



MAKİNE ÖĞRENMESİ TEKNİKLERİNİN SÜRÜŞ STİLİ SINIFLANDIRMASINDA KULLANIMI

Doğan YILDIZ^{1*}, Gülcan YILDIZ², Sercan DEMİRCİ²

¹Ondokuz Mayıs University, Faculty of Engineering, Department of Electrical and Electronics Engineering, 55239, Samsun, Türkiye

²Ondokuz Mayıs University, Faculty of Engineering, Department of Computer Engineering, 55239, Samsun, Türkiye

Özet: Sürücü davranışlarının trafik güvenliğine önemli derecede etkisi vardır. Bu nedenle, sürücülerin davranışsal örüntüleri ve bu örüntüleri etkileyen etmenler tanımlanmalıdır. Sürücüler, araçlarını daha verimli ve kurallara uygun kullanmaya yönlendirilmelidir. Bu bağlamda, sürücünün aracını nasıl kullandığı gözlemlenerek, sürücülerin sürüş risk derecelerine uygun olarak sigorta ya da kasko ücretleri belirlenebilir. Bu çalışmada, Araç İçi Teşhis (On Board Diagnostics-II, OBD-II) ve Küresel Konumlandırma Sistemi (Global Positioning System, GPS) cihazlarından alınan işlenmiş ve etiketlenmiş telemetri verileri ile Makine Öğrenmesi (Machine Learning, ML) algoritmaları kullanılarak sürücülerin risk gruplarının sınıflandırılması gerçekleştirilmiştir. OBD-II ile araçtan elde edilen verilerin işlenerek sürücülerin risk derecesinin belirlenip bu risk derecesine göre sigorta şirketlerinin kişiye özel sigorta ücreti belirlenmesinde yardımcı rol oynaması planlanmaktadır. Çalışmada Destek Vektör Makinesi (Support Vector Machine, SVM, CatBoost, k-NN (k-nearest neighbors, K-En yakın komşuluk) ve Hafif Gradyan Artırma Makinesi (Light Gradient Boosting Machine, LGBM)) olmak üzere dört model kullanılmıştır. Test sonuçları incelendiğinde en iyi tanımayaya SVM modeli ile erişilmiştir.

Anahtar kelimeler: Sınıflandırma, Tanıma, Makine öğrenmesi, Sürücü davranışları

Driving Style Classification Using Machine Learning Techniques

Abstract: Driver behavior has a significant impact on traffic safety. Therefore, drivers' behavioral patterns and the factors affecting these patterns should be identified. Drivers should be directed to use their vehicles more efficiently and in accordance with the rules. In this context, by observing how the driver uses his vehicle, insurance or car insurance fees can be determined in accordance with the drivers' driving risk levels. In this study, the risk groups of drivers are classified using Machine Learning (ML) algorithms with processed and labeled telemetry data obtained from On Board Diagnostics-II (OBD-II) and Global Positioning System (GPS) devices. It is planned to determine the drivers' risk level by processing the data obtained from the vehicle with OBD-II and to play an auxiliary role in determining the personal insurance fee of the insurance companies according to this risk level. Support Vector Machine (SVM), CatBoost, k-NN (k-nearest neighbors), and Light Gradient Boosting Machine (LGBM) were used in the study. The best recognition was achieved with the SVM model.

Keywords: Classification, Recognition, Machine learning, Driver behaviors

*Sorumlu yazar (Corresponding author): Ondokuz Mayıs University, Faculty of Engineering, Department of Electrical and Electronics Engineering, 55239, Samsun, Türkiye

E mail: dogan.yildiz@omu.edu.tr (D. YILDIZ)

Doğan YILDIZ <https://orcid.org/0000-0001-9670-4173>

Gülcan YILDIZ <https://orcid.org/0000-0001-8631-8383>

Sercan DEMİRCİ <https://orcid.org/0000-0001-6739-7653>

Gönderi: 24 Mart 2024

Kabul: 10 Temmuz 2024

Yayınlanma: 15 Temmuz 2024

Received: March 24, 2024

Accepted: July 10, 2024

Published: July 15, 2024

Cite as: Yıldız D, Yıldız G, Demirci S. 2024. Driving style classification using machine learning techniques. BSJ Eng Sci, 7(4): 756-763.

1. Giriş

Sürücülere kendi sürüş davranışları hakkında geri bildirimde bulunmak ve dolayısıyla sürücülerin daha doğru ve güvenli sürüş deneyimleri edinmelerine katkı sağlamak amacıyla, araç kullanımı sırasında araç içerisinde oluşan veriler kullanılabilir. Veri sınıflandırma için literatürde çeşitli sınıflandırma modelleri geliştirilmiştir. Ancak, hangi sınıflandırıcının sürüş stillerini belirlemede daha iyi performans gösterdiği konusunda fikir birliği yoktur. Bu nedenle, sınıflandırma modellerinin bahsedilen amaç doğrultusunda değerlendirilebilmesi için farklı performans metriklerini kullanan daha fazla araştırmaya ihtiyaç olduğu söylenebilir (Silva ve Eugenio Naranjo, 2020).

Dikkatsiz ve agresif sürüş, trafik kazalarının en yaygın nedenlerindedir. Riskli sürüş davranışları, önemli

maddi ve manevi zararlara yol açan kazalara sebebiyet verebilir. Bu bağlamda, trafik kazalarının azaltılması ve yol güvenliğinin sağlanması için sürücü davranışlarının izlenmesi gerekmektedir. Ayrıca, sürücülerin sürüş tarzlarının tanınması amacıyla güvenilir ve etkili tekniklere ihtiyaç duyulmaktadır. Örneğin, Nesnelerin İnterneti (Internet of Things, IoT) teknolojilerindeki gelişmelerle birlikte, uygun sensörler kullanılarak sürücünün sürüş özelliklerinin uzaktan izlenmesi, tanınması ve sürüş davranışının modellenmesi mümkün hale gelmiştir (Mohammed ve ark., 2023).

Literatürde, sürücü davranışlarının izlenmesi problemine önerilen bazı çalışmalar mevcuttur. Örneğin Warren ve ark. (2019) çalışmasında, akıllı telefon verileri kullanılarak sürüş davranışları için kümeleme işlemi gerçekleştirilmiştir. San Francisco metropol bölgesindeki



trafik hızının ve farklı bölgelerdeki sürücülerin manevralarının modellenmesinde, şehirdeki beş yüz sürücünden alınan cep telefonu sensör verileri kullanılmıştır. Sürücüler, sürüş davranışlarına göre gruplandırılmıştır. Araştırmacılar San Francisco şehrindeki sürücü davranışının; agresif sürücüler, temkinli sürücüler ve agresif dönüşler yapan temkinli sürücüler gibi farklı gruplara ayrılabilirliğini bulmuşlardır. Zhang ve ark. (2019) çalışmasında, otonom sürüş ve yol güvenliği konusunda Derin Öğrenme (Deep Learning, DL) kullanımının, sürücü davranışlarını tanıma ve doğruluğunu iyileştirme konusunda umut verici olduğu öne sürülmüştür. Bunun için altı eksenli bir hareket işlemcisi kullanan araç sürüş davranışı tanıma yöntemi önerilmiştir. Bu yöntemde, yerleşik bir sensör tarafından toplanan örnek verileri öğrenmek için DL teknolojisi kullanılmıştır. Küçük örneklem boyutu ve kolay fazla uydurma problemini çözmek için, ortak bir veri artırma (Joint Data Augmentation, JDA) şeması önerilmiş ve yeni birçok-görünümlü evrişimli sinir ağı modeli (Multi-View Convolutional Neural Network, MV-CNN) tasarlanmıştır. CNN modeli ile bilgi yayılımının yönü iyileştirilmiş ve sürüş davranışının eğitimi, öğrenilmesi ve tanınması için yeni bir MV-CNN modeli geliştirilmiştir. Araç hareketleri hızlanma, sola dönüş, sağa dönüş, çarpışma ve normal sürüş olmak üzere beş sınıfa ayrılmıştır. MV-CNN'nin performansı deneysel olarak CNN, Özyinelemeli Sinir Ağları (Recurrent Neural Network, RNN), Uzun Kısa Süreli Bellek (Long Short-Term Memory, LSTM), CNN+LSTM ve üç boyutlu CNN ile karşılaştırılmıştır. Sonuçlar, bu yöntemlerle karşılaştırıldığında MV-CNN'nin en iyi hatırlama, kesinlik ve F1 puanını elde ettiğini göstermektedir.

Zantalis ve ark. (2019) çalışmasında, akıllı ulaşım sistemleri içerisindeki Makine Öğrenmesi (Machine Learning, ML) ve IoT uygulamalarının bağımsız bir incelemesi yapılmıştır. Akıllı ulaşım sistemi uygulama kategorileri için IoT yaklaşımlarının; rota optimizasyonu, park etme ve kaza önleme gibi konularda oldukça kullanışlı olduğu sonucuna varılmıştır. Buna ek olarak bu çalışmada, IoT ve ML yardımıyla akıllı ulaşım alanında şimdiden çok büyük ilerlemeler gözlemlendiği vurgulanmıştır. Silva ve Eugenio Naranjo (2020), iki farklı şehirden 50 sürücünün verilerini içeren bir çalışmadır. Ham verilerden beş özellik çıkarılmıştır. Veri kümesinin gerçek değerini türetmek için veri etiketlemesine 15 uzman katılmıştır. Elde edilen veri kümesi, beş farklı modelle (Destek Vektör Makineleri, (Support Vector Machine, SVM), Yapay Sinir Ağları (Artificial Neural Networks, ANN), Bulanık Mantık (Fuzzy Logic, FL), k-En Yakın Komşu (k-Nearest Neighbor, k-NN) ve Rastgele Ormanlar (Random Forests, RF)) işlenmiştir. Bu modeller, bir dizi performans metriği ve istatistiksel testlerle değerlendirilmiştir. Deneysel sonuçlarına göre, değerlendirmeye konu performans metrikleriyle SVM'nin diğer dört modele göre daha iyi performans gösterdiği tespit edilmiştir. SVM, ortalama doğrulukta %0.96, F1-Skorunda %0.9595, Eğri Altında Kalan Alan (Area Under

Curve, AUC) değerinde %0.9730 ve Kappa değerinde %0.9375 başarı elde etmiştir. Ayrıca, Wilcoxon testleri, ANN'nin diğer dört modele göre farklı tahminler yaptığını göstermiştir.

Bao ve ark. (2020) çalışması, sürücülerin otobanda şerit değiştirmeleri sırasında ortaya çıkan kişiselleştirilmiş öznel sürüş riskinin modellenmesini amaçlayan yeni bir yaklaşım sunmuştur. Dolayısıyla, sürüş esnasındaki öznel riski etkileyen faktörleri belirlemek bu çalışmadaki temel amaçtır. Buna yönelik olarak sürüş sinyalleri ve çevredeki araç konumları belirtilen amaç için kullanılmıştır. Ayrıca, belirli bir bireyin öznel durumunu tahmin etmek için şerit değiştirmeden hemen önceki risk seviyesi belirlenmeye çalışılmıştır. Risk faktörlerinin analizinde, Genetik Algoritmalar (Genetic Algorithms, GA) tarafından optimize edilmiş RF tabanlı bir yöntem kullanılmıştır. Öznel riski çift yönlü bir LSTM ile tahmin etmek amacıyla birleşik çerçeveli kişiselleştirilmiş öznel sürüş riski modeli önerilmiştir. Sonuçlar, öznel risk faktörlerinin bireysel farklılıklarının çıkarılabilirliğini ve bireyselleştirilmiş risk faktörlerinin tanımlanmasının, kişiselleştirilmiş öznel sürüş riskinin daha iyi modellenmesine yol açtığını göstermektedir. Öznel risk algısının bireylerin kişisel faktörlerinden etkilendiği doğrulanmış ve önerilen entegre bireysel modelin daha iyi risk değerlendirme performansına ulaştığı gösterilmiştir. Marafie ve ark. (2021) çalışmasında akıllı bir Yapay Zeka (Artificial Intelligence, AI) aracı olan AutoCoach sunulmaktadır. AutoCoach ayrıca, ML algoritmaları kullanarak sürücülerini farklı sürücü kişilik gruplarına göre sınıflandırabilmektedir. Buna göre kişiselleştirilmiş ve yapıcı bir geri bildirim sağlamak amacıyla, bir sürücünün geçmiş sürüş verileri toplanmış, analiz edilmiş ve bu verilerden öğrenme uygulaması yapılmıştır. Bu işlemler için bulut tabanlı bir Android uygulaması oluşturulmuştur. Çalışmada yeterli veriye sahip 36 sürücü, güvenlik sırasına göre en iyi, iyi, ortalama ve en kötü sürücüler olmak üzere dört gruba ayrılmıştır. Bu gruplar, modeldeki en yüksek riskli dönüş özelliğine dayanarak etiketlenmiştir. Çalışma sonuçları tasarımın, kişilik temelli bir sürüş aracı kullanılırken kullanıcı deneyimini geliştirmede etkili olduğunu kanıtlamaktadır.

Yuksel ve Atmaca (2021) çalışmasında, ivmeölçer ve jiroskop sensörlerinden elde edilen veriler kullanılarak dört ana riskli sürüş davranışı modellenmesi yapılmış ve sürücü profilleri oluşturulmuştur. C4.5 Karar Ağacı, RF, ANN, SVM, k-NN, Naif Bayes ve K-Star algoritmaları gibi yaygın kullanılan ML algoritmaları davranış analizinde kullanılmıştır. Riskli sürüş davranışları ve risk seviyeleri trafik memurlarının uzman görüşleri doğrultusunda değerlendirilmiş ve sürücü riski, FL yöntemiyle modellenmiştir. Deneysel sonuçlarda, K-Star algoritması %100 doğrulukla en başarılı algoritma olmuştur. Sonuç olarak, sürücünün kara kutusu olarak işlev gören yüksek doğruluklu ve düşük maliyetli bir sistem geliştirilmiştir. Bu sistem araçlara entegre edilebilir, sürücünün davranışlarını kaydedebilir ve riskli davranışları

belirleyebilir. Ayrıca, sigorta şirketlerine kullanıma göre politikalar belirleme imkanı sunar. Örneğin sigorta şirketleri, güvenli sürüş yapan müşterilerini düşük araç sigortası primleri ile ödüllendirebilir. Escottá ve ark. (2022) çalışmasında 1D ve 2D CNN'lere dayanan denetimli DL modelleri incelenmiştir. Bu modeller, aracın gösterge paneline yerleştirilmiş bir akıllı telefonun Atalet Ölçüm Birimi (Inertial Measurement Unit, IMU) sensörlerinden elde edilen doğrusal ivme ve açısal hız sinyallerini kullanarak sürüş etkinliklerini sınıflandırmak amacıyla kullanılmıştır. Hızlanma, frenleme, şerit değiştirme ve dönme gibi sürüş olaylarının izlenmesiyle agresif ve agresif olmayan davranışlar tespit edilebilmektedir. Elde edilen deneysel sonuçlarda en iyi sınıflandırma modeli ile %82.40'a kadar doğruluk değerlerine ve sırasıyla %75.36 ve %82.40 olan makro ve mikro ortalama F1 puanlarına ulaşılmıştır. Bu da sürüş etkinliklerinin sınıflandırmasındaki elde edilen yüksek performansı göstermektedir.

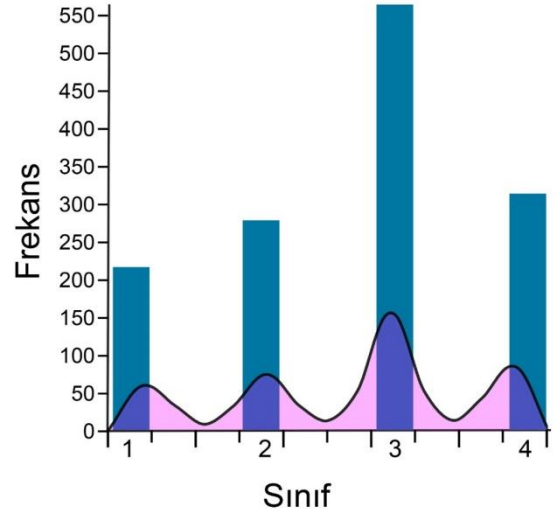
Malik ve Nandal (2023) çalışmasında Araca Yerleşik Teşhis-II (On-Board Diagnostic, OBD-II) araç teşhis sistemi ve ML algoritmaları ile elde edilen bilgilere dayanarak sürücü davranışlarının etkinliği ve iyileştirilme potansiyelleri incelenmiştir. OBD-II'den elde edilen belirli verilerin kullanılması ve belli ML tekniklerinin uygulanmasıyla, sürücülerin neredeyse tüm davranışları modellenilebilmektedir. Sürücü-Araç-Çevre sistemine bağımlı üçlü parametrelili sistem ve bunların etkileri detaylı bir şekilde ele alınmıştır. Daha sonra OBD-II, Elektronik Kontrol Birimleri (Electronics Control Units, ECU) ve sürüş davranışı üzerindeki etkileri en iyi ve etkili sürüş teknikleri ve parametreleriyle birlikte ele alınmıştır. Analizde kullanılan en iyi dokuz ML algoritması, tek tek tartışılmıştır. Mohammed ve ark. (2023) çalışmasında temel amaç, bir elektronik kart sistemi kullanarak gerçek zamanlı sürüşü izlemektir. Bu amaçla, Kontrol Alan Ağı (Control Area Network, CAN) ve OBD-II standartlarına sahip her araca takılabilen veri kaydedici bir cihaz kullanılmıştır. Bu cihaz, aracın mevcut hareketlerini izleyebilen hız, motor hızı, soğutma suyu sıcaklığı ve aracın konum koordinatları gibi ana faktörleri toplamaktadır. Araç takibi ve sürücü davranışının bilinmesi için dikkate alınan tüm parametreler Dakikadaki Devir Sayısı (Revolutions per Minute, RPM), motor soğutma, sıcaklık, enlem, boylam, zaman, tarih ve hızdır. Bu bilgiler, araca takılacak olan izleme sistemi tarafından toplanmıştır. Kurulan sistem, PIC Mikrodenetleyici, MCP2551 verici alıcı, GPS modülü ve WIFI alıcı-vericisi gibi bileşenler içermektedir. Toplanan bilgiler, sabit bir IP adresi ile HTTP protokolünü kullanarak bir web sunucusuna iletilir. Daha sonra geliştirilen grafik arayüz, sunucuya bağlı web sayfasından alınan verileri okumaktadır. Tüm veriler, uzaktan izleme bloğu olarak işlev gören sanal bir gösterge tablosunda görüntülenir ve aracın sürüş ve hareket geçmişi olarak depolanır. Deney sonuçlarına göre önerilen yaklaşım, sürüşü etkili bir şekilde izleyebilmektedir.

Bu çalışmada, ML algoritmaları kullanılarak OBD-II ve

GPS cihazlarından alınan işlenmiş ve etiketlenmiş telemetri verilerinin sürücü risk gruplarına göre sınıflandırılması gerçekleştirilmiştir. OBD-II ile araçtan elde edilen verilerin işlenerek sürücülerin risk derecesinin belirlenmesi ve sigorta şirketlerinin bu risk derecesine göre kişiye özel sigorta ücreti belirlemede yardımcı rol oynanması planlanmaktadır. Çalışmada SVM, CatBoost, k-NN ve Hafif Gradyan Artırma Makinesi (Light Gradient Boosting Machine, LGBM) olmak üzere 4 model kullanılmıştır. Test sonuçları incelendiğinde en iyi tanıma performansına SVM modeli ile erişildiği söylenebilmektedir. Elde edilen sonuçlar, araştırmacılara sürüş stilleri sınıflandırması için hangi ML modelinin uygun olabileceği konusunda önemli fikirler vermektedir.

2. Materyal ve Yöntem

Bu çalışmada, Github tarafından yayınlanan bir csv dosyasında bulunan 1366 telemetri verisini içeren bir veri seti kullanılmıştır (Judson, 2015). Kullanılan veri setindeki veriler etiketlenerek risk derecelerine göre 4 gruba ayrılmıştır. Sürücüler için tanımlanan dört kategori Yüksek, Orta, İyi, Düşük olacak şekilde düşük risk grubundan en yüksek risk grubuna doğru 1'den 4'e doğru sıralanmıştır (Judson, 2015). Şekil 1 ile veri seti içindeki 4 sınıf etiketinin dağılımı gösterilmiştir.



Şekil 1. Sınıf etiketlerinin dağılımı.

Orijinal veri setinde birçok öznelik olmasıyla birlikte risk kategorisinin belirleyicisi olan çok fazla öznelik yoktur. Belirleyiciliği yüksek öznelikleri tespit etmek amacıyla farklı deneyler yapılmıştır. Çalışmada kullanılan öznelikler hızlanma, şerit değiştirme, dönme ve frenlemedir (Judson, 2015).

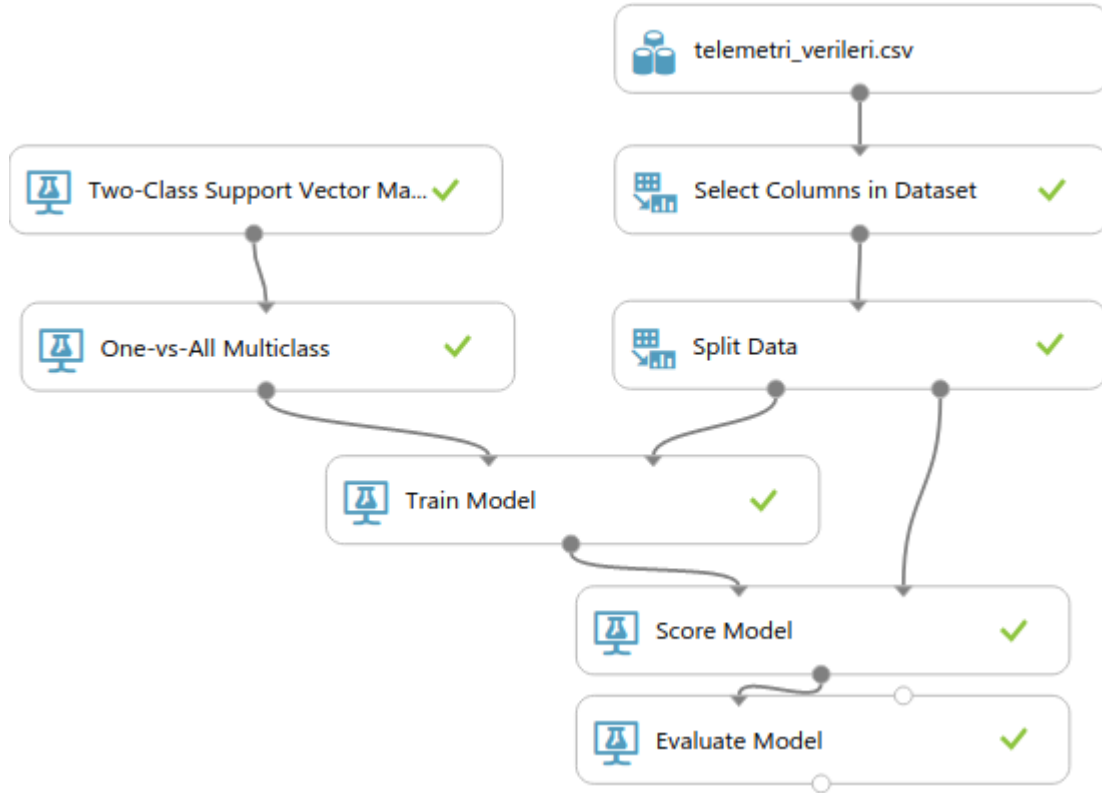
ML modelinde hızlanma, şerit değiştirme, dönme ve frenleme olmak üzere 4 farklı parametre kullanılmıştır. Tablo 1 ile ilgili veri setinden bir kesit gösterilmiştir.

2.1. Kullanılan Yöntemler

Bu çalışmada kullanılan modellerin oluşturulması ve eğitimi için Python programlama ve geliştirme dili kullanılmıştır. ML modeli eğitiminde izlenen adımlar Şekil 2 ile verilmiştir.

Tablo 1. Veri setine ait bir kesit

Sınıf Etiketi	Hızlanma	Frenleme	Dönme	Şerit Değişirme
1	0,24258265	0,41338630	0,45	0,41731518
3	0,16233399	0,12206573	0,44	0,19747082
3	0,03532071	0,05429156	0,91	0,24319066
3	0,07332580	0,09365595	0,33	0,33657588
4	0,15710653	0,06693150	0,21	0,27042802
1	0,12249223	0,32791622	0,55	0,48151751
2	0,08222662	0,15998555	0,55	0,32295720
4	0,09522464	0,06488504	0,36	0,39299611



Şekil 2. ML modeli eğitiminde izlenen adımlar.

Veri ayrıştırma (data decomposition) yoluyla çok etiketli (multi-label) bir veri seti bir ya da daha fazla tek etiketli alt kümelere bölünür. Her bir alt küme için de tercih edilen bir sınıflandırma tekniği ile ayrı bir alt sınıflandırıcı (sub-classifier) oluşturulur. Bu alt sınıflandırıcıların hepsi birleştirilerek tek birçok sınıflı sınıflandırıcı oluşturulur. Yaygın olarak kullanılan dört adet veri ayrıştırma tekniği vardır: etiket güç seti (label powerset, LP), bire karşı kalanlar (one-versus-rest, OVR), birer birer (one-by-one, OBO) ve bire karşı bir (one-versus-one, OVO).

OVR tekniğinde her sınıf için ayrı bir ikili (binary) sınıflandırıcı eğitilir. Bu eğitimde hedef sınıfın verileri pozitif, diğer tüm veriler negatif olarak kabul edilir. Yeni bir verinin sınıflandırması yapılırken eğitilen bütün sınıflandırıcılar bu verinin sınıfını tahmin eder ve en yüksek olasılıklı sınıf bu verinin sınıfı olarak seçilir. Her bir veri sınıfı için veri seti negatif ve pozitif olmak üzere iki sınıfa ayrılır ve her bir sınıfı tahmin etmek için bir ikili sınıflandırıcı eğitilir.

Bu çalışmada kullanılan yöntemler, SVM, CatBoost, k-NN ve LGBM'dir. Sınıflandırma işlemleri için ise, "scikit-learn" kütüphanesine ait OneVsRestClassifier fonksiyonu kullanılmıştır. Kullanılan parametreler için "default" seçeneği kullanılmıştır.

2.1.1. Destek vektör makinesi (SVM)

SVM'ler sınıflandırma ve regresyon problemlerinde çeşitli algoritmalar ile birlikte kullanılan denetimli öğrenme modelleridir. SVM'lerdeki amaç eldeki veri setini iki sınıfa bölen bir hiper-düzlem elde edebilmektir.

2.1.2. CatBoost

"CatBoost" ismi, "Category" (Kategori) ve "Boosting" (Arttırma) kelimelerinin birleşiminden gelen bir ML kütüphanesidir. CatBoost (Prokhorenkova ve ark., 2018), daha az parametreye sahip bir Gradyan Arttırıcı Karar Ağaçları (Gradient Boosting Decision Trees, GBDT) çerçevesidir ve kategori tabanlı değişkenleri destekler. Temel öğrenici olarak simetrik karar ağaçlarını kullanarak yüksek doğruluk sağlar. CatBoost, Gradient Bias ve Prediction Shift sorunlarını çözer. Böylece aşırı

uyumun (overfitting) ortaya çıkma olasılığını azaltır ve bu sayede algoritmanın doğruluğunu ve genelleme yeteneğini artırır.

Gradient Bias, özellikle düşük örneklemlili veri kümesi durumunda, gradyan tabanlı algoritmaların doğruluğunu etkileyen bir sorundur. CatBoost, bu tür problemleri ele alırken simetrik karar ağaçlarını temel öğrenici olarak kullanarak daha dengeli bir yaklaşım benimser. Prediction Shift, modelin eğitim ve test verileri arasında farklı performans göstermesine yol açan bir durumdur. CatBoost, kategori tabanlı değişkenleri etkili bir şekilde ele alarak ve düzenlemeye (regularization) odaklanarak bu tür problemleri önlemeye yardımcı olur.

Tüm bu özellikler, CatBoost'un daha güvenilir ve doğru tahminler yapmasına ve aşırı uyum, genelleme gibi problemleri azaltmasına yardımcı olur. Sonuç olarak, CatBoost'un, çeşitli ML uygulamalarında yüksek performans ve geniş bir kullanım alanı sunan etkili bir algoritma olduğu söylenebilir (Chen ve Wu, 2022).

2.1.3. k-en yakın komşular (k-NN)

k-NN yöntemi, sınıflandırma ve tahmin yapmak için kullanılan popüler bir denetimli ML tekniğidir. Potansiyel sorunların kategorize edilmesi veya gruplandırılması gibi işlemler ile endüstride sıkça kullanılmaktadır. Özellikle benzerliği, k-NN ML modelinde yeni veri noktalarının değerlerinin belirlenmesinde kullanılır. Her yeni veri noktası, eğitim kümesindeki noktalara ne kadar yakın olduğuna bağlı olarak bir değer almaktadır (Akhiladevi ve ark., 2022).

2.1.4. Hafif gradyan artırma makinesi (LGBM)

LGBM (Ke ve ark., 2017), Microsoft tarafından 2017 yılında önerilen ve ağaç tabanlı öğrenme algoritmalarını kullanan bir gradyan artırma çerçevesidir. LGBM, derin ve bağımsız ağaçların (Rastgele Orman) birleştirilmesi yerine, sık ve zayıf ardışık ağaçların birleştirilmesiyle bir birleşik model oluşturur. Her ağaç, önceki sonuçları öğrenir ve geliştirir. LGBM, diğer tüm gradyan artırma algoritmalarından daha hızlı eğitim hızı ve doğruluk sağlar (Huang ve ark., 2022).

3. Bulgular ve Tartışma

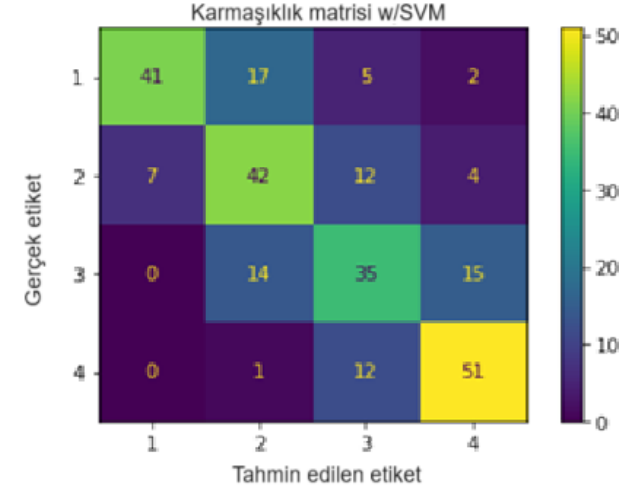
Bu çalışmada kullanılan telemetri veri seti; 1. sınıftan 215, 2. sınıftan 275, 3. sınıftan 561 ve 4. sınıftan 315 veri içermektedir. Eşit sayıda veri kullanabilmek amacıyla en düşük sayıda veri içeren sınıf etiketinin (1.) boyutu baz alınmıştır. Dolayısıyla, her bir veri sınıfından 215 adet veri alınarak toplamda 860 veri elde edilmiştir. Bu veri %70 eğitim ve %30 test olmak üzere iki gruba ayrılmıştır.

Eğitim seti üzerinde SVM, ANN, CatBoost, k-NN ve LGBM gibi algoritmalar kullanılmış ve one-vs-rest tekniğiyle beş model eğitilip test edilmiştir. Model geliştirilirken model performansını belirleyici olarak AUC metriği seçilmiş ve maliyet fonksiyonu olarak ortalama mutlak hata fonksiyonu seçilmiştir. Bu modellerin test seti üzerinde denemesi sonucu elde edilen karmaşıklık matrisleri ve Alıcı İşlem Karakteristikleri (Receiver Operating Characteristic, ROC) eğrileri de sonuçlara eklenmiştir.

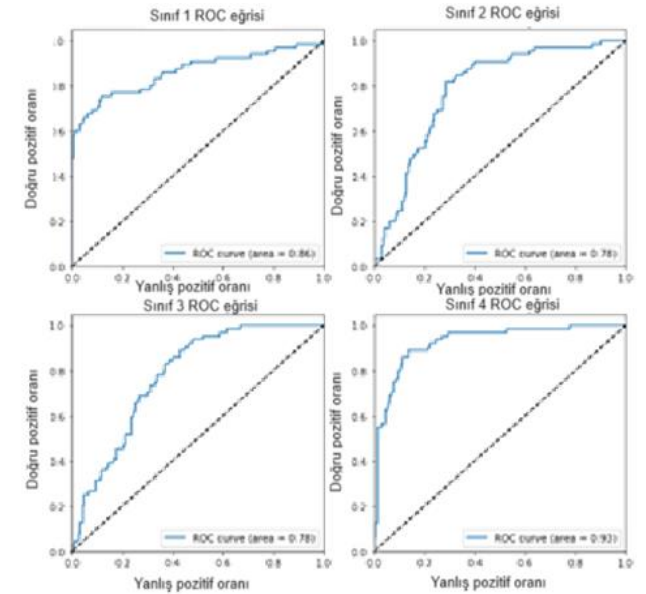
Ayrıca, tüm modellere ilişkin her bir sınıfa ait kesinlik, duyarlılık ve F1-skor sonuçları da verilmiştir.

3.1. SVM İle Elde Edilen Sonuçlar

SVM ile eğitilen modelin test seti üzerinden elde edilen karmaşıklık matrisi ve ROC eğrisi sırasıyla Şekil 3 ve Şekil 4 ile gösterilmiştir. Ayrıca, modelin test seti üzerinde her sınıfa ait kesinlik, duyarlılık ve F1-skor sonuçları Tablo 2 ile verilmiştir. Sonuçlara göre Sınıf 1 ve Sınıf 4, SVM'de verilen öznelikler ile diğer iki sınıfa göre daha iyi tahmin edilmiştir.



Şekil 3. SVM ile eğitilen modelin test seti üzerinden elde edilen karmaşıklık matrisi.



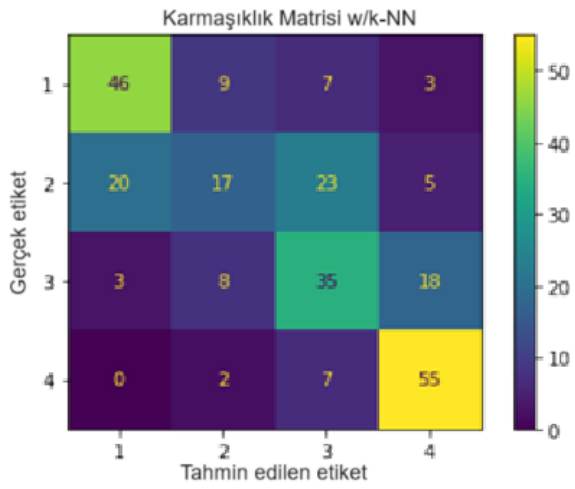
Şekil 4. SVM ile eğitilen modelin test seti üzerinden elde edilen ROC eğrisi.

Tablo 2. SVM ile eğitilen modelin test seti üzerinde her sınıfa ait kesinlik, duyarlılık ve F1-skor sonuçları

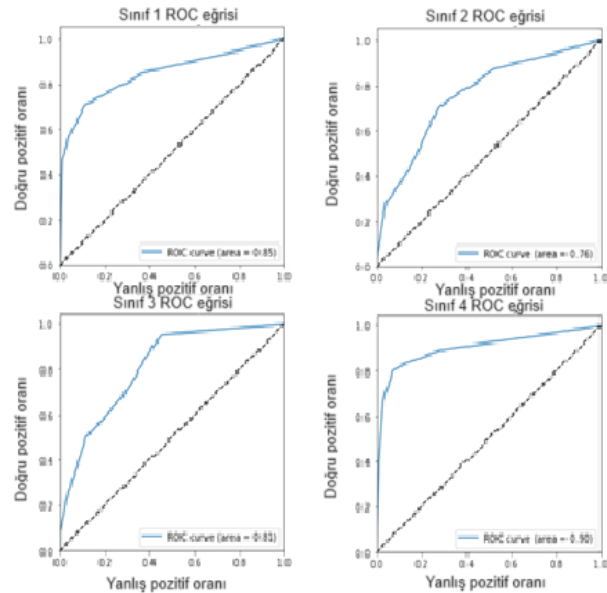
Sınıf	Kesinlik	Duyarlılık	F1-skor
1	0,85	0,63	0,73
2	0,57	0,65	0,60
3	0,55	0,55	0,55
4	0,71	0,80	0,75

3.2. k-NN İle Elde Edilen Sonuçlar

Benzer şekilde, k-NN ile eğitilen modelin test seti üzerinden elde edilen karmaşıklık matrisi ve ROC eğrisi sırasıyla Şekil 5 ve Şekil 6 ile gösterilmiştir. Ayrıca modelin test seti üzerinde her sınıfa ait kesinlik, duyarlılık ve F1-skor sonuçları Tablo 3 ile verilmiştir. Sonuçlara göre Sınıf 4, en yüksek F1-skor değeri ile en iyi tanıma oranına sahiptir. Bu sınıfı sırasıyla Sınıf 1, 3 ve 2 takip etmektedir. SVM ile benzer olarak 2 ve 3 etiketine sahip sınıflar diğer iki sınıfa göre daha düşük tanıma doğruluğuna sahiptir.



Şekil 5. k-NN ile eğitilen modelin test seti üzerinden elde edilen karmaşıklık matrisi.



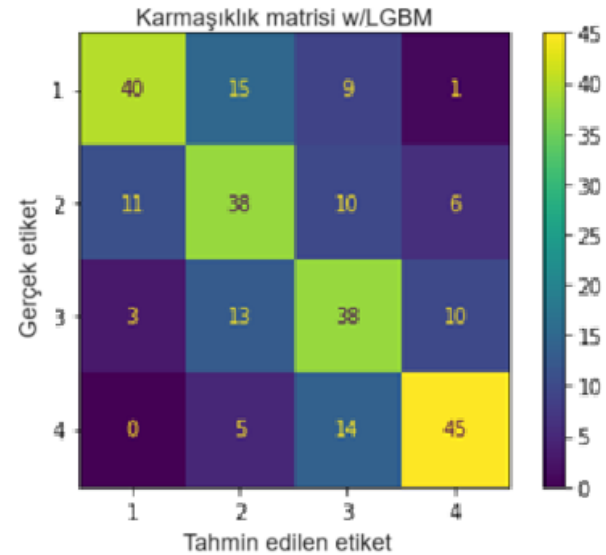
Şekil 6. k-NN ile eğitilen modelin test seti üzerinden elde edilen ROC eğrisi.

Tablo 3. k-NN ile eğitilen modelin test seti üzerinde her sınıfa ait kesinlik, duyarlılık ve F1-skor sonuçları

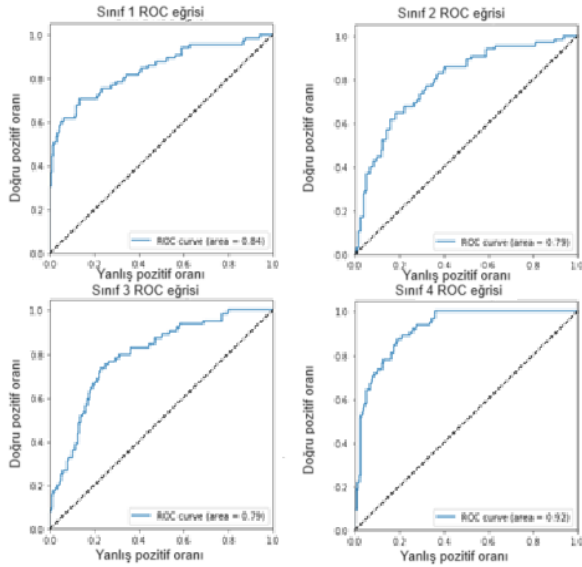
Sınıf	Kesinlik	Duyarlılık	F1-skor
1	0,67	0,71	0,69
2	0,47	0,26	0,34
3	0,49	0,55	0,51
4	0,68	0,86	0,76

3.3. LGBM ile elde edilen sonuçlar

LGBM ile eğitilen modelin test seti üzerinden elde edilen karmaşıklık matrisi ve ROC eğrisi Şekil 7 ve Şekil 8 ile sırasıyla gösterilmiştir. Ayrıca, modelin test seti üzerinde her sınıfa ait kesinlik, duyarlılık ve F1-skor sonuçları Tablo 4 ile verilmiştir. Sonuçlar incelendiğinde, daha düşük olmakla birlikte SVM'ye yakın değerlerin elde edildiği görülmektedir. 1 ve 4 etiketli sınıflar, 2 ve 3 etiketli sınıflardan daha iyi tahmin edilmiştir.



Şekil 7. LGBM ile eğitilen modelin test seti üzerinden elde edilen karmaşıklık matrisi.



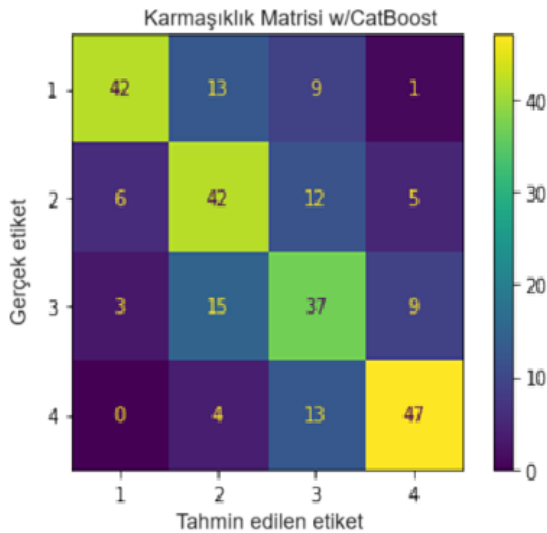
Şekil 8. LGBM ile eğitilen modelin test seti üzerinden elde edilen ROC eğrisi.

Tablo 4. LGBM ile eğitilen modelin test seti üzerinde her sınıfa ait kesinlik, duyarlılık ve F1-skor sonuçları

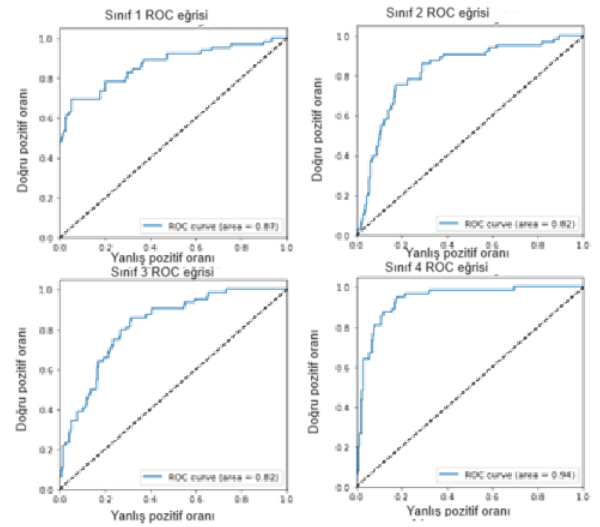
Sınıf	Kesinlik	Duyarlılık	F1-skor
1	0,74	0,62	0,67
2	0,54	0,58	0,56
3	0,54	0,59	0,56
4	0,73	0,70	0,71

3.4. CatBoost ile elde edilen sonuçlar

CatBoost ile eğitilen modelin test seti üzerinden elde edilen karmaşıklık matrisi ve ROC eğrisi Şekil 9 ve Şekil 10 ile sırasıyla gösterilmiştir. Ayrıca, modelin test seti üzerinde her sınıfa ait kesinlik, duyarlılık ve F1-skor sonuçları Tablo 5 ile verilmiştir. Bu model, uygulanan modeller arasında ikinci en yüksek doğruluk değerine sahiptir. LGBM yönteminden daha yüksek olmak üzere, benzer şekilde 1 ve 4 etiketli sınıflar, 2 ve 3 etiketli sınıflardan daha iyi tahmin edilmiştir.



Şekil 9. CatBoost ile eğitilen modelin test seti üzerinden elde edilen karmaşıklık matrisi.



Şekil 10. CatBoost ile eğitilen modelin test seti üzerinden elde edilen ROC eğrisi.

Tablo 5. CatBoost ile eğitilen modelin test seti üzerinde her sınıfa ait kesinlik, duyarlılık ve F1-skor sonuçları

Sınıf	Kesinlik	Duyarlılık	F1-skor
1	0,82	0,65	0,72
2	0,57	0,65	0,60
3	0,52	0,58	0,55
4	0,76	0,73	0,75

Tüm modellerin test doğruluk sonuçları Tablo 6 ile verilmiştir. Sonuçlar incelendiğinde en yüksek doğruluğa %66 değeriyle SVM ile erişildiği görülmektedir. Bu modeli, Catboost algoritması %65 ile takip etmiştir. Tüm modeller arasında en düşük tanıma doğruluğuna k-NN ile ulaşılmıştır. Ayrıca verilen ROC eğrileri, karmaşıklık matrisleri ve diğer karşılaştırma metriklerine bakıldığında eğitilen beş modelin düşük risk ve yüksek risk grubundaki sürücülerini başarı ile ayırt edebildiği, normal risk ve orta risk grubundaki sürücülerin ise, özellikle k-NN algoritması ile aynı başarı ile ayırt edilemediği görülmüştür. Bunun sebebinin de veri setinde bulunan normal risk ve orta risk grubundaki verilerin birbirine yakın olmasından kaynaklı olduğu düşünülmektedir.

Tablo 6. Tüm modellerin test doğruluk sonuçları

Model	Doğruluk
SVM	0,66
KNN	0,59
LGBM	0,62
CatBoost	0,65

4. Sonuç

Çoğu kaza, tehlikeli sürüş davranışlarının sonucudur ve bunlar büyük bedensel ve maddi zararlara yol açar. Ayrıca, kaza sayısındaki artışı durdurmak için tehlikeli sürücü davranışlarını tanımak ve insanları sürüş davranışlarına göre sınıflandırmak gerekir. Sensör

teknolojisindeki gelişmeler ve bu teknolojilerin gömülü sistemlere veya OBD sistemine entegre edilebilmesi sayesinde, güvenilir bir sürücü davranışı geliştirmek mümkün hale gelmiştir. Bu çalışmada, OBD-II ve GPS cihazlarından alınan telemetri verileri kullanılarak ML yöntemleriyle sürüş riskleri sınıflandırması yapılmıştır. Bu sınıflandırma için veriye dayalı ML yöntemlerinin karşılaştırması yapılmıştır. Dört farklı model eğitilmiş ve test sonuçları karşılaştırılmıştır. Sonuçlar incelendiğinde, en yüksek sınıflandırma doğruluğuna SVM ile ulaşıldığı görülmüştür. Ayrıca, genel olarak tüm modellerde yüksek ve düşük risk grubu, orta ve normal risk grubuna göre daha iyi tahmin edilmiştir. Orta ve normal risk verileri arasında önemli farkların olmaması, tanımının az olmasına neden olmuş olabilir. Ayrıca, ikisi de ara seviye riskleri temsil ettiği için benzer sonuçlar üretebilmektedir. Ek olarak dikkat edilecek diğer bir husus, tüm modellerde yüksek riskli veriler, hiçbir zaman düşük risk olarak değerlendirilmemiştir. Bu da modelin gerçek zamanlı uygulanması durumunda önemlidir. Gelecek çalışmalarda ANN gibi farklı modeller karşılaştırmaya dâhil edilebilir. Ek olarak modellerin performanslarını karşılaştırmak için farklı metrikler çalışmaya eklenebilir. Örneğin model karmaşıklığı, model eğitim ve test hızı eklenebilir. Ayrıca, gerçek zamanlı uygulama yapılarak çalışma genişletilebilir. Literatürde bulunan veri setlerine bu çalışmada kullanılan modeller uygulanarak performans karşılaştırması yapılabilir.

Katkı Oranı Beyanı

Yazar(lar)ın katkı yüzdesi aşağıda verilmiştir. Tüm yazarlar makaleyi incelemiş ve onaylamıştır.

	D.Y.	G.Y.	S.D.
K	30	30	40
T	35	30	35
Y	25	25	50
VTI	25	35	40
VAY	30	40	30
KT	35	30	35
YZ	45	35	20
KI	35	35	30
GR	50	25	25
PY	35	20	45
FA			100

K= kavram, T= tasarım, Y= yönetim, VTI= veri toplama ve/veya işleme, VAY= veri analizi ve/veya yorumlama, KT= kaynak tarama, YZ= Yazım, KI= kritik inceleme, GR= gönderim ve revizyon, PY= proje yönetimi, FA= fon alımı.

Çatışma Beyanı

Yazarlar bu çalışmada hiçbir çıkar ilişkisi olmadığını beyan etmektedirler.

Etik Onay Beyanı

Bu çalışmada hayvanlar ve insanlar üzerinde herhangi bir çalışma yapılmadığı için etik kurul onayı alınmamıştır.

Kaynaklar

- Akhiladevi M, Anitha K, Amrutha K, Amrutha M, Chandanashree K. 2022. Accident prediction using KNN algorithm. Fourth International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT): 26-27 December, Karnataka, India, pp: 1-5.
- Bao N, Carballo A, Miyajima C, Takeuchi E, Takeda K. 2020. Personalized subjective driving risk: Analysis and prediction. J Robot Mechatron, 32(3): 503-519.
- Chen H, Wu T. 2022. An improved CatBoost algorithm for red fox optimization in the field of anomaly detection. 2nd International Conference on Computer Science and Blockchain (CCSB), 28-30 October, Wuhan, China, pp: 148-153.
- Escottá Á T, Beccaro W, Ramírez M A. 2022. Evaluation of 1D and 2D deep convolutional neural networks for driving event recognition. Sensors, 22(11): 4226.
- Huang Y, Jafari M, Jin P. 2022. Driving safety prediction and safe route mapping using in-vehicle and roadside data. arXiv, 2022: 2209.05604.
- Judson I R. 2015. Assigning driver risk. <https://github.com/irjudson/AssigningDriverRisk> (accessed date: May 12, 2023).
- Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu T-Y. 2017. Lightgbm: A highly efficient gradient boosting decision tree. Adv Neural Inform Process Systems, 30: 1-9.
- Malik M, Nandal R. 2023. A framework on driving behavior and pattern using On-Board diagnostics (OBD-II) tool. Materials Today: Proc 80: 3762-3768.
- Marafie Z, Lin K-J, Wang D, Lyu H, Liu Y, Meng Y, Ma J. 2021. AutoCoach: an intelligent driver behavior feedback agent with personality-based driver models. Electronics, 10(11): 1361.
- Mohammed K, Abdelhafid M, Kamal K, Ismail N, Ilias A. 2023. Intelligent driver monitoring system: An Internet of Things-based system for tracking and identifying the driving behavior. Comput Stand Interfaces, 84: 103704.
- Prokhorenkova L, Gusev G, Vorobev A, Dorogush A V, Gulin A. 2018. CatBoost: unbiased boosting with categorical features. Adv Neural Inform Process Systems, 31: 1-11.
- Silva I, Eugenio Naranjo J. 2020. A systematic methodology to evaluate prediction models for driving style classification. Sensors, 20(6): 1692.
- Warren J, Lipkowitz J, Sokolov V. 2019. Clusters of driving behavior from observational smartphone data. IEEE Intelligent Transport Systems Magazine, 11(3): 171-180.
- Yuksel A, Atmaca S. 2021. Driver's black box: A system for driver risk assessment using machine learning and fuzzy logic. J Intelligent Transport Systems, 25(5): 482-500.
- Zantalis F, Koulouras G, Karabetos S, Kandris D. 2019. A review of machine learning and IoT in smart transportation. Future Internet, 11(4): 94.
- Zhang Y, Li J, Guo Y, Xu C, Bao J, Song Y. 2019. Vehicle driving behavior recognition based on multi-view convolutional neural network with joint data augmentation. IEEE Transact Vehicular Technol, 68(5): 4223-4234.