

T.C.
MUNZUR ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ



**MAKİNE ÖĞRENME YÖNTEMLERİ İLE KAHRAMANMARAŞ İLİNDEKİ TRAFİK
KAZASI SONUÇLARININ ANALİZİ VE TAHMİN EDİLMESİ**

Hüseyin BOLAT

YÜKSEK LİSANS TEZİ
MÜHENDİSLİK YÖNETİMİ ANABİLİM DALI

DANIŞMAN
Doç. Dr. Melih YÜCESAN

TUNCELİ-2022

T.C.
MUNZUR ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

**MAKİNE ÖĞRENME YÖNTEMLERİ İLE KAHRAMANMARAŞ İLİNDEKİ
TRAFİK KAZASI SONUÇLARININ ANALİZİ VE TAHMİN EDİLMESİ**

Hüseyin BOLAT
(190080023)

YÜKSEK LİSANS TEZİ
MÜHENDİSLİK YÖNETİMİ ANABİLİM DALI

DANIŞMAN
Doç. Dr. Melih YÜCESAN

TUNCELİ-2022

T.C.
MUNZUR ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

**MAKİNE ÖĞRENME YÖNTEMLERİ İLE KAHRAMANMARAŞ İLİNDEKİ
TRAFİK KAZASI SONUÇLARININ ANALİZİ VE TAHMİN EDİLMESİ**

Hüseyin BOLAT
YÜKSEK LİSANS TEZİ
MÜHENDİSLİK YÖNETİMİ ANABİLİM DALI

Bu tez 23.05.2022 tarihinde aşağıdaki jüri üyeleri tarafından ~~oybirliği/oyçokluğu~~ ile kabul edilmiştir.

İmza	İmza	İmza
Doç. Dr. Melih YÜCESAN (Munzur Üniversitesi)	Doç. Dr. Özal YILDIRIM (Fırat Üniversitesi)	Dr. Öğr. Üyesi Ümit CAN (Munzur Üniversitesi)
DANIŞMAN	ÜYE	ÜYE

Bu seminer, Enstitümüz Mühendislik Yönetimi Anabilim Dalı'nda hazırlanmıştır.

Doç. Dr. Murat KORUNUR
Enstitü Müdürü

NOT: Bu tezde kullanılan özgün ve başka kaynaktan yapılan bildirişlerin, çizelge, şekil ve fotoğrafların kaynak gösterilmeden kullanımı, 5846 sayılı “Fikir ve Sanat Eserleri Kanunu”ndaki hükümlere tabidir.

23/05/2022

ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ

Bu tezin bana ait, özgün bir çalışma olduğunu; çalışmamın hazırlık, veri toplama, analiz ve bilgilerin sunumu olmak üzere tüm aşamalarında bilimsel etik ilke ve kurallara uygun davrandığımı; bu çalışma kapsamında elde edilen tüm veri ve bilgiler için kaynak gösterdiğimi ve bu kaynaklara kaynakçada yer verdiğimi ve hiçbir şekilde “intihal içermediğini” beyan ederim. Herhangi bir zamanda, çalışmamla ilgili yaptığım bu beyana aykırı bir durumun saptanması durumunda, ortaya çıkacak tüm ahlaki ve hukuki sonuçları kabul ettiğimi bildiririm.

İmza
Hüseyin BOLAT

Danışman
Doç. Dr. Melih YÜCESAN

TEŞEKKÜR

Bu tez çalışmasında makine öğrenme teknikleri ile Kahramanmaraş ilindeki trafik kazası sonuçlarının analizleri yapılmıştır. Tez çalışması boyunca beni destekleyen eşime ve kızıma, vermiş olduğu fikirlerle tezi geliştirmeme yardımcı olan danışman hocam Sayın Doç. Dr. Melih YÜCESAN'a teşekkür eder, saygılarımı sunarım.

Hüseyin BOLAT
TUNCELİ - 2022

ÖZET

Trafik kazaları insan yaralanmaları ve ölümlerinin en büyük nedenlerindendir. Dünya üzerinde her yıl 1 milyonun üzerinde ölümlü, 50 milyona yakın yaralanmalı trafik kazası meydana gelmektedir. Bu kazalarda ortalama her 2 saniyede bir trafik kazası yaşanırken her 50 saniyede bir ölümcül kaza ile sonuçlanmaktadır. Bu çalışmada, 2017 ile 2020 yılları arasında Kahramanmaraş il merkezi ve ilçelerinde gerçekleşen 28 adet öznitelikten oluşan 7929 adet trafik kazası WEKA aracında makine öğrenme teknikleri ile analiz edildi. Trafik kazalarının analizinde sınıflandırma, kümeleme, birliktelik kuralı ve görsel karşılaştırma yöntemleri kullanılmaktadır. Sınıflandırma yönteminde Naive Bayes, Destek Vektör Makinesi, K En Yakın Komşu, PART, AdaboostM1, J48 ve Rastgele Orman algoritmaları kullanıldı. Kümeleme yönteminde SimpleKMeans ve EM kümeleme algoritmaları kullanıldı. Birliktelik yönteminde Apriori algoritmaları kullanıldı. Sonuç olarak trafik kazalarının tahmininde sınıflandırma algoritmalarının performans ölçütlerine bakıldığında karar ağaçlarından J48 algoritması en başarılı sonuçları elde ederken, NB algoritmasının başarı düzeyi kötü olmasa da diğer algoritmalarla oranla daha düşük kalmıştır. Tüm yöntemlerin ortak sonucu gelecekte yaşanabilecek ölümle ve yaralanmalı potansiyel kazaların daha çok sabah saat 04:00 ile 08:00 ve akşam saat 16:00 ile 20:00 arasında, Kahramanmaraş il merkezi ve Onikişubat ilçesinde yerleşim yeri içerisinde, Göksun ilçesindeki yerleşim yeri dışındaki eğimli ve virajlı yollarda ve Pazarcık ilçesindeki yerleşim yeri dışındaki bölünmüş yollarda yaşanma riskinin yüksek olduğu belirlenmiştir. Ayrıca meydana gelen kazalarda sürücülerin trafik ve hız kurallarına uymadığı, trafik lambası, trafik levhası, yaya yolu, emniyet şeridi ve banket gibi yol, çevre ve güvenlik ekipmanlarının yetersiz kaldığı belirlenmiştir.

Anahtar Kelimeler: Makine öğrenmesi, WEKA, trafik kazaları, sınıflandırma, kümeleme, birliktelik.

ABSTRACT

Analysis And Estimating Of Traffic Accidents Results In Kahramanmaraş With Machine Learning Methods

Traffic accidents are one of the biggest causes of human injuries and deaths. Every year, more than 1 million deaths and 50 million injuries occur in traffic accidents around the world. While there is a traffic accident every 2 seconds in these accidents, it results in a fatal accident every 50 seconds. In this study, 7929 traffic accidents, consisting of 28 features, that took place in Kahramanmaraş city center and its districts between 2017 and 2020 were analyzed by machine learning techniques in the WEKA tool. Classification, clustering, association rule and visual comparison methods are used in the analysis of traffic accidents. Naive Bayes, Support Vector Machine, K Nearest Neighbor, PART, AdaboostM1, J48 and Random Forest algorithms were used in the classification method. SimpleKMeans and EM clustering algorithms were used in the clustering method. Apriori algorithms were used in the association method. As a result, when the performance criteria of the classification algorithms in the prediction of traffic accidents are examined, the J48 algorithm from the decision trees obtained the most successful results, while the success level of the NB algorithm was lower than the other algorithms, although it was not bad. The common result of all methods is that potential accidents with death and injury that may occur in the future are mostly between 04:00 and 08:00 in the morning and between 16:00 and 20:00 in the evening, in the settlement area of Kahramanmaraş city center and Onikisubat district, on slopes outside the settlement in Göksun district. It has been determined that the risk of living is high on winding and winding roads and on divided roads outside the settlement in Pazarcık district. In addition, it was determined that the drivers did not comply with the traffic and speed rules in the accidents that occurred, and the road, environment and safety equipment such as traffic lights, traffic signs, pedestrian ways, safety lanes and banquettes were insufficient.

Keywords: Machine learning, WEKA, traffic accidents, classification, clustering, association.

İÇİNDEKİLER

ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ	I
TEŞEKKÜR.....	II
ÖZET	III
ABSTRACT	IV
İÇİNDEKİLER.....	V
ŞEKİLLER LİSTESİ	VII
TABLOLAR LİSTESİ	IX
SEMBOLLER LİSTESİ	X
KISALTMALAR LİSTESİ	XI
1. GİRİŞ.....	1
1.1. Problemin Tanımı.....	4
1.2. Çalışmanın Amacı ve Önemi	4
1.3. Literatür Araştırması	5
1.4. Makine Öğrenmesi	10
1.5. Makine Öğrenmesi Yöntemleri	11
1.5.1. Denetimli Öğrenme	11
1.5.2. Denetimsiz Öğrenme	12
1.5.3. Yarı Denetimli Öğrenme	13
1.5.4. Pekiştirmeli Öğrenme	13
2. MATERYAL VE YÖNTEM.....	15
2.1. Weka.....	15
2.2. Weka Explorer.....	17
2.2.1. Preprocess Menüsü	17
2.2.2. Classify Menüsü	18
2.2.3. Cluster Menüsü.....	19
2.2.4. Associate Menüsü.....	20
2.2.5. Select Attributes Menüsü.....	20
2.2.6. Visualize Menüsü	21
2.3. WEKA Veri Seti Özellikleri.....	22
2.4. Çalışmada Kullanılan Veri Seti	23
2.4.1. Trafik Veri Seti	23
2.5. Çalışmada Kullanılan Yöntemler	25
2.6. Sınıflandırma Yöntemi	26
2.6.1. Naive Bayes Algoritması.....	27
2.6.2. Destek Vektör Makinası	28
2.6.3. KNN Algoritması	29
2.6.4. PART Algoritması	30
2.6.5. AdaBoostM1 Algoritması.....	30
2.6.6. J48 Algoritması.....	31
2.6.7. Rastgele Orman Algoritması	33
2.7. Sınıflandırma Algoritmalarının Performans Ölçütleri.....	34
2.8. Kümeleme Yöntemi	36
2.8.1. SimpleKMeans Algoritması	36
2.8.2. EM Algoritması	37
2.9. Birlikte Kuralı Çıkarma Yöntemi.....	38
2.9.1. Apriori Algoritması	38
3. BULGULAR VE TARTIŞMA	41

3.1. Önişlem ve Görsel Karşılaştırma Bulguları	41
3.2. Sınıflandırma Yöntemi Bulguları	54
3.2.1. Yerleşim Yeri Sınıfına Göre Algoritmaların Tahminleri	54
3.2.2. Yolun Tipi (Yönü) Sınıfına Göre Algoritmaların Tahminleri	56
3.2.3. Yolun Yatay Geometrisi Sınıfına Göre Algoritmaların Tahminleri	57
3.2.4. Gün Durumu Sınıfına Göre Algoritmaların Tahminleri	59
3.2.5. Kazaya Karışan Araç Sayısına Göre Algoritmaların Tahminleri	60
3.2.6. Kaza Sonucu Sınıfına Göre Algoritmaların Tahminleri	61
3.3. Kümeleme Yöntemi Bulguları	63
3.4. Birliktelik Yöntemi Bulguları	66
4. SONUÇ VE ÖNERİLER	70
4.1. Kısıtlama ve Zorluklar	72
4.2. Potansiyel Gelecek Çalışmalar	72
5. KAYNAKLAR	73
ÖZGEÇMİŞ	79

ŞEKİLLER LİSTESİ

Şekil 1.1. Makine öğrenme yöntemleri	11
Şekil 1.2. Denetimli öğrenme.....	12
Şekil 1.3. Denetimsiz öğrenme	12
Şekil 1.4. Yarı denetimli öğrenme	13
Şekil 1.5. Pekiştirmeli öğrenme	14
Şekil 2.1. Veri madenciliği alanında kullanılan programlar	16
Şekil 2.2. Weka önışlem menü ekranı.....	17
Şekil 2.3. Weka sınıflandırma menüsü ekranı.....	18
Şekil 2.4. Weka kümeleme menüsü ekranı	19
Şekil 2.5. Weka birliktelik menüsü ekranı	20
Şekil 2.6. Weka nitelik seçimi menüsü ekranı	21
Şekil 2.7. Weka görsel karşılaştırma menüsü ekranı	22
Şekil 2.8. CSV dosyası örneği.....	22
Şekil 2.9. Arff dosyası örneği.....	23
Şekil 2.10. Çalışmanın akış şeması	25
Şekil 2.11. Sınıflandırma yöntemi şeması.....	26
Şekil 2.12. Çapraz doğrulama süreci.....	27
Şekil 2.13. Destek vektörü ile sınıflandırma modeli	28
Şekil 2.14. K en yakın komşu algoritması sınıflandırma modeli	30
Şekil 2.15. AdaBoostM1 sınıflandırma modeli.....	31
Şekil 2.16. J48 algoritması sınıflandırma modeli.....	32
Şekil 2.17. Rasgele orman algoritması sınıflandırma modeli	33
Şekil 2.18. Karmaşıklık matrisi örneği.....	34
Şekil 2.19. ROC eğrisi grafiği	35
Şekil 2.20. SimpleKMeans algoritması kümeleme modeli	37
Şekil 2.21. EM algoritması kümeleme modeli	37
Şekil 3.1. Yıllara göre yaşanan trafik kazalarının önışlem verileri	42
Şekil 3.2. Yıl ile kaza sonuçları arasındaki görsel karşılaştırma sonuçları	42
Şekil 3.3. Aylara göre yaşanan trafik kazalarının önışlem verileri	43
Şekil 3.4. Ay ve kaza sonuçları arasındaki görsel karşılaştırma sonuçları.....	43
Şekil 3.5. Haftanın günlerine göre yaşanan trafik kazalarının önışlem verileri	44
Şekil 3.6. Haftanın günleri ile kaza sonuçlarının görsel karşılaştırma sonuçları	44
Şekil 3.7. Saat dilimlerine göre yaşanan trafik kazalarının önışlem verileri.....	45
Şekil 3.8. Saat dilimi ve kaza sonuçları arasındaki görsel karşılaştırma sonuçları	45
Şekil 3.9. İlçelerde yaşanan trafik kazalarının önışlem verileri	46
Şekil 3.10. İlçe ile kaza sonuçları arasındaki görsel karşılaştırma sonuçları	46
Şekil 3.11. Kaza yerine göre yaşanan trafik kazalarının önışlem verileri	47
Şekil 3.12. Kaza yeri ile kaza sonuçları arasındaki görsel karşılaştırma sonuçları.....	47
Şekil 3.13. Yolun sınıfına göre yaşanan trafik kazalarının önışlem verileri	48
Şekil 3.14. Yolun sınıfı ve kaza sonuçlarının görsel karşılaştırma sonuçları	48
Şekil 3.15. Yolun düşey geometrisine göre yaşanan trafik kazalarının önışlem verileri	49
Şekil 3.16. Yolun düşey geometrisi ile kaza sonuçlarının görsel karşılaştırma sonuçları ..	49
Şekil 3.17. Gün durumuna göre yaşanan trafik kazalarının önışlem verileri.....	50
Şekil 3.18. Gün durumu ile kaza sonuçları arasındaki görsel karşılaştırma sonuçları.....	50
Şekil 3.19. Araç sayısına göre yaşanan trafik kazalarının önışlem verileri	51
Şekil 3.20. Araç sayısı ile kaza sonuçları arasındaki görsel karşılaştırma sonuçları	51
Şekil 3.21. Kaza türüne göre yaşanan trafik kazalarının önışlem verileri.....	52

Şekil 3.22. Kaza türü ile kaza sonuçlarının görsel karşılaştırma sonuçları	52
Şekil 3.23. Yolun tipine göre yaşanan trafik kazalarının önişlem verileri	53
Şekil 3.24. Yolun tipi ve kaza sonuçları arasındaki görsel karşılaştırma sonuçları	53
Şekil 3.25. Saat dilimi niteliğine göre SimpleKMeans ve EM kümeleme sonuçları	63
Şekil 3.26. Kaza yeri niteliğine göre SimpleKMeans ve EM kümeleme sonuçları	64
Şekil 3.27. Yolun tipi niteliğine göre SimpleKMeans ve EM kümeleme sonuçları	64
Şekil 3.28. Gün durumu niteliğine göre SimpleKMeans ve EM kümeleme sonuçları	64
Şekil 3.29. Araç sayısı niteliğine göre SimpleKMeans ve EM kümeleme sonuçları.....	65
Şekil 3.30. Yolun tipi niteliğine göre SimpleKMeans ve EM kümeleme sonuçları	65
Şekil 3.31. Kaza sonucu niteliğine göre SimpleKMeans ve EM kümeleme sonuçları	65

TABLÖLAR LİSTESİ

Tablo 2.1. Trafik veri seti özellikleri	24
Tablo 3.1. Sınıflandırma yönteminde kullanılan öznitelik başlıkları	54
Tablo 3.2. Yerleşim yeri sınıfına göre karmaşıklık matrisleri	55
Tablo 3.3. Yerleşim yeri sınıfına göre performans ölçütleri.....	55
Tablo 3.4. Yolun yönü sınıfına göre karmaşıklık matrisleri	56
Tablo 3.5. Yolun yönü sınıfına göre performans ölçütleri.....	57
Tablo 3.6. Yolun yatay geometrisi sınıfına göre karmaşıklık matrisleri	57
Tablo 3.7. Yolun yatay geometrisi sınıfına göre performans ölçütleri	58
Tablo 3.8. Gün durumu sınıfına göre karmaşıklık matrisleri.....	59
Tablo 3.9. Gün durumu sınıfına göre performans ölçütleri	59
Tablo 3.10. Kazaya karışan araç sayısına göre karmaşıklık matrisi	60
Tablo 3.11. Kazaya karışan araç sayısına göre performans ölçütleri	61
Tablo 3.12. Kaza sonucuna göre karmaşıklık matrisi.....	62
Tablo 3.13. Kaza sonucuna göre performans ölçütleri	62
Tablo 3.14. Apriori algoritması ile oluşturulan kurallar	66

SEMBOLLER LİSTESİ

$\mathcal{E}$	: Entropi değeri
h_{man}	: En yakın komşu manhattan uzaklığı
h_{min}	: En yakın komşu minkowski uzaklığı
$h_{öklid}$	: En yakın komşu öklid uzaklığı
K_d	: Doğruluk ölçütü
K_f	: F ölçütü
K_i	: Kappa istatistiği oranı
K_p	: Kesinlik ölçütü
K_r	: Duyarlılık ölçütü
L_b	: Beklenen doğruluk oranı
L_q	: Gözlemlenen doğruluk oranı
N	: Örnek sayısı
P	: Olasılık değeri
P_d	: Destek değeri
P_g	: Güven değeri
P_k	: Kaldıraç değeri

KISALTMALAR LİSTESİ

ARFF	: Attribute Relation File Format
CART	: Classification and Regression Tree
CBS	: Coğrafi Bilgi Sistemleri
CSV	: Comma Separated Values
DSÖ	: Dünya Sağlık Örgütü
DVM	: Destek Vektör Makinesi
EM	: Expectation-Maximization
FN	: False Negative
FP	: False Pozitive
FT	: Fonksiyonel Ağaç Algoritması
IBk	: K En Yakın Komşu Algoritması
MAE	: Mean Absolute Error
NB	: NaiveBayes Algoritması
PART	: Partial Decision Trees
RF	: Rasgele Orman Algoritması
ROC	: Receiver Operating Characteristics
SMO	: Sequential Minimal Optimization
TN	: True Negative
TP	: True Pozitive
TÜİK	: Türkiye İstatistik Kurumu
WEKA	: Waikato Environment for Knowledge Analysis
WHO	: World Health Organization

1. GİRİŞ

Trafik kazası, yerleşim yeri içinde veya dışındaki karayolları üzerinde hareket eden bir ya da daha fazla aracın karıştığı, yaralanma, ölüm ve maddi hasarla sonuçlanan olaylar olarak tanımlanabilir (Anderson, 2009). Trafik kazalarına, genelde sürücüler, yolcular, taşıtlar ile yol ve hava koşulları sebebiyet vermektedir (Oña ve ark., 2011).

Dünyada ülkeler gelişmeye devam etmekte ve buna bağlı olarak insan sayısı ile birlikte araç sayısı da artış göstermektedir. Ülkelerde artan araç sayıları şehir yollarında trafiğin yoğunlaşmasına neden olmaktadır (Zengin ve ark., 2018). Bu yoğunluk trafikte geçirilen zamanı artırmakta ve bunun sonucu kaza oranlarının yükselmesine neden olmaktadır (AlMamlook ve ark., 2019). Yükselen trafik kaza sayıları tüm dünya ülkeleri arasına en ciddi sorunlardan biri olup, günümüzde ortalama 2 saniyede bir trafik kazası yaşanırken 50 saniyede bir ölümcül kaza ile sonuçlanmaktadır (Goniewicz ve ark., 2016). Trafik kazaları önemli sağlık sorunlarını da beraberinde getirerek insan vücudunda kalıcı hasarlara ve erken ölümlerin yaşanmasına sebebiyet vermektedir (Akomolafe ve Olutayo, 2013). Her yıl 1 milyonun üzerinde ölümlü, 50 milyona yakın yaralanmalı trafik kazası meydana gelmektedir. Bu sonuçlara bakıldığında trafik kazaları küresel alanda dikkat edilmesi gereken bir sorun olarak görünmektedir (Taamneh ve ark., 2017).

Dünya Sağlık Örgütü'nün (DSÖ) trafik güvenliği 2015 raporda Dünyada 2013 yılında 1,25 milyon kişinin ölümü ile sonuçlanan trafik kazası meydana geldiği, bu kazaların %90'ı az ve orta gelişmiş ülkelerde yaşandığı ortaya koyulmuştur. Nitekim az ve orta gelişmiş bu ülkeler Dünya nüfusunun %82'lik kısmını oluştursa da %54 civarındaki motorlu araç kullanımına sahiptirler. Bu durumda bu ülkelerde yaşanan ölümlü trafik kaza oranlarının oldukça yüksek olduğunu göstermektedir. Söz konusu raporda yıllık geliri yüksek olan ülkelerde trafik kazalarında kişi başı ölümlü kaza geçirme riski yüzbinde 17,4 olurken, az ve orta gelişmiş ülkelerde bu oranın ikiye katlandığı görülmüştür. Örnek vermek gerekirse Afrika'daki bir ülkede kişi başı ölümlü kaza geçirme riski 26,6 iken, Avrupa ülkelerinde bu risk 9,3 olarak görülmektedir. Yine DSÖ'nün 2019 yılında yayınladığı trafik güvenliği ile ilgili raporda Dünyada 2016 yılında 1,35 milyon kişinin trafik kazalarında hayatını kaybettiği ve 2013 ile 2016 yılları arasında geçen 3 yılda Dünyada trafik kazalarında ölen insan sayısının 100 bin arttığını göstermektedir (World Health Organization (WHO), 2017, 2019).

Ülkemizde de karayolu ulaşımı diğer ulaşım kanallarına göre en önde gelmektedir. Yapılan çalışmalar sonucu elde edilen veriler doğrultusunda Türkiye’de ulaşım %95’e yakın bir oranda karayolları üzerinden sağlanmaktadır. Türkiye’de yaşanan trafik kazaları nedenlerinin insandan, araçtan, iklimden, yoldan, çevresel faktörlerden kaynaklandığı sıralanabilir. Yaşanan bu kazalar telafisi olabildiği gibi telafisi olmayan sonuçlar doğurabilmektedir. Dünya ülkeleri şehirlerinde yaşanan trafik kaza sayılarını en aza indirmek konusunda bir takım stratejiler geliştirerek kazaya neden olan unsurları bulmaya ve oluşan maddi manevi zararlarını azaltmak için çalışmalar yapmaktadır (Yavuz ve ark., 2021). Ülkemizde gelişmekte olan ülkeler arasında yer aldığından ölümlü ve yaralanmalı trafik kazaları sonuçlarına bakıldığında maalesef ki Dünyada üst sıralarda yer almaktadır (Karamaşa ve Erdoğan, 2018). Her ne kadar trafiği azaltmaya yönelik ülkemizde yapılan ulaştırma yatırımları sonucu yeni yollar yapılsa da artan motorlu taşıt sayısı nedeni ile trafik kazalarının yaşanmasının önü alınılamamaktadır (Balcı ve ark., 2017).

Türkiye İstatistik Kurumu’nun (TÜİK) 2021 haber bülteninde, ülkemizde 2020 yılında karayollarında 983.808 trafik kazası meydana gelirken bunların; 833.533’ü maddi hasarlı, 150.275’i yaralanmalı ve ölümlü olarak gerçekleşmiştir. Ölümlü yaralanmalı kaza sonucunda maalesef 2197 kişi kaza yerinde, 2669 kişide kaldırıldıkları sağlık kurumunda hayatını yitirdi. Trafik kazaları en çok trafik yerleşim yerlerinde gerçekleşirken, il bazında trafik kaza sayılarına bakıldığında bin taşıt için en az ölümlü-yaralanmalı kaza İstanbul’da yaşanırken en çok Hakkâri’de yaşanmıştır. Yaşanan trafik kazalarında ölen kişiler sıralamasında sürücüler ilk sırayı alırken, sonrasında yolcu ve yayalar gelmiştir. Cinsiyet dağılımına baktığımızda yaşanan ölümlerden %79,8’i erkek, %20,3’ü kadın olurken, yaralanmalı kazalarda ise %28,6’sı kadın %71,4’ü erkek olmuştur. Trafik kazalarının kusur unsurlarına bakıldığında açık ara ile en çok sürücülerin kazaya sebebiyet verdiği görülmüştür. Sürücülerini sırasıyla yaya, taşıt, yolcu ve yol kusurları takip etmiştir. Trafik kazaları en çok gündüz yaşanırken, taşıt sayılarına göre bakıldığında %49 ile en çok iki araç, %45,2 tek araç ve %5,8 ile çok aracın kazaya karıştığı görülmektedir. Kazaya karışan taşıtlara bakıldığında en çok otomobillerin kazaya karıştığı, aylara göre bakıldığında Ağustos en fazla kaza yaşanan ay olurken Nisan ayı en az kaza yaşanan ay olmuştur. Haftanın günleri olarak bakıldığında Cuma en fazla kaza yaşanan gün, Pazar ise en az kaza yaşanan gün oldu (URL-1, 2021).

Araştırmamıza konu olan Kahramanmaraş, Akdeniz bölgesinde yer alan ve Adana, Osmaniye, Kayseri, Sivas, Malatya, Adıyaman ve Gaziantep illerine komşu bir ilimizdir. Eskiden adı Maraş olarak geçen il, kurtuluş savaşı sırasında halkın göstermiş olduğu direniş sebebiyle 1925 yılında Türkiye Büyük Millet Meclisi tarafından istiklal madalyası verilerek 1973 yılında adı Kahramanmaraş olarak değiştirilmiştir. Ülkemizin büyük şehirleri arasında yer alan Kahramanmaraş, Afşin, Andırın, Çağlayancerit, Dulkadiroğlu, Ekinözü, Elbistan, Göksun, Nurhak, Onikişubat, Pazarcık ve Türkoğlu olmak üzere 11 ilçeye sahiptir (URL-3, 2022).

Kahramanmaraş ilinde TÜİK'in 2021 yılı verilerine göre 1.171.298 kişi yaşamakta olup trafiğe kayıtlı 258.832 adet motorlu kara taşıtı ve 147.627 adet otomobil bulunmaktadır. Ayrıca bin kişi başına düşen otomobil sayısının 126 olduğu görülmektedir. Bu bağlamda Kahramanmaraş'ta nüfusun %22'sini aşan bir kesimin motorlu kara taşıtı kullandığı görülmektedir. 2021 yılında 2.566 adet ölümlü, yaralanmalı trafik kazası yaşanmış olup bu kazaların sonucunda 4.230 kişi yaralanırken 37 kişi hayatını kaybetmiştir (URL-2, 2022).

Bu tez çalışmasının organizasyonu aşağıdaki gibi planlanmıştır.

Giriş bölümünde, trafik kazaların genel tanımı ile trafik kazalarına nedenleri ve sonuçları hakkında bilgiler verilerek, çalışmanın neden yapıldığı problemin tanımı, amacı ve önemine değinilerek literatürde, makine öğrenme teknikleri kullanılarak trafik kazaları hakkında yapılmış çalışmalar yer almaktadır.

Materyal ve yöntem bölümünde, Trafik kazaları ile ilgili elde edilen veri seti ve özellikleri, çalışmada kullanılacak WEKA programı hakkında genel bilgiler ve makine öğrenmesi sınıflandırma, kümeleme ve birliktelik kuralı çıkarma algoritmalarından hangilerini nasıl uygulanacağı ve bu algoritmaların performans metriklerinin neler olduğu anlatılmaktadır.

Bulgular ve tartışma bölümünde, yöntem kısmında anlatılan tüm teknikler bu bölümde uygulanacak olup, Trafik kazaları ile ilgili önışlem, sınıflandırma, kümeleme, birliktelik ve görsel karşılaştırma sonuçları verilmektedir.

Sonuç ve öneriler bölümünde; kullanılan yöntemlerin özet bilgiler ve elde edilen sonuçlar neticesinde öneriler verilerek diğer literatür çalışmaları ile farklılıklarına, çalışmadaki kısıtlamalar ve gelecekte yapılabilecek potansiyel çalışmalara değinilmektedir.

1.1. Problemin Tanımı

Dünyada olduğu gibi ülkemizde de yaşanan trafik kazaları insan yaralanmaları ve ölümlerinin en büyük nedenidir. Ayrıca artan nüfus ve araç sayıları paralelinde trafik kaza sayıları da artmakta ve neticede ölümlü, yaralanmalı ve maddi hasarla sonuçlanan trafik kazalarının da hızla arttığı görülmektedir. Bu durumda trafik kazalarının bir problem olduğu ve bunu önlemek adına Dünyada daha fazla çalışma yapmak gerektiği açıkça görülmektedir. Birçok ülkede artış gösteren trafik kazaları hakkında kazaların önlenmesi adına araştırmalar yapılmaktadır. Bu araştırmaların çoğu istatistiki olurken, veri madenciliği, yapay zekâ ve makine öğrenmesi gibi tekniklerle kazaların tahmin edilmesine yönelik çalışmalara da son zamanlarda sıkça yer verilmektedir. Bu bağlamda problemin önlenmesi adına makine öğrenme teknikleri ile trafik kazası verilerinden yararlı bilgiler çıkarmak doğru bir yaklaşım olacaktır. Trafik kazası verilerini analiz etmek ve değerlendirmek için makine öğrenme yöntemlerinden sınıflandırma, kümele ve birliktelik kurallı algoritmalar en sık kullanılan teknikler arasındadır (Taamneh ve ark., 2017). Yine son yıllarda pek çok alanda kullanılan makine öğrenme yöntemleri gelecekte de trafik kazalarının analizinde etkili bir yöntem olacaktır (Li ve ark., 2019).

1.2. Çalışmanın Amacı ve Önemi

Bu çalışmada, 2017 ile 2020 yılları arasında Kahramanmaraş il merkezi ve ilçelerinde, gerçekleşen dört yıla ait ölümlü ve yaralanmalı trafik kazası verileri bir makine öğrenme tekniği uygulama aracı olan WEKA (Waikato Environment for Knowledge Analysis) ile değerlendirilip analiz edilmektedir. Analiz edilecek veri seti 7929 adet trafik kazası ve 28 adet öznitelikten oluşmaktadır. Söz konusu veriler Emniyet Genel Müdürlüğü Trafik Başkanlığı'nın izni dâhilinde, trafik polislerinin trafik kazalarında Ölümlü/Yaralanmalı Trafik Kazası Tespit Tutanaklarına işlediği bilgilerden elde edilmiştir. Çalışmada yapılan değerlendirme ve analiz işlemleri, WEKA'da ön işlem (Preprocess) ve görsel karşılaştırma (visualize), sınıflandırma (classify), kümeleme (clustering) ve birliktelik (associate) olmak üzere dört farklı yöntem ile gerçekleştirilmektedir. Bu tezde geçmişte yaşanan trafik kazaları yukarıda bahsedilen yöntemlerle analiz edilip daha anlamlı bilgilere dönüştürülerek gelecekte yaşanabilecek potansiyel kazaların önüne geçilmesi amaçlanmaktadır (AlMamlook ve ark., 2019; Yavuz ve ark., 2021).

1.3. Literatür Araştırması

Literatürdeki trafik kazalarının makine öğrenme teknikleri ile analiz edildiği çalışmalar incelendiğinde sınıflandırma, kümeleme ve birliktelik analizi yöntemlerin kullanıldığı görülmektedir. Bu yöntemlerin genel olarak tek tek kullanıldığı, bir veya daha fazla yöntemin bir arada kullanıldığı çalışma sayısı çok az olduğu belirlenmiştir. Ayrıca literatürdeki çalışmaların makine öğrenme tekniklerinden yanında istatistiki ve Coğrafi Bilgi Sistemleri (CBS) kullanılarak çok sayıda çalışma yapıldığı gözlenmiştir. Literatürdeki trafik kazaları ile ilgili yapılan çalışmaların araştırma konularına genel manada aşağıdaki paragraflarda değinilmektedir.

Bu tez çalışmada kullanılan yöntemlere en yakın çalışma Li ve ark., (2017) tarafından, Amerika Birleşik Devletleri eyaletlerinde yaşanan ölümcül trafik kazalarını WEKA programında sınıflandırma modeli NaiveBayes, birliktelik ölçütü Apriori ile kümeleme ölçütü K-Means algoritmaları kullanılarak yapılmıştır. Çalışma sonucunda, elde edilen verilerle en çok ve en az ölümlü kaza yaşanan eyaletler hakkında bilgiler verilmiş ve daha emniyetli sürüş yapılabilmesi adına önerilerinde bulunulmuştur.

WEKA programında kümeleme ve birliktelik algoritmaları kullanarak yapılan iki çalışmanın ilkinde; Nafie ve Mohamed (2018), Sudi Arabistan'ın Alghat Provence şehrinde yaşanan 946 örnek ve 8 nitelikten oluşan trafik kazası veri setini birliktelik algoritması Apriori ve kümele algoritması EM ile analiz etmişlerdir. Çıkan sonuçta, Apriori algoritmasının EM kümeleme algoritmasından daha iyi olduğu belirlenmiştir. İkinci çalışmada, Nandurge ve Dharwadkar (2017), Hindistan'ın Maharashtra bölgesindeki 2015 ile 2016 yılları arasında yaşanan trafik kazalarını analiz etmek için WEKA'da KMeans kümeleme algoritması ve Apriori birliktelik algoritması kullanmışlardır. KMeans kümeleme algoritması ile yol tipi, kaza tipi, gün durumu ve yol özelliklerine göre beş küme oluşturulmuştur. Birliktelik kuralı tüm veri setinin yanı sıra her kümeye uygulanmıştır. Kümeleme ve birliktelik kuralı sonuçları tek tek açıklanmıştır.

WEKA'da sınıflandırma algoritmaları kullanarak yaşanan kazalarda yaralanma veya kaza şiddetini tahmin etmek için yapılan çalışmalardan; Krishnaveni ve Hemalatha (2011), Hong Kong'ta trafik kazalarında yaşanan yaralanma şiddetini tahmin etmek için WEKA'da NaiveBayes, AdaBoostM1, J48, PART, Random Forest (Rastgele Orman) sınıflandırıcıları ile test edilmiş, sonuç olarak, bir karar ağacı algoritması olan Rastgele Ormanın algoritmasının diğerlerine göre daha iyi sonuç verdiği gösterilmiştir. Yine

Taamneh ve ark., (2017) tarafından yapılan çalışmada, Birleşik Arap Emirlikleri'nin Abu Dabi şehrinde 2008 ile 2013 yılları arasında gerçekleşen 5.973 trafik kazası sonucu yaralanma şiddetini tahmin etmek adına WEKA programı sınıflandırma teknikleri NaiveBayes, J48, PART, Multilayer Perceptron kullanmışlardır. Sonuç olarak, en düşük başarıyı NaiveBayes gösterirken, diğer sınıflandırıcılar benzer başarı göstermişlerdir. Beshah ve Hill (2010), tarafından Etiyopya'da yaşanan 18.288 trafik kaza verisi WEKA'da J48, NaiveBayes ve K-en yakın komşu algoritmaları ile analiz edilmiş ve kaza şiddetleri tahmin edilmiştir. Çalışma sonucunda, kaza şiddetlerini en doğru tahmin eden K-en yakın komşu algoritması olurken J48 ve NaiveBayes algoritması da ona yakın değerler vermiştir. Zhang ve Fan (2013) tarafından Kanada'nın Saskatchewan şehrinde son 20 yılda şehir merkezi ve kırsal yollarda yaşanan trafik kazalarının başlıca nedenleri ve kaza şiddetleri, WEKA yazılımı içinde yer alan ID3 ve C4.5 (J48) algoritmaları kullanılarak yaş, mevsim ve cinsiyete göre üç farklı başlıkta analiz edilmiştir. Bahiru ve ark., (2018) tarafından yapılan çalışmada, trafik kazalarını etkileyen faktörler ve kaza şiddetini tahmin etmek için WEKA sınıflandırma araçları NaiveBayes, ID3, J48 ve CART algoritmaları kullanılmıştır. Çıkan sonuçta, J48 algoritmasının doğruluk performansının diğerlerinde daha iyi olduğu vurgulanmıştır. Mateus ve ark., (2019) tarafından Kolombiya Cartagena otoyolunda 2016 ile 2017 yılları arasında 7.894 trafik kaza verisi üzerinde yapılan bir çalışmada makine öğrenme yöntemleriyle kaza şiddetleri analiz edilmiştir. WEKA programı kullanılarak karar ağaçları DT, J48, Destek Vektör Makinesi, NaiveBayes, kural çıkarma algoritması PART ve Çok Katmanlı Algılayıcı değerlendirme yapılmıştır. Çalışma sonucunda 20 ile 39 yaş arası erkek ve kadın motosiklet sürücülerinin kaza yapma olasılıkları yüksek olduğu, bir bisiklet ile motosikletin çarpışma olasılığının düşük olduğu belirtilmiştir. Son olarak AlMamlook ve ark., (2019) tarafından trafik kaza verileri denetimli makine öğrenme algoritmaları, AdaBoost, Logistic Regression, NaiveBayes ve Random Forest ile analiz edilmiş veri düzensizliği için SMOTE algoritması kullanılmıştır. Çalışma sonucunda Random Forest algoritmasının kaza şiddetini tahmin etmekte daha başarılı olduğu vurgulanmıştır.

WEKA aracı kullanılarak yaşanan kazaların nedenlerini, risk faktörlerini ve sonuçlarını analiz etmek için yapılan çalışmalardan; Oña ve ark., (2011) tarafından İspanya'nın daha kırsal yollarında 1.536 trafik kazası sonuçları, Bayes ağları kullanılarak, bir biri ile ilişkileri sınıflandırılmaya çalışılmıştır. Yaşanan kazaların veri setinde 18 farklı değişken bulunurken, değerlendirmeye alınacak sınıflar; ölü, yaralı ve ağır yaralı olmak

üzere 3 farklı sınıftan oluşturulmuştur. Sonuçta, trafik kazalarında ölümlerin veya ağır yaralanmaların kaza tipi, sürücünün yaşı gibi faktörlerle ilişkili olduğu ortaya konmuştur. Akomolafe ve Olutayo (2013), Nijerya'nın Lagos - İbadan ekspres karayolları üzerinde yaşanan trafik kazalarının sonuçları ve etkileri incelemek adına yaptıkları çalışmada, verilerin analizi ve tahmini WEKA'da Fonksiyonel Ağaç Algoritması (FT) ve ID3 algoritmaları ile yapılmıştır. Çalışma sonucunda, kaza verileri analiz edilmiş ve ID3 algoritmasının FT algoritmasına göre daha doğru sonuçlar verdiği belirlenmiştir. Mujalli ve ark., (2016) tarafından Ürdün'de banliyö yollarında 2009-2011 yılları arasında yaşanan trafik kazası sonuçlarının analizi için farklı Bayes sınıflandırıcıları kullanılmıştır. Çalışma sonucunda, kazanın türü, kazanın oluş şekli, kazaya karışan araç sayısı, yolun yüzeyi, hız sınırı, yön sayısı gibi unsurların ölümlü veya ciddi yaralanmalı kazaların meydana gelmesine neden olduğu ortaya konulmuştur. Perez ve ark. (2016), tarafından WEKA programı sınıflandırma algoritmaları kullanılarak Meksika'nın Mexico City şehrindeki trafik kazalarının yoğunlukları incelenmiş ve sonuçları Coğrafi Bilgi Sistemlerinde (CBS) veri görüntüleme ve düzenleme yetenekleri sağlayan QGIS programı ile önlem alınması gereken yollar görsel hale getirilmiştir. Castro ve Kim (2016), Birleşik Krallık ülkelerinde 2010 ile 2012 yılları arasında yaşanan ölümlü ve yaralanmalı trafik kazaları üç WEKA'daki sınıflandırma algoritması Bayes Network, Multilayer Perceptron ve karar ağacı J48 ile risk faktörleri araştırılmıştır. Yapılan çalışmada en başarılı Bayes Network algoritması ona yakın J48 ve Multilayer Perceptron algoritmalarının verdiği belirtilmiştir.

WEKA aracı kullanılarak yaşanan kaza sonuçlarını sınıflandırma algoritmalarının performansını ölçmek için yapılan çalışmalardan; Yavuz ve ark., (2021) tarafından Antalya ili ve çevresinde 2012-2016 yılları arası meydana gelen ölümlü ve yaralanmalı 3.181 trafik kazası WEKA'da makine öğrenme teknikleri ile sınıflandırılarak analiz edilmiştir. Çalışma sonucunda en başarılı doğruluk oranını veren sınıflandırıcının NaiveBayes olduğu belirlenmiştir. Kuşkapan ve Çodur (2022) tarafından 2002 ile 2018 yılları arasında ülkemizde yaşanan trafik kazaları, Çok Katmanlı Algılayıcı, En Yakın Komşu ve Regresyon Algoritmaları kullanılarak WEKA programında hata ve sınıflandırma performansları incelenmiştir. Çalışma sonucunda, bu üç algoritmanın hata ve sınıflandırma performansları analiz edildiğinde en iyi sonucu En Yakın Komşu algoritmasının verdiği değerlendirilmiştir. Ayrıca son yıllarda yaşanan ölümlü ve yaralanmalı trafik kaza oranlarının 2000'li yıllardaki gibi yüksek risk taşıdığı ortaya konmuştur.

WEKA ve Orange aracında sınıflandırma ve birliktelik kurallı algoritmalar kullanılarak Hussain ve ark., (2019) tarafından yapılan çalışmada, WEKA’da 150 trafik kazası verisi üzerinde BayesNet, Multilayer Perceptron, J48, sınıflandırma modelleri ile değerlendirildi. Sonuçta, en başarılı sınıflandırıcının Multilayer Perceptron sonrasında BayesNet ve J48’in olduğu gösterilmiştir. Ayrıca birliktelik kuralı sonucu olarak en başarılı sonuçları Apriori algoritmasının verdiği belirlenmiştir.

Kümeleme algoritmaları kullanılarak trafik kazalarının benzerlikleri veya özelliklerini belirlemek için yapılan çalışmalardan; Güner ve ark., (2018), Sakarya ilinde meydana gelen trafik kazalarının analizi ile ilgili bir çalışmada, WEKA programının K-means kümeleme algoritması kullanılmış ve kazaların benzerlikleri incelenmiştir. Çalışma sonucunda, kazaların en çok açık havalarda, kuru zeminlerde, kavşak olmayan yollarda, sonbahar aylarında, hafta içleri ve öğlen öncesi yaşandığı belirtilmiştir. Atalay ve Tortum (2010) tarafından 1997 ile 2006 yıllarında Türkiye illerinde yaşanan ölümlü ve yaralanmalı trafik kazalarının analizi geleneksel k-ortalama ve bulanık c-ortalama kümeleme yöntemleri beşli guruplara ayrılarak analiz edilmiştir. Çalışma sonucunda, kaza yaşanma potansiyeli en yüksek olan iller belirlenmiştir.

WEKA aracı kullanılmadan yaşanan trafik kazalarını analiz etmek için yapılan çalışmalarda; Chang ve Chen (2005) tarafından 2001 ile 2002 yılları arasında Tayvan ulusal karayolu üzerinde yaşanan kaza verilerini incelemek için trafik özellikleri, karayolu geometrik değişkenleri ve çevresel faktörler arasındaki deneysel ilişkiler Classification and Regression Tree (CART) ve Negatif Binom Regresyon modeliyle kaza sıklıklarını tahmin etme performansları karşılaştırılmıştır. Çalışma sonucunda CART modelinin otoyolda yaşana kaza sıklıklarını analiz etmek daha başarılı bir yöntem olduğu belirlenmiştir. Bayata ve Hattatoğlu (2014) tarafından 1985 ile 2010 yılları arasında Erzincan karayollarında yaşanan trafik kazaları, Yapay Sinir Ağları, Ortalama Karasel Hata, Belirlilik Katsayı ve Akaike Bilgi Kriteri yöntemleri kullanılarak istatistiki bir çalışma yapılmıştır. Çalışma sonucunda Yapay Sinir Ağları yönteminin en başarılı sonucu verdiği belirtilmiştir. Özden ve Acı (2016) tarafından 2005 ve 2014 arasında Adana’da yaşanan yaralanmalı trafik kazaları ve aynı yıllar arasındaki aylık meteoroloji verilerinden oluşan bir veri seti üzerinde yaralı sayıları ve yaralanmalı kaza sayılarını tahmin eden bir çalışma yapılmıştır. Çalışmada Regresyon Ağacı, Çoklu Doğrusal Regresyon, Destek Vektör Makinesi, Fonksiyon Uydurma Yapay Sinir Ağı, İleri Beslemeli Çok Katmanlı Yapay Sinir Ağı ve Genelleştirilmiş Regresyon Yapay Sinir Ağı yöntemleri kullanılmıştır. Çalışma

neticesinde, destek vektör makinesinin hem yaralı sayısı hem de yaralanmalı kaza sayısını tahmin etmede en başarılı yöntem olduğu belirlenmiştir. Regresyon ağacı yöntemi yaralı sayısı tahminini kaza sayısını tahmininden daha başarılı yapmıştır. Ayrıca daha önceki yıllarda meydana gelen trafik kazalarının yol ve hava verilerine bakılarak sonraki yıllarda bu veriler ışığında önlem alınmasının gerektiği vurgulanmıştır. Geymen ve Dedeoğlu (2016) tarafından Emniyet Genel Müdürlüğü kayıtlarında 2012 ile 2013 yıllarında sadece Kahramanmaraş il merkezinde yaşanan kazalar CBS kullanılarak incelenmiş, trafik kazalarının yoğunlaştığı bölgeler ve kaza yerleri analiz edilmiştir. Yaşanan kazalar; arazinin tomografik yapısı, kazaların yaşadığı zaman parametrelerine göre incelenmiş ve bu kazaların sebepleri CBS analizleriyle ortaya konmuştur. Özcan ve Küçükönder (2020) tarafından 2008-2015 yılları arasında Kahramanmaraş merkezinde meydana gelen 14.317 trafik kazaları CBS ile analiz edilmiştir. Bu çalışmada Kahramanmaraş merkezinde yaşanan trafik kazalarını önlemek adına bölgesel analiz, kaza yoğunluğu yüksek alanların tespiti ve risk taşıyan bölgelerin tahmin edilmeye çalışılmıştır.

Bu tez çalışmasını diğer çalışmalardan ayıran özelliği, yeterli büyüklükte veri setine sahip olmasıyla beraber iki boyutlu görsel karşılaştırması, sınıflandırma, kümeleme ve birliktelik kuralı olmak üzere dört farklı yöntem kullanılması ve bu dört yöntemin kullanılması neticesinde diğer çalışmalara göre daha fazla sonuç elde edilecek olmasıdır.

1.4. Makine Öğrenmesi

İnsanlar, çevresinde yaşanan olayları sürekli deneyimleyerek öğrenir. Yani tüm duygu, düşünce ve davranışlar tecrübe edilerek kazanılır. Öğrenme, deneyim, bilgi, beceri, anlayış ve çalışma gibi anlamlar taşır. Kısaca öğrenme, davranışları deneyimleyerek bir eğilimi değiştirmek veya gerçekleştirmek olarak tanımlanabilir. Çok kapsamlı ve geniş olduğundan öğrenme kapasitesini tahmin edebilmek veya bir açıklama yapmak zordur (Nilsson, 1996).

Bilgisayarların insanlar gibi düşünüp davranabileceği geçmişten günümüze kadar hep araştırma konusu olmuştur. Ancak gelinen noktada tam olarak insan gibi düşünüp davranabilen bir bilgisayar sistemi olmasa da yapılan çalışmalara bakıldığında gelecekte bu mümkün görünmektedir (Çoban, 2011). Öte yandan bilgisayar teknolojisinin gelişmesi ile yazılım ve donanım sistemleri de gelişmiş ve buna paralel çok büyük veriler bu sistemlerde depolanmaya başlanmıştır. Bu büyük verilerin işlenerek anlamlı hale getirilmesi için makine öğrenimi önemli bir araç haline gelmiştir. Makine öğreniminin ortaya çıkması ile yalnız o zamanki ana akım olan yapay zeka alanında çalışan araştırmacılar değil onlarla beraber diğer alanlardaki uzman araştırmacılar da bu akıma katılmış ve önemli çalışmalar yapmışlardır (Nikam, 2015; Sandıkçı ve Aydılek. 2018; Wei ve ark., 2019).

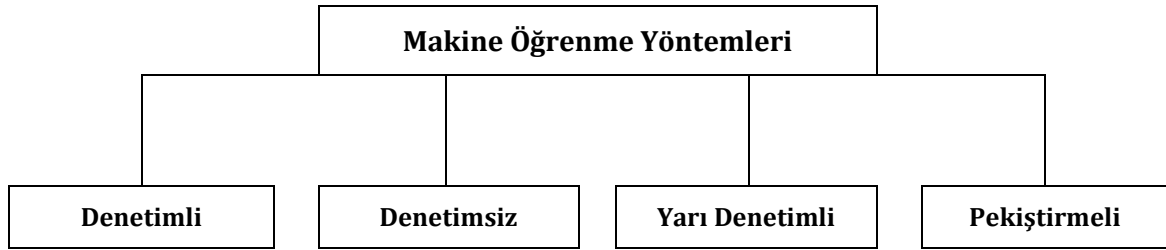
Makine öğrenimi, veri madenciliği, yapay zeka ve istatistiğin alt dalı olarak görülebilir. İlk olarak 1950'li yıllarda yapay zeka ve yapay sinir ağı araştırmalarında adı geçmeye başladı. 1959 yılında ise Arthur Lee Samuel, ünlü makine öğrenimi tanımını; "bilgisayarların programlanmaya ihtiyaç duymadan kendi kendine öğrenmesi." olarak yaptı (Schuld ve ark., 2015; Zhang, 2020). Diğer bir deyişle makine öğrenmesi, çevreden öğrendiği bilgilerle insan aklını taklit etmesi için tasarlanan ve durmadan gelişen zeki hesaplama araçlarıdır. Makine öğrenme teknikleri doğrudan bir veri kümesine uygulanır. Bu veri kümesi içindeki örneklerden yararlanarak mantıksal hesaplamalar ve belirli örüntüler çıkararak sonuca ilişkin tahminlerde bulunur. 1959 yılında önerildikten sonra, bilgisayarla görüntüleme, oyun, nesne tanıma, uzay araçları, finans, eğlence, biyoloji, biyomedikal ve tıp uygulamalarına kadar çok geniş bir alanda etkin olarak kullanılmaya başlandı (Naqa ve Murphy, 2015; Kiranmai ve Laxmi, 2018).

En çok kullanılan Makine öğrenimi yöntemleri; Bayes Olasılık Sınıfından NaiveBayes, Karar Ağaçlarından J48 (C4.5), Sinir Ağlarından Yapay Sinir Ağı, Merkez Uzaklık Tabanlı K-Ortalamlar, Karar Destek Sınıflandırmalarından Destek Vektör

Makinesi, Komşuluk İlişkisine Göre Sınıflandırma Yapan K-En Yakın Komşu Algoritmalarıdır (Atalay ve Çelik, 2017).

1.5. Makine Öğrenmesi Yöntemleri

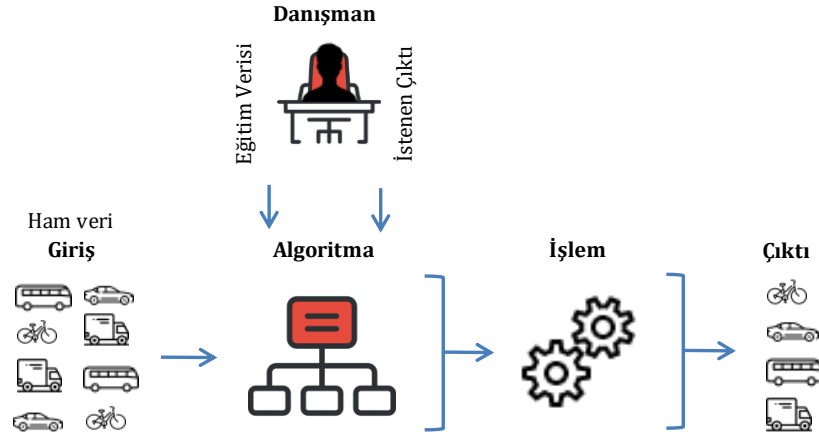
Makine öğrenme yöntemleri, denetimli, denetimsiz, yarı denetimli ve pekiştirmeli olarak dört ana kategoride gruplandırılır. Şekil 1.1’de gösterilmektedir.



Şekil 1.1. Makine öğrenme yöntemleri

1.5.1. Denetimli Öğrenme

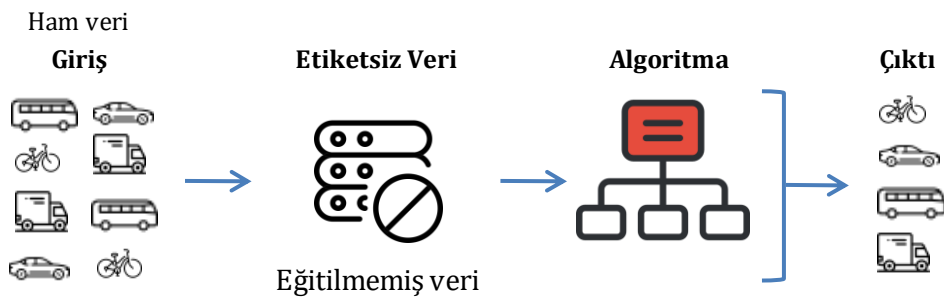
Denetimli öğrenmede, öğrenme işleminin gerçekleşmesi için sistemde bir danışman bulunur. Bu danışmanın görevi, eğitim verilerinin ve istenen çıktıların sisteme vermesidir. Denetimli öğrenmede, model içindeki bilgisayar algoritmasına giriş ve çıkış örnekleri verilir ve bu örneklerden ilgili algoritmanın sonucu tahmin etmesi beklenir. Denetimli öğrenmede veriler etiketlenir yani giriş verilerinin tüm özellikleri verilerek çıkışın ne olacağı öğretilir ve bilgisayar algoritmasının girdi verilerinden sonucu tahmin eder (Nilsson, 1996; Schuld ve ark., 2015). Örneğin, bisiklet, otomobil, kamyon ve otobüsten oluşan bir taşıt kümesinin makine öğrenme algoritmaları ile sınıflandırılmak istendiğinde; girişe taşıtların etiket bilgileri (ağırlık, boyut, kasa tipi vb.) çıkışa da taşıtların isimleri verilerek giriş bilgilerine bakılarak gelen aracın hangisi olduğunu tahmin edilmesi beklenir. Denetimli öğrenme algoritmalarına; basit bayes, destek vektör makinesi, karar ağaçları, lojistik regresyon, k-en yakın komşu, rasgele orman ve yapay sinir ağlarını örnek verilebilir. Şekil 1.2’de gösterilmektedir.



Şekil 1.2. Denetimli öğrenme

1.5.2. Denetimsiz Öğrenme

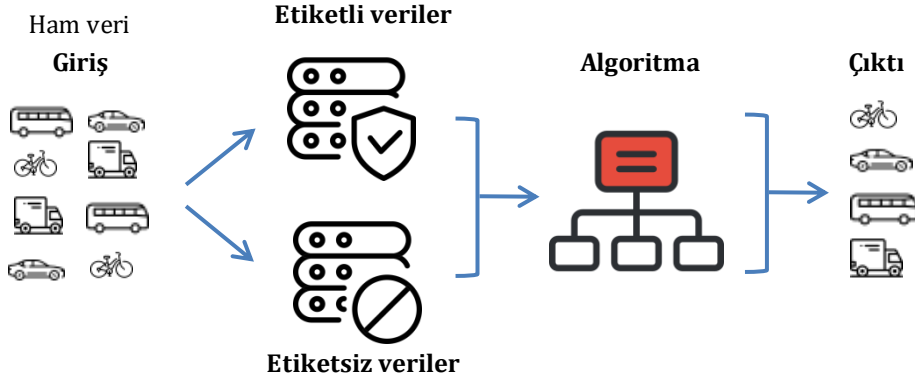
Denetimsiz öğrenmede, denetimli öğrenmedeki gibi sistemde öğrenme işlemini gerçekleştirecek bir danışmanın bulunmamaktadır. Denetimsiz öğrenmede, sisteme sadece giriş verileri verilir ve çıkış verilerine bakılmaz. Bu sistemde veriler etiketsizdir ve kullanılan bilgisayar algoritması sadece giriş verilerinden sonucu öğrenmeye çalışır (Nilsson, 1996). Diğer bir deyişle denetimsiz öğrenme, herhangi bir deneyim olmadan veri kalıplarından sonucu bulma süreci olarak tanımlanabilir (Schuld ve ark., 2015). Örneğin denetimsiz öğrenme ile bisiklet, otomobil, kamyon ve otobüsün sınıflandırılması istendiğinde algoritmaya sadece taşıtların görüntüsü gösterilir ancak etiket bilgileri verilmez ve bu bilgiler ışığında doğru sınıflandırılmayı yapması beklenir. Denetimsiz öğrenmeye, birliktelik ve kümeleme algoritmaları ile olasılık yoğunluk tahmini, öznelilikler arasındaki ilişkilerin bulunması ve boyutsallık azaltma örnek verilebilir (Brunton ve ark., 2020). Şekil 1.3'te gösterilmektedir.



Şekil 1.3. Denetimsiz öğrenme

1.5.3. Yarı Denetimli Öğrenme

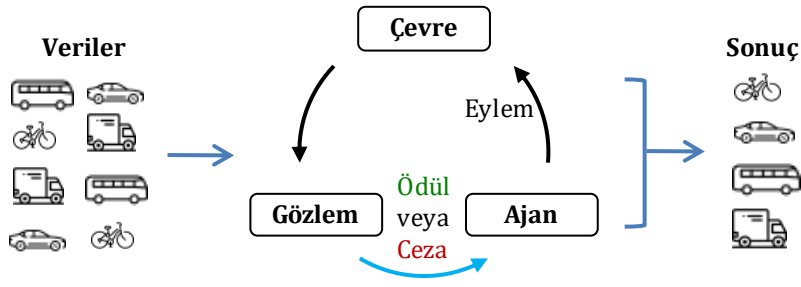
Yarı gözetimli öğrenme denetimli ve denetimsiz öğrenmenin bir arada kullanıldığı öğrenme türüdür. Yarı denetimli öğrenmede, etiketlenmiş verilerle etiketlenmemiş verilerin aynı sistem içinde bulunur ve algoritma sınırlı bir eğitim verisiyle ya da çevreden gelen bilgilerle doğal olarak öğrenme işlemini gerçekleştirir (Brunton ve ark., 2020). Örneğin, taşıt sınıflandırmasında otomobil ve kamyonun etiket bilgileri verilirken otobüs ve bisikletin etiket bilgisi verilmeden sınıflandırma algoritmasının aracı doğru tahmin etmesi beklenir. Şekil 1.4’de gösterilmektedir.



Şekil 1.4. Yarı denetimli öğrenme

1.5.4. Pekiştirmeli Öğrenme

Diğer bir adıyla takviyeli öğrenme, öğrenme denilen ifadeye en yakın olan yöntemdir. Bu öğrenme türünde, sisteme giriş verileri verilirken çıkış verileri verilmez. Sisteminde giriş verilerine bakarak sonuç verileri deneyimleyerek üretilir. Pekiştirmeli öğrenmede, sistem bir ajan (genellikle bir bilgisayar programı), çevre ve gözlemciden oluşmaktadır. Burada bahsedilen ajan gerçekleştirdiği eylemi çevreye iletir. Gözlemci bu eylemin sonucunda her doğru için bir ödül verirken yapılan yanlışlar içinde cezalandırır. Bu durumda ajan, her ödül aldığı anda mevcut stratejiyi güçlendirirken ceza aldığı anda mevcut stratejiyi yeniden gözden geçirir. Bu geri bildirimler sayesinde aracı sürekli öğrenmeye devam ederek öğrenme işlemini pekiştirmiş olur (Schuld ve ark., 2015). Şekil 1.5’de gösterilmektedir. Pekiştirmeli öğrenme, bilgisayar oyunları, Q Learning algoritması, LVQ ağı, Temporal difference learning alanlarında kullanılır (Çoban, 2011).



Şekil 1.5. Pekiştirmeli öğrenme

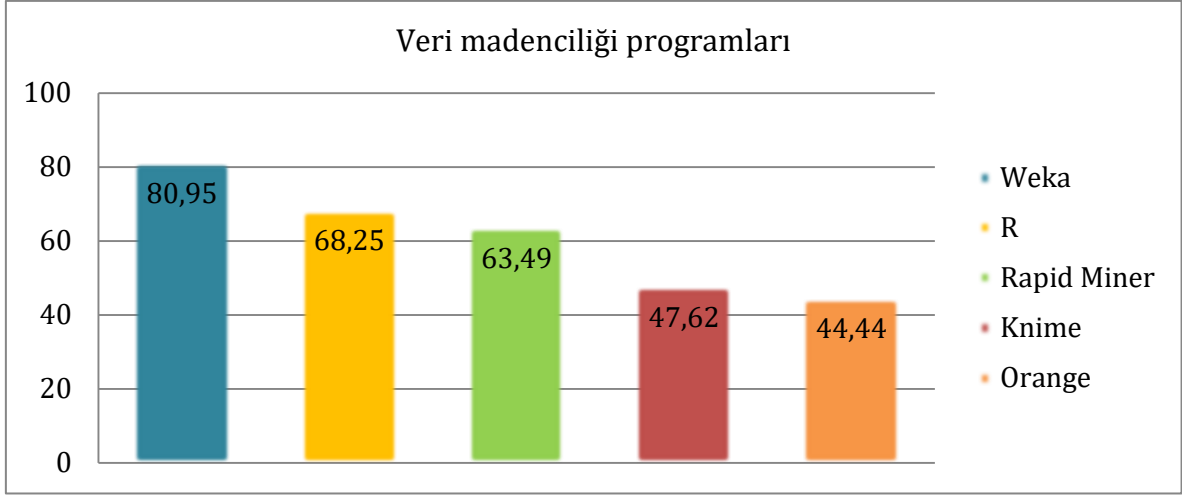
2. MATERYAL ve YÖNTEM

2.1. Weka

WEKA, Dünya üzerinde akla gelen hemen hemen birçok veri ile makine öğrenme tekniklerini kullanımına olanak sağlayan yazılımdır. WEKA, içerisinde çok sayıda makine öğrenme yöntemlerini barındıran modern bir veri işleme aracı olarak tanımlanır (Zhang ve Tsai, 2003).

Büyük veri kümelerinin bir program olmadan analiz edilmesi ve değerlendirmesi zordur. Bu büyük verilerin analiz edilmesi için makine öğrenimi ve veri madenciliği yöntemlerini kullanan araçlar gereklidir. Bu bağlamda Yeni Zelanda'daki Waikato Üniversitesinde makine öğrenmesi ile veri madenciliği alanlarında kullanılmak üzere, Java dilinde yazılan açık kaynak kodlu ve kolay bir kullanım ara yüzüne sahip WEKA yazılımını geliştirdi. İlk olarak tarım alanlarında yaşanan sorunları gidermek ve daha faydalı tarım uygulamalarına katkı sağlamak için kullanılsa da artık eğitim ve araştırma başta olmak üzere birçok alanda kullanılmaktadır. WEKA'nın güçlü yönleri genel kamu lisansına sahip ücretsiz bir program olması ve Java dilinde yazıldığından herhangi bir platformda taşınması ve çalıştırmasının kolay olmasıdır. Ayrıca kullanılabilirlik açısından Sınıflandırma, kümeleme, birliktelik kuralı çıkarma gibi birçok makine öğrenmesi algoritmasını bünyesinde barındırması, iki boyutlu grafikleri ile verilerin karşılaştırmasına imkân vermesi ve kullanıcıların internet üzerinden veya kendi verilerini kendi üretmesine olanak sağlaması ve verilerin güçlü olduğu yanlardır. WEKA'nın desteklediği dosya türleri; Arff, Arff.Gz, Bsi, Csv, Dat, Data, Libsvm, Names, Xrff, Xrff.Gz'dir. Ancak en çok tercih edilen ARFF (Attribute Relation File Format) ve CSV (Comma Separated Values) formatlarındaki veri dosyalarıdır (Holmes ve ark., 1994; Witten ve ark., 1999; Krishnaveni ve Hemalatha, 2011; Mhetre, 2017).

Dokuz Eylül Üniversitesinde ücretsiz veri madenciliği programlarından bilinirliği ile ilgili yapılan istatistiki bir araştırmada WEKA programının, R, Rapid Miner, Knime ve Orange programlarına oranla bilinirlikleri ortaya konmuştur. Çalışma sonucunda en çok bilinen ve haberdar olunan programın WEKA olduğu tespit edilmiştir. Araştırma sonuçları Şekil 2.1'deki gibidir (Doğan, 2017).



Şekil 2.1. Veri madenciliği alanında kullanılan programlar

Weka uygulamasında giriş menüsü beş ana başlıktan oluşur. Bunlar; Explorer, Experimenter, KnowledgeFlow, Workbench ve Simple CLI'dir (Bouckaert ve ark., 2016).

Explorer: kullanıcının verileri analiz edeceği ilk menüdür. Veri ön işleme ve görselleştirmeye izin verir. Bu menü yalnızca bir sınıflandırma, kümeleme, birliktelik kuralı çıkarma algoritmasının ve karşılaştırmalı görsellerin kullanılmasına izin verir (Baumgartner ve Serpen, 2009).

Experimenter: Veri setleri ile deneyler, istatistiksel değerlendirmeler ve öğrenme şemaları oluşturmaya olanak veren menüdür. Kullanıcıların benzetim oluşturmaya ve çalıştırmasına izin verir. Verilerde ön işleme ve görselleştirme yoktur. Eğitim ve test verileri yalnızca çapraz doğrulama ve yüzde dağılımı ile sınırlıdır (Baumgartner ve Serpen, 2009).

KnowledgeFlow: Explorer menüsü ile aynı işlemleri destekler. Farkı ise grafiksel bir ara yüze sahip olmasıdır. Aşamalı öğrenmeyi desteklemesi avantajlarından (Baumgartner ve Serpen, 2009).

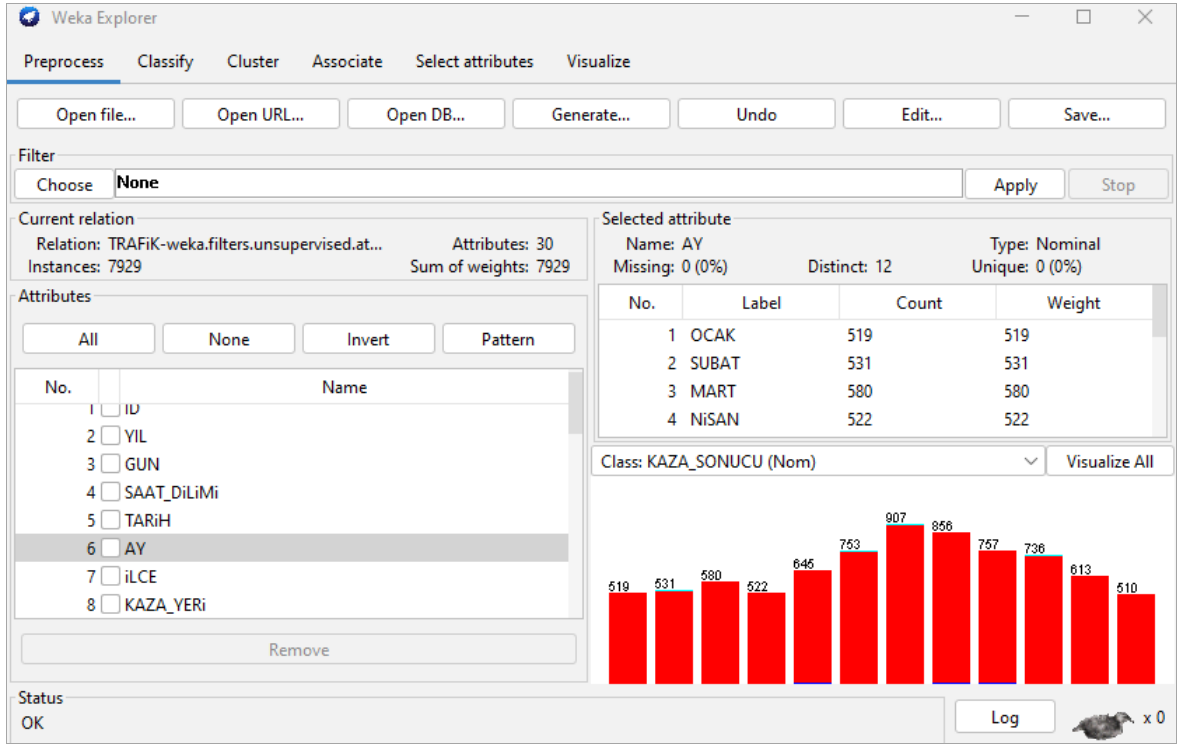
Workbench: Diğer tüm menülerdeki özellikleri içinde birleştiren hepsi bir arada sunan bölümdür (Bouckaert ve ark., 2016).

Simple CLI: komut satırı arabirimini ile WEKA komutlarının doğrudan çalıştırılmasına izin veren basit bir arabirimdir (Bouckaert ve ark., 2016).

2.2. Weka Explorer

2.2.1. Preprocess Menüsü

Veri yüklemek, düzenlemek, incelemek ve veriler üzerinde birçok işlem yapmaya izin veren bölümdür. Şekil 2.2’de gösterilmektedir.



Şekil 2.2. Weka önışlem menü ekranı

Open file: Bilgisayarınızda bulunan bir dosyayı yüklemenizi sağlar.

Open URL: Bir web sayfasından veri yüklemenizi sağlar.

Open DB: Bir veri tabanından veri yüklemenizi sağlar.

Generate: Kullanıcının yapay veriler oluşturmaya olanak sağlar.

Edit: Veri setinde düzenleme yapmayı sağlar.

Save: veri setinde yapılan değişiklikleri farklı kaydetmeyi sağlar.

Choose: Veri setine çeşitli filtrelerin uygulanmasına ve dönüştürülmesine izin verir.

2.2.2. Classify Menüsü

Bu bölüm içerisinde birçok sınıflandırma algoritmasını barındırmaktadır. Sınıflandırma algoritmaları bu bölümde test edilir ve eğitilir. Choose butonu ile WEKA'da bulunan sınıflandırıcılardan birini seçilerek performans değerlendirmesi yapılır. Şekil 2.3'te gösterilmektedir.

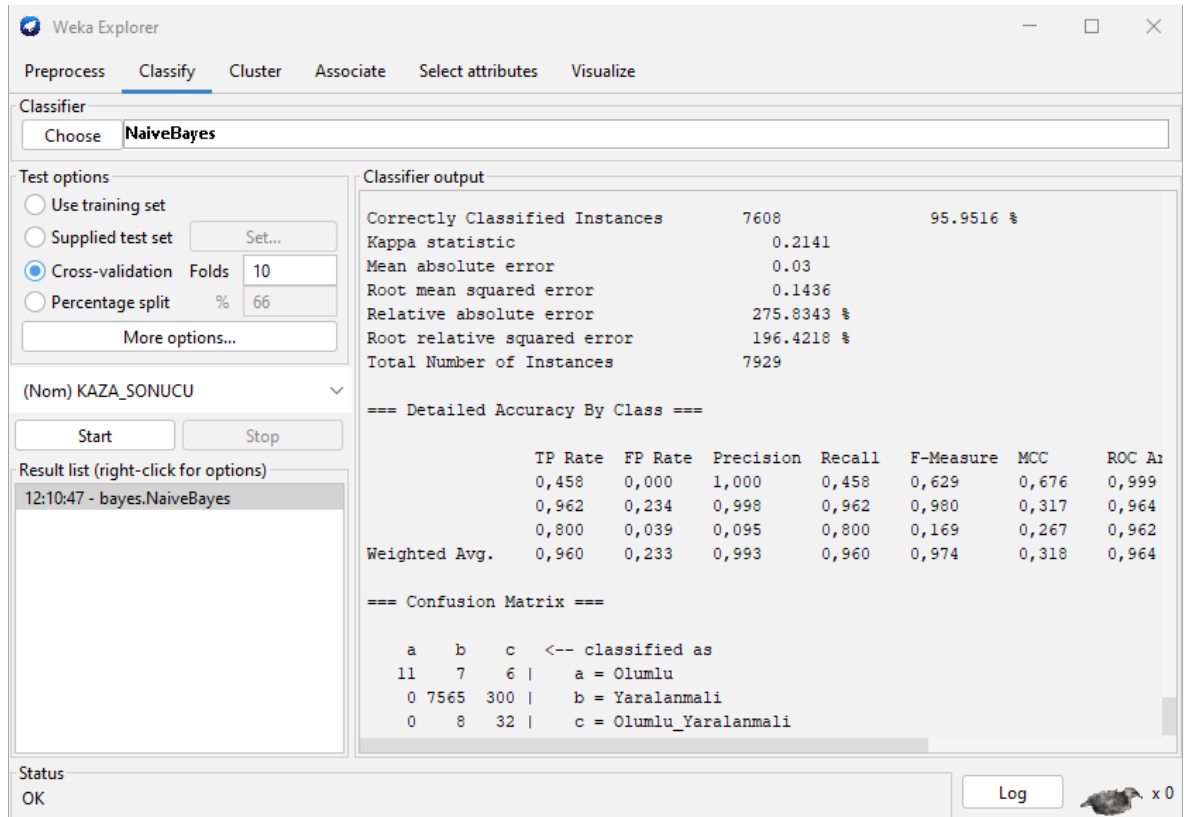
Use training set: Verilerin tamamı eğitim için kullanılır. Sınıflandırıcı, üzerinde eğitim aldığı örneklerin sınıfını ne kadar iyi tahmin ettiğine göre değerlendirilir.

Supplied test set: Daha önce eğitilmiş bir veri seti dosyasından yüklenen bir dizi örneği ile sınıflandırma işleminin yalıtılmasını sağlar.

Cross-validation: Çapraz doğrulama veri setini kümelere ayırarak sınıflandırma işlemini girilen k sayısı kadar kümeler ve çapraz doğrulama ile değerlendirilir.

Percentage split: Veri seti belli yüzdelerle iki bölünür. Bir kısmı eğitim bir kısmı test için kullanılır.

More options: Daha fazla seçenek eklemek için kullanılır.



Şekil 2.3. Weka sınıflandırma menüsü ekranı

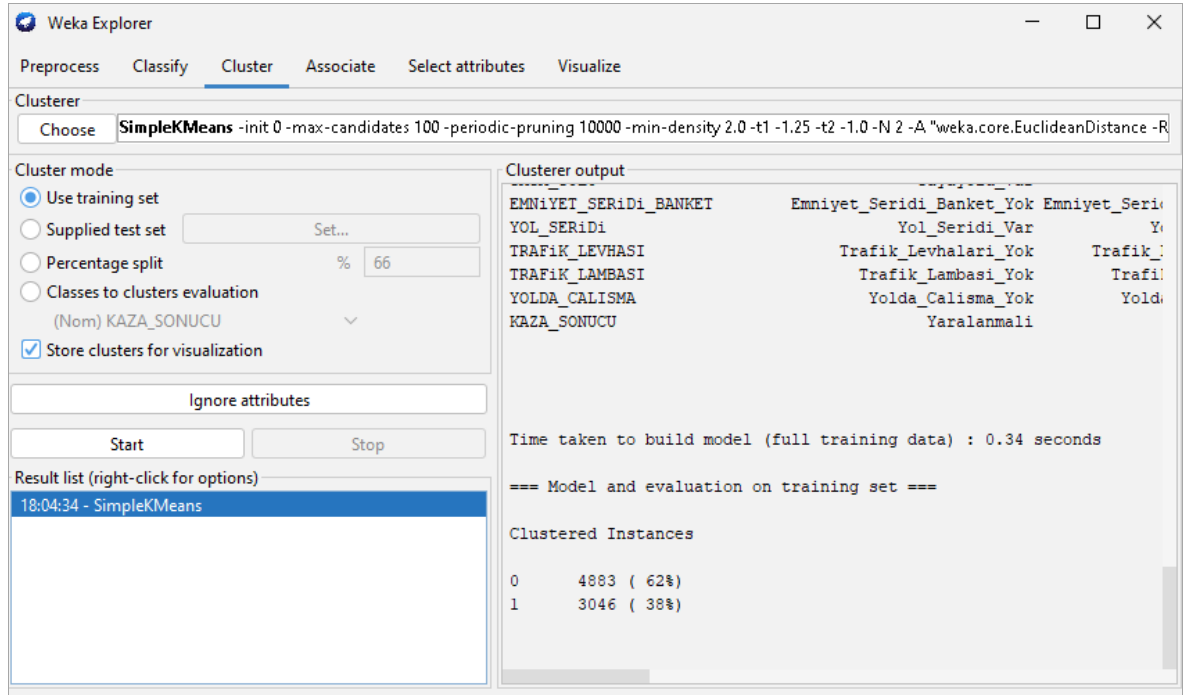
2.2.3. Cluster Menüsü

Verilerin kümeleme algoritmaları ile analiz edildiği bölümdür. Choose butonu yardımıyla kümeleme algoritmalarından biri seçilir. Şekil 2.4'te gösterilmektedir.

Cluster mode (Kümeleme modu): Kümelemeye tabi tutulacak verilerin nasıl değerlendirileceği ile ilgili ayarlar buradan yapılır. Beş seçenekten oluşur. İlk üç seçenek sınıflandırma bölümündeki ile aynıdır.

Classes to clusters evaluation: Bir özniteliğin daha önceden belirlenmiş bir sınıfla ne kadar iyi eşleştiğini gösterir.

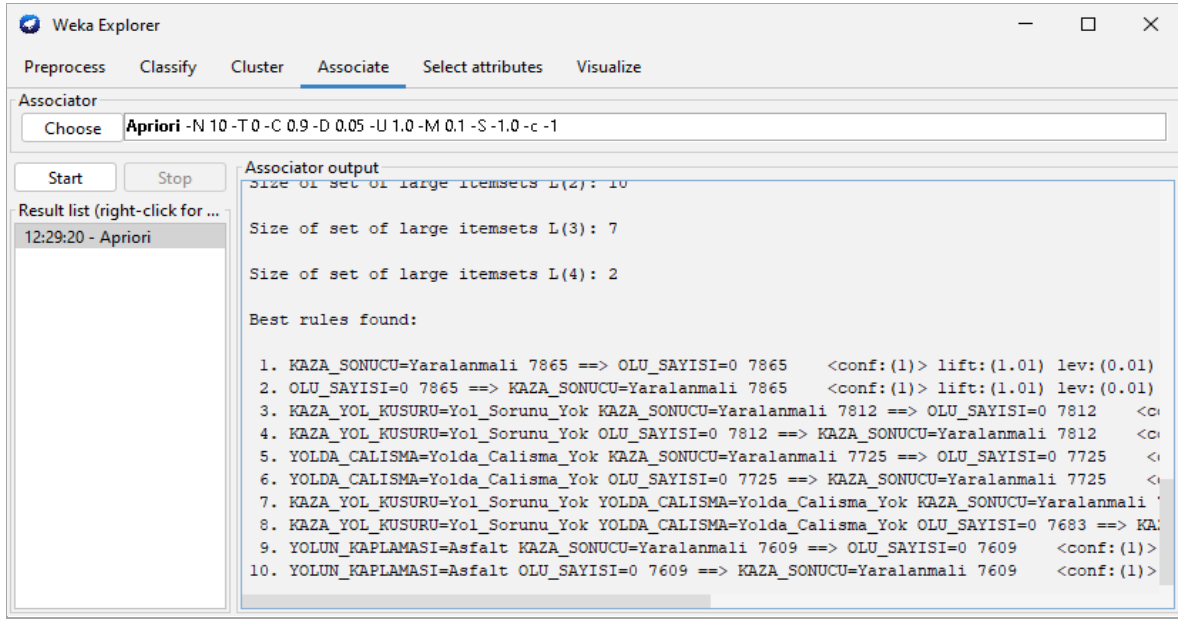
Ignore attribute: Kümeleme yapılırken bazı özniteliklerin dikkate alınmaması için kullanılır.



Şekil 2.4. Weka kümeleme menüsü ekranı

2.2.4. Associate Menüü

Associate, verilerin bir birleri ile olan ilişkilerinin değerlendirildiği bölümdür. Birliktelik kuralı ile değerlendirme yapmak için istenen algoritma seçimi yapıldıktan sonra Start düğmesine basılarak değerlendirme sonuçlarına görüntülenir. Şekil 2.5'te gösterilmektedir.



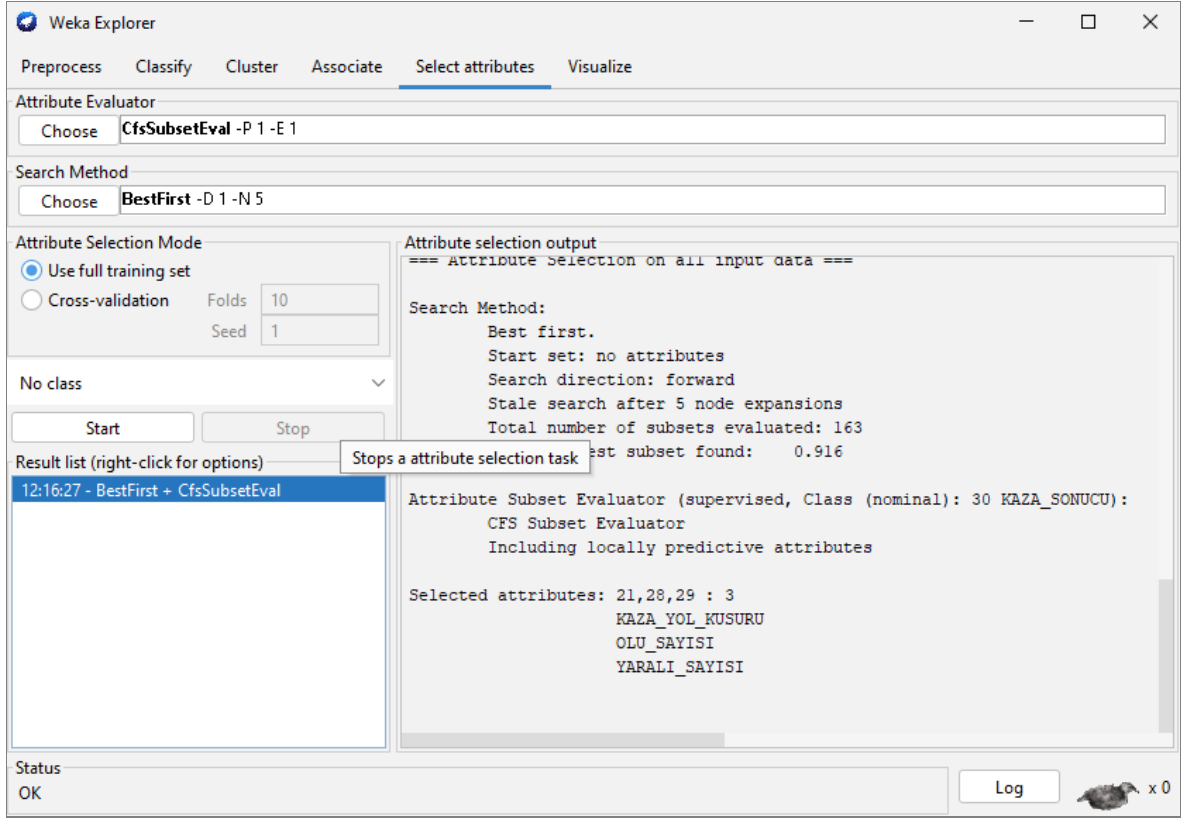
Şekil 2.5. Weka birliktelik menüsü ekranı

2.2.5. Select Attributes Menüü

Veri seti içindeki en alakalı özniteliklerin seçimi bu bölümde yapılır. Burada en alakalı öznitelik seçimi, tüm öznitelik kombinasyonları içinde arama yapılır. Bu arama iki adımda gerçekleşir. İlk adımda bir öznitelik değişkenine bir değer atamak için kullanılacak yöntem belirlenir. İkinci ve son adımda hangi arama yönteminin kullanılacağı belirlenir. Şekil 2.6'da gösterilmektedir. Nitelik Seçim Modu (Attribute Selection Mode) kutusunun iki seçeneği vardır:

Use full training set: Tam eğitim veri seti kullanılarak özniteliklerin alt küme değerleri belirlenir.

Cross-validation (Çapraz doğrulama): Özniteliklerin alt küme değerleri bir çapraz doğrulama yapılarak belirlenir.



Şekil 2.6. Weka nitelik seçimi menüsü ekranı

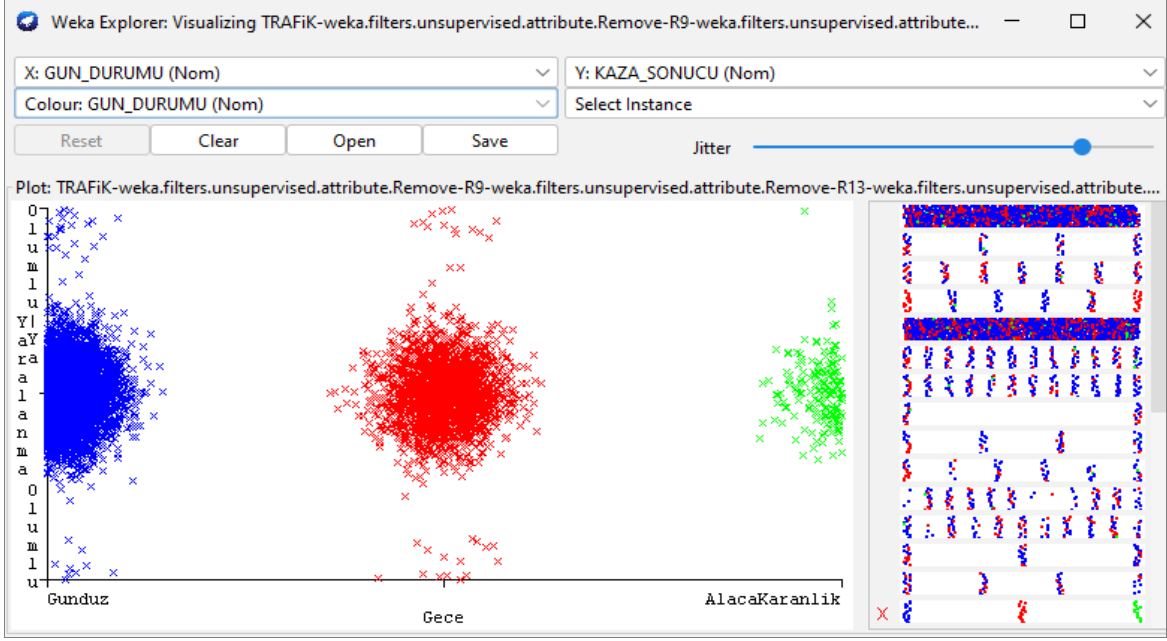
2.2.6. Visualize Menüsü

Veri setindeki özniteliklerin birbirleriyle olan etkileşimlerinin iki boyutlu bir grafik olarak gösterildiği bölümdür. Görselleştirme paneli seçildiğinde, seçilen öznitelik sınıfları farklı renk gruplarına ayırarak bir dağılım grafiği matrisi oluştur. Bir hücreye tıklandığında iki özniteliği karşılaştırmaya yarayan bir pencere açılır. Açılan bu pencerenin en üstünde iki açılır liste sekmesi bulunur. Soldaki sekme x ekseninde gösterilecek özniteliği, sağdaki sekme y ekseninde gösterilecek özniteliği seçmeyi sağlar. Şekil 2.7’de gösterilmektedir.

Jilter: Grafikteki nokta yoğunluğunun tespit edilmesi için kullanılır. Bu sekme en solda olduğu sürece çok büyük veriler grafikte tek bir çizgi gibi görünür. Bu nedenle sağa sürükleyerek nokta yoğunluğunu azaltarak özniteliklerin birbirleriyle olan ilişkilerini tespit etmekte fayda sağlar.

PlotSize: Grafikteki noktaların kapladığı alanın büyüklüğünü ifade eder.

PointSize: Grafikteki noktaların boyutunu ifade eder.



Şekil 2.7. Weka görsel karşılaştırma menüsü ekranı

2.3. WEKA Veri Seti Özellikleri

WEKA en çok tercih edilen ARFF ve CSV formatlarındaki veri dosyalarıdır. ARFF dosyaları içerisine değişken tanımlamaya izin veren ASCII tabanlı metin dosyalarıdır. CSV dosyaları içerisine değişken değeri tanımlamaya gerek olmayan virgülle karaktere ayrılmış dosyalardır. Yani değişken başlıklarının verilmesi yeterlidir. Bir CSV dosyalarının veri seti görünümü Şekil 2.8’de verilmiştir.

```
YolunTipi,Kaplaması,GeoYatay,GeoDüsey,Gündurumu,Havadurumu,YolunYüzeyi,ölü,Yarali,Sonuç
Bolunmus_Yol,Asfalt,Duz_Yol,Egimli,Gece,Acik,Kuru,0,1,1
Bolunmus_Yol,Asfalt,Duz_Yol,Egimli,Gunduz,Acik,Islak_Nemli,0,2,1
Bolunmus_Yol,Asfalt,Duz_Yol,Egimsiz,Gece,Yagmur,Islak_Nemli,0,1,1
Bolunmus_Yol,Asfalt,Duz_Yol,Egimli,Gece,Yagmur,Islak_Nemli,0,2,1
Bolunmus_Yol,Asfalt,Duz_Yol,Egimli,Gunduz,Yagmur,Islak_Nemli,0,2,1
```

Şekil 2.8. CSV dosyası örneği

ARFF dosyalarında öznitelikler, sürekli değişkenlerde numeric, real, integer veya date olarak tanımlanırken, nominal değerlere sahip öznitelikler ise süslü parantez içerisinde virgüllerle ayrılarak tanımlanırlar. Bir ARFF dosyalarının WEKA’da doğru çalışabilmesi için önceden bazı eklemelerin yapılması gerekir. Eklenecek bu değişiklikler şunlardır:

@Relation: Çalışılan veri setinin adı tanımlanır.

@Attribute: Veri setindeki özniteliklerin tanımlanmasını sağlar

@Data: Verilerin program tarafından okunması sağlar.

Örnek bir ARFF dosyasının veri seti görünümü Şekil 2.9’da gösterilmektedir.

```
@Relation TRAFİK

@Attribute YOLUN_TiPi {Bolunmus_Yol, Diger, iki_Yonlu_Yol, Tek_Yonlu_Yol}
@Attribute YOLUN_KAPLAMASI {Asfalt, Beton, Parke, Sati_Kaplama, Stabilize, Toprak}
@Attribute YATAY_GEO {Duz_Yol, Tehlikeli_Viraj, Viraj}
@Attribute DUSEY_GEO {Egimsiz, Egimli, Tehlikeli_Egim, Tepe_Ustu}
@Attribute GUN_DURUMU {Gunduz, Gece, AlacaKaranlik}
@Attribute HAVA_DURUMU {Acik, Bulutlu, Dolu, Kar, Kuvvetli_Ruzgar, Sis_Duman, Sulusepken, Ti}
@Attribute YOLUN_YUZEYi {Buzlu, Diger, Islak_Nemli, Karli, Kuru, Sel_Su_Birikintili}
@Attribute OLU_SAYISI Numeric
@Attribute YARALI_SAYISI Numeric
@Attribute KAZA_SONUCU {1, 2, 3}

@Data

Bolunmus_Yol,Asfalt,Duz_Yol,Egimli,Gece,Acik,Kuru,0,1,1
Bolunmus_Yol,Asfalt,Duz_Yol,Egimli,Gunduz,Acik,Islak_Nemli,0,2,1
```

Şekil 2.9. Arff dosyası örneği

2.4. Çalışmada Kullanılan Veri Seti

2.4.1. Trafik Veri Seti

Çalışmada Kahramanmaraş il merkezi ve ilçelerinde, 2017 ile 2020 yılları arasında gerçekleşen dört yıla ait ölümlü yaralanmalı trafik kazalarının verileri bulunmaktadır. Kullanılan veri seti 28 adet öznitelik ve 7929 veri satırından (trafik kaza sayısı) oluşmaktadır. Trafik veri seti ARFF veya CSV formatlarına dönüştürülerek WEKA programında çalışmaya uygun hale getirilmiştir. Bu veri seti içinde yer alan öznitelik başlıkları şunlardır; Kaza Yılı, Hafta Günü, Kaza Saat Dakika, Ay, Kazanın Yaşandığı İlçe, Kaza Yerleşim Yeri, Yolun Tipi, Yolun Kaplaması, Yolun Sınıfı, Yasal Hız Limiti, Yol Adı/Kesim Numarası, Yatay Geometrisi, Düşey Geometrisi, Gün Durumu, Hava Durumu, Yolun Yüzeyi Durumu, Kaza Oluş Türü, Kazaya Karışan Araç Sayısı, Çarpışma Yeri, Kaza Yol Kusuru, Yaya Yolu, Emniyet Şeridi Banket, Yol Şerit Çizgisi, Trafik İşaret Levhası, Trafik lambası, Yolda Çalışma, Ölü ve Yaralı Sayısı ile Kazanın Sonucundan oluşmaktadır. Trafik veri setinin özellikleri Tablo 2.1’de gösterilmektedir.

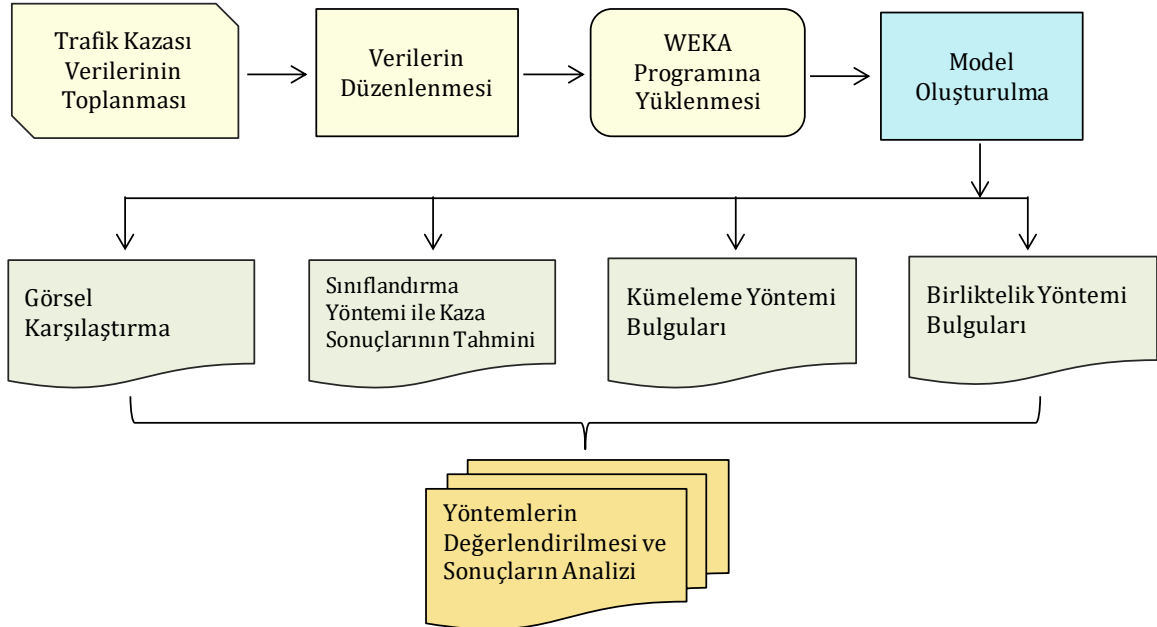
Tablo 2.1. Trafik veri seti özellikleri

Öznitelik Adı	Değişken Tipi	Alt Nitelikler
Yıl	Nominal	2017, 2018, 2019, 2020
Hafta Günü	Nominal	Pazartesi, Salı, Çarşamba, Perşembe, Cuma, Cumartesi, Pazar
Kaza Saati Dilimi	Nominal	0-4, 4-8, 8-12, 12-16, 16-20, 20-24
Ay	Nominal	Ocak, Şubat, Mart, Nisan, Mayıs, Haziran, Temmuz, Ağustos, Eylül, Ekim, Kasım, Aralık
Kazanın Yaşandığı İlçe	Nominal	Afşin, Andırın, Çağlayancerit, Dulkadiroğlu, Ekinözü, Elbistan, Göksun, K.Maraş/Merkez, Nurhak, Onikişubat, Pazarcık, Türkoğlu
Kaza Yerleşim Yeri	Nominal	İçi, Dışı
Yolun Tipi	Nominal	Tek Yönlü Yol, İki Yönlü Yol, Bolünmüş Yol, Diğer
Yolun Kaplaması	Nominal	Asfalt, Beton, Parke, Sati Kaplama, Stabilize, Toprak
Yolun Sınıfı	Nominal	Cadde, Sokak, Bağlantı Yolu, Devlet Yolu, Diğer, İl Yolu, Koy Yolu, Orman Yolu, Otoyol, Park Alanı, Servis Yolu, Tesis Önü veya İçi, Diğer
Yasal Hız Limiti	Numeric	-
Yatay Geometrisi	Nominal	Düz Yol, Viraj, Tehlikeli Viraj
Düşey Geometrisi	Nominal	Eğimsiz, Eğimli, Tehlikeli Eğim, Tepe Üstü
Gün Durumu	Nominal	Gündüz, Gece, Alacakaranlık
Hava Durumu	Nominal	Acık, Bulutlu, Dolu, Kar, Kuvvetli Rüzgâr, Sis Duman, Sulusepken, Tipi, Yağmur
Yolun Yüzeyi Durumu	Nominal	Buzlu, Islak Nemli, Karlı, Kuru Sel Su Birikintili, Diğer
Kaza Oluş Türü	Nominal	Araçtan Cisim Düşmesi, Araçtan insan Düşmesi Arkadan Çarpma, Çoklu Çarpışma Devrilme Savrulma Takla, Duran Araca Çarpma Engel Cisme Çarpma, Hayvana Çarpma Karşılıklı Çarpışma, Park Etmiş Araca Çarpma Yan Yana Çarpışma, Yandan Çarpma Yayaya Çarpma, Yoldan Çıkma, Zincirleme Çarpışma
Kazaya Karışan Araç Sayısı	Nominal	Tek Araç, İki Araç, Çok Araçlı
Çarpışma Yeri	Nominal	Banket Üzerinde, Hemzemin Geçit Okul Geçidi, Orta Refüjde Tespit Edilemedi, Yaya Geçidi Yaya Kaldırımında, Yol Kenarında Yol Üzerinde, Diğer
Kaza Yol Kusuru	Nominal	Düşük Banket, Kısmi Münferit Çökme Şerit Çökmesi, Yol Sorunu Yok Yolda Gevşek Malzeme, Yolda Münferit Çukur Diğer Yol Sorunu
Yaya Yolu	Nominal	Var, Yok, Uygun Değil
Emniyet Şeridi Banket	Nominal	Var, Yok, Uygun Değil
Yol Şerit Çizgisi	Nominal	Var, Yok, Uygun Değil
Trafik İşaret Levhası	Nominal	Var, Yok, Uygun Değil
Trafik Lambası	Nominal	Var, Yok, Var Bozuk
Yolda Çalışma	Nominal	Var, Yok
Ölü Sayısı	Numeric	-
Yaralı Satısı	Numeric	-
Kazanın Sonucu	Nominal	Ölümlü, Yaralanmalı, Ölümlü/Yaralanmalı

2.5. Çalışmada Kullanılan Yöntemler

Çalışmada Kahramanmaraş il merkezi ve ilçelerinde 2017 ile 2020 yılları arasında meydana gelen 7929 adet ölümlü ve yaralanmalı trafik kazası verisi WEKA’da görsel karşılaştırma, sınıflandırma, kümeleme ve birliktelik kuralı olmak üzere dört yöntem kullanılmaktadır.

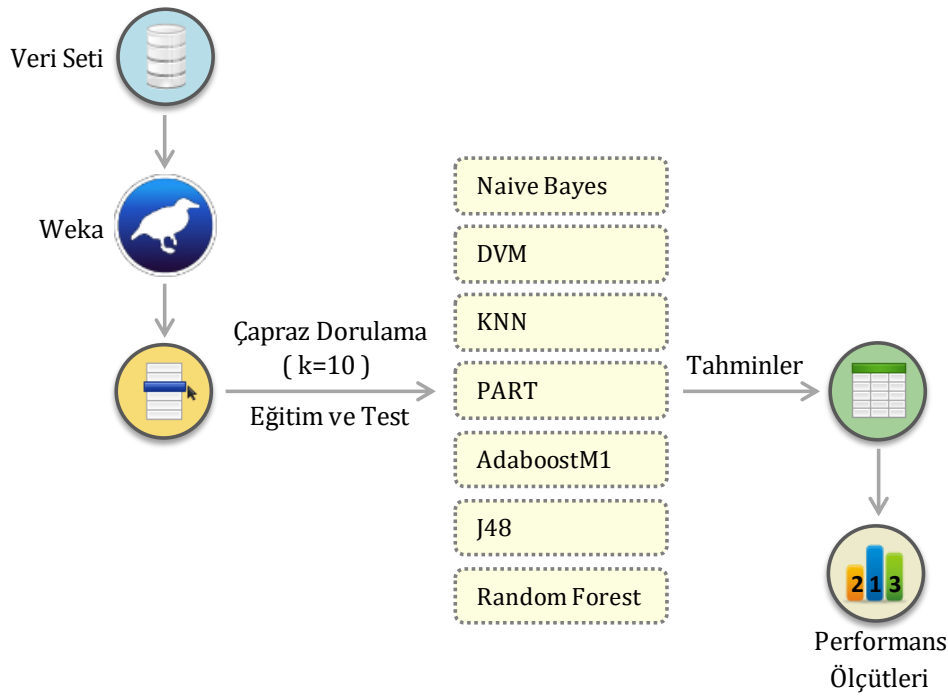
Ön işlem ve görsel karşılaştırma yönteminde veri setindeki özniteliklerin birbirleriyle olan etkileşimleri iki boyutlu grafiklerde gösterilerek değerlendirilmektedir. Sınıflandırma yöntemde, trafik kazası verileri on kat çapraz doğrulama (cross-validation) ile hem eğitim hem de test işlemlerine tabi tutularak NaiveBayes, Destek Vektör Makinesi, K En Yakın Komşu, PART, AdaboostM1, J48 ve Rastgele Orman algoritmalarının gerçek verilere karşı tahminleri ve performans metrikleri değerlendirilmektedir. Kümeleme yöntemde, SimpleKMeans ve EM kümeleme algoritmaları kullanılarak trafik kazalarının kümelenme özellikleri gözlemlenerek benzerlikleri veya farklılıkları incelenmektedir. Birliktelik kuralı çıkarma yönteminde Apriori algoritması ile trafik kaza unsurlarının bir birleriyle gerçekleşme olasılıkları değerlendirilmektedir. Çalışmada kullanılan yöntemlere ait akış şeması Şekil 2.10’da gösterilmektedir.



Şekil 2.10. Çalışmanın akış şeması

2.6. Sınıflandırma Yöntemi

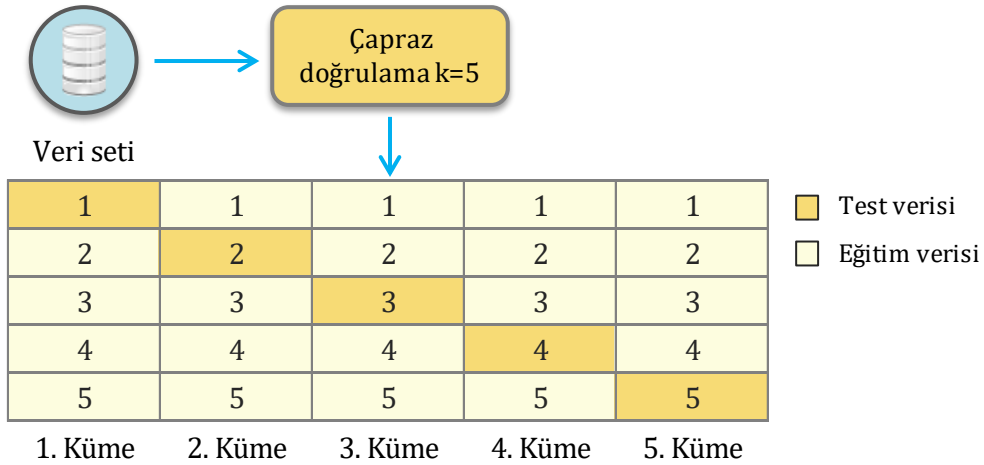
Trafik kazaları sonuçlarını içeren veri seti WEKA programında yer alan 7 sınıflandırma algoritması ile çapraz doğrulama yöntemi kullanılarak bir sınıflandırma modeli oluşturuldu. Bu model içinde; NaiveBayes (NB), Destek Vektör Makinesi (DVM), K En Yakın Komşular (KNN), PART, AdaboostM1, J48 (C4.5) ve Rastgele Orman (RF) sınıflandırma algoritmaları kullanıldı. Sınıflandırma yöntemi şekil 2.11’de gösterildiği gibidir.



Şekil 2.11. Sınıflandırma yöntemi şeması

Kullanılan sınıflandırma algoritmaları WEKA’da çapraz doğrulama ayarı seçilerek test edildi. Bu yöntem yapılacak sınıflandırmanın öğrenmesini artırmak için kullanılmaktadır. Çapraz doğrulama veri setini k kadar kümeye ayrılarak öğrenme işlemi gerçekleştirilmektedir. Sırasıyla bir küme test edilirken kalan k-1 kadar küme eğitimde kullanılır. Bu k değeri WEKA’da standart olarak 10’dur ancak kullanım amacına göre farklı bir değer de verilebilir. Yapılan çalışmalarda yaygın olarak çapraz doğrulama k değeri on olarak tercih edilmektedir. Bu çalışmada da bu değer 10 olarak belirlenerek veri seti on farklı kümeye ayrılarak sınıflandırma işlemleri gerçekleştirilmiştir. Çapraz doğrulama sürecini örnekle açıklamak gerekirse;

Bir A veri seti sınıflandırmasında 25 verinin bulunsun ve çapraz doğrulama değeri $k=5$ seçilsin. Burada k değeri 5 olduğundan veri seti 5 kümeye ve 5’li alt kümeye bölünür. İlk kümeden başlanarak kümeler sırasıyla test edilir ve kalan 4 küme ise eğitimde kullanılır. Bu işlem 5 kez tekrarlanarak öğrenme işlemi tamamlanır. Çapraz doğrulama süreci Şekil 2.12’de gösterilmektedir.



Şekil 2.12. Çapraz doğrulama süreci

2.6.1. Naive Bayes Algoritması

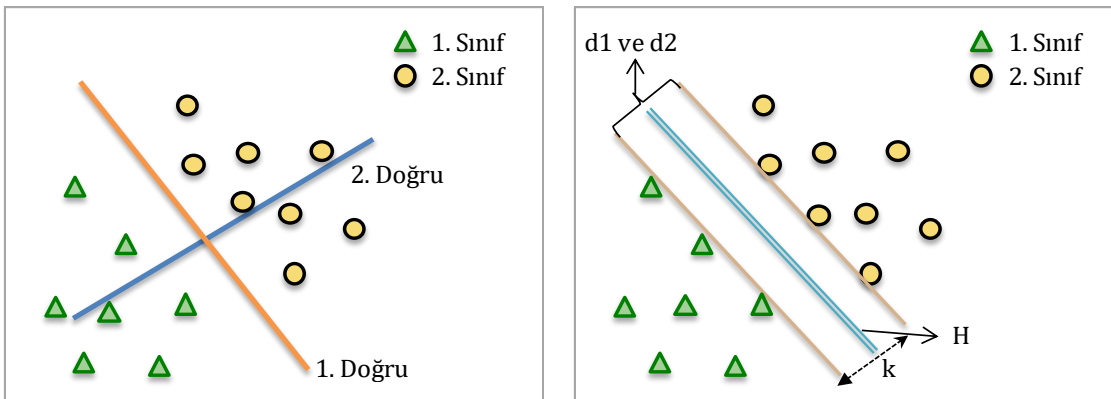
Naive Bayes (NB) algoritması, WEKA’da istatistiğe dayalı öğrenme Bayes kategorisi içinde yer almaktadır. Bir denetimli öğrenme türü olan NB, temeli Bayes teoremine dayanan basit bir olasılık sınıflandırıcısıdır. Adını ünlü matematik uzmanı Thomas Bayes’in Bayes teoreminden alır. Bu teoremden; bir veri grubunda yer alan her bir nitelik için gerçekleşme olasılıkları hesaplanır ve gerçekleşme olasılığı en yüksek olana göre sınıflandırma yapılır. NB sınıflandırıcılar girdi verilerine bakarak olası en iyi çıktıyı hesaplarlar. Örneğin bir meyvenin yuvarlak ve turuncu olduğu biliniyorsa bu meyvenin bir portakal olduğu NB tarafından tahmin edilebilir (Nikam, 2015; Bahiru, 2018). NB algoritmasının avantajlarını şu şekilde sıralayabiliriz; yorumlaması kolay ve alakasız niteliklere karşı dayanıklı olması, hızlı ve yüksek düzeyde bir model oluşturabilmesi, öngörücü, satır sayısı ile doğrusal ölçeklenebilir olmasıdır (Krishnaveni ve Hemalatha, 2011). NB sınıflandırma işlemleri aşağıda gösterilen (2.1) formülü ile hesaplanır.

$$P(x|y) = \frac{P(y|x) * P(x)}{P(y)} \quad (2.1)$$

$P(x|y)$: y olayına karşı x olayının olması olasılığıdır. $P(y|x)$: x olayına karşı y olayının olması olasılığıdır. $P(x)$: x durumunun olma olasılığı, $P(y)$: y durumunun olma olasılığıdır.

2.6.2. Destek Vektör Makinası

WEKA'daki adı ile SMO algoritması, Function kategorisi içinde yer almakta olup bu algortmada Destek Vektör Makinası (DVM) kullanılmaktadır. Son yıllarda aktif olarak nesne, resim, ses ve tıbbi tanıma alanlarında kullanılmaktadır. V.N. Vapnik'in geliştirdiği DVM istatistiksel öğrenme teorisine dayanmaktadır. Çalışma prensibi, temel olarak veri sınıfları arasında sonsuz sayıda doğru çizerek bunların içinden ayırımı en iyi yapan doğrular seçilerek sınıflandırma işlemi yapılmasıdır. Burada çizilen doğrulara hiperdüzlem denilmektedir. Veri kümesi içindeki iki sınıfı birbirinden ayırmak için kullanılan iki hiperdüzlem Şekil 2.13'ün sol bölümünde gösterilmiştir. Birinci doğruda veriler doğru sınıflanırken, ikinci doğruda veriler yanlış sınıflandırılmıştır. Şeklin sol tarafına verileri sınıflarına ayırmak için birbirine paralel hiperdüzlemler çizilir. Çizilen hiperdüzlemler arasında Destek Vektör denilen bir paralel doğru çizilerek en büyük marjine sahip olan seçilir. İki sınıfın birbirinden ayrılması Şekil 2.13'ün sağ bölümünde gösterildiği gibidir. Burada H hiperdüzlem çizgisi, d1 ve d2 destek vektörlerini k ise iki düzlem arasındaki marjini ifade etmektedir (Nikam, 2015; Alan ve Yeşilyurt, 2018).



Şekil 2.13. Destek vektörü ile sınıflandırma modeli

2.6.3. KNN Algoritması

WEKA programında Ibk ismi ile kullanılan ve mesafeye dayalı algoritmalar sınıfında yer alan en yakın komşu algoritması, verilen verinin hangi sınıfa ait olduğunu belirlemek için en yakın komşuluk ilişkisinden faydalanır ve KNN olarak adlandırılır. Bir denetimli öğrenme türü olan KNN algoritması, 1970 yıllarda bu yana örüntü ve istatistiki tahmin gibi birçok alanda kullanılmaktadır. Bu algoritmanın çalışma prensibi, veri sınıflarının birbirleri arasındaki yakınlık ve mesafeye göre kümeler oluşturularak sınıflandırma işlemine dayanır. KNN algoritmasında bir sınıf ayırt edilirken verilerin k değeri kadar en yakın komşu uzaklığına bakılır. Buradaki k kümenin kaç elemandan oluşacağını ifade eder. KNN algoritması k-ortalamalar (k-means) algoritması ile benzerlik gösterse de aslında büyük farklılıklar bulunmaktadır. Bu farklardan en büyüğü KNN algoritmasında eğitim verisi kullanırken k-means algoritmasında kullanılmamasıdır (Nikam,2015; Taşçı ve Şamlı, 2020).

En yakın komşu uzaklığı hesap edilirken bazı hesaplamaları kullanılır. En çok tercih edilen uzaklık hesaplamaları Öklid, Minkowski ve Manhattan'dır. Bu uzaklıklar aşağıda gösterilen (2.2), (2.3) ve (2.4), formülleriyle hesaplanır.

$$h_{öklid} = \sqrt{\sum_{i=1}^n (X_i - Y_i)^2} \quad (2.2)$$

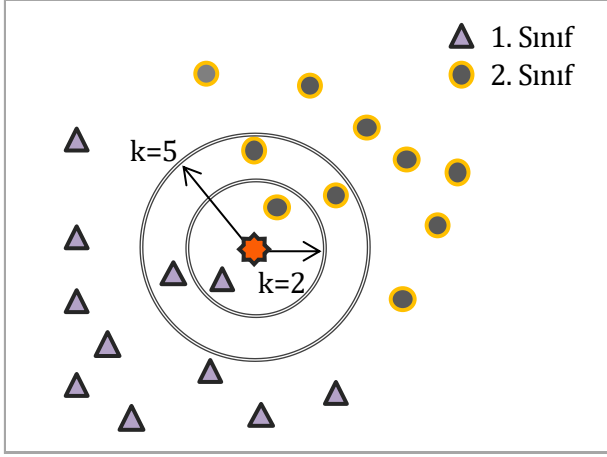
$$h_{man} = \sum_{i=1}^n |X_i - Y_i| \quad (2.3)$$

$$h_{min} = (\sum_{i=1}^n |X_i - Y_i|^t)^{1/t} \quad (2.4)$$

Özetle bir KNN (Ibk) algoritması aşağıdaki adımları izleyerek çalışmaktadır.

- Yakınlık derecesini ölçmek için bir k değeri seçilir.
- Veriler arasındaki mesafeyi ölçmek için bir uzunluk formülü belirlenir.
- Veriler arasındaki uzaklıklar tek tek hesaplanır.
- Hesaplanan uzaklıklara bakılarak en yakından uzak olana doğru sıralama yapılır.
- Birbirine en yakın k tanesi bir kümeye ayrılarak sınıflandırılma tamamlanır.

KNN sınıflandırma modeli Şekil 2.14'te verilmiştir.



Şekil 2.14. K en yakın komşu algoritması sınıflandırma modeli

2.6.4. PART Algoritması

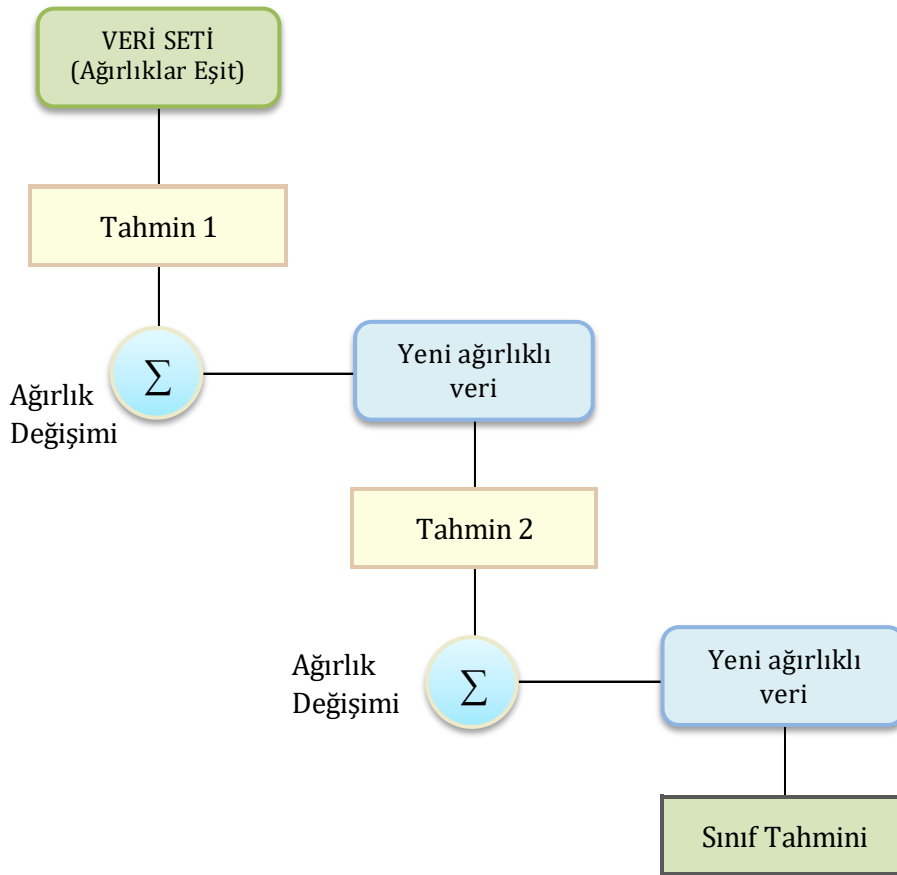
PART (Partial Decision Trees) algoritması WEKA’da “rules” kategorisi altında yer almaktadır. Kural tabanlı algoritmalarından olan PART algoritması, C4.5 ile RIPPER algoritmalarının farklı stratejilerini kullanan basit ama güçlü bir sınıflayıcıdır. C4.5 algoritmasının karar ağacı yaklaşımını, RIPPER algoritmalarının böl ve yönet yaklaşımını kullanır. Temel olarak, bir öğrencini eğitim verisi için bir dizi kural oluşturur ve bu kuralları daha sonra sınıflandırma amacıyla kullanılır. PART algoritmasında, bir kural oluşturmak için veri kümesinde kısmi olarak bir ağaç oluşturulur ve yeni bir kural oluşturmak için geniş kapsamlı olan yaprak seçilir. Mevcut veri kümesi üzerinde böl ve yönet yaklaşımı ile bir dizi kural oluşturur ve bu kurallar örnekleri eğitim kümesinden kaldırılarak hiçbir örnek kalmayana dek yinelemeye devam eder (Krishnaveni ve Hemalatha, 2011). Ayrıca PART, bir karar listesine sahip planlanmış kural kümesi üretebilen bir algoritmadır. Sırası ile her kural için yeni bir veri eşleştirilir ve ilk eşleşen örneklere bakılarak hangi sınıfa ait oldukları belirlenir (Parsania ve Ark., 2014).

2.6.5. AdaBoostM1 Algoritması

AdaBoostM1 algoritması WEKA’da “meta” kategorisi altında yer almaktadır. Robert E. Schapire ve Yoav Freund tarafından 1995 senesinde geliştirilen bu algoritmanın çalışma prensibi; Öncelikle eşit ağırlıklar kullanılarak sınıflandırma işlemine başlanır, daha sonra hatalı tahminler için ağırlık artırılarak tekrar tekrar eğitime alınır ve böylece

sınıflandırmanın başarısı arttırılmaya çalışılır. Yani bir sınıflandırıcının çıktısı bir sonraki sınıflandırıcının girdisi olarak gerçekleşir. Böylece hatalı veriler üzerinde daha fazla durularak başarılı tahminlerin yapılması sağlanır (Soylu, 2018).

Bu algoritmanın avantajları; yüksek tahmin etme olasılığı, daha az hafızaya ihtiyaç duyması, yüksek uygulanabilirliğe sahip olması, basit ve tam eğitilmiş verilerle çalışabilmesi, karmaşık ve sınırlı sayıdaki verilerden daha güvenli veriler elde etmesidir (Krishnaveni ve Hemalatha, 2011). AdaBoostM1 sınıflandırma modeli Şekil 2.15'te gösterilmektedir.



Şekil 2.15. AdaBoostM1 sınıflandırma modeli

2.6.6. J48 Algoritması

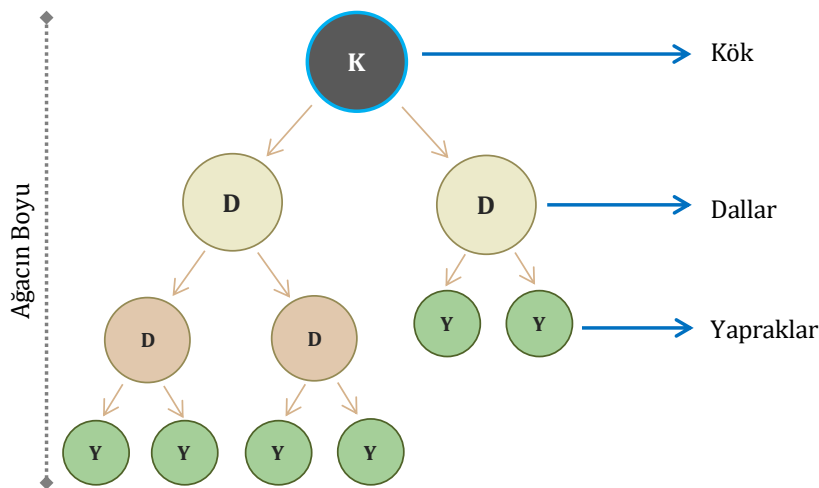
WEKA'da "trees" kategorisi altında yer alan J48 algoritması aslında C4.5 algoritmasının WEKA'daki uzantısıdır. İlk olarak 1993 yılında Quinlan tarafından geliştirilen bir karar ağacı algoritmasıdır. Yine bu algoritma Quinlan tarafından geliştirilen

ID3 algoritmasının daha gelişmiş halidir. ID3 algoritmasından farklı olarak kayıp verilerde değerleri tahmin etmekte kullanılmaktadır. J48 algoritması, örnek bir veri kümesi üzerinde sınıflandırma yaparken bir ağaç yapısı oluşturup, yapraklara bağlı düğümler ve düğümlere bağlı yapraklar oluşturur (Cihan ve Kalıpsız, 2015). J48 algoritması bir denetimli bir öğrenme algoritması olduğundan giriş verisi ile bir çıkış verisi gerektirir. Veri seti analiz edilerek hem test hem de eğitim verilerine bakılarak bir sınıflandırma yapılır (Bahiru, 2018). J48 algoritmasında veri kümesinde belirsizlikler bulunabilir bu belirsizlikler entropi hesabı ile giderilmekte olup aşağıdaki (2.5) formülüyle hesaplanır (Aksu ve Karaman, 2017).

$$\mathcal{L} = \sum_{i=1}^n p_i \log_2 p_i \quad (2.5)$$

J48 algoritması ile sınıflandırma yapılırken aşağıdaki adımlar izlenir:

Öncelikler veriler eğitim ve test olmak üzere iki guruba ayrılır. Eğitim verilerindeki belirsizlikler giderilerek öğrenme aşamasına geçilir. Bir karar ağacı modeli oluşturulur. Oluşturulan bu model test verileri ile karşılaştırılarak güvenilirliği test edilir. Elde edilen veriler ile sınıflandırma yapılır. Bir ağaç yapısı içinde en üst kısım kök en son kısım yaprak ve bunların arasında kalan bölüm ise dal olarak adlandırılır (Çoban, 2011). J48 ağaç yapısı Şekil 2.16'da gösterilmektedir.

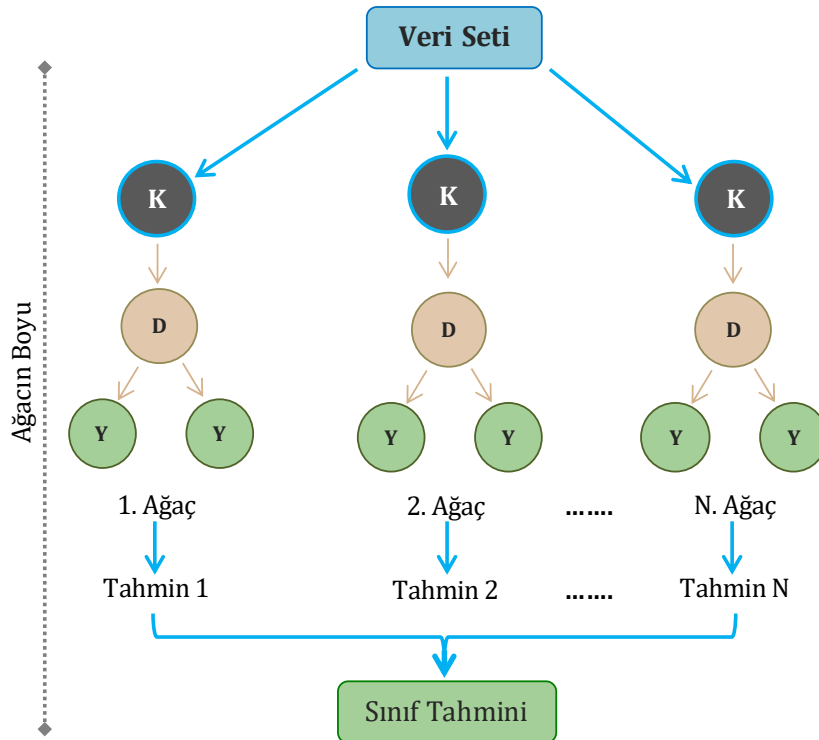


Şekil 2.16. J48 algoritması sınıflandırma modeli

2.6.7. Rastgele Orman Algoritması

Rastgele Orman (RF) algoritması WEKA’da "trees" kategorisinde yer almaktadır. Leo Breiman ve Adele Cutler tarafından geliştirilmiştir. Adından da anlaşılacağı gibi sınıflandırma yaparken rasgele birçok ağaç oluşturduğundan rasgele orman olarak adlandırılır. RF algoritmasının çalışma prensibi; veri kümesi sınıflandırılırken birçok ağaç oluşturulur ve her düğümde rastgele belirlenen değişkenler içinden en iyisi seçilerek her bir düğüm dallara ayrılarak bunların sonucunda bir sınıf tahmininde bulunulur. Son olarak tüm tahminler değerlendirilerek verinin hangi sınıfa ait olduğu belirlenir. RF algoritmasının diğer karar ağacı algoritmalarından farkı oluşturulan ağaç yapısı ile birlikte kök ve düğüm kısımlarının da rastgele çalışmasıdır. Ayrıca oluşturulan ağaçlarda budama yapılmaz ve böylece RF algoritmasının doğruluğu artırılmış olur (Breiman, 2001).

Bu algoritmanın avantajları; hızlı ve doğruluk açısından benzersiz olması, büyük veri setleri ile çalışması, eksik verileri doğru tahmin etmekte etkilidir. Başka verilerde kullanılmak için kaydedilebilir (Krishnaveni ve Hemalatha, 2011). RF ağaç yapısı Şekil 2.17’de gösterilmektedir.



Şekil 2.17. Rasgele orman algoritması sınıflandırma modeli

2.7. Sınıflandırma Algoritmalarının Performans Ölçütleri

Karmaşıklık Matrisi (Confusion Matrix): sınıflandırma algoritmalarının yanlış ve doğru tahminlerin gösterildiği bir sonuç tablosudur. Yapılan tahminler sonrası test ve eğitim verileri karşılaştırılır. Karmaşıklık matrisinin örneği Şekil 2.18’de gösterilmektedir.

==Karmaşıklık Matrisi==

Pozitif	Negatif	Tahmin Sınıfı
TP	FN	Pozitif
FP	TN	Negatif

Şekil 2.18. Karmaşıklık matrisi örneği

TP Oranı: Doğru pozitiflerdir. Pozitif olarak tahmin edilen pozitif doğru örneklerdir.

FP Oranı: Yanlış pozitiflerdir. Pozitif olarak tahmin edilen negatif yanlış örneklerdir.

TN Oranı: Doğru negatiflerdir. Negatif olarak tahmin edilen doğru örneklerdir.

FN Oranı: Yanlış negatiflerdir. Negatif olarak tahmin edilen yanlış örneklerdir (Umadevi ve Marseline, 2017).

Doğruluk (Accuracy): Doğruluk, veri kümesinde sadece doğru sınıflandırılmış örnek sayısının veri kümesindeki toplam örnek sayısına bölümüdür. Doğruluk değeri (2.6)’da gösterilen formül ile hesaplanır (Umadevi ve Marseline, 2017).

$$K_d = \frac{TP+TN}{TP+FP+TN+FN} \quad (2.6)$$

Kesinlik (Precision): Kesinlik, doğru olarak sınıflandırılan doğru pozitif sayısının kendisi ile birlikte yanlış pozitif sayısının toplamına oranıdır. Kesinlik değeri (2.7)’de gösterilen formül ile hesaplanır.

$$K_p = \frac{TP}{TP+FP} \quad (2.7)$$

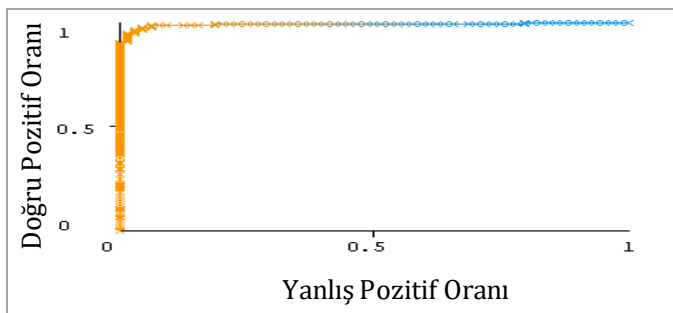
Duyarlılık (Recall): Duyarlılık, doğru pozitif sayının yine doğru pozitif ile yanlış negatiflerin toplam sayısına bölünmesi olarak tanımlanır. Duyarlılık değeri (2.8)'de gösterilen formül ile hesaplanır.

$$K_r = \frac{TP}{TP+FN} \quad (2.8)$$

F-Ölçütü (F-Measure): F-Ölçütü, duyarlılık ile kesinliğin bir birleşimidir. F-Ölçütü, iki kat duyarlılık ve kesinlik çarpımının yine duyarlılık ve kesinlik toplamına bölünmesiyle hesaplanır. F-Ölçütü değeri (2.9)'da gösterilen formül ile hesaplanır.

$$K_f = \frac{2 \times \text{Duyarlılık} \times \text{Kesinlik}}{\text{Duyarlılık} + \text{Kesinlik}} \quad (2.9)$$

ROC Eğrisi (Receiver Operating Characteristics): ROC eğrisi, alıcı çalışma alanı olarak bilinen yapılan sınıflandırmada doğru pozitif ve yanlış pozitif oranlarını karşılaştırmak için çizilen grafikdir. Şekil 2.19'da gösterildiği gibi yanlış pozitif örnek sayısı x ekseninde, doğru pozitif örnek sayısı y ekseninde gösterilir. ROC grafiğinde eğrinin sol üst köşeye yaklaşması sınıflamaların başarılı olduğunu gösterir. ROC değeri için 0-0,60 arası başarısız, 0,61-0,70 arası zayıf, 0,71-0,80 arası orta, 0,81-0,90 arası iyi ve 0,91-1,00 arası çok iyi olarak değerlendirilir (Umadevi ve Marseline, 2017, Bharati ve Ark., 2018).



Şekil 2.19. ROC eğrisi grafiği

Kappa İstatistiği: Sayısal olarak beklenen değer ile gözlemlenen değerlerin karşılaştırılmasıdır. Karmaşıklık matrisine bakılarak hesaplanır. Satırlardaki sayılar gerçek değerleri gösterirken, sütunlardaki sayılar tahmin edilen değerleri gösterir. Kappa istatistiği (K_i) gözlemlenen doğruluk oranından (L_g) beklenen doğruluk oranının (L_b) çıkartılarak

bulunan değerin, beklenen doğruluk oranının bir rakamından çıkarılarak elde edilen değere bölünmesi ile (2.10)'da gösterilen formülle hesaplanır. K_i değeri için 0,00-0,20 arası başarısız, 0,21-0,40 arası zayıf, 0,41-0,60 arası orta; 0,61-0,80 arası iyi ve 0,81-1,00 arası çok iyi olarak tanımlanmaktadır (Aksu ve Doğan, 2019).

$$K_i = \frac{L_{(g)} - L_{(b)}}{1 - L_{(b)}} \quad (2.10)$$

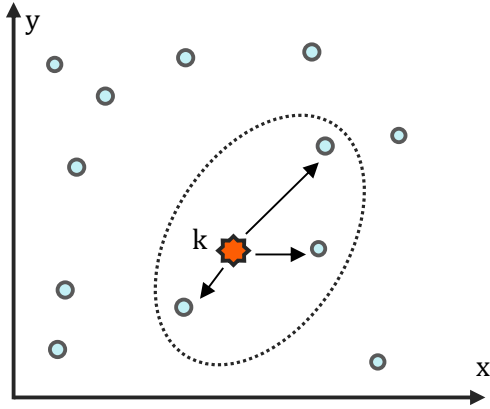
Ortalama Mutlak Hata (Mean absolute error, MAE): MAE, bir veri dizinindeki tahminlerde ortalama hata değerlerini bulmaya yarar. MAE değeri 0 ile sonsuz arasında olabilir. Ancak performans anlamında sıfıra yakın olması istenmektedir.

2.8. Kümeleme Yöntemi

Denetimsiz öğrenme türlerinden olan kümeleme algoritmaları, trafik kazalarının analiz edilmesinde sıklıkla kullanılan bir yöntemdir. Kümeleme algoritmaları ile kazaların benzerlik ve farklılıkları ortaya çıkarılabilmektedir. Böylece kazaların karakteristik özellikleri belirlenerek trafik kazaların önlenmesi hedeflenir (Güner ve Ark., 2018). Bu çalışmada SimpleKMeans ve EM kümeleme algoritmaları kullanılarak, trafik kazalarının özellikleri incelenmiştir.

2.8.1. SimpleKMeans Algoritması

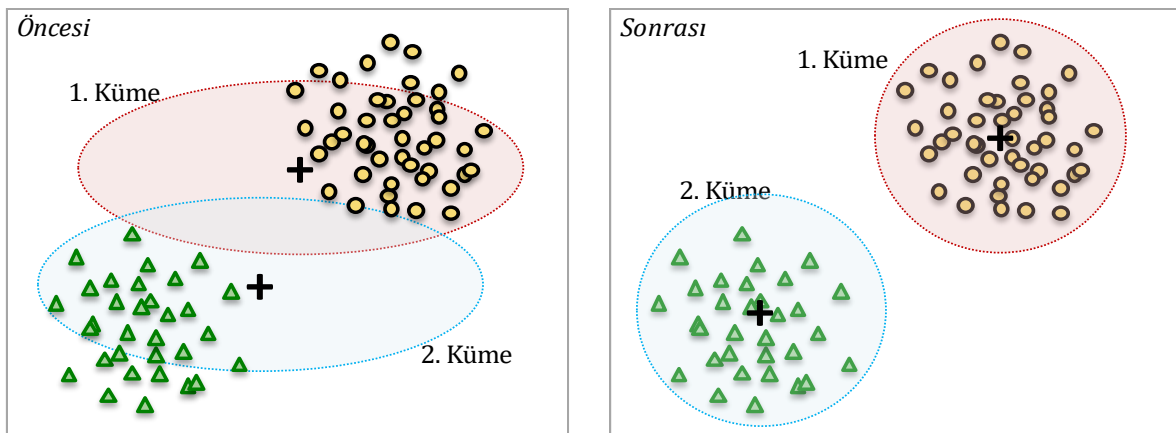
KMeans algoritması, kümeleme algoritmaları arasında yaygın olarak kullanılan ve ilk olarak 1967 yılında MacQueen tarafından önerilen en basit Denetimsiz öğrenme algoritmalarındandır. Bu algoritma bölümlere yaparak çalışmaktadır ve iki adımda oluşur. İlk adımda, önceden sabitlenen bir merkezi k değeri rasgele seçilir. İkinci adımda, genellikle Öklid mesafesi kullanılarak söz konusu k noktası ile tüm noktalar arasındaki uzaklıklar tek tek hesaplanır ve en yakın örnekler seçilir. Tüm veri örnekleri dâhil edilerek öncelikli kümeleme tamamlanır. Daha sonra öncelikli kümelemenin ortalaması yeniden hesaplanır. Bu yenileme süreci istenen kriter minimum olana kadar tekrarlanır (Na ve Ark., 2010). SimpleKMeans algoritması modeli Şekil 2.20'de verilmiştir.



Şekil 2.20. SimpleKMeans algoritması kümeleme modeli

2.8.2. EM Algoritması

Beklenti (expectation) ve maksimizasyon (maximization) kelimelerinin birleşimi olan EM algoritması, ilk olarak 1977 yılında Arthur Dempster, Nan Laird ve Donald Rubin tarafından yayınlanan bir makalede açıklanmıştır. EM algoritması, eksik ve belirsiz olan veri kümesi dağılımlarında maksimum olabilirlik tahmini yapabilmenin genel bir yöntemidir. Bu kümeleme algoritmasının çalışması iki adımdan oluşur (URL-4, 2022). Şekil 2.21’de gösterildiği gibi; başlangıçta kümeleme işlemi rasgele bir noktadan başlayarak tahmin yapılmakta ve küme çapları geniş ve küme ayrımları belirsiz görünmekteyken, sonrasında beklenti ve maksimizasyonular yapılarak küme çapları küçülmekte ve iki küme birbirinden ayrılarak belirgin hale dönüşmektedir.



Şekil 2.21. EM algoritması kümeleme modeli

2.9. Birliktelik Kuralı Çıkarma Yöntemi

Birliktelik kurallı algoritmalar, bir veri setindeki örneklerin birbirleri ile olan ilişkilerini eş zamanlı olarak inceleyen algoritmalarlardır. Veriler arasındaki ilişkiler eş zamanlı olarak incelenirken bir olasılık korelasyonu kullanılır (Karamaşa ve Erdoğan, 2018). Bu çalışmada Apriori algoritması kullanılarak, trafik kazalarının birbiri ile olan ilişkileri incelenmiştir.

2.9.1. Apriori Algoritması

İlk olarak 1994 yılında Agrawal ve Srikant tarafından öne sürülen apriori algoritması, Birliktelik temelli kurallar üretilmekte kullanılan en yaygın yöntemlerdendir. Apriori Algoritmasının hesaplanmasında destek, güven ve kaldırma değeri olmak üzere üç önemli adım bulunmaktadır (Karamaşa ve Erdoğan, 2018).

Destek (Support): Bir kural oluşturulurken veri setindeki örneklerin (N) ne kadar tekrarlandığını hesaplamak için kullanılır. Formül (2.11)'deki gibi hesaplanır.

$$P_d (X \rightarrow Y) = \frac{\text{Sayı } (X, Y)}{N} \quad (2.11)$$

Güven (Confidence): Bir kural oluşturulurken veri setindeki örneklerden X değişkeninin Y değişkeni ile birlikte bulunma olasılığıdır. Formül (2.12)'deki gibi hesaplanır.

$$P_g (X \rightarrow Y) = \frac{\text{Sayı } (X, Y)}{\text{Sayı } (X)} \quad (2.12)$$

Kaldırma (Lift): X ve Y değerlerinin birbirine bağımlı olup olmadıklarını gösteren bir destek değeridir. Lift değeri 1'den ne kadar büyükse çıkan birliktelik kuralıda o derece ilginç olarak yorumlanır ve formül (2.13)'deki gibi hesaplanır.

$$P_k (X \rightarrow Y) = \frac{\text{Destek } (X, Y)}{\text{Destek } (X) * \text{Destek } (Y)} \quad (2.13)$$

Apriori algoritması çalışması şu adımlarda gerçekleşir: İlk olarak bir minimum bir destek ve güven oranı belirlenir. Veri setindeki örnekler için destek oranları hesaplanır. Belirlenen destekten daha düşük olan veriler devre dışı bırakılır. Birli, ikili, üçlü veya daha fazla birliktelik kümeleri oluşturulur. Son olarak oluşturulan birliktelik kümelerinden destek ve güven oranını sağlayanlar değerlendirilir.

Örneğin, Bir manava gelen müşterilerin almış olduğu meyvelerin (A, B, C, D, E, F) apriori algoritması ile destek oranın %20, güven oranın %60 seçildiği durumda birliktelik durumlarını aşağıdaki gibi hesaplanır.

Müşteri ID	Alınan ürünler
1	A, B, C, D
2	A, B, C
3	C, F
4	E, F
5	A, B
6	A, B, F

İlk adımda alınan ürünlerin tek tek destek oranları hesaplanır ve bu oranın altında kalanlar elenir.

1-Adım

Ürün adı	Sayı	Destek oranı (%)
A	4	67
B	4	67
C	3	50
D	1	17
E	1	17
F	3	50

İkinci adımda ürünlerin ikili kümeler oluşturularak destek oranları hesaplanır ve bu oranın altında kalanlar elenir.

2-Adım

Ürün adı	Sayı	Destek oranı (%)
A, B	4	67
A, C	2	33
A, F	1	17
B, C	2	33
B, F	1	17
C, F	1	17

Üçüncü adımda ürünlerin üçerli kümeler oluşturularak destek oranları hesaplanır ve bu oranın altında kalanlar tekrar elenir. Ancak aşağıdaki tabloda olduğu gibi başka küme oluşumu kaldığından işlem tamamlanır ve birliktelik kuralları çıkarılarak güven oranları hesaplanır.

3-Adım

Ürün adı	Sayı	Destek oranı
A, B, C	2	%33

Üçüncü adım sonunda destek oranı %20'nin üzerinde olan birbiri ile ilişkili beş farklı kural oluşmaktadır.

Sayı	Kurallar	Destek (%)	Güven (%)
1	A meyvesini alanlar (4), B ve C'de (2) alır.	%33	%50 (2/4)
2	B meyvesini alanlar (4), A ve C'de (2) alır.	%33	%50 (2/4)
3	C meyvesini alanlar (3), A ve B'de (2) alır.	%33	%67 (2/3)
4	A ve B meyvesini alanlar (4), C'de (2) alır.	%33	%50 (2/4)
5	A ve C meyvesini alanlar (2), B'de (2) alır.	%33	%100 (2/2)

Yukarıdaki veriler doğrultusunda, destek oranı %20, güven oranı %60 şartını sağlamakta olan iki birliktelik kuralı oluşmuştur. Buna göre;

- C meyvesini alan müşteriler aynı zamanda A ve B meyvelerini almıştır.
- A ve C meyveleri alan müşteriler aynı zamanda B meyvesini almıştır.

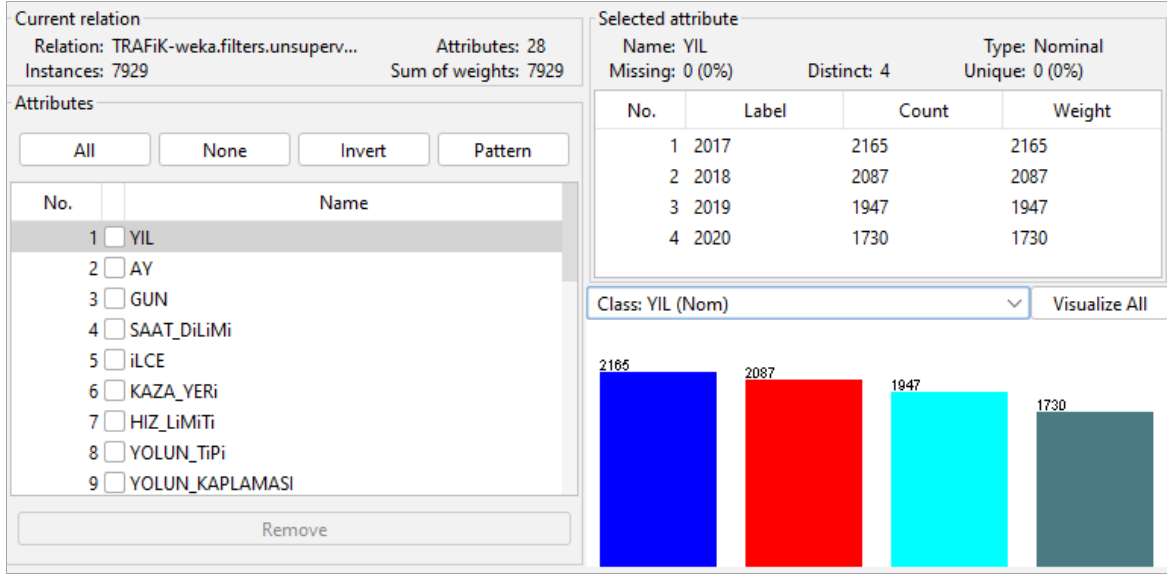
3. BULGULAR ve TARTIŞMA

Kahramanmaraş ili genelinde yaşanan trafik kazaları WEKA’da önışlem ve görsel karşılaştırma, sınıflandırma, kümeleme, birliktelik olmak üzere dört yöntem ile analiz edildi. İlk yöntemde veri setindeki özniteliklerin birbirleriyle olan etkileşimleri iki boyutlu grafiklerde gösterilerek değeriendirildi. Denetimli öğrenme tekniklerinden sınıflandırma algoritmaları ile trafik veri seti hem eğitim hem de test işlemlerine tabi tutularak algoritmaların sonuçları tahmin etme durumuna göre başarı oranları karşılaştırıldı. Denetimsiz öğrenme tekniklerinden kümeleme algoritmaları ile trafik kazaların yoğunlaştığı alanlar ve dikkat edilmesi kaza unsurları belirlenmektedir. Yine Denetimsiz öğrenme tekniklerinden birliktelik kuralı ile trafik kaza unsurlarının birbirleriyle benzerlik ilişkileri değeriendirilmektedir.

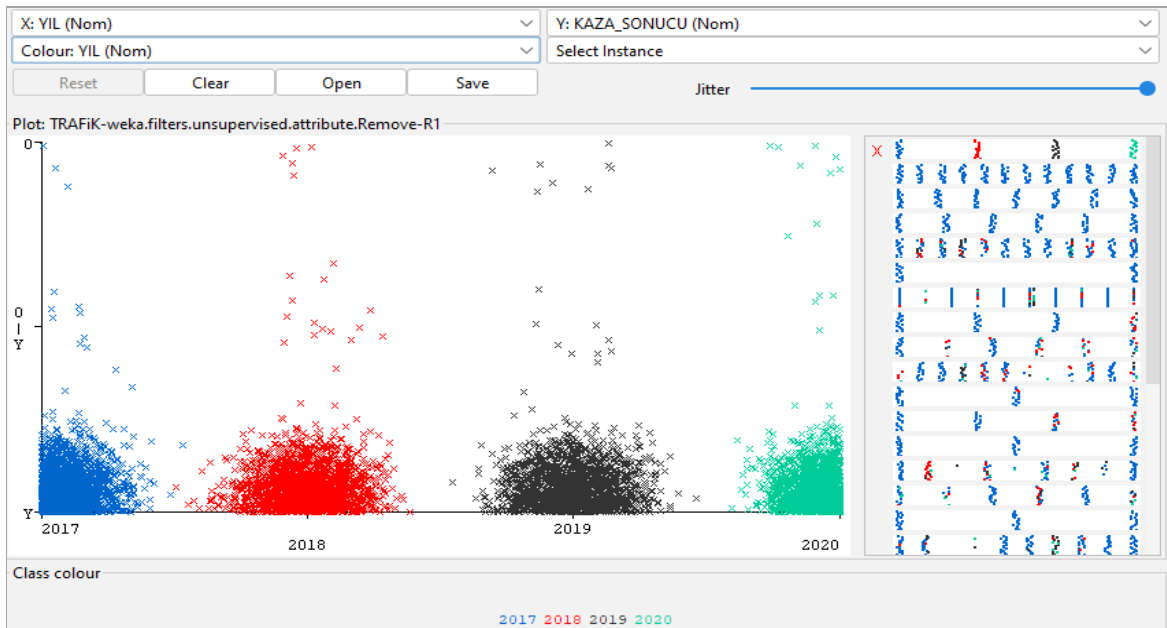
3.1. Önışlem ve Görsel Karşılaştırma Bulguları

Trafik kazası verileri WEKA’ya Preprocess (Önışlem) menüsünden yüklendikten sonra veri setine ait ön inceleme bu menüden, iki boyutlu görsel karşılaştırma ve değeriendirme işlemi Visualize menüsünden yapılmaktadır. Bu kapsamda yaşanan trafik kazalarının WEKA’daki önışlem ve görsel karşılaştırma bulguları aşağıdaki gibidir.

Kahramanmaraş ili ve ilçelerinde 2017 ile 2020 yılları arasında meydana gelen Şekil 3.1’deki trafik kazaları incelediğinde; en çok kaza yaşanan yılın 2.165 adet ile 2017, en az kaza yaşanan yılın 1730 adet ile 2020 olduğu görülmektedir. Meydana gelen trafik kazaları “Kazanın Sonucu” ve “Yıl” niteliklerine göre Şekil 3.2’deki gibi görsel olarak karşılaştırıldığında; sadece yaralanmalı sonuçlanan kazalar en çok 2017 yılında yaşanırken 2020 yılına kadar azalış göstermiştir. Ölümlü ve yaralanmalı sonuçlanan kazalar en çok 2018 yılında yaşanırken 2020 yılına kadar azalış göstermiştir. Çok sayıda olmasa da sadece ölümle sonuçlanan kazalar 2020 yılı sonuna kadar artış gözlenmiştir.



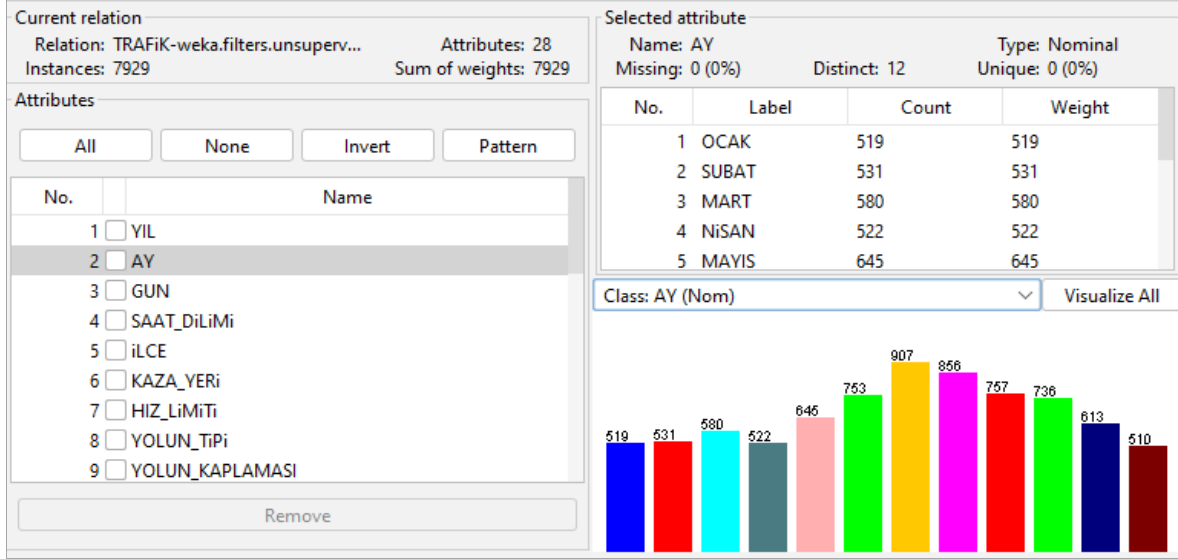
Şekil 3.1. Yıllara göre yaşanan trafik kazalarının önışlem verileri



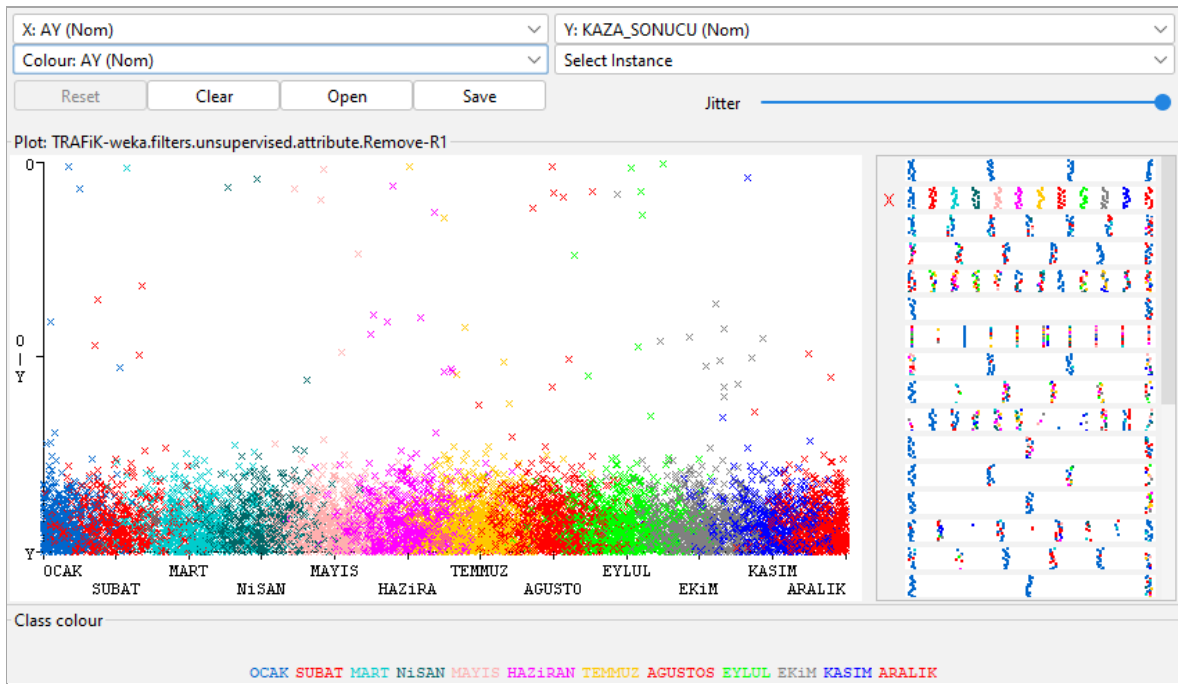
Şekil 3.2. Yıl ile kaza sonuçları arasındaki görsel karşılaştırma sonuçları

2017 ile 2020 yılları arasında meydana gelen trafik kazaları aylara göre incelediğinde; Temmuz ayı 907 adet kaza ile ilk sırada yer alırken, Aralık ayı 510 adet kaza ile en az kaza yaşanan ay olmuştur. Yaşanan trafik kazaları “Kazanın Sonucu” ve “Ay” niteliklerine göre görsel olarak karşılaştırıldığında; sadece yaralanmalı kazaların en çok Temmuz ayında yaşandığı, ölümlü ve yaralanmalı kazaların en çok Ekim ayında yaşandığı ve sadece ölümlü sonuçlanan kazaların Ağustos ayında yaşandığı görülmüştür.

Kaza yaşanan aylara ait önışlem ve görsel karşılaştırma verileri Şekil 3.3 ve 3.4'te gösterilmektedir.



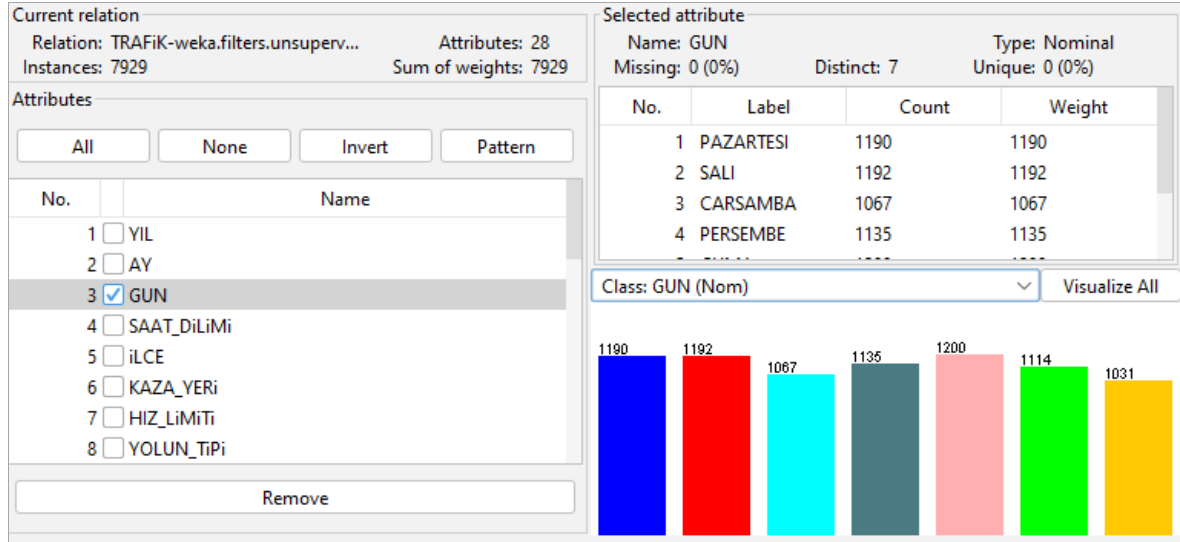
Şekil 3.3. Aylara göre yaşanan trafik kazalarının önışlem verileri



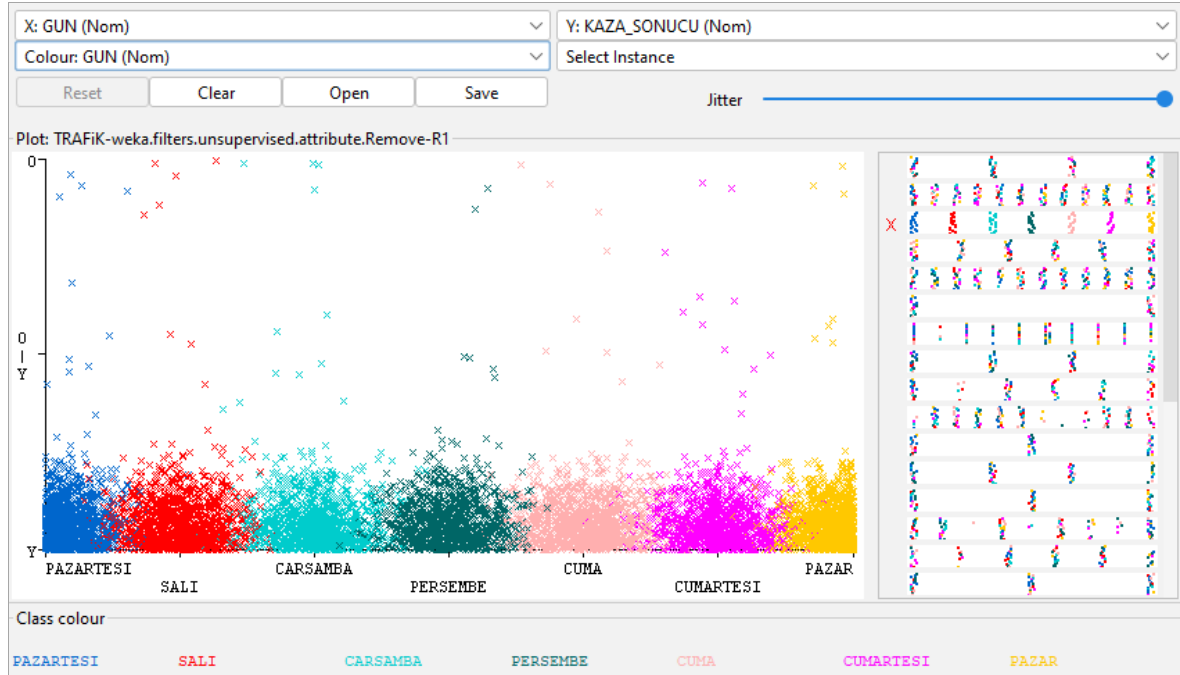
Şekil 3.4. Ay ve kaza sonuçları arasındaki görsel karşılaştırma sonuçları

Haftanın günlerine göre trafik kazaları incelendiğinde; Cuma günleri 1200 kaza ile ilk sırayı alırken, Pazar günleri 1031 kaza ile en az kaza yaşanan gün olmuştur. Yaşanan

trafik kazaları “Kazanın Sonucu” ve “Gün” niteliklerine göre görsel olarak karşılaştırıldığında; sadece yaralanmalı kazaların en çok Cuma günü yaşandığı, ölümlü ve yaralanmalı kazaların en çok Cumartesi günü yaşandığı ve sadece ölümle sonuçlanan kazaların çok sayıda olmasa da Salı günleri yaşandığı görülmüştür. Kaza yaşanan günlere ait önilem ve görsel karşılaştırma verileri Şekil 3.5 ve 3.6’da gösterilmektedir.

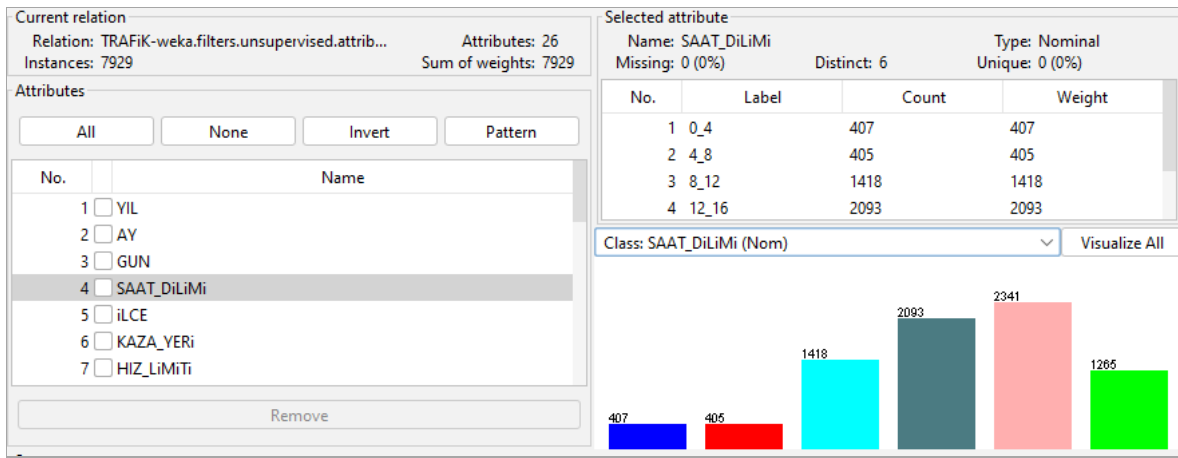


Şekil 3.5. Haftanın günlerine göre yaşanan trafik kazalarının önilem verileri

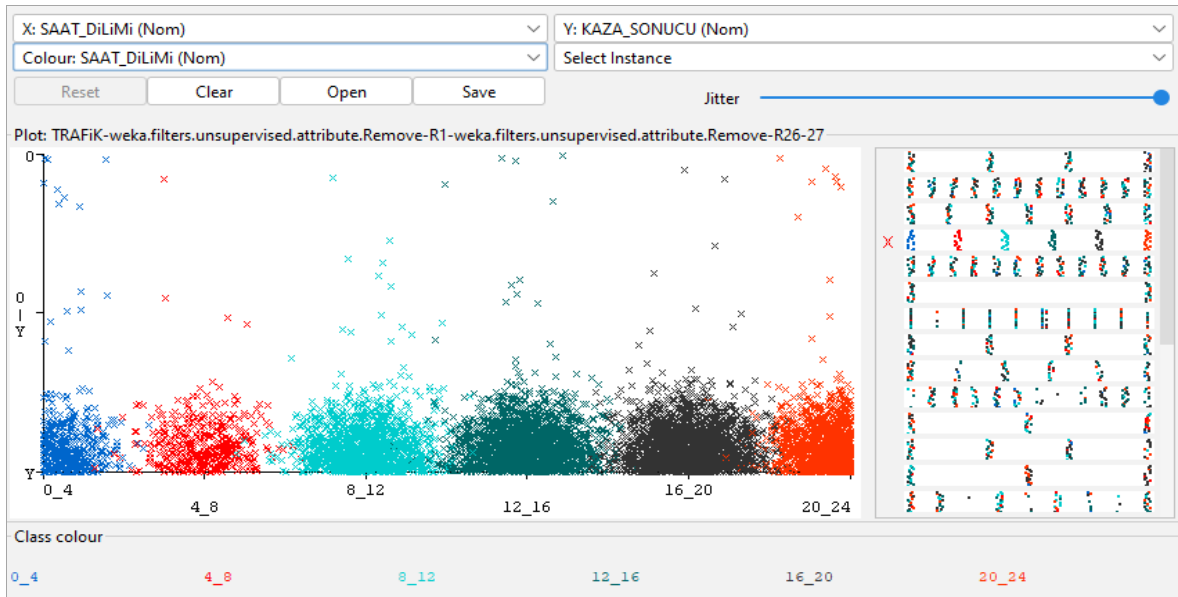


Şekil 3.6. Haftanın günleri ile kaza sonuçlarının görsel karşılaştırma sonuçları

Saat dilimine göre meydana gelen trafik kazaları incelediğinde; Kazalar 405 adetle en az gece 04.00-08.00 arasında yaşanırken, 2341 adetle en çok genelde iş çıkış saatleri olan 16.00-20.00 arasına yaşandığı görülmektedir. “Kazanın Sonucu” ve “Saat Dilimi” nitelikleri görsel olarak karşılaştırıldığında; sadece yaralanmalı kazaların en çok saat 16.00-20.00 aralığında yaşandığı, ölümlü ve yaralanmalı kazaların en çok saat 08.00-12.00 aralığında yaşandığı ve sadece ölümle sonuçlanan kazaların gece saat 04.00-08.00 aralığında yaşandığı görülmüştür. Kaza yaşanan saatlere ait önışlem ve görsel karşılaştırma verileri Şekil 3.7 ve 3.8’de gösterilmektedir.

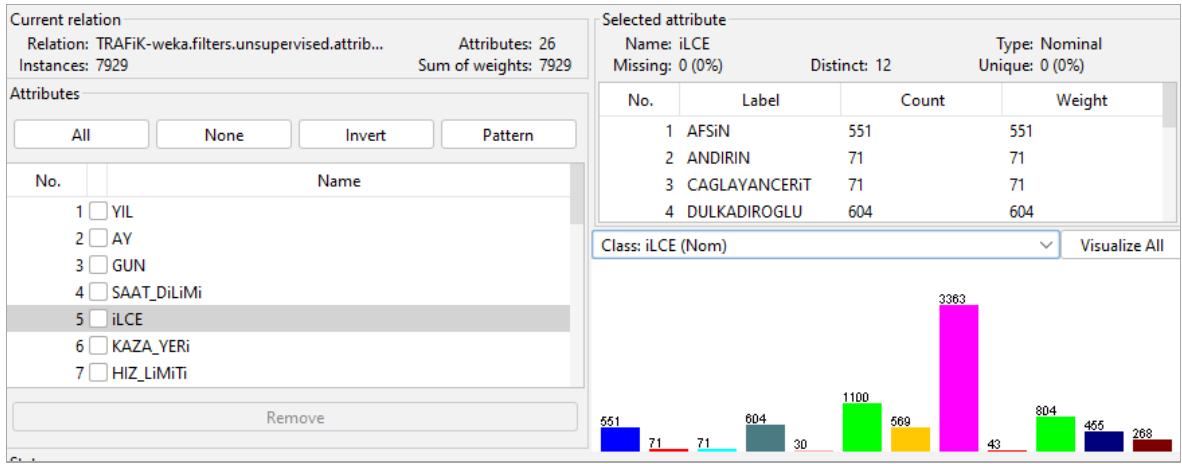


Şekil 3.7. Saat dilimlerine göre yaşanan trafik kazalarının önışlem verileri

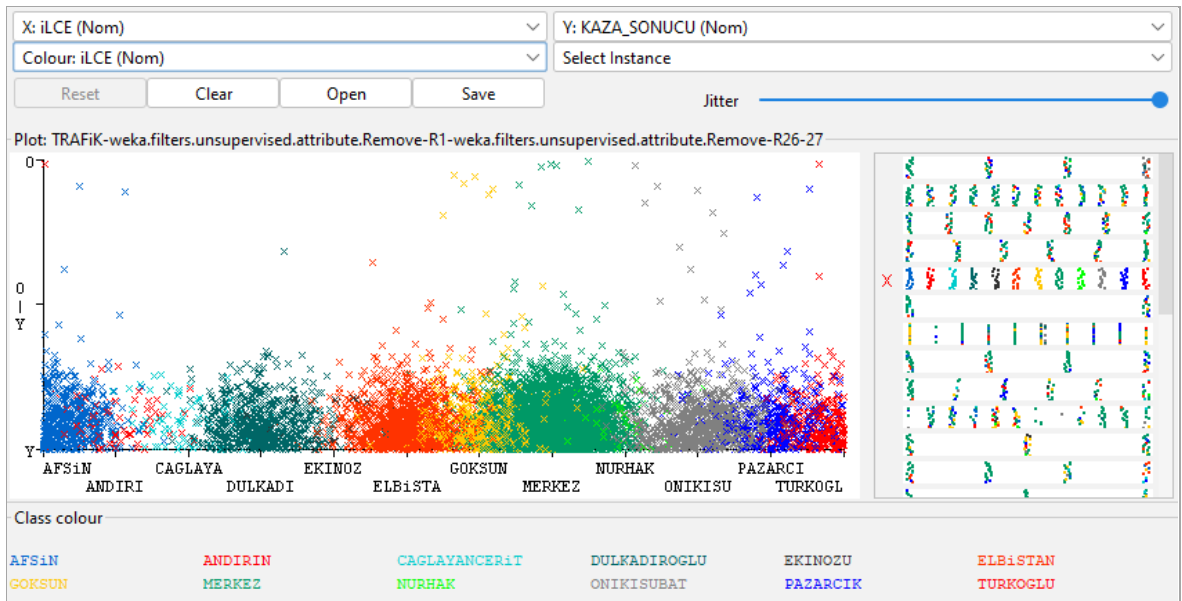


Şekil 3.8. Saat dilimi ve kaza sonuçları arasındaki görsel karşılaştırma sonuçları

Kahramanmaraş merkezi ve ilçelerinde yaşanan trafik kazaları incelendiğinde; İl merkezi 3363 kaza ile ilk sırayı alırken, onu 1100 kazayla Elbistan ilçesi takip etmektedir. Ekinözü 30 kaza, Nurhak 43 kaza ile en az kaza yaşanan ilçeler olmuştur. “Kazanın Sonucu” ve “İlçe” nitelikleri görsel olarak karşılaştırıldığında; sadece yaralanmalı kazaların en çok Kahramanmaraş merkezinde yaşandığı, ölümlü ve yaralanmalı kazaların en çok il merkezi ile Onikişubat ilçesinde yaşandığı ve sadece ölümle sonuçlanan kazaların il merkezi ile Göksun ilçesinde yaşandığı görülmüştür. Kaza yaşanan ilçelere ait önışlem ve görsel karşılaştırma verileri Şekil 3.9 ve 3.10’da gösterilmektedir.

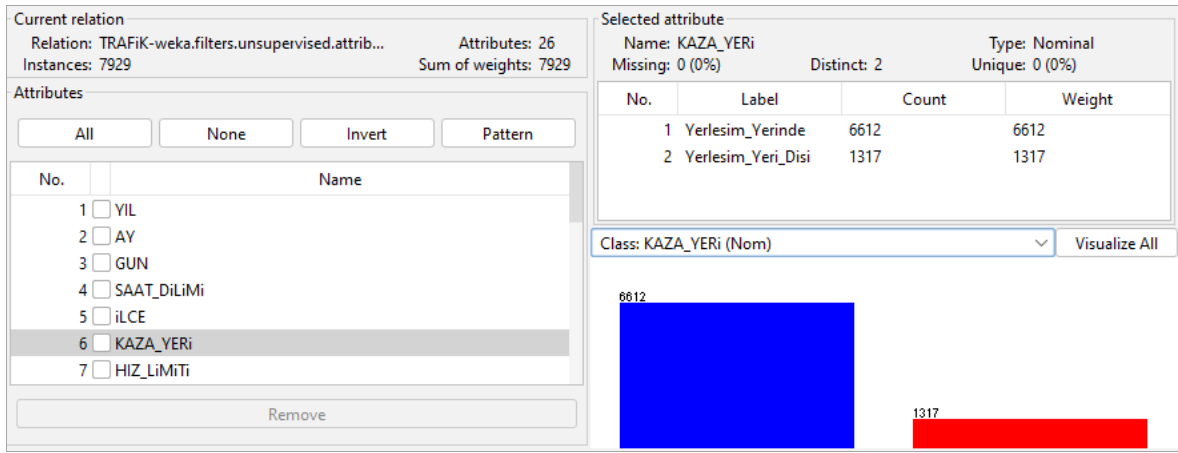


Şekil 3.9. İlçelerde yaşanan trafik kazalarının önışlem verileri

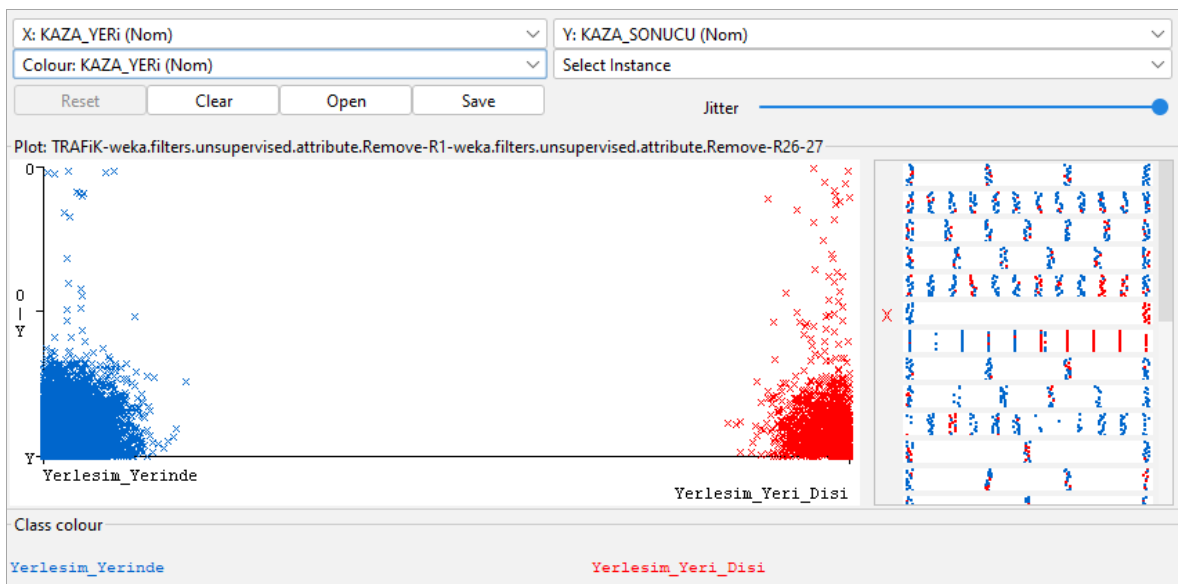


Şekil 3.10. İlçe ile kaza sonuçları arasındaki görsel karşılaştırma sonuçları

Meydana gelen trafik kazaları yerleşim yeri bilgisi göre incelediğinde; kazalarının 6612'si yerleşim yeri sınırlarında olurken 1317'si yerleşim yeri dışında meydana gelmiştir. “Kazanın Sonucu” ve “Kaza Yeri” niteliklerine göre görsel olarak karşılaştırıldığında; sadece yaralanmalı kazaların en çok yerleşim yeri içinde yaşandığı, ölümlü ve yaralanmalı kazaların en çok yerleşim yeri dışında yaşandığı ve sadece ölümle sonuçlanan kazaların yerleşim yeri içi ve dışında eşit yaşandığı görülmüştür. Ancak yerleşim yeri dışında yaşanan kazalar yerleşim yeri içinde yaşananlardan az olsa da ölüm oranının yüksek olduğu gözlenmiştir. Kazaların yerleşim yeri bilgisine ait önışlem ve görsel karşılaştırma verileri Şekil 3.11 ve 3.12’de gösterilmektedir.

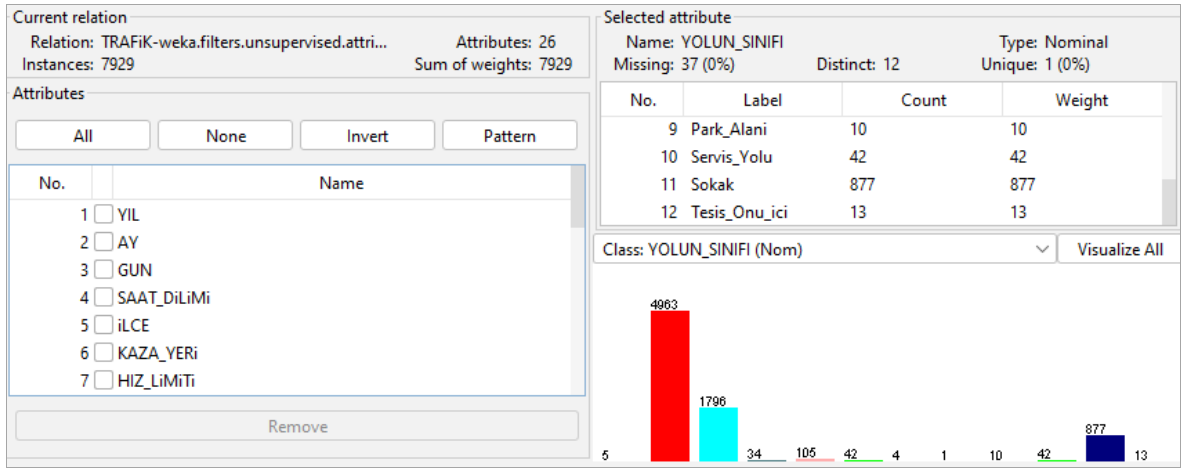


Şekil 3.11. Kaza yerine göre yaşanan trafik kazalarının önışlem verileri

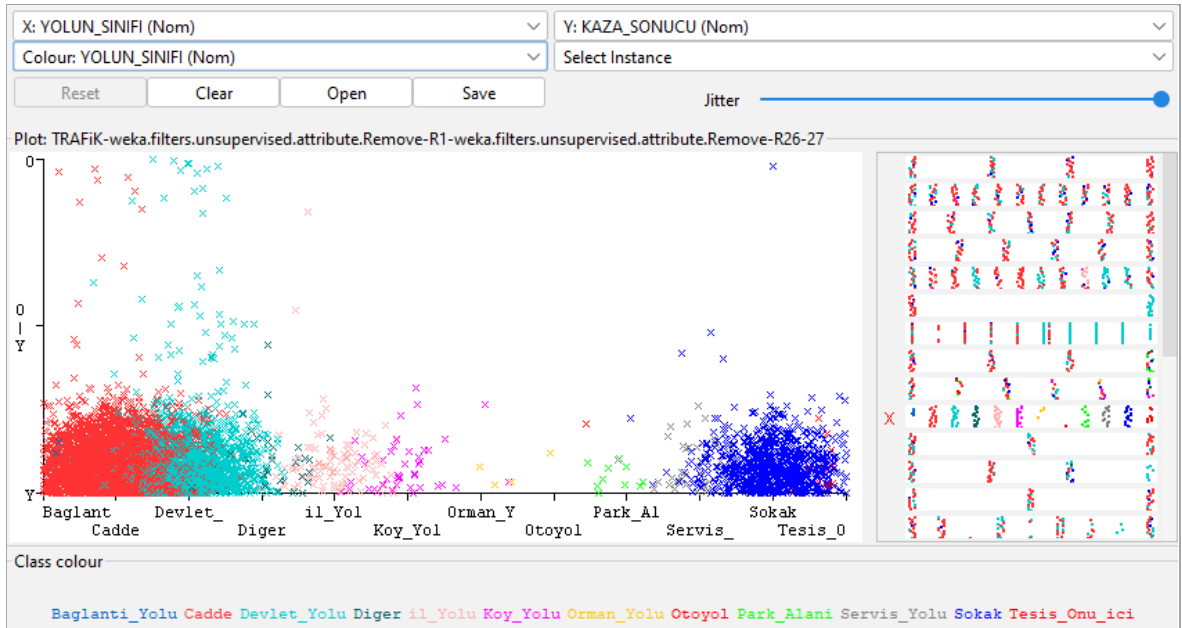


Şekil 3.12. Kaza yeri ile kaza sonuçları arasındaki görsel karşılaştırma sonuçları

Yolun sınıfı niteliğine göre trafik kazaları incelendiğinde; caddeler ve sokaklar kazaların en çok yaşandığı yol türleri olurken, otoyol ve orman yolları en az kaza yaşanan yol türleri olmuştur. Yaşanan trafik kazaları “Kazanın Sonucu” ve “Yolun Sınıfı” niteliklerine göre görsel olarak karşılaştırıldığında; sadece yaralanmalı kazaların en çok caddelerde yaşandığı, ölümlü/yaralanmalı ve sadece ölümle sonuçlanan kazaların devlet yollarında yaşandığı görülmüştür. Devlet yollarında yaşanan kazalar cadde üzerinde yaşanan kazalardan az olsa da ölüm oranının yüksek olduğu gözlenmiştir. Yolun sınıfı niteliğine ait önışlem ve görsel karşılaştırma verileri Şekil 3.13 ve 3.14’te gösterilmektedir.

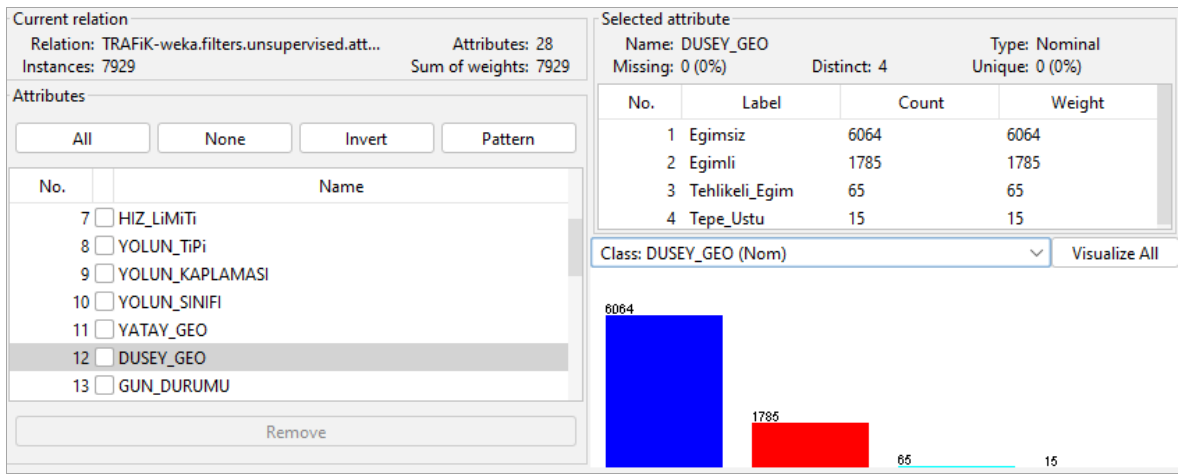


Şekil 3.13. Yolun sınıfına göre yaşanan trafik kazalarının önışlem verileri

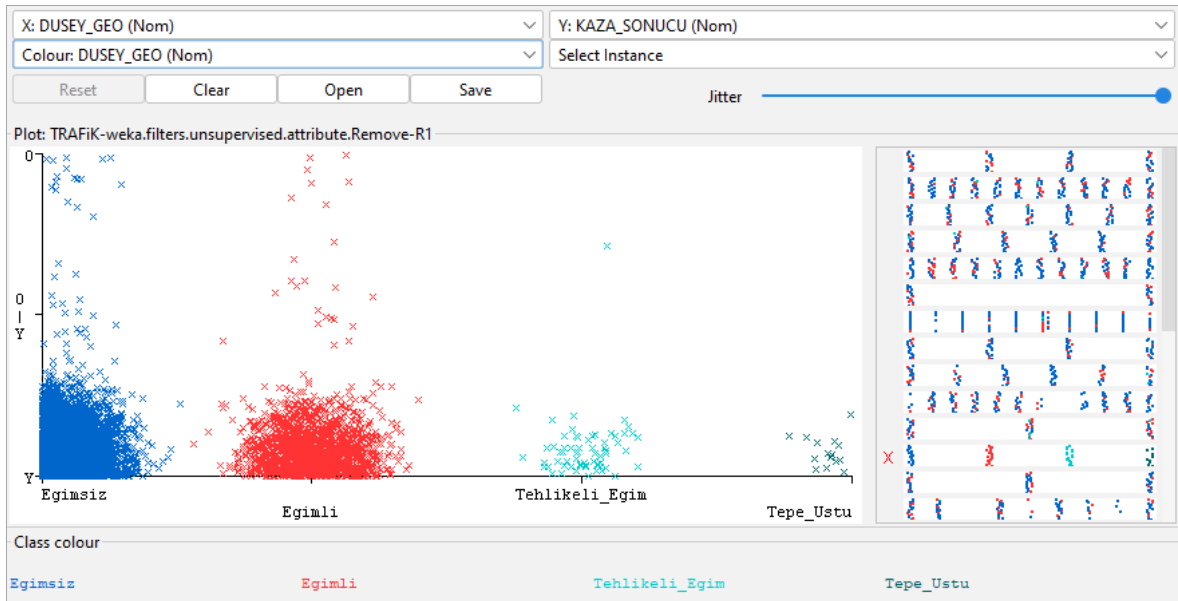


Şekil 3.14. Yolun sınıfı ve kaza sonuçlarının görsel karşılaştırma sonuçları

Yolun düşey geometrisi niteliğine göre trafik kazaları incelendiğinde; en çok kaza eğimsiz yollarda yaşanırken, en az kaza tepe üstü olan yollarda yaşanmaktadır. Yaşanan kazalar “Kazanın Sonucu” ve “Düşey Geometri” niteliklerine göre görsel olarak karşılaştırıldığında; eğimsiz yollar yaralanmalı, ölümlü/yaralanmalı ve ölümlü olmak üzere 3 sınıfta da en çok kaza yaşanan düşey geometriye sahip yol olarak görülse de eğimli yollarda meydana gelen ölümlü ve ölümlü/yaralanmalı kazaların oranı da yüksektir. Yolun düşey geometri niteliğine ait önilem ve görsel karşılaştırma verileri Şekil 3.15 ve 3.16’da gösterilmektedir.

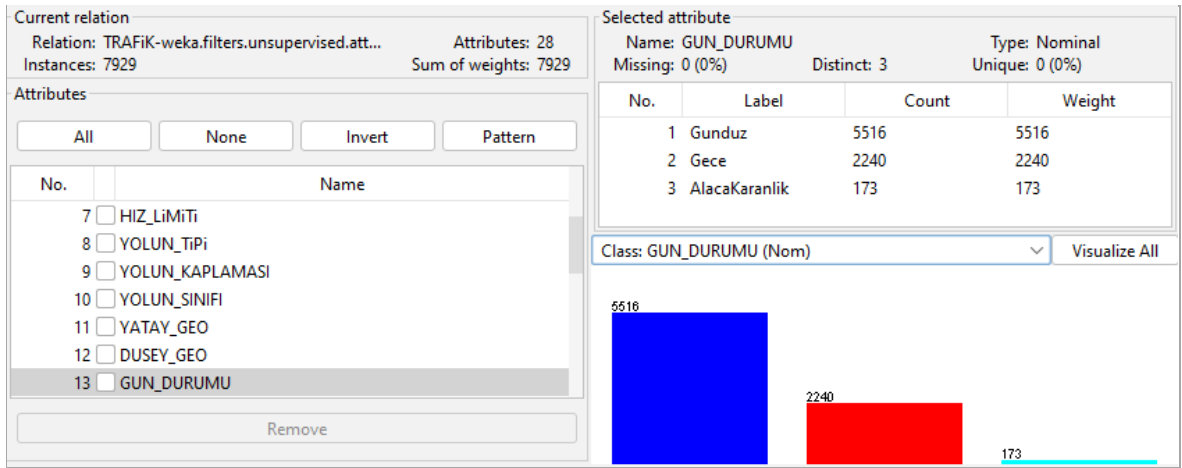


Şekil 3.15. Yolun düşey geometrisine göre yaşanan trafik kazalarının önilem verileri

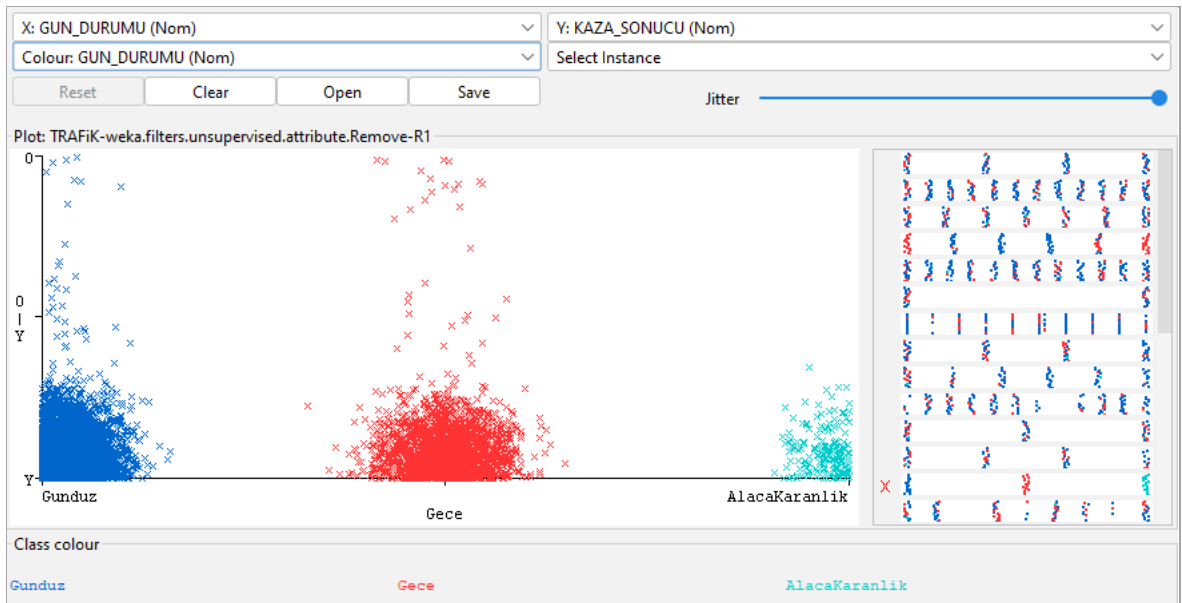


Şekil 3.16. Yolun düşey geometrisi ile kaza sonuçlarının görsel karşılaştırma sonuçları

Gün durumu niteliğine göre trafik kazaları incelendiğinde; Gündüzleri 5516, geceleri 2240, alacakaranlıkta ise 173 kaza yaşanmıştır. Yaşanan kazalar “Kazanın Sonucu” ve “Gün Durumu” niteliklerine göre görsel olarak karşılaştırıldığında; gündüzleri yaralanmalı, ölümlü/yaralanmalı ve ölümlü olmak üzere 3 sınıfta da en çok kaza yaşanan gün durumu olarak görülse de gece meydana gelen ölümlü veya ölümlü/yaralanmalı kazaların oranı da yüksektir. Gün durumu niteliğine ait önışlem ve görsel karşılaştırma verileri Şekil 3.17 ve 3.18’de gösterilmektedir.

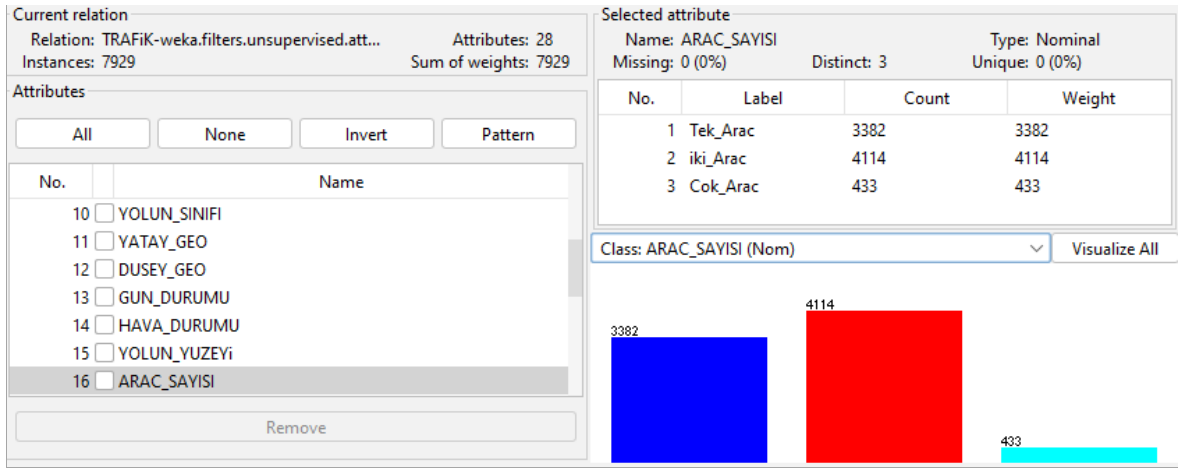


Şekil 3.17. Gün durumuna göre yaşanan trafik kazalarının önışlem verileri

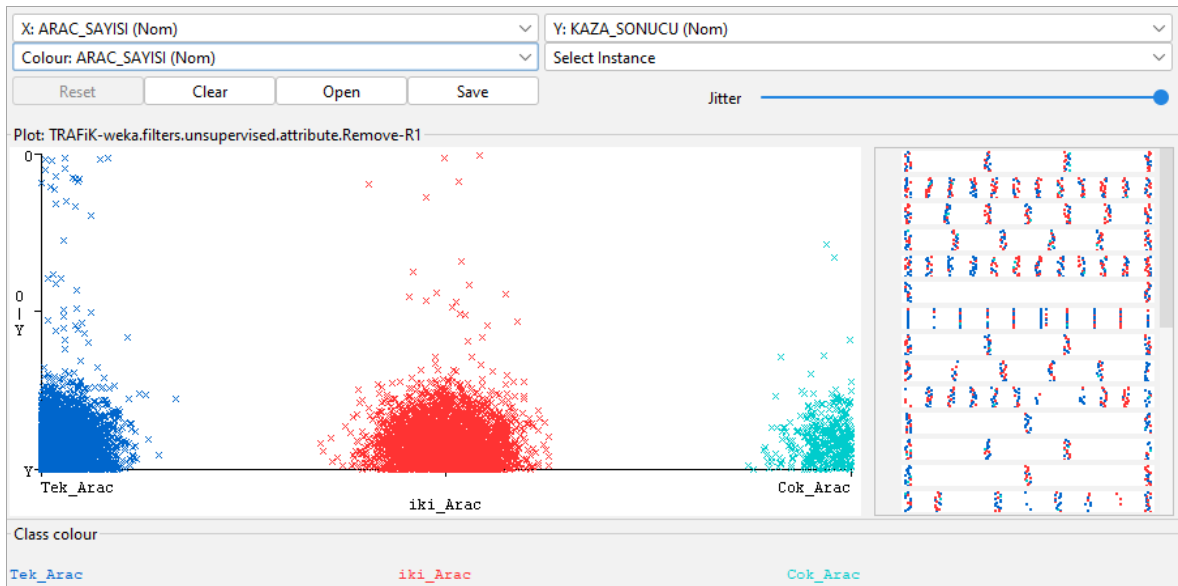


Şekil 3.18. Gün durumu ile kaza sonuçları arasındaki görsel karşılaştırma sonuçları

Kazaya karışan araç sayısına göre trafik kazaları incelendiğinde; kazaların en çok iki araçlı, tek araçlı ve çok araçlı olarak sıralandığı görülmektedir. Yaşanan trafik kazaları “Kazanın Sonucu” ve “Araç Sayısı” niteliklerine göre görsel olarak karşılaştırıldığında; yaralanmalı, ölümlü/yaralanmalı ve ölümlü olmak üzere 3 sınıfta da en çok kaza iki aracın karıştığı kazalarda yaşansa da tek araçla yaşanan kazalarda da ölümlü/yaralanmalı kazaların oranı yüksektir. Kazaya karışan araç sayısı niteliğine ait önişlem ve görsel karşılaştırma verileri Şekil 3.19 ve 3.20’de gösterilmektedir.

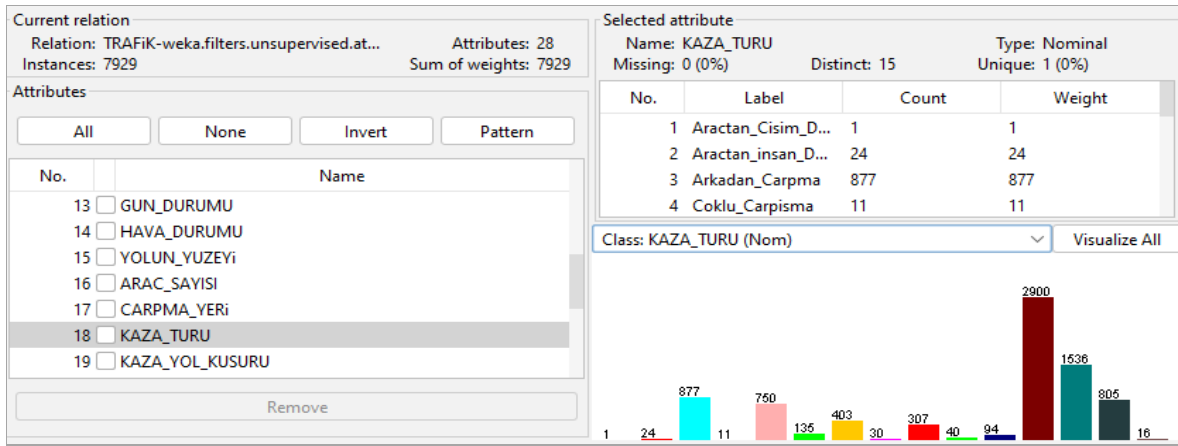


Şekil 3.19. Araç sayısına göre yaşanan trafik kazalarının önişlem verileri

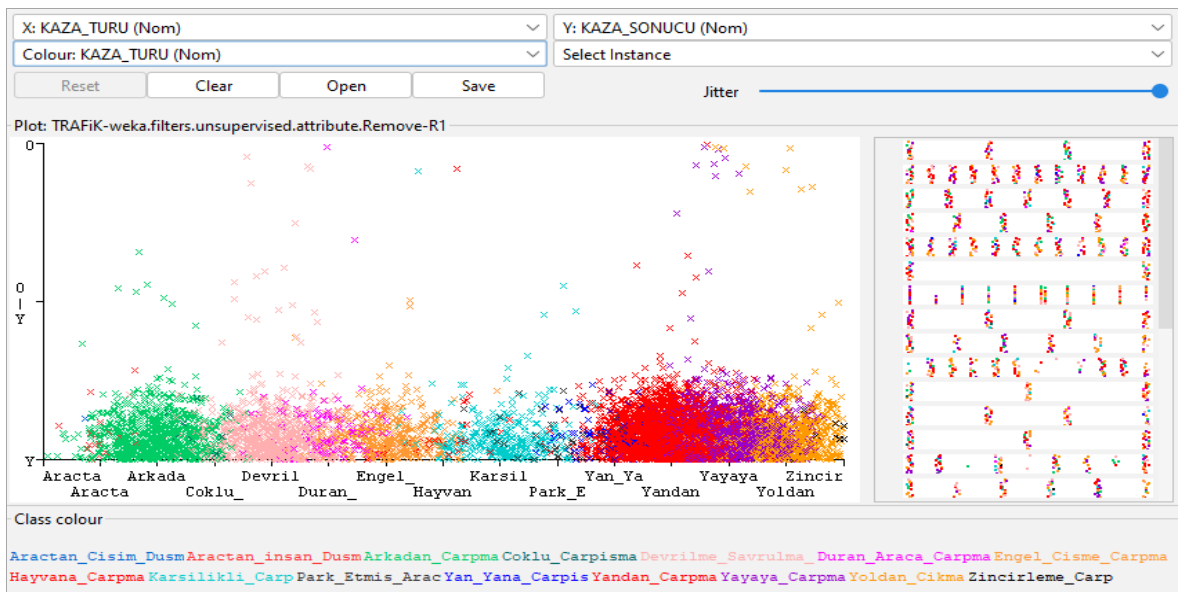


Şekil 3.20. Araç sayısı ile kaza sonuçları arasındaki görsel karşılaştırma sonuçları

Kaza türüne göre trafik kazaları incelendiğinde; en çok yandan çarpma, yaya çarpma, arkadan çarpma ve yoldan çıkma gibi kaza türleri olurken, araçtan cisim düşmesi, çoklu çarpışma ve zincirleme çarpışma en az görülen kaza türleri olmuştur. Yaşanan trafik kazaları “Kazanın Sonucu” ve “Kaza Türü” nitelikleri görsel olarak karşılaştırıldığında; sadece yaralanmalı kazalar en çok yandan çarpmalı kazalarda yaşanmıştır. Ölümlü/yaralanmalı kazalar en çok arkadan çarpma ve devrilme savrulmalı kazalarda yaşanmıştır. Sadece ölümle sonuçlanan kazalar ise yayaya çarpma ve yoldan çıkma olarak gözlenmiştir. Kaza türü niteliğine ait önışlem ve görsel karşılaştırma verileri Şekil 3.21 ve 3.22’de gösterilmektedir.

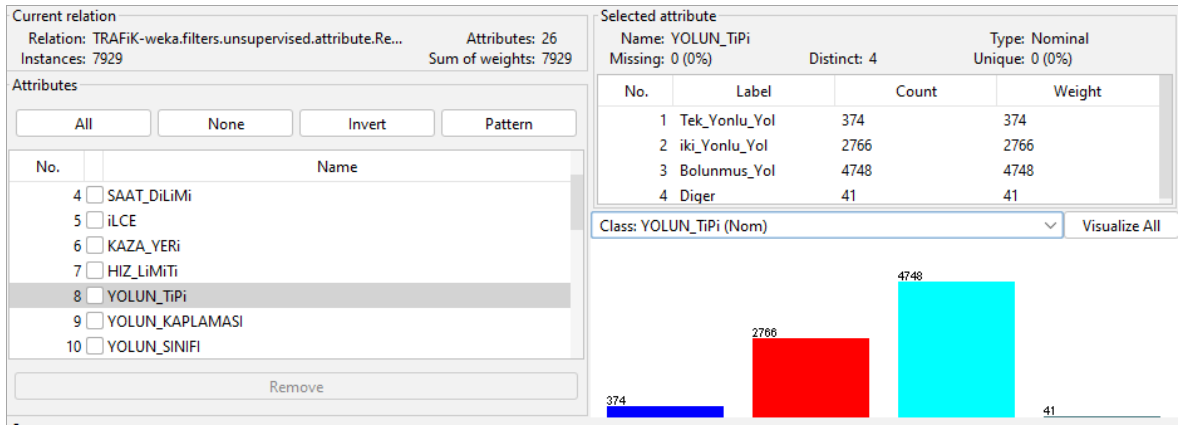


Şekil 3.21. Kaza türüne göre yaşanan trafik kazalarının önışlem verileri

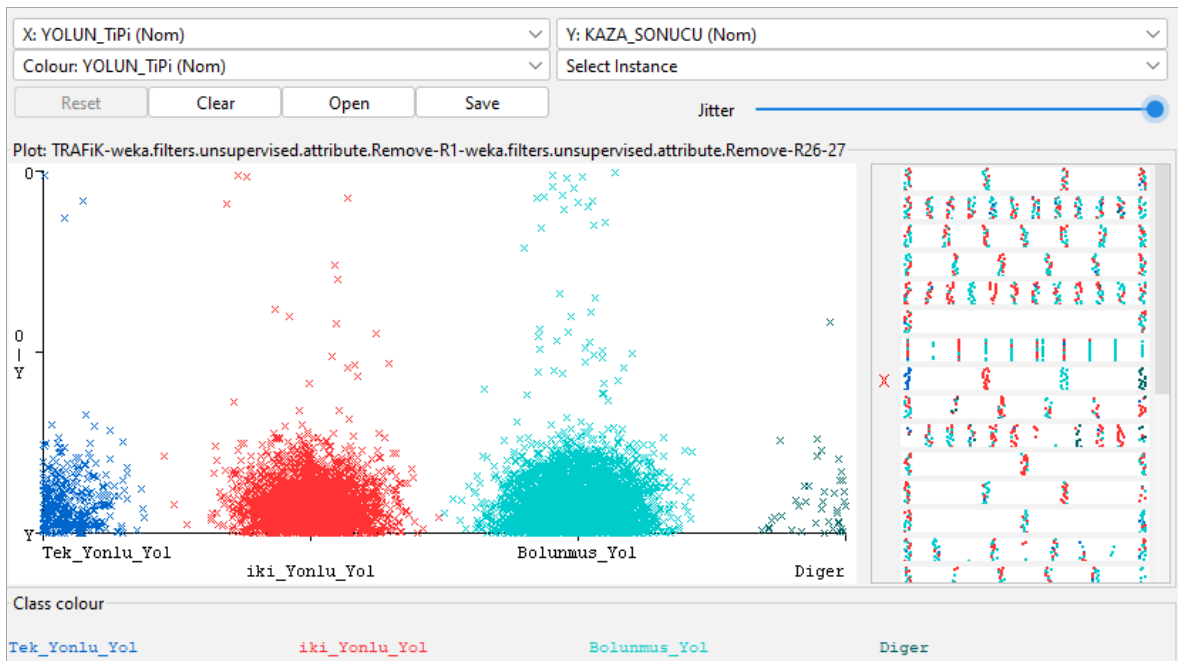


Şekil 3.22. Kaza türü ile kaza sonuçlarının görsel karşılaştırma sonuçları

Yolun tipi niteliğine göre trafik kazaları incelendiğinde; 4748 adet kaza ile bölünmüş yollar en çok kaza yaşanan nitelik olmuştur. Yaşanan kazalar “Kazanın Sonucu” ve “Yolun Tipi” niteliklerine göre görsel olarak karşılaştırıldığında; bölünmüş ve iki yönlü yollar, yaralanmalı, ölümlü/yaralanmalı ve ölümlü olmak üzere kaza sonucuna ait 3 sınıfta da en çok kaza yaşanan yol tipleri olmuştur. Yolun tipi niteliğine ait önişlem ve görsel karşılaştırma verileri Şekil 3.23 ve 3.24’te gösterilmektedir.



Şekil 3.23. Yolun tipine göre yaşanan trafik kazalarının önişlem verileri



Şekil 3.24. Yolun tipi ve kaza sonuçları arasındaki görsel karşılaştırma sonuçları

3.2. Sınıflandırma Yöntemi Bulguları

Sınıflandırma yöntemiyle yaşanan gerçek trafik kazası sonuçları WEKA programında 10 kat çapraz doğrulama yöntemi kullanılarak NaiveBayes, Destek Vektör Makinesi, K En Yakın Komşular, PART, AdaboostM1, J48 ve Rastgele Orman sınıflandırma algoritmalarının tahmin sonuçları ve performans değerleri verilmektedir. Algoritmaların başarıları Karmaşıklık Matrisi, Kappa İstatistiği (K_i), Ortalama Mutlak Hata (MAE), Doğruluk (Başarı Oranı), Kesinlik, Duyarlılık, F-Ölçütü ve ROC Eğrisi gibi performans ölçütleri ile belirlendi. Bu bağlamda gerçek trafik kazası verilerine göre algoritmaların kendi tahminlerini yapacağı 6 sınıflandırma başlığı aşağıdaki Tablo 3.1’de verilmiştir.

Tablo 3.1. Sınıflandırma yönteminde kullanılan öznitelik başlıkları

S.No	Öznitelik Adı	Alt Nitelik
1	Kaza Yerleşim Yeri	İçi, Dışı
2	Yolun Tipi	Tek Yönlü Yol, İki Yönlü Yol, Bölünmüş Yol, Diğer
3	Yolun Yatay Geometrisi	Düz Yol, Viraj, Tehlikeli Viraj
4	Gün Durumu	Gündüz, Gece, Alacakaranlık
5	Araç Sayısı	Tek Araç, İki Araç, Çok Araçlı
6	Kazanın Sonucu	Ölümlü, Yaralanmalı, Ölümlü/Yaralanmalı

3.2.1. Yerleşim Yeri Sınıfına Göre Algoritmaların Tahminleri

WEKA’da 10 kat çapraz doğrulama yöntemi ile NB, DVM, KNN, AdaBoostM1, PART, J48 ve RF sınıflandırma algoritmalarının Yerleşim yeri içinde ve dışında meydana gelen gerçek trafik kazası verilerine karşı algoritmaların tahminleri ve elde edilen karşılaştırmalı performans ölçütleri (Karmaşıklık Matrisi, Duyarlılık, Kesinlik, F-ölçütü, ROC Oranı, K_i , MAE ve Başarı Oranı) Tablo 3.2 ve 3.3’de gösterilmektedir.

Tablo 3.2. Yerleşim yeri sınıfına göre karmaşıklık matrisleri

=== Karmaşıklık Matrisleri ===											
a	b		a	b		a	b		a	b	
6339	273	a	6484	128	a	6485	127	a	6528	84	a
85	1232	b	111	1206	b	199	1118	b	220	1097	b
NB			DVM			KNN			AdaBoostM1		
a	b		a	b		a	b				
6496	116	a	6516	96	a	6529	83	a			
125	1192	b	113	1204	b	127	1190	b			
PART			J48			RF					

Tablo 3.3. Yerleşim yeri sınıfına göre performans ölçütleri

Algoritma	Kesinlik Precision	Duyarlılık Recall	F-Ölçütü F-Measure	ROC Oranı	K _i	MAE	Başarı Oranı(%)
J48	0,974	0,974	0,974	0,974	0,904	0,034	97,36
RF	0,973	0,974	0,973	0,994	0,903	0,063	97,35
DVM	0,970	0,970	0,970	0,948	0,891	0,030	96,98
PART	0,970	0,970	0,970	0,976	0,890	0,037	96,96
AdaBoost	0,961	0,962	0,961	0,989	0,855	0,051	96,16
KNN (k=10)	0,958	0,959	0,959	0,987	0,848	0,063	95,88
NB	0,959	0,955	0,956	0,989	0,845	0,048	95,48

Yerleşim yeri içinde ve dışında meydana gelen gerçek trafik kazası verilerine karşı sınıflandırma algoritmalarının yaptığı tahminler yani Tablo 3.2 ve Tablo 3.3'deki performans ölçütleri analiz edildiğinde; tüm algoritmaların başarı düzeyinin çok iyi seviyede ve aralarına çok küçük farklar olduğu görülmektedir. MAE oranına göre en düşük hata ile DVM algoritması en başarılı sonucu verirken, NB Karmaşıklık matrisindeki TP ve FP oranlarını karşılaştıran ROC eğrisine göre yaklaşık %99 oranında bir başarı sağlasa da genel başarıya göre en sonda yer almıştır. Performans ölçütlerinin tamamı göz önüne alındığında, %97,36 başarı puanı ile gerçek verilere en yakın tahminleri yapan J48 algoritmasının ilk sırayı aldığı ve onu %97,35 oranında sadece onbinde bir puan farkla RF algoritmasının izlediği görülmektedir.

3.2.2. Yolun Tipi (Yönü) Sınıfına Göre Algoritmaların Tahminleri

WEKA’da 10 kat çapraz doğrulama yöntemi ile NB, DVM, KNN, AdaBoostM1, PART, J48 ve RF sınıflandırma algoritmalarının yolun yönüne bağlı (Tek Yönlü Yol, İki Yönlü Yol, Bölünmüş Yol ve Diğer) meydana gelen gerçek trafik kazası verilerine karşı algoritmaların tahminleri ve elde edilen karşılaştırmalı performans ölçütleri Tablo 3.4 ve 3.5’te gösterilmektedir.

Tablo 3.4. Yolun yönü sınıfına göre karmaşıklık matrisleri

=== Karmaşıklık Matrisleri ===									
a	b	c	d		a	b	c	d	
6	231	135	2	a	1	158	215	0	a
13	2027	707	19	b	2	1791	969	4	b
22	1152	3573	1	c	1	583	4161	3	c
1	16	8	16	d	2	17	8	14	d
NB					DVM				
0	158	216	0	a	0	154	220	0	a
3	1691	1072	0	b	0	1362	1404	0	b
2	600	4146	0	c	0	725	4023	0	c
0	30	11	0	d	0	34	7	0	d
KNN					AdaboostM1				
12	186	176	0	a	1	199	174	0	a
56	1946	758	6	b	26	2126	609	5	b
52	731	3964	1	c	23	713	4011	1	c
2	17	10	12	d	0	14	9	18	d
PART					J48				
	a	b	c	d		a	b	c	d
	1	161	212	0		a			
	3	1905	857	1		b			
	3	576	4159	0		c			
	0	27	10	4		d			
RF									

Tablo 3.5. Yolun yönü sınıfına göre performans ölçütleri

Algoritma	Kesinlik Precision	Duyarlılık Recall	F-Ölçütü F-Measure	ROC Ort.	K _i	MAE	Başarı Oranı(%)
J48	0,748	0,776	0,761	0,835	0,555	0,156	77,63
RF	0,736	0,767	0,745	0,857	0,520	0,181	76,66
DVM	0,722	0,753	0,731	0,742	0,488	0,275	75,25
PART	0,727	0,748	0,736	0,815	0,498	0,155	74,83
KNN (k=10)	-	0,736	-	0,820	0,450	0,186	73,61
NB	0,699	0,709	0,698	0,787	0,435	0,185	70,90
AdaBoost	-	0,679	-	0,661	0,321	0,321	67,91

Yolun yönüyle ilişkili meydana gelen gerçek trafik kazası verilerine karşı sınıflandırma algoritmalarının yaptığı tahminler Tablo 3.4 ve Tablo 3.5’deki performans ölçütleri doğrultusunda analiz edildiğinde; en başarısız %67,91 başarı oranı ile AdaBoostM1 algoritması olurken, %77,63 başarı puanı ile gerçek verilere en yakın tahminleri yapan J48 algoritması ilk sırayı alırken onu %76,66 oranında RF ve %75,25 oranında DVM algoritmalarının izlediği görülmektedir.

3.2.3. Yolun Yatay Geometrisi Sınıfına Göre Algoritmaların Tahminleri

WEKA’da 10 kat çapraz doğrulama yöntemi ile NB, DVM, KNN, AdaBoostM1, PART, J48 ve RF sınıflandırma algoritmalarının Yolun yatay geometrisine bağlı (Düz Yol, Viraj ve Tehlikeli Viraj) meydana gelen gerçek trafik kazası verilerine karşı algoritmaların tahminleri sonucu elde edilen karşılaştırmalı performans ölçütleri Tablo 3.6 ve 3.7’de gösterilmektedir.

Tablo 3.6. Yolun yatay geometrisi sınıfına göre karmaşıklık matrisleri

=== Karmaşıklık Matrisleri ===											
a	b	c		a	b	c		a	b	c	
6422	297	356	a	7066	7	2	a	7065	0	10	a
29	36	30	b	84	10	1	b	89	1	5	b
498	84	177	c	754	4	1	c	732	2	25	c
NB				DVM				KNN			

a	b	c		a	b	c		a	b	c	
7075	0	0	a	6807	35	233	a	7012	6	57	a
95	0	0	b	62	11	22	b	78	8	9	b
759	0	0	c	601	11	142	c	695	5	59	c
AdaBoostM1				PART				J48			
a	b	c		a	b	c		a	b	c	
7055	0	20	a	7055	0	20	a	7055	0	20	a
94	0	1	b	94	0	1	b	94	0	1	b
732	0	27	c	732	0	27	c	732	0	27	c
				RF							

Tablo 3.7. Yolun yatay geometrisi sınıfına göre performans ölçütleri

Algoritma	Kesinlik Precision	Duyarlılık Recall	F-Ölçütü F-Measure	ROC Ort.	K _i	MAE	Başarı Oranı(%)
KNN (k=10)	0,833	0,894	0,849	0,776	0,056	0,100	89,43
RF	-	0,893	-	0,818	0,050	0,111	89,31
J48	0,854	0,893	0,857	0,715	0,124	0,112	89,27
DVM	0,827	0,893	0,844	0,511	0,027	0,248	89,25
AdaBoost	-	0,892	-	0,799	0,000	0,132	89,22
PART	0,849	0,878	0,860	0,760	0,208	0,106	87,77
NB	0,856	0,837	0,845	0,797	0,224	0,129	83,68

Yolun yatay geometrisine ilişkin düz yol, viraj ve tehlikeli virajlı yollarda meydana gelen gerçek trafik kazası verilerine karşı sınıflandırma algoritmalarının yaptığı tahminler Tablo 3.6 ve Tablo 3.7’deki performans ölçütleri doğrultusunda analiz edildiğinde; KNN algoritması en yakın 10 komşu değerine göre %89,43 oranla en başarılı sonucu verdiği onu sırası ile %89,31 oranla RF ve %89,27 oranla J48 algoritmasının izlediği görülmektedir. Diğer algoritmalara kıyasla en başarısız tahminleri ise %83,68 oranla NB algoritmasının yaptığı görülmektedir. Genel olarak bakıldığında algoritmaların başarı oranları yüksek olsa da karmaşıklık matrisindeki tahminlere göre gözlemlenen değer ile beklenen değer arasındaki farkı gösteren kappa istatistiği oranlarının sıfır veya sıfıra yakın olması başarı oranlarının pek tutarlı olmadığını göstermektedir. Algoritmaların başarısında kappa istatistiğinin de bire yakın olması istenir. Çünkü Bir algoritmanın başarısı sadece başarı puanına bağlı değil Tablo 3.7’de gösterilen diğer performans ölçütleriyle de paralel olmalıdır.

3.2.4. Gün Durumu Sınıfına Göre Algoritmaların Tahminleri

WEKA’da 10 kat çapraz doğrulama yöntemi ile NB, DVM, KNN, AdaBoostM1, PART, J48 ve RF sınıflandırma algoritmalarının günün durumuna bağlı (Gündüz, Gece, Alacakaranlık) meydana gelen gerçek trafik kazası verilerine karşı algoritmaların tahminleri sonucu elde edilen karşılaştırmalı performans ölçütleri (Karmaşıklık Matrisi, Duyarlılık, Kesinlik, F-ölçütü, ROC Oranı, K_i , MAE ve Başarı Oranı) Tablo 3.8 ve 3.9’da gösterilmektedir.

Tablo 3.8. Gün durumu sınıfına göre karmaşıklık matrisleri

=== Karmaşıklık Matrisleri ===											
a	b	c		a	b	c		a	b	c	
5256	155	105	a	5272	244	0	a	5442	74	0	a
488	1697	55	b	389	1851	0	b	1279	961	0	b
133	30	10	c	123	50	0	c	185	15	0	c
NB				DVM				KNN			
a	b	c		a	b	c		a	b	c	
5474	42	0	a	5163	316	37	a	5277	233	6	a
1031	1209	0	b	366	1845	29	b	366	1866	8	b
159	14	0	c	107	61	5	c	121	51	1	c
AdaBoostM1				PART				J48			
a	b	c		a	b	c					
5391	123	2	a								
525	1712	3	b								
146	26	1	c								
				RF							

Tablo 3.9. Gün durumu sınıfına göre performans ölçütleri

Algoritma	Kesinlik Precision	Duyarlılık Recall	F-Ölçütü F-Measure	ROC Ort.	K_i	MAE	Başarı Oranı(%)
J48	0,884	0,901	0,891	0,945	0,762	0,091	90,09
DVM	-	0,899	-	0,872	0,756	0,249	89,86
RF	0,882	0,896	0,884	0,951	0,741	0,154	89,59
PART	0,873	0,884	0,879	0,934	0,728	0,092	88,44
NB	0,878	0,878	0,876	0,952	0,707	0,108	87,81
AdaBoost	-	0,843	-	0,931	0,575	0,182	84,28
KNN (k=10)	-	0,808	-	0,885	0,464	0,199	80,75

Gün durumuna göre gündüz, gece, alacakaranlıkta meydana gelen gerçek trafik kazası verilerine karşı sınıflandırma algoritmalarının yaptığı tahminler Tablo 3.8 ve Tablo 3.9'daki performans ölçütleri doğrultusunda analiz edildiğinde; J48 algoritmasının %90,09 oranla ilk sırayı aldığı görülürken onu %89,86 başarı oranıyla DVM ve % 89,59 başarı oranı ile RF algoritmalarının takip ettiği görülmektedir. Diğer algoritmalara kıyasla en başarısız tahminleri %80,75 oranla KNN (en yakın 10 komşu değerine göre) algoritması ve %84,28 oranla AdaBoostM1 algoritmasının yaptığı görülmektedir.

3.2.5. Kazaya Karışan Araç Sayısına Göre Algoritmaların Tahminleri

WEKA'da 10 kat çapraz doğrulama yöntemi ile NB, DVM, KNN, AdaBoostM1, PART, J48 ve RF sınıflandırma algoritmalarının kazaya karışan araç sayısına bağlı (Tek araç, İki araç, Çok araç) meydana gelen gerçek trafik kazası verilerine karşı algoritmaların tahminleri sonucu elde edilen karşılaştırmalı performans Tablo 3.10 ve 3.11'de gösterilmektedir.

Tablo 3.10. Kazaya karışan araç sayısına göre karmaşıklık matrisi

=== Karmaşıklık Matrisleri ===											
a	b	c		a	b	c		a	b	c	
3358	7	17	a	3379	3	0	a	2929	453	0	a
146	3874	94	b	146	3967	1	b	271	3839	4	b
27	370	36	c	23	383	27	c	38	395	0	c
NB				DVM				KNN			
a	b	c		a	b	c		a	b	c	
3382	0	0	a	3365	14	3	a	3379	3	0	a
1385	2729	0	b	151	3871	92	b	147	3965	2	b
262	171	0	c	23	364	46	c	23	383	27	c
AdaBoostM1				PART				J48			
	a	b	c		a	b	c		a	b	c
	3375	5	2		3375	5	2		3375	5	2
	148	3956	10		148	3956	10		148	3956	10
	24	383	26		24	383	26		24	383	26
				RF							

Tablo 3.11. Kazaya karışan araç sayısına göre performans ölçütleri

Algoritma	Kesinlik Precision	Duyarlılık Recall	F-Ölçütü F-Measure	ROC Ort.	K_i	MAE	Başarı Oranı(%)
DVM	0,916	0,928	0,907	0,950	0,862	0,142	92,98
J48	0,930	0,930	0,908	0,949	0,865	0,084	92,96
RF	0,916	0,928	0,907	0,950	0,862	0,142	92,78
PART	0,896	0,918	0,904	0,943	0,846	0,084	91,84
NB	0,892	0,917	0,902	0,950	0,843	0,091	91,66
KNN (k=10)	0,811	0,854	0,830	0,922	0,717	0,205	85,35
AdaBoost	-	0,771	-	0,860	0,572	0,253	77,07

Kazaya karışan araç sayısına bağlı olarak meydana gelen gerçek trafik kazası verilerine karşı sınıflandırma algoritmalarının yaptığı tahminler Tablo 3.10 ve Tablo 3.11'deki performans ölçütleri doğrultusunda analiz edildiğinde; DVM algoritmasının %92,98 oranla ilk sırayı aldığı görülürken onu %92,96 başarı oranıyla J48 ve % 92,78 başarı oranı ile RF algoritmalarının takip ettiği görülmektedir. Genel olarak en başarısız tahmini %77,07 oranla AdaBoostM1 algoritmasının yaptığı görülmektedir.

3.2.6. Kaza Sonucu Sınıfına Göre Algoritmaların Tahminleri

WEKA'da 10 kat çapraz doğrulama yöntemi ile NB, DVM, KNN, AdaBoostM1, PART, J48 ve RF sınıflandırma algoritmalarının Kazaların sonucuna bağlı (Ölümlü, Yaralanmalı, Ölümlü/Yaralanmalı) meydana gelen gerçek trafik kazası verilerine karşı algoritmaların tahminleri sonucu elde edilen karşılaştırmalı performans ölçütleri (Karmaşıklık Matrisi, Duyarlılık, Kesinlik, F-ölçütü, ROC Oranı, Kappa İstatistiği, Ortalama Mutlak Hata ve Başarı Oranı) Tablo 3.12 ve 3.13'te gösterilmektedir.

Tablo 3.12. Kaza sonucuna göre karmaşıklık matrisi

=== Karmaşıklık Matrisleri ===											
a	b	c		a	b	c		a	b	c	
7559	306	0	a	7864	1	0	a	7865	0	0	a
8	32	0	b	34	5	1	b	24	0	0	b
7	4	13	c	24	0	0	c	40	0	0	c
NB				DVM				KNN			
a	b	c		a	b	c		a	b	c	
7865	0	0	a	7865	0	0	a	7865	0	0	a
0	40	0	b	0	40	0	b	0	40	0	b
0	0	24	c	0	0	24	c	0	0	24	c
AdaBoostM1				PART				J48			
a	b	c		a	b	c		a	b	c	
7865	0	0	a	7865	0	0	a	7865	0	0	a
0	40	0	b	0	40	0	b	0	40	0	b
0	0	24	c	0	0	24	c	0	0	24	c
RF											
a	b	c		a	b	c		a	b	c	
7865	0	0	a	7865	0	0	a	7865	0	0	a
24	16	0	b	24	16	0	b	24	16	0	b
0	1	23	c	0	1	23	c	0	1	23	c

Tablo 3.13. Kaza sonucuna göre performans ölçütleri

Algoritma	Kesinlik Precision	Duyarlılık Recall	F-Ölçütü F-Measure	ROC Ort.	K _i	MAE	Başarı Oranı(%)
J48	1,000	1,000	1,000	1,000	1	0	100
PART	1,000	1,000	1,000	1,000	1	0	100
AdaBoost	1,000	1,000	1,000	1,000	1	0,0002	100
RF	0,997	0,997	0,996	1,000	0,758	0,004	99,68
DVM	0,989	0,992	0,989	0,555	0,153	0,224	99,24
KNN (k=10)	-	0,992	-	0,721	0	0,008	99,19
NB	0,993	0,959	0,974	0,964	0,215	0,030	95,90

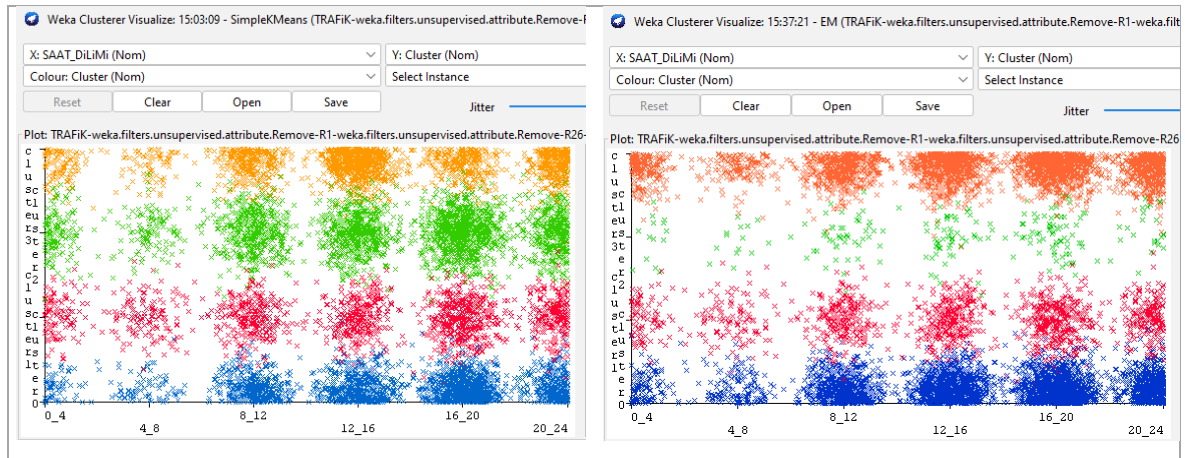
Ölümlü, yaralanmalı ve hem ölümlü hem de yaralanmalı meydana gelen gerçek trafik kazası verilerine karşı sınıflandırma algoritmalarının yaptığı tahminler Tablo 3.12 ve Tablo 3.13'teki performans ölçütleri doğrultusunda analiz edildiğinde; J48, PART ve AdaBoostM1 tüm verileri doğru sınıflayarak %100 başarı elde ettikleri görülmektedir. NB algoritması %95,90 oranla kötü bir sonuç almasa da diğerlerine oranla geride kaldığı görülmektedir. Öte yandan trafik kazası sonuçlarına göre veri setindeki 7929 kazadan 7865'nin yaralanmalı olması ve sadece yaralanmalı sonuçlanan trafik kazalarını doğru tahmin eden algoritmanın %99'un üzerinde bir başarı elde ettiği göz önüne alındığında DVM ve KNN algoritmalarının da başarılarının buna bağlı olduğu görülmektedir. Ancak

kappa istatistiği değerlerinin sıfır veya sıfıra yakın olması nedeniyle genel bir başarı sağladıkları söylenemez.

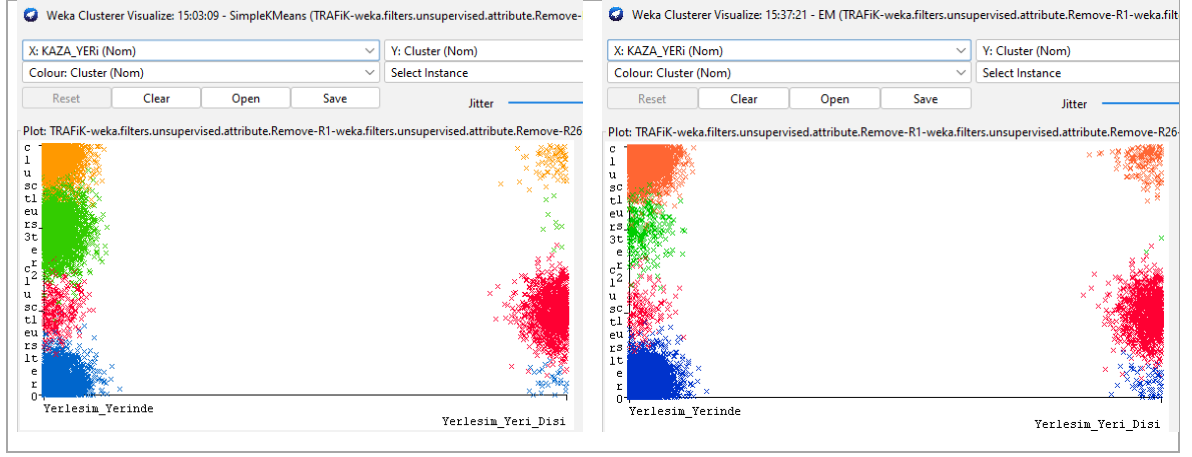
3.3. Kümeleme Yöntemi Bulguları

Çalışmanın bu bölümünde, SimpleKMeans ve EM kümeleme algoritmaları ile veri集中的 trafik kazaları saat dilimi, kaza yerleşim yeri, yolun tipi, gün durumu, araç sayısı, kaza türü ve kazanın sonucu niteliklerine göre yaşanması muhtemel kazaların öngörülerini ve benzerliklerini ortaya çıkarmak için belli sayıda kümeler oluşturularak çıkan sonuçlar değerlendirilmektedir.

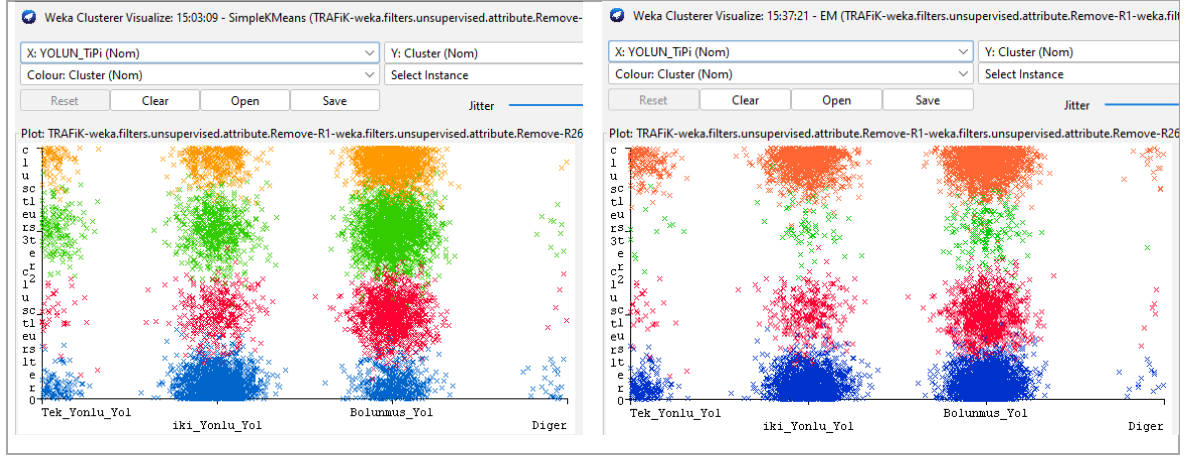
WEKA’da yapılan Kümeleme analizinde dört farklı küme oluşturulmuş ve elde edilen küme dağılımları aşağıdaki gibi oluşmuştur; 1. Kümede SimpleKMeans 2132 (%27), EM 3533 (%45) veriden oluşan bir küme oluşturmuştur. 2. Kümede SimpleKMeans 1435 (%18), EM 1241 (%16) veriden oluşan bir küme oluşturmuştur. 3. Kümede SimpleKMeans 2525 (%32), EM 228 (%3) veriden oluşan bir küme oluşturmuştur. 4. Kümede SimpleKMeans 1827 (%23), EM 2927 (%37) veriden oluşan bir küme oluşturmuştur. Bu dört küme genel olarak değerlendirildiğinde kazaların saat 16.00 ile 20.00 arasında, yerleşim yeri içinde, bölünmüş veya iki yönlü yollarda iki aracın yandan çarpıştığı ve yaralanmalı olarak sonuçlandığı görülmektedir. Dört kümenin saat dilimi, kaza yerleşim yeri, yolun tipi, gün durumu, araç sayısı, kaza türü ve kazanın sonucu niteliklerine göre dağılımları Şekil 3.25, Şekil 3.26, Şekil 3.27, Şekil 3.28, Şekil 3.29, Şekil 3.30 ve Şekil 3.31’de gösterilmektedir.



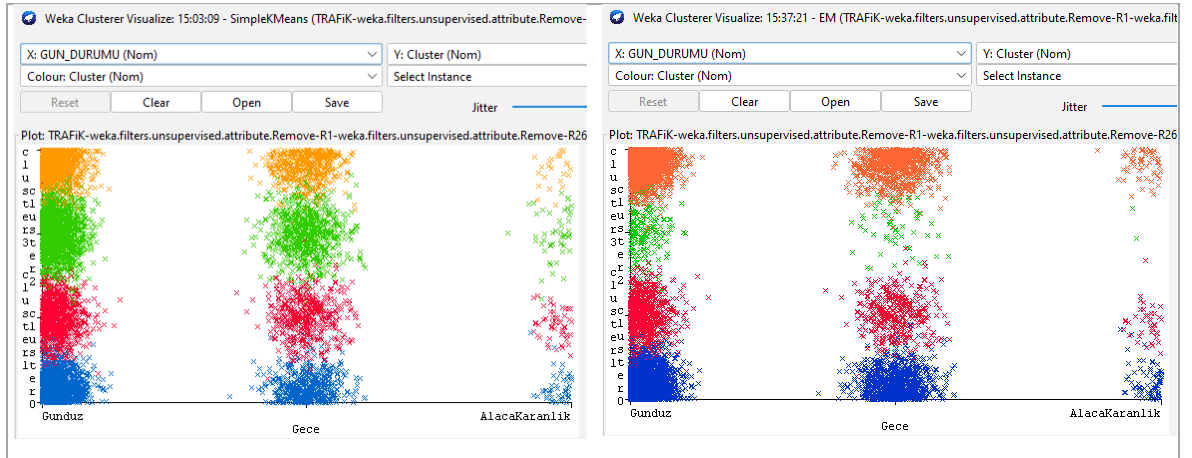
Şekil 3.25. Saat dilimi niteliğine göre SimpleKMeans ve EM kümeleme sonuçları



Şekil 3.26. Kaza yeri niteliğine göre SimpleKMeans ve EM kümeleme sonuçları



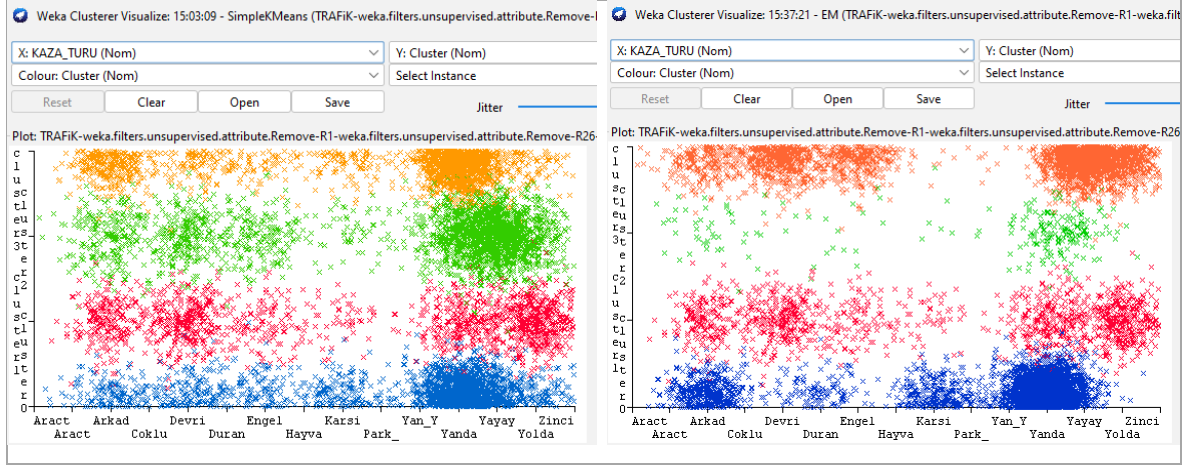
Şekil 3.27. Yolum tipi niteliğine göre SimpleKMeans ve EM kümeleme sonuçları



Şekil 3.28. Gün durumu niteliğine göre SimpleKMeans ve EM kümeleme sonuçları



Şekil 3.29. Araç sayısı niteliğine göre SimpleKMeans ve EM kümeleme sonuçları



Şekil 3.30. Yolun tipi niteliğine göre SimpleKMeans ve EM kümeleme sonuçları



Şekil 3.31. Kaza sonucu niteliğine göre SimpleKMeans ve EM kümeleme sonuçları

3.4. Birliktelik Yöntemi Bulguları

Bu bölümde WEKA’da Associate menüsünde yer alan birliktelik kuralı Apriori algoritması kullanılarak, trafik kazalarının birbiri ile olan ilişkileri incelenmektedir. Apriori Algoritmasında çıkan kural sayıları belirli bir destek (support), güven (confidence) ve lift oranına bağlıdır. Buradaki lift değeri sonucun teorik olarak ne kadar ilginç olduğunu gösterir. Mevcut 7929 trafik kazası verisi Apriori algoritmasında minimum destek değeri 0.8 minimum güven değeri 0.9 olarak belirlendiğinde 493 tane kural oluşurken, destek değeri 0.6 minimum güven değeri 0.8 olarak değiştirildiğinde 18.157 kural oluştuğu görülmektedir. Bu çalışmada daha fazla kural oluşturmak ve değerlendirmek adına minimum destek değeri 0.4 minimum güven değeri 0.7 olarak belirlendiğinde 291.815 adet birliktelik kuralı oluşurken, lift değerinin en çok 1,43 olduğu ve 10 adet büyük öge kümesi oluştuğu görülmüştür.

Oluşan ilk kural ve son kural;

1. KAZA_YERi=Yerlesim_Yerinde ARAC_SAYISI=iki_Arac CARPMA_YERi=Yol_Uzerinde 3486
==> KAZA_SONUCU=Yaralanmali 3485 <conf:(1)> lift:(1.01) lev:(0) [27]
conv:(14.07)

291815. EMNiyET_SERiDi_BANKET=Emniyet_Seridi_Banket_Yok 4600 ==> ATAY_GEO=Duz_Yol
HAVA_DURUMU=Acik KAZA_YOL_KUSURU=Yol_Sorunu_Yok TRAFiK_LAMBASI=Trafik_Lambasi_Yok
YOLDA_CALISMA=Yolda_Calisma_Yok KAZA_SONUCU=Yaralanmali 3220 <conf:(0.7)>
lift:(1.06) lev:(0.02) [184] conv:(1.13)

Oluşturulan birliktelik kurallarının bir bölümü aşağıdaki Tablo 3.22’de verilmiştir. Tablodaki güven değeri kuralın birliktelik oranını gösterirken, lift değeri kuralın ilginç bir sonuç olup olmadığını gösterir.

Tablo 3.14. Apriori algoritması ile oluşturulan kurallar

Kurallar	Güven (Confidence)	Lift
1 YOLUN_SINIFI=Cadde DUSEY_GEO=Egimsiz HAVA_DURUMU=Acik KAZA_YOL_KUSURU=Yol_Sorunu_Yok KAZA_SONUCU=Yaralanmali 3437 ==> KAZA_YERi=Yerlesim_Yerinde 3435	100%	1,20
2 KAZA_YERi=Yerlesim_Yerinde YATAY_GEO=Duz_Yol ARAC_SAYISI=iki_Arac CARPMA_YERi=Yol_Uzerinde KAZA_YOL_KUSURU=Yol_Sorunu_Yok 3258 ==> KAZA_SONUCU=Yaralanmali 3257	100%	1,01
3 YOLUN_KAPLAMASI=Asfalt GUN_DURUMU=Gunduz HAVA_DURUMU=Acik YOL_SERiDi=Yol_Seridi_Var 3271 ==> KAZA_SONUCU=Yaralanmali 3249	99%	1,00

Tablo 3.14'ün Devamı.

4	YOLUN_YUZEYi=Kuru TRAFİK_LEVHASI=Trafik Levhaları Yok TRAFİK_LAMBASI=Trafik Lambası Yok 3599 ==> HAVA_DURUMU=Acik YOLDA_CALISMA=Yolda Calisma Yok KAZA_SONUCU=Yaralanmali 3507	97%	1,15
5	GUN_DURUMU=Gunduz HAVA_DURUMU=Acik KAZA_SONUCU=Yaralanmali 4849 ==> YOLUN_KAPLAMASI=Asfalt 4677	96%	1,00
6	YOLUN_SINIFI=Cadde DUSEY_GEO=Egimsiz YOLUN_YUZEYi=Kuru KAZA_YOL_KUSURU=Yol_Sorunu_Yok KAZA_SONUCU=Yaralanmali 3363 ==> KAZA_YERi=Yerlesim_Yerinde YATAY_GEO=Duz_Yol HAVA_DURUMU=Acik 3179	95%	1,41
7	YOLUN_YUZEYi=Kuru YAYA_YOLU=Yayayolu_Var 3547 ==> KAZA_YERi=Yerlesim_Yerinde YOLUN_KAPLAMASI=Asfalt HAVA_DURUMU=Acik KAZA_YOL_KUSURU=Yol_Sorunu_Yok YOLDA_CALISMA=Yolda Calisma Yok 3349	94%	1,34
8	YOLUN_KAPLAMASI=Asfalt CARPMA_YERi=Yol_Uzerinde 6400 ==> KAZA_YERi=Yerlesim_Yerinde KAZA_SONUCU=Yaralanmali 5672	89%	1,07
9	YOLUN_KAPLAMASI=Asfalt YATAY_GEO=Duz_Yol 6845 ==> CARPMA_YERi=Yol_Uzerinde KAZA_YOL_KUSURU=Yol_Sorunu_Yok KAZA_SONUCU=Yaralanmali 5787	85%	1,03
10	YOLUN_SINIFI=Cadde YATAY_GEO=Duz_Yol YOLUN_YUZEYi=Kuru 3907 ==> KAZA_YERi=Yerlesim_Yerinde DUSEY_GEO=Egimsiz HAVA_DURUMU=Acik KAZA_YOL_KUSURU=Yol_Sorunu_Yok KAZA_SONUCU=Yaralanmali 3179	81%	1,43
11	KAZA_YERi=Yerlesim_Yerinde GUN_DURUMU=Gunduz KAZA_SONUCU=Yaralanmali 4632 ==> YOLUN_KAPLAMASI=Asfalt CARPMA_YERi=Yol_Uzerinde TRAFİK_LAMBASI=Trafik Lambası Yok 3554	77%	1,11
12	YATAY_GEO=Duz_Yol CARPMA_YERi=Yol_Uzerinde 6040 ==> KAZA_YERi=Yerlesim_Yerinde YOLUN_KAPLAMASI=Asfalt HAVA_DURUMU=Acik YOLUN_YUZEYi=Kuru KAZA_SONUCU=Yaralanmali 4487	74%	1,08
13	KAZA_YERi=Yerlesim_Yerinde YOLUN_KAPLAMASI=Asfalt YATAY_GEO=Duz_Yol HAVA_DURUMU=Acik KAZA_YOL_KUSURU=Yol_Sorunu_Yok TRAFİK_LAMBASI=Trafik Lambası Yok YOLDA_CALISMA=Yolda Calisma Yok 4372 ==> YOLUN_SINIFI=Cadde YOLUN_YUZEYi=Kuru 3190	73%	1,34
14	YOLUN_KAPLAMASI=Asfalt YATAY_GEO=Duz_Yol TRAFİK_LAMBASI=Trafik Lambası Yok 5924 ==> GUN_DURUMU=Gunduz 4199	71%	1,02
15	KAZA_YERi=Yerlesim_Yerinde YOLUN_YUZEYi=Kuru CARPMA_YERi=Yol_Uzerinde 5132 ==> YOLUN_SINIFI=Cadde YATAY_GEO=Duz_Yol KAZA_SONUCU=Yaralanmali 3596	70%	1,24

1. kuralda, yol sorunu olmayan, hava durumu açık ve eğimsiz caddelerde meydana gelen trafik kazaları %100 yerleşim yeri içerisinde meydana gelmiştir. Apriori algoritması lift değerine göre 1,2 oranında ilginç olarak değerlendirmiştir.

2. kuralda, yerleşim yeri içinde, yol sorunu olmayan, düz yollarda ve yol üzerinde iki aracın karıştığı trafik kazaları %100 yaralanmalı sonuçlanmıştır. Apriori algoritması lift değerine göre 1,01 oranında beklenen bir durumdur.

3. kuralda, asfalt yollarda, yol şeridi olan, hava durumunun açık ve gündüz yaşanan trafik kazaları %99 yaralanmalı sonuçlanmıştır. Apriori algoritması lift değerine göre 1,00 oranında beklenen bir durumdur.

4. kuralda, yol yüzeyini kuru olduğu, trafik levhası ve lambasının olmadığı yollarda meydana gelen trafik kazaları %97 yaralanmalı sonuçlanmıştır. Apriori algoritması lift değerine göre 1,15 oranında ilginç olarak değerlendirmiştir.

5. kuralda, hava durumunun açık, gündüz ve yaralanmalı sonuçlanan trafik kazaları %96 asfalt yollarda yaşanmıştır. Apriori algoritması lift değerine göre 1,00 oranında beklenen bir durumdur.

6. kuralda, yol sorunu olmayan, eğimsiz yol yüzeyinin kuru olduğu yollarda, yaralanmalı olarak sonuçlanan kazalar, %95 hava durumunun açık, düz yollarda ve yerleşim yeri içerisinde meydana gelmiştir. Apriori algoritması lift değerine göre 1,41 oranında ilginç olarak değerlendirmiştir.

7. kuralda, yaya yolu olan ve yüzeyi kuru olan yollarda meydana gelen kazalar, %94 yerleşim yeri içinde, yol sorunu ve çalışması olmayan asfalt yollarda ve hava durumu açıkken meydana gelmiştir. Apriori algoritması lift değerine göre 1,34 oranında ilginç olarak değerlendirmiştir.

8. kuralda, asfalt ve yol üzerinde yaşanan trafik kazaları %89 yaralanmalı sonuçlanmıştır. Apriori algoritması lift değerine göre 1,07 oranında beklenen bir durumdur.

9. kuralda, asfalt ve düz yollarda meydana gelen trafik kazaları %85 yol üzerinde, sorunu olmayan yollarda ve yaralanmalı sonuçlanmıştır. Apriori algoritması lift değerine göre 1,03 oranında beklenen bir durumdur.

10. kuralda, düz ve yol yüzeyi kuru caddelerde yaşanan kazalar, %81 yerleşim yeri içinde, hava durumu açık, sorunu olamaya eğimsiz yollarda ve yaralanmalı sonuçlanmıştır. Apriori algoritması lift değerine göre 1,43 oranında ilginç olarak değerlendirmiştir.

11 kuralda, yerleşim yerinde, gündüz ve yaralanmalı sonuçlanan trafik kazaları, %77 trafik lambası olmayan, asfalt yolların üzerinde yaşanmıştır. Apriori algoritması lift değerine göre 1,11 oranında beklenen bir durumdur.

12. kuralda, düz yollarda yaşanan trafik kazaları, %74 yerleşim yerlerinde, asfalt ve kuru yüzeyli yollarda, yaralanmalı sonuçlanmıştır. Apriori algoritması lift değerine göre 1,08 oranında beklenen bir durumdur.

13. kuralda, yerleşim yeri içinde, düz asfalt yollarda, trafik lambası olmaya, hava durumunun açık, yol sorunu ve yol çalışması olmayan yollardaki kazalar, %73 yol yüzeyinin kuru olduğu caddelerde meydana gelmiştir. Apriori algoritması lift değerine göre 1,34 oranında ilginç olarak değerlendirmiştir.

14. kuralda, düz, asfalt ve trafik lambası olmayan yollarda yaşanan trafik kazaları, %71 gündüz yaşanmıştır. Apriori algoritması lift değerine göre 1,02 oranında beklenen bir durumdur.

15. kuralda, yerleşim yeri içinde, yol yüzeyini kuru olduğu yollarda meydana gelen kazalar, %70 düz caddelerde yaralanmalı sonuçlanmıştır. Apriori algoritması lift değerine göre 1,24 oranında ilginç olarak değerlendirmiştir.

4. SONUÇ ve ÖNERİLER

Trafik kazaları her gün binlerce insanın yaralanmasına veya hayatını kaybetmesine neden olan en büyük etmenlerdendir. Trafik kazalarının analizinde makine öğrenme teknikleri son yıllarda sıklıkla kullanılan yöntemlerdendir. Bu çalışmada, Emniyet Genel Müdürlüğü Trafik Başkanlığı'ndan alınan 2017 ile 2020 yılları arasında Kahramanmaraş il merkezi ve ilçelerinde, gerçekleşen 7929 adet trafik kazası ve 28 adet öznitelikten oluşan ölümlü ve yaralanmalı trafik kazaları WEKA aracı kullanılarak makine öğrenme teknikleri ile analiz edilmiştir. Çalışmadaki analiz işlemleri, önışlem ve görsel karşılaştırma, sınıflandırma, kümeleme ve birliktelik algoritmaları kullanılarak dört farklı yöntem ile gerçekleştirilmiş ve elde edilen sonuçlara değinilmiştir.

Önişlem ve görsel karşılaştırma sonuçlarına göre; trafik kazaların 2017 den 2020 yılına kadar azalış gösterdiği ancak ölümlü sonuçlanan kazaların az da olsa 2020 yılı sonuna kadar artış gösterdiği görülmüştür. Temmuz ayları, Cuma günleri, saat 16:00 ile 20:00 arası, sabah saat 04:00 ile 08:00 arası dikkat edilmesi gereken zamanlardır. En çok trafik kazası Kahramanmaraş il merkezinde yaşanırken ölümlü sonuçlanan kazalarda sırası ile il merkezi, Göksun, Onikişubat ve Pazarcık ilçelerinin öne çıktığı görülmektedir. Yerleşim yeri dışında meydana gelen kazalar yerleşim yeri içinde meydana gelen kazaların %20 kadar olsa da ölümlü sonuçlanan kaza sayılarına göre ölüm oranının yüksek olduğu gözlenmiştir. Kazalar daha çok gündüz vakti iki aracın bir biriyle yandan çarpıldığı, kuru zeminli, açık havalarda, cadde üzerinde, eğimsiz, düz, asfalt, trafik lambası ve levhası olmayan yollarda yaşandığı görülmüştür. Bunun yanında gece, eğimli yollarda, tek aracın yoldan çıkması veya yaya çarpması ile meydana gelen kazalarda ölümlü sonuçlanma olasılığının yüksek olduğu tespit edilmiştir.

Sınıflandırma yönteminde kullanılan NB, DVM, KNN, AdaBoostM1, PART, J48 (C4.5) ve RF algoritmaları ile trafik kazaları analiz edildiğinde; gelecekte yaşanacak kazaların; daha çok pazartesi günleri, saat 12:00 ile 16:00 arasında, yerleşim yeri içinde, bölünmüş veya iki yönlü, kuru yüzeyli asfalt caddelerde, düz ve eğimsiz yollarda, açık havalarda, yaya yolu olan, emniyet şeridi ve banket olmayan, yol şeridi olan, trafik lambası ve levhası olmayan, yol çalışması olmayan, yol üzerinde iki aracın yandan çarpıştığı ve kazaların yaralanmalı olarak sonuçlanacağı tahmin edilmektedir. Algoritmaların performans ölçütlerine bakıldığında karar ağaçlarından J48 algoritması en başarılı

sonuçları elde ederken, NB algoritmasının başarı düzeyi kötü olmasa da diğer algoritmalarla oranla daha düşük olduğu görülmüştür.

Kümeleme algoritmaları SimpleKMeans ve EM ile yaşanan trafik kazaları analizi edildiğinde; kazaların genellikle saat 16.00 ile 20.00 arasında, yerleşim yeri içinde, bölünmüş veya iki yönlü yollarda iki aracın yandan çarpıştığı ve yaralanmalı olarak sonuçlanan kazalar şeklinde kümelendiği görülmektedir.

Birliktelik kuralı algoritması Apriori ile Kahramanmaraş genelinde meydana gelen trafik kazalarının birbiriyle olan ilişkileri incelendiğinde; kazalar yerleşim yeri içinde, bölünmüş ve asfalt yollarda, düz ve eğimsiz yollarda, gündüz, açık havalarda, yol yüzeyinin kuru olduğu ve yol sorununun olmadığı yollarda, yaya yolu olan, yol çalışması olmayan, yol şeridi olan, emniyet şeridi/banket olmayan, trafik levhası ve lambasının olmadığı yollarda araçların yandan çarpışması ile yaralanmalı olarak sonuçlanan kaza unsurlarının birbiri ile daha çok ilişkilendirildiği görülmüştür.

Çalışmada kullanılan yöntemlerin ortak sonucu; hem ölümle hem de yaralanmalı sonuçlanma olasılığı yüksek ve dikkat edilmesi gereken kazaların daha çok sabah saat 04:00 ile 08:00 ve akşam saat 16:00 ile 20:00 arasında, Kahramanmaraş il merkezi ve Onikişubat ilçesinde yerleşim yeri içerisinde, Göksun ilçesindeki yerleşim yeri dışındaki eğimli ve virajlı yollarda ve Pazarcık ilçesindeki yerleşim yeri dışındaki bölünmüş yollarda yaşanma riskinin yüksek olduğu belirlenmiştir. Ayrıca meydana gelen kazalarda genel sebepleri; sürücülerin trafik ve hız kurallarına uymaması, trafik lambası, trafik levhası, yaya yolu, emniyet şeridi ve banket gibi yol, çevre ve güvenlik ekipmanlarının yetersiz olduğu belirlenmiştir.

Bu çalışmada ortaya konulan sonuçlar neticesinde gelecekte yaşanabilecek potansiyel trafik kazaları hakkında bilgiler verilerek bu kazaların önüne geçmek ve gerekli önlemleri almak için sürücü, yolcu, yaya ve devlet otoritelerine yardımcı olmak istenmektedir.

Makine öğrenme algoritmaları ile trafik kazalarının analiz edildiği literatür çalışmalarında sınıflandırma, kümeleme ve birliktelik analizi yöntemlerinin son yıllarda kullanıldığını görülmektedir. Bu bağlamda önceden yapılan çalışmalar bu tez çalışmasına referans niteliği taşımaktadır. Bu tez çalışmasında diğerlerinden farklı olarak, görsel karşılaştırma, sınıflandırma, kümeleme ve birliktelik kuralı yöntemleri bir arada kullanılarak bir birini destekleyen daha fazla ve daha farklı sonuçlar elde edildiği gösterilmiştir.

4.1. Kısıtlama ve Zorluklar

Çalışmada karşılaşılan zorluklar veya eksiklikler, meydana gelen trafik kazası verilerinden bazı sınıfların diğerlerine oranla çok sayıda tekrar etmesidir. Örneğin ölümlü ve yaralanmalı yaşanan 7929 trafik kazasından 7865'nin yaralanmalı sonuçlanması, bir algoritmanın sadece yaralanmalı sonuçlanan kazalarını doğru tahmin ederek %99'un üzerinde başarı elde etmesine neden olmaktadır. Bu durumda sadece bir veri sınıfını doğru tahmin ederek büyük başarı elde edildiğini göstermekte ve veri setine bağlı bir eksiklik olarak değerlendirilmektedir.

4.2. Potansiyel Gelecek Çalışmalar

Daha farklı veya başarılı sonuçlar alabilmek için WEKA programına ek olarak R, Rapid Miner, Knime, Orange gibi makine öğrenmesi ve veri madenciliği alanında kullanılan programlarla daha farklı sonuçlar elde etmek mümkündür. Ayrıca farklı makine öğrenme teknikleri kullanılarak veya tüm bu önerileri harmanlayarak hibrit yapıda çalışmalar yapılabilir.

5. KAYNAKLAR

- Akomolafe, T.D., Olutayo, A.,** 2013. Using Data Mining Technique to Predict Cause of Accident and Accident Prone Locations on Highways. *American Journal of Database Theory and Application*, 1(3):26-38.
- Aksu, G., Doğan, N.,** 2019. Veri Madenciliğinde Kullanılan Güncel Bir Analiz Programı: WEKA. *Journal of Measurement and Evaluation in Education and Psychology*, 10(1):80-95.
- Aksu, M., Karaman, E.,** 2017. Karar Ağaçları ile Bir Web Sitesinde Link Analizi ve Tespiti. *Acta Infologica*, 1(2):84-91.
- Alan, M.A., Yeşilyurt, C.,** 2018. Farklı Veri Setleri Üzerinde SMO ve J48 Algoritmalarının Sınıflandırma Sonuçlarının Karşılaştırılması. *İşletme Bilimi Dergisi*, 6(3):199-213.
- 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*, Amman, Jordan, April 9-11.
- Anderson, T.K.,** 2009. Kernel density estimation and K-means clustering to profile road accident hotspots. *Accident Analysis & Prevention*, 41(3):359-364.
- Atalay, A., Tortum, A.,** 2010. Türkiye'deki İllerin 1997-2006 Yılları Arası Trafik Kazalarına Göre Kümeleme Analizi. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, 16(3):1997-2006.
- Atalay, M., Çelik, E.,** 2017. Büyük Veri Analizinde Yapay Zekâ ve Makine Öğrenmesi Uygulamaları. *Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 9(22):155-172.
- Bahiru, T.K., Kumar Singh, D., Tessfaw, E.A.,** 2018. Comparative Study on Data Mining Classification Algorithms for Predicting Road Traffic Accident Severity. In: *2018 Second International Conference on Inventive Communication and Computational Technologies*, Coimbatore, India, April 20-21.
- Balcı, M., Gölcük, A., Kahramanli, H.,** 2017. İstatistiksel Yaklaşımla Trafik Kazalarındaki Ölüm ve Yaralanma Durumlarının Kusurlu Unsurlarla İlişkilerinin İncelenmesi, *Selçuk-Teknik Dergisi*, 16(3):16.
- Baumgartner, D., Serpen, G.,** 2009. Large Experiment and Evaluation Tool for WEKA Classifiers, *DMIN*, 16:340-346.
- Bayata, H., Hattatoğlu, F.,** 2014. Erzincan İli İçin Farklı Yöntemlerle Trafik Kaza Tahmin Modellemesi. *Erzincan Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 4(1):31-46.

- Beshah, T., Hill, S.,** 2010. Mining Road Traffic Accident Data to Improve Safety: Role of Road-related Factors on Accident Severity in Ethiopia. *AAAI Spring Symposium Series*, 6.
- Bharati, S., Rahman, M.A., Podder, P.,** 2018. Breast Cancer Prediction Applying Different Classification Algorithm with Comparative Analysis using WEKA. In: *2018 4th International Conference on Electrical Engineering and Information & Communication Technology*, Dhaka, Bangladesh, September 13-15.
- Bouckaert, R.R., Frank, E., Hall, M., Kirkby, R., Reutemann, P., Seewald, A., Scuse, D.,** 2016. *WEKA Manual for Version 3-8-0*. Hamilton, New Zealand, 342s.
- Breiman, L.,** 2001. Random Forests. *Machine Learning*, 45(1):5-32.
- Brunton, S.L., Noack, B.R., Koumoutsakos, P.,** 2020. Machine Learning for Fluid Mechanics. *Annu Rev Fluid Mech*, 52(1):477-508.
- Castro, Y., Kim, Y.J.,** 2016. Data mining on road safety: factor assessment on vehicle accidents using classification models. *International Journal of Crashworthiness*, 21(2):104-111.
- Chang, L.Y., Chen, W.C.,** 2005. Data mining of tree-based models to analyze freeway accident frequency. *Journal of Safety Research*, 36(4):365-375.
- Cihan, P., Kalıpsız, O.,** 2016. Öğrenci Proje Anketlerini Sınıflandırmada En İyi Algoritmanın Belirlenmesi. *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*, 8(1):41-49.
- Çoban, T.,** 2011. Makine Öğrenme Algoritmaları ile Web Siteleri Tıklamalarının Analizi. *Yüksek Lisans Tezi*, Beykent Üniversitesi, İstanbul, 102s.
- Doğan, O.,** 2017. Ücretsiz Veri Madenciliği Araçları ve Türkiye’de Bilinirlikleri Üzerine Bir Araştırma. *Ege Stratejik Araştırmalar Dergisi*, 8(1):17.
- Geymen, A., Dedeoğlu, O.K.,** 2016. Reducing Traffic Accidents Using Geographical Information Systems: Kahramanmaraş Case Study. *Journal of the Institute of Science and Technology*, 6(2):79-88.
- Goniewicz, K., Goniewicz, M., Pawłowski, W., Fiedor, P.,** 2016. Road accident rates: strategies and programmes for improving road traffic safety. *Eur J Trauma Emerg Surg*, 42(4):433-438.
- Güner, S., Codal, K.S., Geçer, H.S., Coşkun, E.,** 2018. Trafik Kaza Desenlerinin Tanımlanmasında K-Means Kümeleme Algoritmasının Kullanılması: Sakarya İli Uygulaması. *İşletme Bilimi Dergisi*, 6(3):89-105.
- Holmes, G., Donkin, A., Witten, I.H.,** 1994. WEKA: a machine learning workbench. In: *Proceedings of ANZIIS '94 - Australian New Zealand Intelligent Information Systems Conference*. Brisbane, QLD, Australia, August November 29 -December 2.

- Hussain, S., Muhammad, L.J., Ishaq, F.S., Yakubu, A., Mohammed, I.A.,** 2019. Performance Evaluation of Various Data Mining Algorithms on Road Traffic Accident Dataset. In: Satapathy SC, Joshi A, eds. *Information and Communication Technology for Intelligent Systems*. Vol 106. Smart Innovation, Systems and Technologies. Springer, Singapore, 67-78.
- Karamaşa, Ç., Erdoğan, N.K.,** 2018. Bayramlarda Gerçekleşen Trafik Kazalarının Birlikte Kuralları İle Analiz Edilmesi. *Karadeniz Uluslararası Bilimsel Dergi*, (40):386-411.
- Kiranmai, A.K., Laxmi, J.A.,** 2018. Data mining for classification of power quality problems using WEKA and the effect of attributes on classification accuracy. *Protection and Control of Modern Power Systems*, 3(1):29.
- Krishnaveni, S., Hemalatha, M.,** 2011. A Perspective Analysis of Traffic Accident using Data Mining Techniques. *International Journal of Computer Applications*, 23(7):40-48.
- Kuşkapan, E., Çodur, M.Y.,** 2022. Trafik Kazalarının Sınıflandırılmasında Çok Katmanlı Algılayıcı, Regresyon ve En Yakın Komşuluk Algoritmalarının Performans Analizi. *Journal of Polytechnic*, 25(1).
- Li, J., He, J., Liu, Z., Zhang, H., Zhang, C., Elkamel, A.,** 2019. Traffic accident analysis based on C4.5 algorithm in WEKA. Elkamel A, ed. *MATEC Web of Conferences*, Jiangsu, China, March 8.
- Li, L., Shrestha, S., Hu, G.,** 2017. Analysis of road traffic fatal accidents using data mining techniques. In: *2017 IEEE 15th International Conference on Software Engineering Research, Management and Applications*, London, UK, June 7-9.
- Mateus, O.H., Quintana Jiménez, L.A., López-Valdés, F.J., Morales-Londoño, N., Salas-Navarro, K.,** 2019. Using Data-Mining Techniques for the Prediction of the Severity of Road Crashes in Cartagena, Colombia. In: Figueroa-García JC, Duarte-González M, Jaramillo-Isaza S, Orjuela-Cañon AD, Díaz-Gutiérrez Y, eds. *Applied Computer Sciences in Engineering*. Vol 1052. Communications in Computer and Information Science. Springer, Cham, Bogota, 309-320.
- Mhetre, V., Nagar, M.,** 2017. Classification based data mining algorithms to predict slow, average and fast learners in educational system using WEKA. In: *2017 International Conference on Computing Methodologies and Communication*, Erode, India, July 18-19.
- Mujalli, R.O., López, G., Garach, L.,** 2016. Bayes classifiers for imbalanced traffic accidents datasets. *Accident Analysis & Prevention*, 88:37-51.
- Na, S., Xumin, L., Yong, G.,** 2010. Research on k-means Clustering Algorithm: An Improved k-means Clustering Algorithm. In: *2010 Third International Symposium*

on *Intelligent Information Technology and Security Informatics*. Jian, China, April 2-4.

- Nafie, A.F.M., Mohamed, H.A.A.**, 2018. Usage Apriori and clustering algorithms in WEKA tools to mining dataset of traffic accidents. *Journal of Information and Telecommunication*, 2(3):231-245.
- Nandurge, P.A., Dharwadkar, N.V.**, 2017. Analyzing road accident data using machine learning paradigms. In: *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)*, Palladam, India, Feburay 10-11.
- Naqa, E.I., Murphy, M.J.**, 2015. What Is Machine Learning? In: Naqa EI, Li R, Murphy MJ, eds. *Machine Learning in Radiation Oncology*. Springer International Publishing, Chem, 3-11.
- Nikam, S.S.**, 2015. A Comparative Study of Classification Techniques in Data Mining Algorithms. *International Journal of Modern Trends in Engineering & Research*, 8(1):13-19.
- Nilsson, N.J.**, 1996. *Introduction to Machine Learning, An Early Draft of a Proposed Textbook*. Stanford, California.
- Oña, J., Mujalli, R.O., Calvo, F.J.**, 2011. Analysis of traffic accident injury severity on Spanish rural highways using Bayesian networks. *Accident Analysis & Prevention*, 43(1):402-411.
- Özcan, M., Küçükönder, M.**, 2020. Investigation of Spatiotemporal Changes in the Incidence of Traffic Accidents in Kahramanmaraş, Turkey, Using GIS-Based Density Analysis. *J Indian Soc Remote Sens*, 48(7):1045-1056.
- Özden, C., Acı, Ç.**, 2018. Makine öğrenmesi yöntemleri ile yaralanmalı trafik kazalarının analizi: Adana örneği. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, 24(2):266-275.
- Parsania, D.V.S., Jani, D.N.N., Bhalodiya, N.H.**, 2014. Applying Naïve bayes, BayesNet, PART, JRip and OneR Algorithms on Hypothyroid Database for Comparative Analysis. *International Journal of Darshan Institute*, 3(1):60-64.
- Perez, G.V.A., Lopez, J.C., Cabello, A.L.R., Grajales, E.B., Espinosa, A.P., Fabian, J.L.Q.**, 2018. Road Traffic Accidents Analysis in Mexico City through Crowdsourcing Data and Data Mining Techniques. *International Journal of Computer and Information Engineering*. 12(8):604-608.
- Sandıkçı, Y., Aydılek, İ.B.**, 2018. Tarımsal Veri Analizlerinin Veri Madenciliği ile Yapılması. *Harran Üniversitesi Mühendislik Dergisi*, 3(2):1-7.
- Schuld, M., Sinayskiy, I., Petruccione, F.**, 2015. An introduction to quantum machine learning. *Contemporary Physics*, 56(2):172-185.

- Soylu, K.**, 2018. Kredi Kartı Sahte işlem Tespiti. *Yüksek Lisans Tezi*, Ankara Üniversitesi Ankara, 48s.
- Taamneh, M., Alkheder, S., Taamneh, S.**, 2017. Data-mining techniques for traffic accident modeling and prediction in the United Arab Emirates. *Journal of Transportation Safety & Security*, 9(2):146-166.
- Taamneh, M., Taamneh, S., Alkheder, S.**, 2017. Clustering-based classification of road traffic accidents using hierarchical clustering and artificial neural networks. *International Journal of Injury Control and Safety Promotion*, 24(3):388-395.
- Taşçı, M.E., Şamlı, R.**, 2020. Veri Madenciliği İle Kalp Hastalığı Teşhisi. *European Journal of Science and Technology*, 88-95.
- Umadevi, S., Marseline, K.S.J.**, 2017. A survey on data mining classification algorithms. In: *2017 International Conference on Signal Processing and Communication*, Coimbatore, India, July 28-29.
- URL-1**, 2021. <https://data.tuik.gov.tr/Bulten/Index?p=Karayolu-Trafik-Kaza-Istatistikleri-2020-37436> 18.10.2021.
- URL-2**, 2022. <https://biruni.tuik.gov.tr/ilgosterge/?locale=tr> 12.01.2022.
- URL-3**, 2022. <https://tr.wikipedia.org/wiki/Kahramanmara%C5%9F> 13.01.2022.
- URL-4**, 2022. https://en.wikipedia.org/wiki/Expectation%E2%80%93maximization_algorithm 23.02.2022
- Wei, J., Chu, X., Sun, X.**, 2019. Machine learning in materials science. *InfoMat*, 1(3):338-358.
- Witten, I.H., Frank, E., Trigg, L.E., Hall, M.A., Holmes, G., Cunningham, S.J.**, 1999. *Weka: Practical Machine Learning Tools and Techniques with Java Implementations*. Hamilton, New Zealand.
- World Health Organization (WHO)**, 2017. *Global Status Report On Road Safety 2015*. WHO Press, Geneva.
- World Health Organization (WHO)**, 2019. *Global Status Report On Road Safety 2018*. WHO Press, Geneva.
- 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.
- Zengin, B., Kaymaz, K., Arslannur, B.**, 2018. Tunceli İlindeki Trafik Kazası Oranlarının İncelenmesi. *Gümüşhane Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 8(2):318-324.

- Zhang, D., Tsai, J.J.**, 2003. Machine Learning and Software Engineering. *Software Quality Journal*, 11(2):87-119.
- Zhang, X.D.**, 2020. Machine Learning. In: *A Matrix Algebra Approach to Artificial Intelligence*. Springer Singapore, 223-440.
- Zhang, X.F., Fan, L.**, 2013. A decision tree approach for traffic accident analysis of saskatchewan highways. In: *2013 26th IEEE Canadian Conference on Electrical and Computer Engineering*, Regina, SK, Canada, May 5-8.