

Trafik Kazalarının Makine Öğrenmesi Yöntemleriyle Analizi ve Tahmini: Kahramanmaraş İçin Örnek Bir Çalışma

Hüseyin Bolat¹, Melih Yücesan², Anıl Utku^{3*}

¹ Munzur Üniversitesi, Lisansüstü Eğitim Enstitüsü, Mühendislik Yönetimi Bölümü, Tunceli, Türkiye

² Munzur Üniversitesi, Sağlık Bilimleri Fakültesi, Acil Yardım ve Afet Yönetimi Bölümü, Tunceli, Türkiye

³ Munzur Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, Tunceli, Türkiye

huseyinbolat@munzur.edu.tr^{ID}, melihyucesan@munzur.edu.tr^{ID}, *anilutku@munzur.edu.tr^{ID}

Makale gönderme tarihi:16.08.2022 Makale kabul tarihi:26.10.2022

Öz

Dünyada 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 makine öğrenme teknikleri ile analiz edilmiştir. Trafik kazalarının analizinde sınıflandırma, birliktelik kuralı ve görsel karşılaştırma yöntemleri kullanılmıştır. Sınıflandırma için NB, SVM, kNN, PART, AdaboostM1, J48 ve RF algoritmaları kullanılmıştır. Trafik kazalarının sınıflandırılmasında J48 algoritması en başarılı sonuçları elde etmiştir. Naive Bayes algoritmasının başarı düzeyi kötü olmasa da diğer algoritmalara 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, trafik kazası, sınıflandırma, birliktelik kuralları

Analysis and Prediction of Traffic Accidents with Machine Learning Methods: A Case Study for Kahramanmaraş

Abstract

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 centre and its districts between 2017 and 2020 were analysed by machine learning techniques. Classification, clustering, association rule and visual comparison methods are used in the analysis of traffic accidents. NB, SVM, kNN, PART, AdaboostM1, J48 and RF algorithms were used for classification. SimpleKMeans and EM clustering algorithms were used for clustering. Apriori algorithm was used in the association rule. The J48 algorithm achieved the most successful results in the classification of traffic accidents. Although the success level of the Naive Bayes algorithm was not bad, it remained lower than other algorithms. 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 centre 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, traffic accident, classification, association rules

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 vd., 2011).

Ülkelerde artan araç sayıları şehir yollarında trafiğin yoğunlaşmasına neden olmaktadır (Zengin vd., 2018). 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. 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 (WHO, 2015; WHO, 2018).

Dünyada olduğu gibi ülkemizde de yaşanan trafik kazaları insan yaralanmaları ve ölümlerinin en büyük nedenidir (Balcı vd., 2017). 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 ve birliktelik kuralı algoritmaları en sık kullanılan teknikler arasındadır (Taamneh vd., 2017). Yine son yıllarda pek çok alanda kullanılan makine öğrenme yöntemleri gelecekte de trafik kazalarının analizinde etkili bir yöntem olacaktır (Liv d., 2019; Chong vd., 2005).

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 öznelikten 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. Bu çalışmada geçmişte yaşanan trafik kazaları yukarıda bahsedilen yöntemlerle analiz edilip daha anlamlı bilgilere dönüştürülerek gelecekte yaşanabilecek potansiyel kazaların önüne geçilmesi amaçlanmaktadır.

Literatür Araştırması

Literatürdeki çalışmalar incelendiğinde sınıflandırma 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.

Li vd., ABD'de yaşanan ölümcül trafik kazalarını Naive Bayes (NB), Apriori ve K-Means algoritmalarını kullanarak analiz etmiştir (Li vd., 2017). 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.

Nafie vd., 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 (Nafie vd., 2018). Çıkan sonuçta, Apriori algoritmasının EM algoritmasından daha iyi olduğu belirlenmiştir.

Nandurge ve Dharwadkar, Hindistan'ın Maharashtra bölgesindeki 2015-2016 yılları arasında yaşanan trafik kazalarını analiz etmek için KMeans ve Apriori algoritması kullanmışlardır (Nandurge ve Dharwadkar, 2017). KMeans 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.

Krishnaveni ve Hemalatha tarafından yapılan çalışmada (Krishnaveni ve Hemalatha, 2011), Hong

Research article/Araştırma makalesi
 DOI:10.29132/ijpas.1163115

Kong'ta trafik kazalarında yaşanan yaralanma şiddetini tahmin etmek için NB, AdaBoostM1, J48, PART, Random Forest (RF) sınıflandırıcıları ile test edilmiş, sonuç olarak, RF'nin diğerlerine göre daha iyi sonuç verdiği gösterilmiştir.

Taamneh ve ark. 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 NB, J48, PART ve Multilayer Perceptron (MLP) kullanılmışlardır (Taamneh vd., 2017). Sonuç olarak, en düşük başarıyı NB gösterirken, diğer sınıflandırıcılar benzer başarı göstermişlerdir.

Beshah ve Hill tarafından yapılan çalışmada, Etiyopya'da yaşanan 18.288 trafik kaza verisi J48, NB ve k Nearest Neighbour (kNN) algoritmaları ile analiz edilmiş ve kaza şiddetleri tahmin edilmiştir (Beshah ve Hill, 2010). Çalışma sonucunda, kaza şiddetlerini en doğru tahmin eden algoritma kNN olurken J48 ve NB de yakın değerler vermiştir.

Zhang ve Fan, Kanada'nın Saskatchewan şehrinde son 20 yılda şehir merkezi ve kırsal yollarda yaşanan trafik kazalarının başlıca nedenleri ve kaza şiddetleri, ID3 ve J48 algoritmaları kullanılarak yaş, mevsim ve cinsiyete göre üç farklı başlıkta analiz etmiştir (Zhang ve Fan, 2013).

Bahiru ve ark., tarafından yapılan çalışmada, trafik kazalarını etkileyen faktörler ve kaza şiddetini tahmin etmek için NB, ID3, J48 ve CART algoritmaları kullanılmıştır (Bahiru vd., 2018). Çıkan sonuçta, J48 algoritmasının doğruluk performansının diğerlerinde daha iyi olduğu vurgulanmıştır.

Mateus ve ark. tarafından yapılan çalışmada, Kolombiya Cartagena otoyolunda 2016 ile 2017 yılları arasında 7.894 trafik kaza verisi üzerinde yapılan bir çalışmada makine öğrenme yöntemleriyle kaza şiddetleri analiz edilmiştir (Mateus vd., 2019). DT, J48, Support Vector Machine (SVM), NB, PART ve MLP ile 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.

AlMamlook ve ark. tarafından yapılan çalışmada, trafik kaza verileri denetimli makine öğrenme algoritmaları, AdaBoost, Logistic Regression (LR), NB ve RF ile analiz edilmiş veri düzensizliği için SMOTE algoritması kullanılmıştır (AlMamlook vd., 2019). Çalışma sonucunda RF'nin

kaza şiddetini tahmin etmekte daha başarılı olduğu vurgulanmıştır.

Oña ve ark. tarafından yapılan çalışmada, İspanya'nın daha kırsal yollarında 1.536 trafik kazası sonuçları, BayesNet kullanılarak, bir biri ile ilişkileri sınıflandırılmaya çalışılmıştır (Oña vd., 2011). 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 tarafından yapılan çalışmada, Nijerya'nın Lagos - İbadan ekspres karayolları üzerinde yaşanan trafik kazalarının sonuçları ve etkileri incelemek adına yaptıkları çalışmada, verilerin analizi ve tahmini Functional Tree (FT) ve ID3 algoritmaları ile yapılmıştır (Akomolafe ve Olutayo, 2013). Çalışma sonucunda, kaza verileri analiz edilmiş ve ID3'ün FT'ye göre daha doğru sonuçlar verdiği belirlenmiştir.

Mujalli ve ark. tarafından yapılan çalışmada, Ürdün'de banliyö yollarında 2009-2011 yılları arasında yaşanan trafik kazası sonuçlarının analizi için farklı Bayes sınıflandırıcıları kullanılmıştır (Mujalli vd., 2016). 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. tarafından yapılan çalışmada, 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 (Perez vd., 2018).

Castro ve Kim tarafından yapılan çalışmada, Birleşik Krallık ülkelerinde 2010 ile 2012 yılları arasında yaşanan ölümlü ve yaralanmalı trafik kazaları BayesNet, MLP ve J48 ile risk faktörleri araştırılmıştır (Castro ve Kim, 2016). Yapılan çalışmada en başarılı BayesNet algoritması ona yakın J48 ve MLP algoritmalarının verdiği belirtilmiştir.

Yavuz ve ark. tarafından yapılan çalışmada, Antalya ili ve çevresinde 2012-2016 yılları arası meydana gelen ölümlü ve yaralanmalı 3.181 trafik kazası sınıflandırılarak analiz edilmiştir (Yavuz vd., 2021). Çalışma sonucunda en başarılı doğruluk

Research article/Araştırma makalesi
 DOI:10.29132/ijpas.1163115

oranını veren sınıflandırıcının NB olduğu belirlenmiştir.

Kuşkapan ve Çodur tarafından yapılan çalışmada, 2002 ile 2018 yılları arasında ülkemizde yaşanan trafik kazaları, MLP, kNN ve Regresyon Algoritmaları kullanılarak sınıflandırma performansları incelenmiştir (Kuşkapan ve Çodur, 2022). Çalışma sonucunda, bu üç algoritmanın hata ve sınıflandırma performansları analiz edildiğinde en iyi sonucu En Yakın Komşu algoritmasının verdiği görülmüştür.

Hussain ve ark., tarafından yapılan çalışmada, 150 trafik kazası verisi üzerinde BayesNet, MLP ve J48 algoritmaları değerlendirilmiştir (Hussain vd., 2019). Sonuçta, en başarılı sınıflandırıcının MLP sonrasında BayesNet ve J48'in olduğu gösterilmiştir. Ayrıca birliktelik kuralı sonucu olarak en başarılı sonuçları Apriori'nin verdiği belirlenmiştir.

Güner ve ark., Sakarya ilinde meydana gelen trafik kazalarının analizi için, WEKA programının K-means kümeleme algoritması kullanılmış ve kazaların benzerlikleri incelenmiştir (Güner vd., 2018). Çalışma sonucunda, kazaların en çok açık havalarda, kuru zeminlerde, kavşak olmayan yollarda, sonbahar aylarında, hafta içleri ve öğlen öncesi yaşandığı belirtilmiştir.

Atalay ve Tortum tarafından yapılan çalışmada, 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 (Atalay ve Tortum, 2010). Çalışma sonucunda, kaza yaşanma potansiyeli en yüksek olan iller belirlenmiştir.

Chang ve Chen tarafından yapılan çalışmada, 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 (Chang ve Chen, 2005). Çalışma sonucunda CART modelinin otoyolda yaşana kaza sıklıklarını analiz etmek daha başarılı bir yöntem olduğu belirlenmiştir.

Bayata ve Hattatoğlu tarafından yapılan çalışmada, 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 (Bayata ve Hattatoğlu, 2014). Çalışma sonucunda Yapay Sinir Ağları yönteminin en başarılı sonucu verdiği belirtilmiştir.

Özden ve Acı tarafından yapılan çalışmada, 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 (Özden ve Acı, 2018). Çalışmada Regresyon Ağacı, Çoklu Doğrusal Regresyon, SVM, Fonksiyon Uydurma Yapay Sinir Ağı, MLP ve Genelleştirilmiş Regresyon Yapay Sinir Ağı yöntemleri kullanılmıştır. Çalışma neticesinde, SVM'in 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.

Geymen ve Dedeoğlu tarafından yapılan çalışmada, 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 (Geymen ve Dedeoğlu, 2016). Yaşanan kazalar; arazinin tomografik yapısı, kazaların yaşandığı zaman parametrelerine göre incelenmiş ve bu kazaların sebepleri CBS analizleriyle ortaya konmuştur.

Özcan ve Küçükönder tarafından yapılan çalışmada, 2008-2015 yılları arasında Kahramanmaraş merkezinde meydana gelen 14.317 trafik kazaları CBS ile analiz edilmiştir (Özcan ve Küçükönder, 2020). Bu çalışmada Kahramanmaraş merkezinde yaşanan trafik kazalarını önlemek adına bölgesel analiz, kaza yoğunluğu yüksek alanların tespiti ve risk taşıyan bölgelerin tahmin edilmeye çalışılmıştır.

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

MAKİNE ÖĞRENMESİ YÖNTEMLERİ

Makine öğrenmesi, veri madenciliği, yapay zekâ ve istatistiğin alt dalı olarak görülebilir. İlk olarak 1950'li yıllarda yapay zekâ ve yapay sinir ağı

Research article/Araştırma makalesi
 DOI:10.29132/ijpas.1163115

araştırmalarında adı geçmeye başladı. Makine öğrenmesi, bilgisayarların programlanmaya ihtiyaç duymadan kendi kendine öğrenmesi olarak tanımlanabilir (Schuld vd., 2015; Zhang, 2020). Makine öğrenmesi, çevreden öğrendiği bilgilerle insan aklını taklit etmek tasarlanan ve durmadan gelişen zeki hesaplama araçlarıdır (Wei vd., 2019; Naqa ve Murphy, 2015; Kiranmai ve Laxmi, 2018; Atalay ve Celik, 2017, Bingol, 2022B).

Makine öğrenme yöntemleri, denetimli, denetimsiz, yarı denetimli ve pekiştirmeli olarak dört ana kategoride gruplandırılır.

Denetimli Öğrenme

Denetimli öğrenmede, model içindeki bilgisayar algoritmasına giriş ve çıkış örnekleri verilir ve bu örneklerden ilgili algoritmanın sonucu sınıflandırması beklenir. Denetimli öğrenmede veriler etiketlenir yani giriş verilerinin tüm özellikleri verilerek çıkışın ne olacağı öğretilir ve sınıf etiketlerinin belirlenmesi sağlanır (Nilsson, 1996; Schuld, 2015, Bingöl, 2022). Denetimli öğrenme algoritmalarına; NB, SVM, karar ağaçları, LR, kNN, RF ve ANN örnek verilebilir.

Denetimsiz Öğrenme

Denetimsiz öğrenmede, sisteme sadece giriş verileri verilir ve çıkış verilerine bakılmaz (Güner vd., 2018). 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 vd., 2015). Denetimsiz öğrenmeye, birliktelik ve kümeleme algoritmaları ile olasılık yoğunluk tahmini, öznitelikler arasındaki ilişkilerin bulunması

ve boyutsallık azaltma örnek verilebilir (Brunton vd., 2020; Na vd., 2010).

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 vd., 2020).

Pekiştirmeli Öğrenme

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, 2015; Çoban, 2011).

MAKİNE ÖĞRENMESİ TABANLI TRAFİK KAZASI ANALİZİ

Çalışmada Kahramanmaraş il merkezi ve ilçelerinde, 2017 ile 2020 yılları arasında gerçekleşen ölümlü yaralanmalı trafik kazalarının verileri bulunmaktadır. Kullanılan veri seti 28 adet öznitelik ve 7929 veri satırından oluşmaktadır. Trafik veri setinin öznitelikleri Tablo 1'de görülmektedir.

Tablo 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

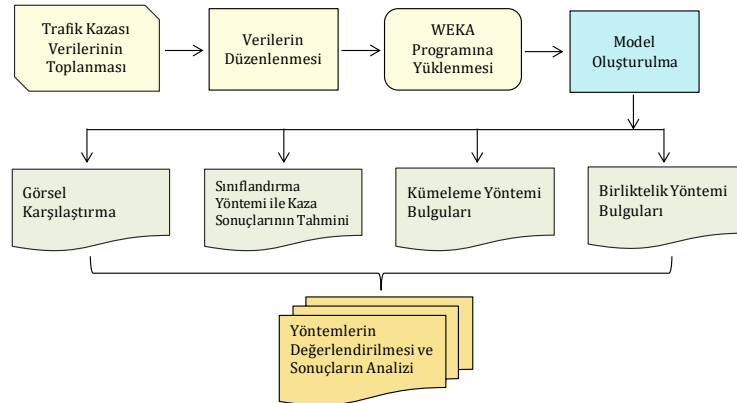
Yolun Tipi	Nominal	Tek Yönlü Yol, İki Yönlü Yol, Bölünmüş Yol, Diğer
Yolun Kaplaması	Nominal	Asfalt, Beton, Parke, Sati Kaplama, Stabilize, Toprak
Yolun Sınıfı	Nominal	Cadde, Sokak, Bağlantı Yolu, Devlet Yolu, Diğer, İl Yolu, Koy Yolu, Orman Yolu, Otoyol, Park Alanı, Servis Yolu, Tesis Önü veya İçi, Diğer
Yasal Hız Limiti	Nümerik	-
Yatay Geometrisi	Nominal	Düz Yol, Viraj, Tehlikeli Viraj
Düsey 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
Kazanın Sonucu	Nominal	Ölümlü, Yaralanmalı, Ölümlü/Yaralanmalı

Kullanılan Yöntemler

Çalışmada trafik kazası verileri görsel karşılaştırma, sınıflandırma ve birliktelik kuralları kullanılarak analiz edilmiştir.

Veri ön işleme 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 NB, SVM, kNN, PART, AdaboostM1, J48 ve RF algoritmalarının gerçek verilere karşı tahminleri ve performans metrikleri değerlendirilmektedir. Birliktelik kuralı çıkarma yönteminde Apriori algoritması ile trafik kaza unsurlarının bir birleriyle gerçekleşme olasılıkları değerlendirilmektedir. Çalışmada kullanılan yöntemlere ait akış şeması Şekil 1'de gösterilmektedir.



Şekil 1. Çalışmanın akış şeması

Sınıflandırma yöntemleri

Trafik kazaları sonuçlarını içeren veri seti NB, SVM, kNN, PART, AdaboostM1, J48 ve RF sınıflandırma algoritmaları kullanılarak sınıflandırılmıştır. Kullanılan sınıflandırma algoritmaları çapraz doğrulama ayarı seçilerek test edilmiştir. Bu yöntem yapılacak sınıflandırmanın öğrenmesini artırmak için kullanılmaktadır. Çapraz doğrulama veri setini k kadar kümeye ayırarak öğrenme işlemi gerçekleştirilmektedir. Sırasıyla bir küme test edilirken kalan k-1 kadar küme eğitimde kullanılır. Bu k değeri 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 ayırarak sınıflandırma işlemleri gerçekleştirilmiştir.

NB, temeli Bayes teoremine dayanan basit bir olasılık sınıflandırıcıdır. Bu teoreme; bir veri grubunda yer alan her bir nitelik için gerçekleşme olasılıkları hesaplanır ve gerçekleşme olasılığı en yüksek olana göre sınıflandırma yapılır (Nikam, 2015, Bahiru vd, 2018). NB sınıflandırma işlemleri Eş.1 kullanılarak hesaplanır.

$$P(x|y) = \frac{P(y|x) * P(x)}{P(y)} \quad (1)$$

$P(x|y)$: 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.

SMO algoritmasında, SVM kullanılmaktadır. SVM istatistiksel öğrenme teorisine dayanmaktadır.

Temel olarak veri sınıfları arasında sonsuz sayıda doğru çizerek bunların içinden ayrımı en iyi yapan doğrular seçilerek sınıflandırma işlemi yapılmaktadır (Nikam, 2015; Alan ve Yesilyurt, 2018).

k En Yakın Komşu (kNN) algoritması, verilen verinin hangi sınıfa ait olduğunu belirlemek için en yakın komşuluk ilişkisinden faydalanır. Bu algoritmanın çalışma prensibi, veri sınıflarının birbirleri arasındaki yakınlık ve mesafeye göre kümeler oluşturularak sınıflandırma işlemine dayanır (Nikam, 2015; Taşçı ve Şamlı, 2020). En yakın komşu uzaklığı hesaplanırken Öklid, Minkowski ve Manhattan metrikleri kullanılır. Bu uzaklıklar Eş.2, Eş.3 ve Eş.4 kullanılarak hesaplanır.

$$h_{\text{öklid}} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2)$$

$$h_{\text{manhattan}} = \sqrt{\sum_{i=1}^n |x_i - y_i|} \quad (3)$$

$$h_{\text{minkowski}} = \left(\sum_{i=1}^n (|x_i - y_i|^q) \right)^{1/q} \quad (4)$$

PART (Partial Decision Trees) algoritması, Böl ve yönet yaklaşımını kullanır. 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

Research article/Araştırma makalesi
 DOI:10.29132/ijpas.1163115

yinelemeye devam eder (Krishnaveni ve Hemalatha, 2011; Parsania vd., 2014).

AdaBoostM1 algoritması öncelikle eşit ağırlıklar kullanılarak sınıflandırma işlemine başlanır, daha sonra hatalı tahminler için ağırlık artırılarak tekrar tekrar eğitime alınır ve böylece sınıflandırmanın başarısı artırılmaya çalışılır. Yani bir sınıflandırıcının çıktısı bir sonraki sınıflandırıcının girdisi olarak gerçekleşir. Böylece hatalı veriler üzerinde daha fazla durularak başarılı tahminlerin yapılması sağlanır (Soylu, 2018).

J48 algoritması, örnek bir veri kümesi üzerinde sınıflandırma yaparken bir ağaç yapısı oluşturup, yapraklara bağlı düğümler ve düğümlere bağlı yapraklar oluşturur (Cihan ve Kalıpsız, 2016; Bahiru vd., 2018). J48 algoritmasında veri kümesinde bulunabilecek belirsizlikler entropi hesabı ile giderilmekte olup Eş. 5 kullanılarak hesaplanır (Aksu ve Karaman, 2017).

$$Entropi = \sum_{i=1}^n p_i \log_2 p_i \quad (5)$$

RF, sınıflandırma yaparken rasgele birçok ağaç oluşturmaktadır. RF, 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 (Breiman, 2001).

Sınıflandırma Algoritmalarının Performans Ölçütleri

Sınıflandırma algoritmalarının yanlış ve doğru tahminlerin gösterilmesi için Karmaşıklık Matrisi (Confusion Matrix) kullanılmaktadır. Yapılan tahminler sonrası test ve eğitim verileri karşılaştırılır. Karışıklık matrisinin örneği Şekil 2'de gösterilmektedir.

		Gerçek değerler	
		Pozitif	Negatif
Tahmin edilen değerler	Pozitif	TP	FN
	Negatif	FP	TN

Şekil 2. Karmaşıklık matrisi

TP, pozitif olarak tahmin edilen pozitif doğru örneklerdir. FP, pozitif olarak tahmin edilen negatif yanlış örneklerdir. TN, negatif olarak tahmin edilen doğru örneklerdir. FN, negatif olarak tahmin edilen yanlış örneklerdir (Umadevi ve Marseline, 2017).

Doğruluk (Accuracy), veri kümesinde sadece doğru sınıflandırılmış örnek sayısının veri kümesindeki toplam örnek sayısına bölümüdür. Doğruluk değeri Eş. 6 kullanılarak hesaplanır (Umadevi ve Marseline, 2017).

$$\text{Doğruluk} = \frac{TP+TN}{TP+FP+FN+TN} \quad (6)$$

Kesinlik (Precision), doğru olarak sınıflandırılan doğru pozitif sayısının kendisi ile birlikte yanlış pozitif sayısının toplamına oranıdır. Kesinlik değeri Eş. 7 kullanılarak hesaplanır.

$$\text{Kesinlik} = \frac{TP}{TP+FP} \quad (7)$$

Duyarlılık (Recall), doğru pozitif sayının yine doğru pozitif ile yanlış negatiflerin toplam sayısına bölünmesi olarak tanımlanır. Duyarlılık değeri Eş.8 kullanılarak hesaplanır.

$$\text{Duyarlılık} = \frac{TP}{TP+FN} \quad (8)$$

F-Ölçütü (F-Measure), duyarlılık ile kesinliğin bir birleşimidir. F-Ölçütü, iki kat duyarlılık ve kesinlik çarpımının yine duyarlılık ve kesinlik toplamına bölünmesiyle hesaplanır. F-Ölçütü değeri Eş.9 kullanılarak hesaplanır.

$$F\text{-Ölçütü} = \frac{2 \cdot \text{Duyarlılık} \cdot \text{Kesinlik}}{\text{Duyarlılık} + \text{Kesinlik}} \quad (9)$$

ROC Eğrisi (Receiver Operating Characteristics), alıcı çalışma alanı olarak bilinen yapılan sınıflandırmada doğru pozitif ve yanlış pozitif oranlarını karşılaştırmak için çizilen grafikdir. 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 vd., 2018; Aksu ve Karaman, 2017).

Birliktelik Kuralı Çıkarma Yöntemleri

Research article/Araştırma makalesi
DOI:10.29132/ijpas.1163115

Birliktelik kurallı algoritmalar, bir veri setindeki örneklerin birbirleri ile olan ilişkilerini eş zamanlı olarak inceleyen algoritmalar. Veriler arasındaki ilişkiler eş zamanlı olarak incelenirken bir olasılık korelasyonu kullanılır (Can ve Alataş, 2017). Bu çalışmada Apriori algoritması kullanılarak, trafik kazalarının birbiri ile olan ilişkileri incelenmiştir.

Apriori algoritmasının hesaplanmasında destek, güven ve kaldırmaç değeri olmak üzere üç önemli adım bulunmaktadır (Can ve Alataş, 2017). Apriori algoritmasında ilk olarak bir minimum bir destek ve güven değeri 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.

Destek (Support): Bir kural oluşturulurken veri setindeki örneklerin (N) ne kadar tekrarlandığını hesaplamak için kullanılır. Destek, Eş.10 kullanılarak hesaplanır.

$$Destek(X \rightarrow Y) = \frac{Sayı(X, Y)}{N} \quad (11)$$

Güven (Confidence): Bir kural oluşturulurken veri setindeki örneklerden X değişkeninin Y değişkeni ile birlikte bulunma olasılığıdır. Güven, Eş.11 kullanılarak hesaplanır.

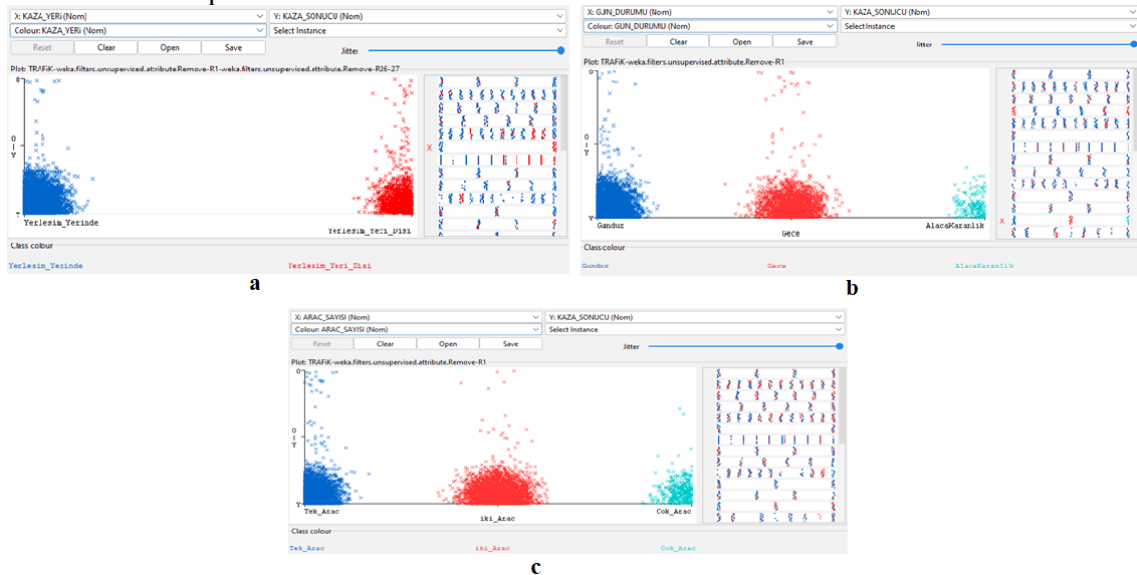
$$Güven(X \rightarrow Y) = \frac{Sayı(X, Y)}{Sayı(X)} \quad (11)$$

Kaldırmaç (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 Eş.12 kullanılarak hesaplanır.

$$Kaldırmaç(X \rightarrow Y) = Destek(X, Y) \quad (12)$$

Deneyisel Sonuçlar

Trafik kazası verileri, yerleşim yeri bilgisi göre incelendiğinde; 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 ölümlü 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 ön işleme ve görsel karşılaştırma verileri Şekil 3.a'da gösterilmektedir.



Şekil 1. Kaza yeri ile kaza sonuçları arasındaki görsel karşılaştırma sonuçları

Gün durumu niteliğine göre; gündüzleri 5516, geceleri 2240, alacakaranlıkta ise 173 kaza

Research article/Araştırma makalesi
 DOI:10.29132/ijpas.1163115

yaşanmıştır. Yaşanan kazalar “Kazanın Sonucu” ve “Gün Durumu” niteliklerine göre görsel olarak karşı-

laştırıldığında; 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 ön işleme ve görsel karşılaştırma verileri Şekil 3.b’de gösterilmektedir.

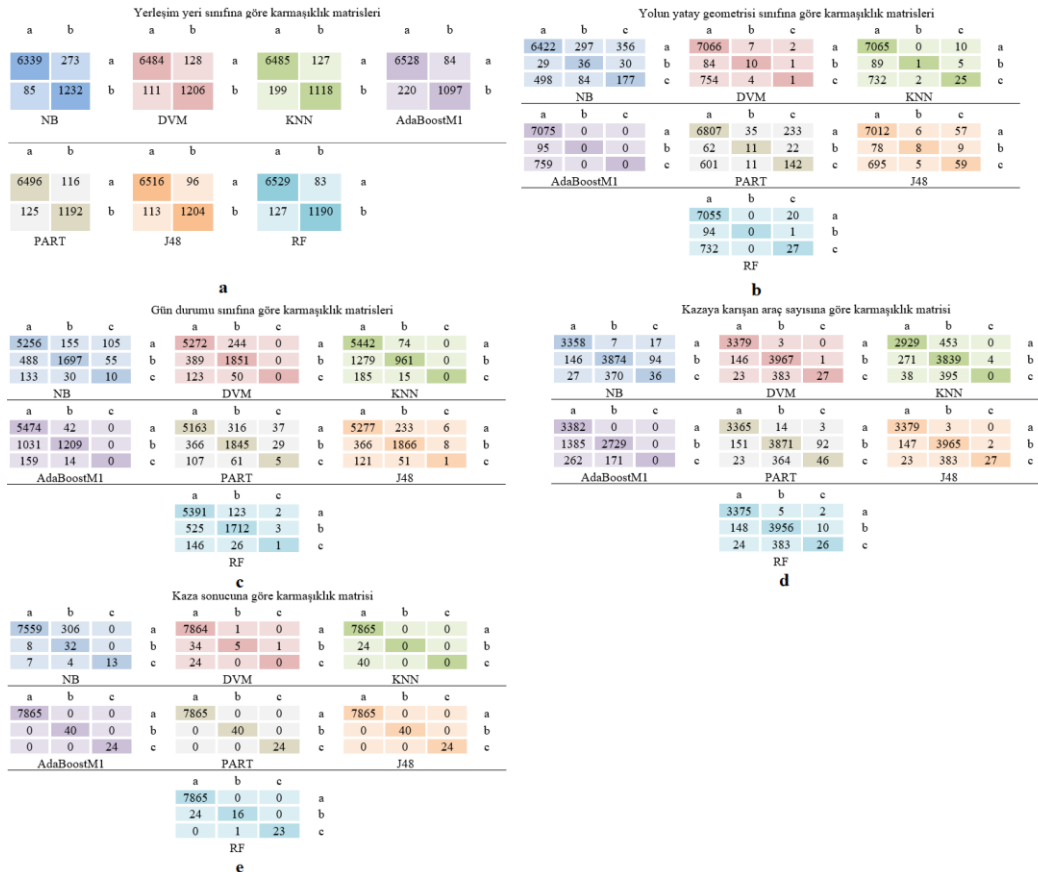
Kazaya karışan araç sayısına göre; kazaların en çok iki araçlı, tek araçlı ve çok araçlı olarak sıralandığı görülmektedir. Yaşanan trafik kazaları “Kazanın Sonucu” ve “Araç Sayısı” niteliklerine göre görsel olarak karşılaştırıldığında; yaralanmalı, ölümlü/yaralanmalı ve ölümlü olmak üzere 3 sınıfta da 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 ön işleme ve görsel karşılaştırma verileri Şekil 3.c’de gösterilmektedir.

Sınıflandırma yöntemiyle yaşanan gerçek trafik kazası sonuçları 10 kat çapraz doğrulama yöntemi kullanılarak NB, SVM, kNN, PART, AdaboostM1, J48 ve RF algoritmalarının tahmin sonuçları ve performans değerleri verilmektedir. Algoritmaların başarıları doğruluk, kesinlik, duyarlılık, F-Ölçütü ve ROC Eğrisi gibi performans ölçütleri ile belirlendi. Bu bağlamda gerçek trafik kazası verilerine göre algoritmalar Kaza Yerleşim Yeri (İçi, Dışı), Yolun Yatay Geometrisi (Düz yol, Viraj, Tehlikeli viraj), Gün Durumu (Gündüz, Gece, Alacakaranlık), Araç Sayısı (Tek araç, İki araç, Çok araçlı), Kazanın sonucu (Ölümlü, Yaralanmalı, Ölümlü ve yaralanmalı) öz niteliklerini kullanmaktadır.

Yerleşim Yeri Sınıfına Göre Deneysel Sonuçlar

10 kat çapraz doğrulama yöntemi ile NB, SVM, KNN, AdaBoostM1, PART, J48 ve RF sınıflandırma algoritmalarının tahminleri ve elde edilen karşılaştırmalı performans ölçütleri (Karmaşıklık matrisi, duyarlılık, kesinlik, F-ölçütü, ROC oranı, ve başarı oranı) Şekil 4 ve Tablo 2’de gösterilmektedir.

Sınıflandırma Yöntemi Bulguları



Şekil 4. Karmaşıklık matrisleri

Tablo 2. Deneysel sonuçlar

	Algoritma	Kesinlik	Duyarlılık	F-Ölçütü	ROC	Doğruluk (%)
Yerleşim yeri sınıfına göre sonuçlar	J48	0,974	0,974	0,974	0,974	97,36
	RF	0,973	0,974	0,973	0,994	97,35
	SVM	0,970	0,970	0,970	0,948	96,98
	PART	0,970	0,970	0,970	0,976	96,96
	AdaBoost	0,961	0,962	0,961	0,989	96,16
	KNN (k=10)	0,958	0,959	0,959	0,987	95,88
	NB	0,959	0,955	0,956	0,989	95,48
	KNN (k=10)	0,833	0,894	0,849	0,776	89,43
Yolun yatay geometrisine göre sonuçlar	RF	-	0,893	-	0,818	89,31
	J48	0,854	0,893	0,857	0,715	89,27
	SVM	0,827	0,893	0,844	0,511	89,25
	AdaBoost	-	0,892	-	0,799	89,22
	PART	0,849	0,878	0,860	0,760	87,77
	NB	0,856	0,837	0,845	0,797	83,68
	J48	0,884	0,901	0,891	0,945	90,09
Gün durumu sınıfına göre sonuçlar	SVM	-	0,899	-	0,872	89,86
	RF	0,882	0,896	0,884	0,951	89,59
	PART	0,873	0,884	0,879	0,934	88,44
	NB	0,878	0,878	0,876	0,952	87,81
	AdaBoost	-	0,843	-	0,931	84,28
	KNN (k=10)	-	0,808	-	0,885	80,75
	SVM	0,916	0,928	0,907	0,950	92,98
Kazaya karışan araç sayısına göre sonuçlar	J48	0,930	0,930	0,908	0,949	92,96
	RF	0,916	0,928	0,907	0,950	92,78
	PART	0,896	0,918	0,904	0,943	91,84
	NB	0,892	0,917	0,902	0,950	91,66
	kNN (k=10)	0,811	0,854	0,830	0,922	85,35
	AdaBoost	-	0,771	-	0,860	77,07
	J48	1,000	1,000	1,000	1,000	100
Kaza sonucuna göre sonuçlar	PART	1,000	1,000	1,000	1,000	100
	AdaBoost	1,000	1,000	1,000	1,000	100
	RF	0,997	0,997	0,996	1,000	99,68
	SVM	0,989	0,992	0,989	0,555	99,24
	kNN (k=10)	-	0,992	-	0,721	99,19
	NB	0,993	0,959	0,974	0,964	95,90

Yerleşim yerine göre deneysel sonuçları tüm algoritmaların başarı düzeyinin çok iyi seviyede ve aralarında çok küçük farklar olduğunu göstermiştir. 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 doğruluk ile gerçek verilere en yakın tahminleri yapan J48 algoritmasının ilk sırayı aldığı ve onu %97,35

Research article/Araştırma makalesi
 DOI:10.29132/ijpas.1163115

oranında sadece on binde bir puan farkla RF algoritmasının izlediği görülmektedir.

Yolun Yatay Geometrisi Sinifına Göre Deneysel Sonuçlar

Yolun yatay geometrisine göre (Düz Yol, Viraj ve Tehlikeli Viraj) deneysel sonuçlar Şekil 4.b ve Tablo 2’de gösterilmektedir.

Yolun yatay geometrisine ilişkin deneysel sonuçlar, kNN algoritmasının en yakın 10 komşu değerine göre %89,43 oranla en başarılı sonucu verdiği, kNN’yi %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.

Gün Durumu Sınıfına Göre Deneysel Sonuçlar

Günün durumuna göre deneysel sonuçlar Şekil 4.c ve Tablo 2’de görülmektedir. J48 algoritmasının %90,09 oranla ilk sırayı aldığı görülürken onu %89,86 başarı oranıyla SVM ve %89,59 başarı oranı ile RF algoritmalarının takip ettiği görülmektedir. Diğer algoritmalara kıyasla en başarısız tahminleri %80,75 oranla kNN (k=10) algoritması ve %84,28 oranla AdaBoostM1 algoritmasının yaptığı görülmektedir.

Kazaya Karışan Araç Sayına Göre Deneysel Sonuçlar

Kazaya karışan araç sayısına göre deneysel sonuçlar Şekil 4.d ve Tablo 2’de gösterilmektedir. SVM 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.

Kaza Sonucuna Göre Deneysel Sonuçlar

Kazaların sonucuna göre deneysel sonuçlar Şekil 4.e ve Tablo 2’de gösterilmektedir. J48, PART ve AdaBoostM1’in tüm verileri doğru sınıflandırarak %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 SVM ve kNN algoritmalarının da başarılarının buna bağlı olduğu görülmektedir.

Birliktelik Yöntemi Bulguları

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 0.4 minimum güven değeri 0.7 olarak belirlendiğinde 291.815 adet birliktelik kuralı oluşurken, lift değerinin en çok 1,43 olduğu olduğu görülmüştür.

Oluşturulan birliktelik kurallarının bir bölümü aşağıdaki Tablo 3’te verilmiştir. Tablodaki güven değeri kuralın birliktelik oranını gösterirken, lift değeri kuralın ilginç bir sonuç olup olmadığını gösterir.

Tablo 1. Apriori algoritması ile oluşturulan kurallar

Kurallar	Güven (Confidence)	Lift
YOLUN_SINIFI=Cadde DUSEY_GEO=Egimsiz HAVA_DURUMU=Acik		
1 KAZA_YOL_KUSURU=Yol_Sorunu_Yok KAZA_SONUCU=Yaralanmali 3437 ==> KAZA_YERi=Yerlesim_Yerinde 3435	100%	1,20
YOLUN_KAPLAMASI=Asfalt GUN_DURUMU=Gunduz		
2 HAVA_DURUMU=Acik YOL_SERiDi=Yol_Seridi_Var 3271 ==> KAZA_SONUCU=Yaralanmali 3249	99%	1,00
YOLUN_YUZEYi=Kuru TRAFiK_LEVHASI=		
3 Trafik_Levhalari_Yok TRAFiK_LAMBASI= Trafik_Lambasi_Yok 3599 ==> HAVA_DURUMU=Acik	97%	1,15

	YOLDA_CALISMA=Yolda_Calisma_Yok KAZA_SONUCU=Yaralanmali 3507		
	YOLUN_SINIFI=Cadde DUSEY_GEO=Egimsiz YOLUN_YUZEYi=Kuru		
4	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
	YOLUN_KAPLAMASI=Asfalt		
5	CARPMA_YERi=Yol_Uzerinde 6400 ==> KAZA_YERi=Yerlesim_Yerinde KAZA_SONUCU=Yaralanmali 5672	89%	1,07
	YOLUN_SINIFI=Cadde YATAY_GEO=Duz_Yol YOLUN_YUZEYi=Kuru 3907		
6	==> KAZA_YERi=Yerlesim_Yerinde DUSEY_GEO=Egimsiz HAVA_DURUMU=Acik KAZA_YOL_KUSURU=Yol_Sorunu_Yok KAZA_SONUCU=Yaralanmali 3179	81%	1,43
	KAZA_YERi=Yerlesim_Yerinde GUN_DURUMU=Gunduz KAZA_SONUCU=Yaralanmali 4632		
7	==> YOLUN_KAPLAMASI=Asfalt CARPMA_YERi=Yol_Uzerinde TRAFiK_LAMBASI=Trafik_Lambasi_Yok 3554	77%	1,11
	YATAY_GEO=Duz_Yol CARPMA_YERi=Yol_Uzerinde 6040 ==> KAZA_YERi=Yerlesim_Yerinde		
8	YOLUN_KAPLAMASI=Asfalt HAVA_DURUMU=Acik YOLUN_YUZEYi=Kuru KAZA_SONUCU=Yaralanmali 4487	74%	1,08
	KAZA_YERi=Yerlesim_Yerinde YOLUN_KAPLAMASI=Asfalt YATAY_GEO=Duz_Yol HAVA_DURUMU=Acik		
9	KAZA_YOL_KUSURU=Yol_Sorunu_Yok TRAFiK_LAMBASI=Trafik_Lambasi_Yok YOLDA_CALISMA=Yolda_Calisma_Yok 4372 ==> YOLUN_SINIFI=Cadde YOLUN_YUZEYi=Kuru 3190	73%	1,34
	KAZA_YERi=Yerlesim_Yerinde YOLUN_YUZEYi=Kuru CARPMA_YERi=Yol_Uzerinde 5132		
10	==> 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, 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.

3. 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.

4. 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

Research article/Araştırma makalesi
 DOI:10.29132/ijpas.1163115

gelmiştir. Apriori algoritması lift değerine göre 1,41 oranında ilginç olarak değerlendirmiştir.

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

6. 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.

7. 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.

8. 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.

9. 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.

10. 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.

SONUÇLAR

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 makine öğrenme teknikleri ile analiz edilmiştir. Çalışmadaki analiz işlemleri, ön işleme ve görsel karşılaştırma, sınıflandırma ve birliktelik algoritmaları kullanılarak dört farklı yöntem ile gerçekleştirilmiş ve elde edilen sonuçlara değinilmiştir.

Veri ön işleme ve görsel karşılaştırma sonuçlarına göre; trafik, 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 çarpıştığı, 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, SVM, 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 algoritmalara oranla daha düşük olduğu görülmüştür.

Birliktelik kuralı algoritması ile 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 ölümlü hem de yaralanmalı sonuçlanma olasılığı yüksek ve dikkat edilmesi gereken kazaların daha çok

Research article/Araştırma makalesi
 DOI:10.29132/ijpas.1163115

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.

Bu çalışmada literatürden farklı olarak, görsel karşılaştırma, sınıflandırma ve birliktelik kuralı yöntemleri bir arada kullanılarak bir birini destekleyen daha fazla ve daha farklı sonuçlar elde edilmiştir.

Trafik kazası verilerinden bazı sınıfların diğerlerine oranla çok sayıda tekrar etmesi, 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.

TEŞEKKÜR

Bu çalışma "Makine öğrenme yöntemleri ile Kahramanmaraş ilindeki trafik kazası sonuçlarının analizi ve tahmin edilmesi" başlıklı ve 733517 tez no'lu yüksek lisans tezinden türetilmiştir.

ÇIKAR ÇATIŞMASI BEYANI

Yazarlar bu makale ile ilgili herhangi bir çıkar çatışması bildirmemektedir.

ARAŞTIRMA VE YAYIN ETİĞİ BEYANI

Yazarlar bu çalışmanın araştırma ve yayın etiğine uygun olduğunu beyan eder.

KAYNALAR

Akomolafe, T.D. ve 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, M. ve Karaman, E. (2017). Karar Ağaçları ile Bir Web Sitesinde Link Analizi ve Tespiti. Acta Infologica, 1(2), 84-91.
- Alan, M.A. ve 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. ve Prefer, 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.
- Atalay, A. ve Tortum, A. (2010). Türkiye'deki İllerin 1997-2006 Yılları Arası Trafik Kazalarına Göre Kümeleme Analizi. Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi, 16(3).
- Atalay, M. ve Çelik, E. (2017). Büyük Veri Analizinde Yapay Zekâ ve Makine Öğrenmesi Uygulamaları. Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 9(22), 155-172.
- Anderson, T.K. (2009). Kernel density estimation and K-means clustering to profile road accident hotspots. Accident Analysis & Prevention, 41(3), 359-364.
- Bahiru, T.K., Kumar Singh, D. ve 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.
- Balci, M., Gölcük, A. ve Kahramanli, H. (2017). İstatistiksel Yaklaşımla Trafik Kazalarındaki Ölüm ve Yaralanma Durumlarının Kusurlu Unsurlarla İlişkilerinin İncelenmesi. Selçuk-Teknik Dergisi, 16(3).
- Bayata, H. ve Hattatoğlu, F. (2014). Erzincan İli İçin Farklı Yöntemlerle Trafik Kaza Tahmin Modellemesi. Erzincan Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 4(1), 31-46.
- Beshah, T. ve 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.
- Bharati, S., Rahman, M.A. ve Podder, P. (2018). Breast Cancer Prediction Applying Different Classification Algorithm with Comparative Analysis using WEKA. 2018 4th International Conference on Electrical Engineering and Information & Communication Technology, Dhaka, Bangladesh.
- Bingol, H. (2022). NCA-based hybrid convolutional neural network model for classification of cervical cancer on gauss-enhanced pap-smear images. International Journal of Imaging Systems and Technology.

Research article/Araştırma makalesi
 DOI:10.29132/ijpas.1163115

- Bingöl, H. (2022). Classification of OME with Eardrum Otoendoscopic Images Using Hybrid-Based Deep Models, NCA, and Gaussian Method. *Traitement du Signal*, 39(4).
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5-32.
- Brunton, S.L., Noack, B.R. ve Koumoutsakos, P. (2020). Machine Learning for Fluid Mechanics. *Annu Rev Fluid Mech*, 52(1), 477-508.
- Castro, Y. ve 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. ve Chen, W.C. (2005). Data mining of tree-based models to analyze freeway accident frequency. *Journal of Safety Research*, 36(4), 365-375.
- Chong, M., Abraham, A. ve Paprzycki, M. (2005). Traffic accident analysis using machine learning paradigms. *Informatica*, 29(1).
- Cihan, P. ve 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.
- Geymen, A. ve 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.
- Güner, S., Codal, K.S., Geçer, H.S. ve Coşkun, E. (2018). Trafik Kaza Desenlerinin Tanımlanmasında K-Means Kümeleme Algoritmasının Kullanılması: Sakarya İli Uygulaması. *İşletme Bilimi Dergisi*, 6(3), 89-105.
- Hussain, S., Muhammad, L.J., Ishaq, F.S., Yakubu, A. ve Mohammed, I.A. (2019). Performance Evaluation of Various Data Mining Algorithms on Road Traffic Accident Dataset. *Information and Communication Technology for Intelligent Systems*. Singapore.
- Can, U. ve Alatas, B. (2017). Automatic mining of quantitative association rules with gravitational search algorithm. *International Journal of Software Engineering and Knowledge Engineering*, 27(03), 343-372.
- Kiranmai, A.K. ve 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).
- Krishnaveni, S. ve Hemalatha, M.A. (2011). Perspective Analysis of Traffic Accident using Data Mining Techniques. *International Journal of Computer Applications*, 23(7),40-48.
- Kuşkapan, E. ve 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. ve Elkamel, A. (2019). Traffic accident analysis based on C4.5 algorithm in WEKA. *MATEC Web of Conferences*, Jiangsu, China.
- Li, L., Shrestha, S. ve Hu, G. (2017). Analysis of road traffic fatal accidents using data mining techniques. 2017 IEEE 15th International Conference on Software Engineering Research, Management and Applications, London, UK.
- Mateus, O.H., Quintana Jiménez, L.A., López-Valdés, F.J., Morales-Londoño, N. ve Salas-Navarro, K. (2019). Using Data-Mining Techniques for the Prediction of the Severity of Road Crashes in Cartagena, Colombia. *Applied Computer Sciences in Engineering*, 309-320.
- Mujalli, R.O., López, G. ve Garach, L. (2016). Bayes classifiers for imbalanced traffic accidents datasets. *Accident Analysis & Prevention*, 88, 37-51.
- Na, S., Xumin, L. ve Yong, G. (2010). Research on k-means Clustering Algorithm: An Improved k-means Clustering Algorithm. 2010 Third International Symposium on Intelligent Information Technology and Security Informatics, Jian, China.
- Nafie, A.F.M. ve 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. ve 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.
- Naqa, E.I. ve Murphy, M.J. (2015). What Is Machine Learning? In: Naqa EI, Li R, Murphy MJ, eds. *Machine Learning in Radiation Oncology*. Springer International Publishing, 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. ve 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. ve 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. ve Acı, Ç. (2018). Makine öğrenmesi yöntemleri ile yaralanmalı trafik kazalarının analizi: Adana

Research article/Araştırma makalesi
DOI:10.29132/ijpas.1163115

- örneği. Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi, 24(2), 266-275.
- Parsania, D.V.S., Jani, D.N.N. ve 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. ve 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. ve Aydılek, İ.B. (2018). Tarımsal Veri Analizlerinin Veri Madenciliği ile Yapılması. Harran Üniversitesi Mühendislik Dergisi, 3(2), 1-7.
- Schuld, M., Sinayskiy, I. ve Petruccione, F. (2015). An introduction to quantum machine learning. Contemporary Physics, 56(2), 172-185.
- Soylu, K. (2018). Kredi Kartı Sahte İşlem Tespiti. Yüksek Lisans Tezi, Ankara Üniversitesi, Ankara.
- Taamneh, M., Alkheder, S. ve 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. ve 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. ve Şamlı, R. (2020). Veri Madenciliği İle Kalp Hastalığı Teşhisi. European Journal of Science and Technology, 88-95.
- Umadevi, S. ve Marseline, K.S.J. (2017). A survey on data mining classification algorithms. In: 2017 International Conference on Signal Processing and Communication, Coimbatore, India.
- Wei, J., Chu, X. ve Sun, X. (2019). Machine learning in materials science. InfoMat, 1(3), 338-358.
- WHO, Global Status Report On Road Safety (2015). https://apps.who.int/iris/bitstream/handle/10665/44122/9789241563840_eng.pdf (Erişim tarihi: 12.03.2022).
- WHO, Global Status Report On Road Safety (2018). <https://apps.who.int/iris/bitstream/handle/10665/276462/9789241565684-eng.pdf> (Erişim tarihi: 12.03.2022).
- Yavuz, A.A., Ergül, B. ve 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. ve Arslannur, B. (2018). Tunceli İlindeki Trafik Kazası Oranlarının İncelenmesi. Gümüşhane Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 8(2), 318-324.
- Zhang, X.F. ve Fan, L. (2013). A decision tree approach for traffic accident analysis of saskatchewan highways. 2013 26th IEEE Canadian Conference on Electrical and Computer Engineering, Regina, SK, Canada.
- Zhang, X.D. (2020). Machine Learning. In: A Matrix Algebra Approach to Artificial Intelligence, Springer Singapore, 223-440.