olarak görmüştür. Bu iki parametrenin doğru tasarlanması halinde hızı düşürücü etkisi olduğu zaten bilinmektedir Karar Ağacı da bize bu sonucu tekrar göstermiştir.

Rastgele Orman özellik önem seviyesine bakıldığında giriş sapma yarıçapı, yaklaşma şerit sayısı ve yaklaşma genişliği sırasıyla en önemli 3 parametreyi göstermektedir. Giriş sapma yarıçapı ve yaklaşma genişliği açısından Karar Ağacında da buna benzer bir sonuç çıkmıştır. Yine dönel kavşaktaki kapasite ile doğrudan ilişkisi bulunan yaklaşma şerit sayısı da Rastgele Orman modelinde önemli bir parametre olarak gösterilmiştir. Gerek Ki-Kare testi gerekse kullanılan makine öğrenmesi algoritmaları bize bu çalışmada kavşak geometrik özelliklerinin ne kadar önemli olduğunu göstermiştir.

Bu çalışmadaki sonuçlar literatür ile karşılaştırıldığında Rastgele Orman algoritması daha düşük sonuç vermiştir. Bu durumun sebebi veri seti kaynaklı olabilir. Veri setinin az olması, modelin daha hassas olmasına neden olabilir ve her seferinde seçilen %70'lik eğitim verisindeki değişim doğruluk oranını etkileyebilir. Rastgele Orman algoritması Karar Ağacı yöntemlerinden türetilmiş ve daha doğru sonuç veren bir algoritma olarak kabul edilsede kaza analizlerinde farklı veri setleri için bu durum kesin kabul değildir. Bu çalışmada çıkan sonuç algoritmaların kesin bir doğruluk hiyerarşisine sahip olamayacağı ve ileride yapılacak çalışmalarda mutlaka farklı algoritmaların analizlerde değerlendirilmesi ve karşılaştırma yapılarak en doğru sınıflandırma ve tahmin modelleri oluşturmasını önermektir.

KAYNAKLAR

Abraham, A., Chong, M., Paprzycki, M. (2005). Traffic Accident Analysis Using Machine Learning Paradigms. *Informatica, 29* (1), 89–98.

Akgün, N. (2023). Testing The Performance of Random Forest and Support Vector Machine Algorithms for Predicting Cyclist Casualty Severity. *Firat University Journal of Experimental and Computational Engineering*, *2*(3), 124-133.

Akman, M., Genç, Y., Ankarali, H. (2011). Random forests yöntemi ve sağlık alanında bir uygulama. *Turkiye Klinikleri Journal of Biostatistics*, *3*(1), 36-48.

Akyiğit, H. E., Taşcı, T. (2022). Sigortacılık Sektöründe Makine Öğrenmesi İle Müşteri Kaybı Analizi. *Tasarım Mimarlık ve Mühendislik Dergisi*, *2*(1), 66-79.

AlMamlook, R. E., Kwayu, K. M., Alkasisbeh, M. R., Frefer, A. A. (2019). Comparison of machine learning algorithms for predicting traffic accident severity. In *2019 IEEE Jordan international joint conference on electrical engineering and information technology (JEEIT),* 272-276)

Al-Moqri, T., Haijun, X., Namahoro, J. P., Alfalahi, E. N., Alwesabi, I. (2020). Exploiting machine learning algorithms for predicting crash injury severity in Yemen: Hospital case study. *Applied and Computational Mathematics, 9*(5), 155–164.

Arslan, İ. (2022). *Veri Madenciliğinde Karar Ağacı Algoritmaları: Ufuk Üniversitesi Öğrenci Verileri İle Bir Çalışma.* (Yüksek lisans Tezi). Ufuk Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.

Bolat, H., Yücesan, M., Utku, A. (2022). Trafik kazalarının makine öğrenmesi yöntemleriyle analizi ve tahmini: Kahramanmaraş için örnek bir çalışma. *International Journal of Pure and Applied Sciences*, *8*(2), 490-506.

Bounsaythip, C., Rinta-Runsala, E. (2001). Overview of data mining for customer behavior modeling. *VTT Information Technology Research Report, Version*, *1*, 1-53.

Breiman, L. (2001). Random forests. *Machine learning*, *45*, 5-32.

Cevahir, E. (2020). *SPSS ile nicel veri analizi rehberi*. Kibele.

Chen, M. M., Chen, M. C. (2020). Modeling road accident severity with comparisons of logistic regression, decision tree and random forest. *Information*, *11*(5), 270.

Cornet, Y., Hook, H. (2025). Resistance to cycling infrastructure: Lessons from The Plain roundabout in Oxford. *European Transport Studies*, *2*, 100018.

Çalış, A., Kayapınar, S., Çetinyokuş, T. (2014). Veri Madenciliğinde Karar Ağacı Algoritmaları ile Bilgisayar ve İnternet Güvenliği Üzerine Bir Uygulama. *Endüstri Mühendisliği*, *25*(3), 2-19.

Çelik, A., Sevli, O. (2022). Predicting traffic accident severity using machine learning techniques. *Türk Doğa ve Fen Dergisi*, *11*(3), 79-83.

Emel, G., G., Taşkın, Ç. (2005). Veri Madenciliğinde Karar Ağaçları ve Bir Satış Analizi Uygulaması. *Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi, 6*(2), 221-239.

Eyduran, E., Özdemir, T., Küçük, M. (2005). Hayvancılıkta Kategorik Verilerde Ki-Kare ve G Testi. *Yüzüncü Yıl Üniversitesi Veteriner Fakültesi Dergisi*, *16*(1), 1-3.

Ersoy, M. (2012). *Çok Şeritli Döner Kavşakların Kapasite Analizleri: HCM 2010 Üzerine Bir Değerlendirme.*(Yüksek lisans tezi).İstanbul Teknik Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul.

Furtado, G. (2004). Accommodating vulnerable road users in roundabout design. In *2004 TAC Annual Conference*.

Gacar, B. K., Kocakoç, İ. D. (2020).Regresyon Analizleri mi Karar Ağaçları mı? *Manisa Celal Bayar Üniversitesi Sosyal Bilimler Dergisi, 2020; 18* (4); 251-260.

Gan, X., Weng, J. (2020). Predicting Crash Injury Severity for the Highways Involving Traffic Hazards and Those Involving No Traffic Hazards. In *CICTP 2020*, 4195-4206

Gökalp, Ö. M. (2022). Makine Öğrenmesi. *Gazi Üniversitesi, Gazi Bilişim Enstitüsü, Adli Bilişim Bölümü,* 1-14.

Güldal, H., Çakıcı, Y.(2017). Ders yönetim sistemi yazılımı kullanıcı etkileşimlerinin sınıflandırma algoritmaları ile analizi. *Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, *21*(4), 1355-1367.

Güngör, M., Bulut, Y. (2008). Ki-kare testi üzerine. *Fırat Üniversitesi Doğu Araştırmaları Dergisi*, *7*(1), 84-89.

Gürsel, M. F., Akdur, H. T. K. (2024). Trafik Kaza Sayılarının Regresyon ve Karar Ağacı Yöntemleri ile Modellenmesi: Ankara Devlet Yolları Örneği. *Gazi Üniversitesi Fen Fakültesi Dergisi*, *5*(2), 149-160.

Hels, T., Orozova-Bekkevold, I. (2007). The effect of roundabout design features on cyclist accident rate. *Accident Analysis & Prevention*, *39*(2), 300-307.

Iranitalab, A., Khattak, A. (2017). Comparison of four statistical and machine learning methods for crash severity prediction. *Accident Analysis & Prevention*, *108*, 27-36.

İnançlı, M. (2012). *Dönel Kavşakların Güvenliği ve Konya'daki Bazı Kavşakların İncelenmesi*. (Yüksek lisans tezi). Bahçeşehir Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul.

Karayolları Genel Müdürlüğü, (2005). *Karayolu Tasarım El Kitabı*. Ankara

Korkem, E. (2013). *Mikroarray Gen Ekspresyon Veri Setlerinde Random Forest ve Naive Bayes Sınıflama Yöntemleri Yaklaşımı*. (Yüksek Lisans Tezi). Hacettepe Üniversitesi, Sağlık Bilimleri Enstitüsü. Ankara.

Korkmaz, G. (2025). Karar Ağaçlarının Aşırı Uyumla Sınavı: Karşılaştırmalı Bir Performans Analizi, *Bilişim Teknolojileri Dergisi*, *18 (3),* 253-268.

Krishnaveni, S., Hemalatha, M. (2011). A perspective analysis of traffic accident using data mining techniques. *International Journal of Computer Applications*, *23*(7), 40-48.

Kubat, A., Ayaşlıgil, T. (2002). Doğa bilimlerinde ki-kare bağımsızlık testi ve tek yönlü varyans analiz testi uygulamasında SPSS programı kullanımı. *Journal of the Faculty of Forestry Istanbul University*, *52*(1), 155-167.

Lee, J., Yoon, T., Kwon, S., Lee, J. (2019). Model evaluation for forecasting traffic accident severity in rainy seasons using machine learning algorithms: Seoul city study. *Applied Sciences*, *10*(1), 129.

McHugh, M. L. (2013). The chi-square test of independence. *Biochemia medica*, *23*(2), 143-149.

Miran, B. (2021). *Temel İstatistik: Genişletilmiş Baskı* (Vol. 1).

Oğuzlar, A. (2004). CART Analizi ile Hanehalkı İşgücü Anketi Sonuçlarının Özetlenmesi, *iktisadi ve İdari Bilimler Dergisi, 18*(3-4), 79-90.

Öğütveren, E. (2019). *Modern Dönel Kavşakların Geometrik Tasarımı ve Kapasite İlişkisi.* (Yüksek lisans tezi). Pamukkale Üniversitesi, Fen Bilimleri Enstitüsü, Denizli.

Özinal, Y., Uz , V. E.(2021). Dönel kavşak geometrik elemanlarının kavşak güvenliği üzerine etkisinin literatür ışığında değerlendirilmesi. *Politeknik Dergisi, 24*(1), 283-297.

Özşahin, M., Yilmaz, K. B., Akcan, S. (2022). Makine öğrenmesi algoritmalari ile ev kirasi tahmini. *Araştırma ve Değerlendirmeler II*, 23.

Panicker, A. K., Ramadurai, G. (2020). Injury severity prediction model for two-wheeler crashes at mid-block road sections. *International Journal of Crashworthiness*, 1–9.

Patel, N., Upadhyay, S. (2012). Study of various decision tree pruning methods with their empirical comparison in WEKA. *International journal of computer applications*, *60*(12).

Robinson, B. W., Rodegerdts, L., Scarborough, W., Kittelson, W., Troutbeck, R., Brilon, W., ... , Mason, J. (2000). *Roundabouts: An informational guide* (No. FHWA-RD-00-067; Project 2425). United States. Federal Highway Administration.

Rodegerdts, L. A. (2010). Roundabouts: An informational guide. *Transportation Research Board, 672.*

Santos, K., Dias, J. P., Amado, C. (2022). A literature review of machine learning algorithms for crash injury severity prediction. *Journal of safety research*, *80*, 254-269.

Sarıcaoğlu., C. (2019). *Sözcüksel Analiz Kulanılarak Kötü Niyetli Url'leri Derin Öğrenme Teknikleri İle Tespit Etme*. (Yüksek lisans Tezi). Gazi Üniversitesi, Fen Bilimler Enstitüsü, Ankara.

Scarano, A., Riccardi, M. R., Mauriello, F., D'Agostino, C., Pasquino, N., Montella, A. (2023). Injury severity prediction of cyclist crashes using random forests and random parameters logit models. *Accident Analysis & Prevention*, *192*, 107275.

Sezer, E. A., Bozkır, A. S., Yağız, S., Gökçeoğlu, C. (2010). Karar ağacı derinliğinin CART algoritmasında kestirim kapasitesine etkisi: bir tünel açma makinesinin ilerleme hızı üzerinde uygulama. *Akıllı Sistemlerde Yenilikler ve Uygulamaları Sempozyumu, Kayseri*.

Taekratok, T. (1998). Modern Roundabouts for Oregon. Oregon Department of Transportation Research Unit, Salem, Oregon.

Topcu, D.,Vupa Çilengiroğlu, Ö. (2021). Futbol Verilerinin Karar Ağaçları ve Lojistik Regresyon Yöntemleri ile İncelenmesi. *Spor Bilimleri Araştırmaları Dergisi, 6*(2), 491-503.

Topyıldız, N. (2019). *Veri Madenciliği Sınıflandırma Yöntemlerinden Karar Ağaçları ve Bir Uygulama.* (Yüksek lisans Tezi). Selçuk Üniversitesi, Fen Bilimleri Enstitüsü, Konya.

Tuncuk, M. (2004). *Coğrafi Bilgi Sistemi Yardımıyla Trafik Kaza Analizi: Isparta Örneği.* (Yüksek lisans Tezi). Süleyman Demirel Üniversitesi, Fen Bilimleri Enstitüsü, Isparta.

Umer, M., Sadiq, S., Ishaq, A., Ullah, S., Saher, N., Madni, H. A. (2020). Comparison Analysis Of Tree Based And Ensembled Regression Algorithms For Traffic Accident Severity Prediction. *ArXiv Preprint. ArXiv*:2010.14921.

Yavuz, A. A., Ergül, B., Aşık, E. G. (2021). Trafik kazalarının makine öğrenmesi yöntemleri kullanılarak değerlendirilmesi. *International Journal of Engineering Research and Development*, *13*(1), 66-73.

Yan, M., Shen, Y. (2022). Traffic accident severity prediction based on random forest. *Sustainability*, *14*(3), 1729.

Yılmaz, H. (2014). *Random Forests Yönteminde Kayıp Veri Probleminin İncelenmesi ve Sağlık Alanında Bir Uygulama.* (Yüksek lisans tezi). Eskişehir Osmangazi Üniversitesi, Sağlık Bilimleri Enstitüsü, Ocak 2014.

Yılmaz, H. (2022). *Eşdüzey Kavşakların Performans Etkilerinin Araştırılması: Bursa Esentepe Kavşağı Örneği.* (Yüksek Lisans Tezi). Bilecik Şeyh Edebali Üniversitesi, Fen Bilimleri Enstitüsü, Bilecik.

Yüzer, A. F. (2009). *İstatistik* (Vol. 1448). Anadolu Universitesi.

Zhang, J., Li, Z., Pu, Z., Xu, C. (2018). Comparing prediction performance for crash injury severity among various machine learning and statistical methods. *IEEe Access*, *6*, 60079-60087.

Zorman, M., Podgorelec, V., Kokol, P., Peterson, M., Šprogar, M., Ojsteršek, M. (2001). Finding the right decision tree's induction strategy for a hard real world problem. *International journal of medical informatics*, *63*(1-2), 109-121.

ÖZGEÇMİŞ

Ad-SOYAD **:**Ebru SADIK

ÖĞRENİM DURUMU:

- **Lisans** **:** 2021, Afyon Kocatepe Üniversitesi, Mühendislik Fakültesi, İnşaat Mühendisliği.
- **Yüksek Lisans** **:** 2023, Bursa Teknik Üniversitesi, İş Sağlığı ve Güvenliği, Tezsiz Yüksek Lisans Programı.

TEZDEN TÜRETİLEN ESERLER, SUNUMLAR VE PATENTLER:

- **Sadık, E., Akgün, N.** (2025). Rastgele Orman ve Karar Ağacı Yöntemlerinin Karşılaştırılması: Bisikletli Kaza Analizi Örnek Çalışması. *5. Uluslararası Lisansüstü Çalışmalar Kongresi*, 25-28 Haziran 2025.