Journal of the Faculty of Engineering and Architecture of Gazi University 41:3 (2026) 1697-1712

Mühendislik Mimarlık Fakültesi Dergisi
Journal of The Faculty of Engineering and Architecture of Gazi University
Elektronik / Online ISSN : 1304 - 4915
Basılı / Printed ISSN : 1300 - 1884

Prediction of wheel flat failures in railway vehicles using machine learning methods with operational and meteorological data

Gökhan Arık[1,2], Gökhan Azizoğlu[3,4,5], Şengül Danışman[6]*, Ahmet Nusret Toprak[3]

[1]Department of Mechanical Engineering, Graduate School of Natural and Applied Sciences, Erciyes University, 38039 Kayseri, Türkiye
[2]Kayseri Transportation Inc., 38070 Kayseri, Türkiye
[3]Department of Computer Engineering, Faculty of Engineering, Erciyes University, 38030 Kayseri, Türkiye
[4]Department of Computer Engineering, Graduate School of Natural and Applied Sciences, Erciyes University, 38039 Kayseri, Türkiye
[5]Department of Software Engineering, Faculty of Engineering, Sivas Cumhuriyet University, 58140 Sivas, Türkiye
[6]Department of Mechanical Engineering, Faculty of Engineering, Erciyes University, 38030 Kayseri, Türkiye

Highlights:

- Wheel flat prediction using integrated operational and meteorological data
- A machine learning framework was developed for early prediction of railway wheel flat failures
- High-accuracy wheel flat prediction using stacking ensemble learning

Keywords:

- Railway systems
- Predictive maintenance
- Wheel flat defect
- Machine learning
- Ensemble learning

Article Info:

Research Article
Received: 30.03.2026
Accepted: 19.07.2026

DOI:

10.17341/gazimmfd.1919320

Correspondence:

Author: Şengül Danışman
e-mail: sdanisman@erciyes.edu.tr
phone: +90 507 125 4266

Graphical/Tabular Abstract

In this research, performance comparisons were made using various machine learning methods shown in Figure A to predict wheel flat failures in rail system vehicles.

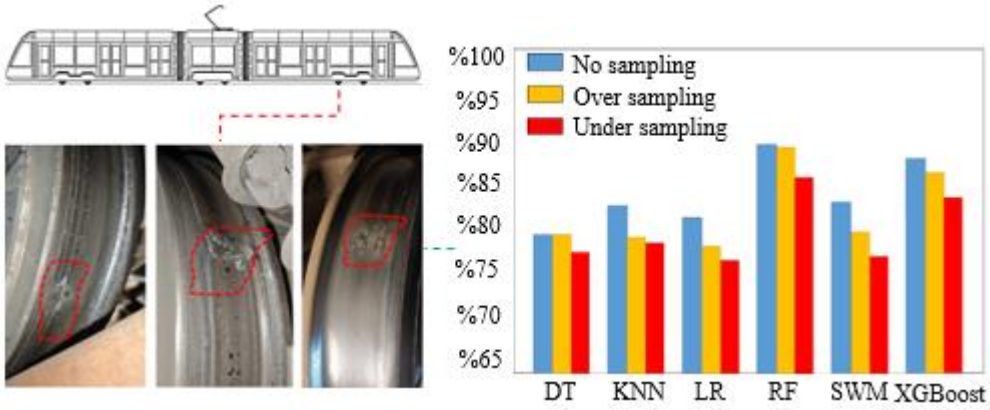


Figure A. Prediction of wheel flat failures using machine learning methods

Purpose:
This study aims to develop a machine learning framework for early prediction of wheel flat failures using operational and meteorological data.

Theory and Methods:
A dataset containing 1,562 records and 19 features was analyzed using six machine learning algorithms. Min–Max normalization, cross-validation, class-balancing methods, ensemble learning, and SHAP analysis were applied.

Results:
Random Forest achieved the best overall individual performance, while XGBoost provided the highest sensitivity. The RF + XGBoost hard-voting model achieved an accuracy of 0.8828, precision of 0.8984, sensitivity of 0.7522, and F1-score of 0.8182.

Conclusion:
The results demonstrate that ensemble machine learning can effectively support early prediction of wheel flat failures. Temporal variables and wheel diameter change were the most influential predictors, while meteorological variables provided complementary information.

Mühendislik Mimarlık Fakültesi Dergisi
Journal of The Faculty of Engineering and Architecture of Gazi University
Elektronik / Online ISSN : 1304 - 4915
Basılı / Printed ISSN : 1300 - 1884

Makine öğrenmesi yöntemleri ile raylı sistem araçlarında apleti arızalarının operasyonel ve meteorolojik veriler kullanılarak tahmini

Gökhan Arık[1,2], Gökhan Azizoğlu[3,4,5], Şengül Danışman[6]*, Ahmet Nusret Toprak[3]
[1]Erciyes Üniversitesi, Fen Bilimleri Enstitüsü, Makine Mühendisliği Anabilim Dalı, 38039, Kayseri, Türkiye
[2]Kayseri Ulaşım A.Ş. 38070, Kayseri, Türkiye
[3]Erciyes Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, 38030, Kayseri, Türkiye
[4]Erciyes Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Anabilim Dalı, 38039, Kayseri, Türkiye
[5]Sivas Cumhuriyet Üniversitesi, Mühendislik Fakültesi, Yazılım Mühendisliği Bölümü, 58140, Sivas, Türkiye
[6]Erciyes Üniversitesi, Mühendislik Fakültesi, Makine Mühendisliği Bölümü, 38030, Kayseri, Türkiye

Ö N E Ç I K A N L A R

- Entegre operasyonel ve meteorolojik veriler kullanılarak tekerlek apleti tahmini
- Demiryolu tekerlek apleti arızalarının erken tahmini için bir makine öğrenmesi yöntemi
- Yığınlama topluluk öğrenmesi tekniği kullanılarak yüksek doğrulukta tekerlek apleti tahmini

Makale Bilgileri
Araştırma Makalesi
Geliş: 30.03.2026
Kabul: 19.07.2026

DOI:
10.17341/gazimmfd.1919320

Anahtar Kelimeler:
Raylı sistemler, kestirimci bakım, apleti arızası, makine öğrenmesi, topluluk öğrenmesi

ÖZ
Raylı sistemlerde tekerlek yüzeyinde oluşan apleti arızaları, dinamik darbe yüklerini artırarak işletme güvenilirliğini ve bakım süreçlerini olumsuz etkilemektedir. Bu çalışmada, operasyonel ve meteorolojik değişkenler kullanılarak apleti arızalarının erken tahminine yönelik açıklanabilir bir makine öğrenmesi yaklaşımı geliştirilmiştir. Kayseri raylı sistem işletmesinden 2015–2024 döneminde elde edilen veriler birleştirilerek 1562 gözlemden oluşan bir veri seti hazırlanmıştır. Lojistik Regresyon, Karar Ağacı, Rastgele Orman, Destek Vektör Makineleri, K-En Yakın Komşu ve XGBoost algoritmaları karşılaştırılmıştır. Sınıf dengesizliğinin etkisini değerlendirmek amacıyla SMOTE, Tomek Bağları ve rastgele aşağı örnekleme yöntemleri uygulanmıştır. Optuna ile hiperparametre optimizasyonu gerçekleştirilmiş, veri sızıntısını önlemek için örnekleme işlemleri iç içe çapraz doğrulamanın yalnızca eğitim katlarında yürütülmüştür. Tekil modeller arasında Rastgele Orman, 0,8758 doğruluk ve 0,8118 F1-skor ile öne çıkmıştır. Rastgele Orman ve XGBoost modellerini birleştiren katı oylama yaklaşımı 0,8828 doğruluk, 0,8984 kesinlik ve 0,8182 F1-skor sağlamıştır. SHAP analizi, yılın günü, teker çapı değişimi ve yılın haftasının tahminlerde belirleyici olduğunu göstermiştir. Bulgular, mevcut işletme kayıtlarının açıklanabilir arıza tahmini ve bakım kararlarının desteklenmesi amacıyla kullanılabileceğini ortaya koymaktadır. Böylece ek sensör altyapısına bağımlı olmayan, saha verilerine dayalı ve bakım planlamasına uyarlanabilecek sürdürülebilir bir karar destek çerçevesi sunulmaktadır.

Prediction of wheel flat failures in railway vehicles using machine learning methods with operational and meteorological data

H I G H L I G H T S

- Wheel flat prediction using integrated operational and meteorological data
- A machine learning framework was developed for early prediction of railway wheel flat failures
- High-accuracy wheel flat prediction using stacking ensemble learning

Article Info
Research Article
Received: 30.03.2026
Accepted: 19.07.2026

DOI:
10.17341/gazimmfd.1919320

Keywords:
Railway systems, predictive maintenance, wheel flat defect, machine learning, ensemble learning

ABSTRACT
Wheel flat defects occurring on the wheel surfaces of railway vehicles increase dynamic impact loads, adversely affecting operational reliability and maintenance processes. This study develops an explainable machine learning approach for the early prediction of wheel flat failures using operational and meteorological variables. A dataset comprising 1,562 observations was constructed by integrating data collected from the Kayseri railway system during the 2015–2024 period. Six machine learning algorithms, namely Logistic Regression, Decision Tree, Random Forest, Support Vector Machines, K-Nearest Neighbors, and XGBoost, were comparatively evaluated. To assess the effects of class imbalance, SMOTE, Tomek Links, and random undersampling techniques were applied. Hyperparameter optimization was performed using Optuna, while resampling procedures were restricted to the training folds within nested cross-validation to prevent data leakage. Among the individual models, Random Forest achieved the best performance, with an accuracy of 0.8758 and an F1-score of 0.8118. The hard-voting ensemble combining Random Forest and XGBoost achieved an accuracy of 0.8828, precision of 0.8984, and an F1-score of 0.8182. SHAP analysis revealed that day of the year, wheel diameter change, and week of the year were the most influential predictors. The findings demonstrate that existing operational records can be utilized for explainable failure prediction and maintenance decision support. Thus, a sustainable, field-data-driven decision support framework is proposed that does not rely on additional sensor infrastructure and can be integrated into maintenance planning processes.

*Sorumlu Yazar / Yazarlar / *Corresponding Author / Authors: 4010640052@erciyes.edu.tr, gazizoglu@erciyes.edu.tr, *sdanisman@erciyes.edu.tr, antoprak@erciyes.edu.tr / Tel: +90 505 915 4220

1. Giriş (Introduction)

Raylı ulaşım sistemleri, sürdürülebilirlik, yüksek taşıma kapasitesi ve enerji verimliliği gibi önemli özellikleri sayesinde günümüzde hem kent içi hem de şehirlerarası ulaşım ağlarının temel bileşenlerinden biri hâline gelmiştir. Bununla birlikte, bu sistemlerde tekerlek–ray etkileşiminden kaynaklanan yüzey düzensizlikleri, güvenlik, işletme sürekliliği ve bakım maliyetleri açısından önemli mühendislik sorunları oluşturmaktadır. Özellikle tekerlek yüzeyinde meydana gelen apleti (wheel flat) arızaları, temas bölgesinde darbe etkilerinin artmasına, ray üst yapısında yorulma hasarının hızlanmasına ve sistem genelinde titreşim seviyelerinin yükselmesine neden olmaktadır [1, 2]. Dairesel formunu kaybetmiş (out-of-round) tekerleklerin dinamik etkilerine yönelik gerçekleştirilen kapsamlı çalışmalar, bu tür geometrik düzensizliklerin tekerlek–ray temasında önemli darbe yükleri oluşturduğunu ve bu durumun ray, travers, balast ile araç bileşenleri üzerinde ilave dinamik yükler meydana getirdiğini ortaya koymaktadır. Bu dinamik etkiler, altyapı ve araç sistemlerinin işletme performansını olumsuz yönde etkileyerek sistem genelinde erken hasar oluşumuna zemin hazırlayabilmektedir [3].

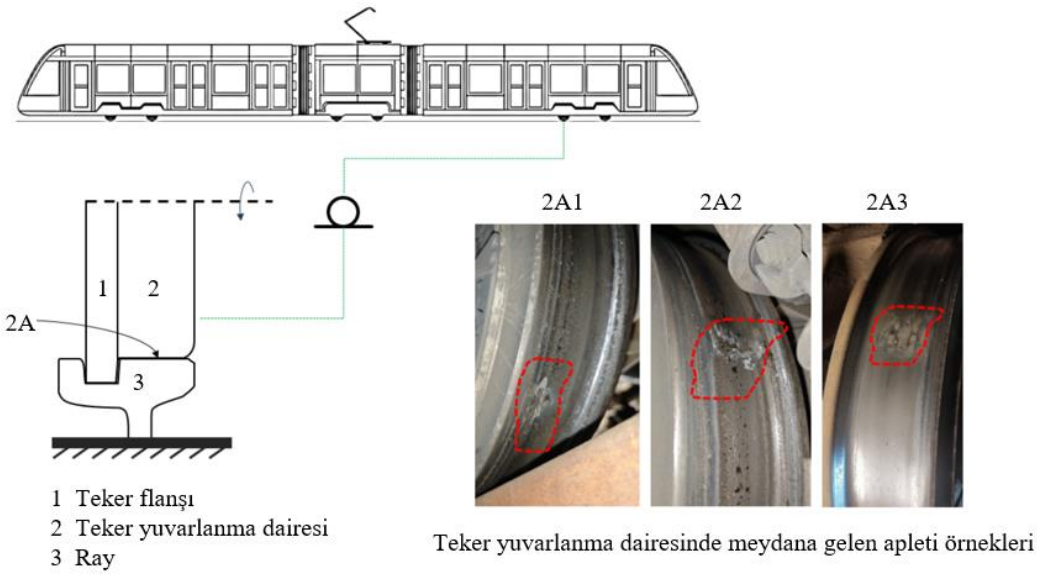
Apleti oluşumu, frenleme sırasında meydana gelen kayma, yüzey dökülmeleri (spalling) ve yüksek temas gerilmeleri gibi karmaşık mekanik süreçlerle ilişkilidir. Tekerlek–ray temas mekaniğine yönelik deneysel ve sayısal analizler, temas bölgesinde oluşan yerel gerilme yoğunlaşmalarının çatlak oluşumunu tetiklediğini ve çatlak ilerleme hızını artırdığını ortaya koymaktadır [4]. Özellikle yüksek hızlı trenler ve metro sistemlerinde, tekerlek düzensizliklerinin dinamik davranış üzerindeki etkileri daha belirgin hâle gelmektedir. Bu durum hem yolcu konforunu hem de altyapı güvenliğini doğrudan etkilemektedir [5]. Literatürde apleti tespitine yönelik yaklaşımlar genel olarak iki ana başlık altında toplanmaktadır: (i) fiziksel model temelli yöntemler ve (ii) veri odaklı makine öğrenmesi yaklaşımları. Fiziksel model temelli çalışmalarda, titreşim analizi, temas kuvveti hesapları ve frekans alanı incelemeleri gibi yöntemlerle yüzey düzensizliklerinin belirlenmesi amaçlanmaktadır [1]. Özellikle ray hattı boyunca yerleştirilen (wayside) sensör sistemleri aracılığıyla gerçekleştirilen titreşim ve spektral analiz çalışmaları, apleti tespitinde belirli başarı düzeylerine ulaşmıştır [6, 7]. Bununla birlikte, raylı sistem aracı üzerine entegre edilen (on-board) ölçüm sistemleri kullanılarak geliştirilen yöntemlerin gerçek zamanlı apleti tespiti sağlayarak bakım planlamasında etkin bir rol üstlenebildiği gösterilmiştir [8, 9]. Ancak hem hat boyunca konumlandırılan hem de araç üzerine entegre edilen sistemler, genellikle yüksek kurulum ve entegrasyon maliyetleri gerektirmektedir. Ayrıca operasyonel ve çevresel değişkenlerin birleşik etkilerini sınırlı ölçüde dikkate almaktadır.

Son yıllarda veri odaklı makine öğrenmesi yaklaşımlarının raylı sistem bakım süreçlerinde kullanımı giderek yaygınlaşmış ve bu alandaki uygulamalar önemli ölçüde artış göstermiştir. Geliştirilen yaklaşımlar arıza tahmini ve risk değerlendirme problemlerinde yüksek doğruluk oranları sunmaktadır [10, 11]. Özellikle ağaç tabanlı algoritmalar, destek vektör makineleri ve derin öğrenme mimarileri tekerlek arızalarının sınıflandırılmasında etkin biçimde kullanılmaktadır [7, 12]. Bunun yanında topluluk öğrenme (ensemble learning) yöntemlerinin tekil modellere kıyasla daha kararlı ve yüksek performanslı sonuçlar ürettiği gösterilmiştir [13]. Literatürdeki çalışmaların önemli bir bölümü ağırlıklı olarak fiziksel sensör verilerine dayanmaktadır [1, 4, 14]. Ancak, operasyonel parametreler (yük durumu, sefer yoğunluğu vb.) ile meteorolojik değişkenlerin arıza oluşum sürecine olan etkisi yeterince ele alınmamıştır.

Raylı sistem arıza tahmin çalışmalarında karşılaşılan temel sorunlardan bir diğeri veri setlerindeki sınıf dengesizliğidir. Gerçek işletme koşullarında arıza kayıtlarının sayısı genellikle olağan durum kayıtlarına kıyasla daha düşüktür. Bu durum, makine öğrenmesi algoritmalarının çoğunluk sınıfa eğilim göstermesine ve azınlık sınıfı yeterince öğrenmemesine yol açabilmektedir [15]. Dengesiz veri problemini gidermek amacıyla SMOTE gibi sentetik yukarı örnekleme teknikleri literatürde yaygın olarak kullanılmaktadır [16-18]. Ayrıca Tomek Bağları ve Rastgele Aşağı Örnekleme gibi yöntemler, sınıflar arasındaki örtüşmeyi azaltmak ve karar sınırını netleştirmek amacıyla önerilmektedir [18]. Bununla birlikte, örnekleme yöntemlerinin model performansına katkısı veri setinin yapısına, kullanılan algoritmaya ve doğrulama stratejisine bağlı olarak değişebildiğinden, bu yöntemlerin etkisinin karşılaştırmalı biçimde değerlendirilmesi önem taşımaktadır.

Örnekleme yöntemleri, sınıf dengesizliğinin azaltılmasında etkili olsa da modelin karar sınırını doğru biçimde öğrenmekte zorlandığı durumlarda tek başına yeterli olmayabilmektedir. Bu nedenle topluluk öğrenme stratejileri, raylı sistem arıza tahmini çalışmalarında önemli bir alternatif olarak değerlendirilmektedir. Topluluk öğrenme, birden fazla temel öğrenicinin çıktılarının birleştirilmesiyle daha güçlü ve genellenebilir modeller elde edilmesini amaçlayan bir makine öğrenmesi yaklaşımıdır [19]. Bu yöntemler literatürde genel olarak Torbalama, Artırma, İstifleme ve Oylama tabanlı yaklaşımlar altında ele alınmaktadır. Torbalama yöntemlerinde modeller paralel biçimde eğitilmekte ve elde edilen tahminler çoğunluk oylaması ya da ortalama alma yoluyla birleştirilmektedir. Artırma yöntemlerinde modeller ardışık olarak eğitilmekte ve her yeni model önceki modelin hatalarına odaklanarak genel hata kademeli biçimde azaltılmaktadır. İstifleme yaklaşımında farklı temel modellerin çıktıları bir üst seviye öğrenici tarafından yeniden değerlendirilirken, oylama tabanlı yaklaşımlarda farklı modellerin tahminleri çoğunluk oylaması veya sınıf olasılıklarının birlikte değerlendirilmesi yoluyla birleştirilmektedir [20]. Raylı sistemlerde makine öğrenmesi tabanlı arıza tahmin modellerinin yalnızca yüksek doğruluk sağlaması yeterli görülmemekte, model kararlarının bakım uzmanları ve karar vericiler tarafından yorumlanabilir olması da önem taşımaktadır. SHAP tabanlı açıklanabilir yapay zekâ yöntemleri, model tahminlerinde etkili olan değişkenlerin katkı düzeylerini ortaya koyarak sonuçların yorumlanmasına olanak sağlamaktadır [21]. Bu kapsamda Liu vd. [22], yüksek hızlı trenlerde arıza erken uyarısı için özellik katkısı ve nedensel çıkarıma dayalı bir makine öğrenmesi modeli önermiştir. Bu çalışma, SHAP yöntemiyle yüksek katkı sağlayan değişkenleri belirleyerek model çıktılarının uzman bilgisiyle birlikte yorumlanabileceğini göstermiştir. Benzer şekilde García-Méndez vd. [23], Porto metro işletmecisine ait veri akışlarını kullanarak raylı sistemlerde kestirimci bakıma yönelik açıklanabilir bir makine öğrenmesi çerçevesi geliştirmiştir. Bu çalışmada, çevrimiçi arıza tahmini, veri ön işleme ve görsel tabanlı açıklanabilirlik bileşenleri birlikte ele alınmıştır. Bahsi geçen çalışmalar, raylı sistemlerde açıklanabilir yapay zekâ yaklaşımlarının model performansını desteklemenin yanında, bakım kararlarının daha anlaşılır ve izlenebilir hale getirilmesi açısından önemli olduğunu göstermektedir. Bu doğrultuda mevcut çalışmada, apleti arızası tahmininde operasyonel ve meteorolojik değişkenlerin model kararına katkısını değerlendirebilmek adına SHAP analizinden yararlanılmıştır.

Bu çalışma, raylı sistem araçlarında görülen apleti arızalarını operasyonel ve meteorolojik verileri birlikte erken tahmin etmeyi amaçlamaktadır. Bu kapsamda, Kayseri Ulaşım A.Ş.'ye ait 2015–2024 dönemini kapsayan gerçek saha verilerinden oluşturulan veri seti üzerinde altı farklı makine öğrenmesi algoritması karşılaştırmalı olarak değerlendirilmiştir. Sınıf dengesizliğinin model performansı üzerindeki etkisini incelemek amacıyla farklı veri dengeleme stratejileri uygulanmış ve tekil modeller ile topluluk öğrenme yaklaşımları aynı deneysel çerçevede analiz edilmiştir. Ayrıca SHAP tabanlı açıklanabilirlik analizi kullanılarak operasyonel ve meteorolojik değişkenlerin model kararlarına katkısı yorumlanmıştır.



Şekil 1. Raylı sistem tekerleklerinde apleti oluşumunun şematik gösterimi ve saha örnekleri
(Schematic illustration of wheel flat formation on a railway wheel and field examples observed on the rolling surface)

Bu çalışmanın literatüre sağladığı temel katkılar aşağıda özetlenmiştir:

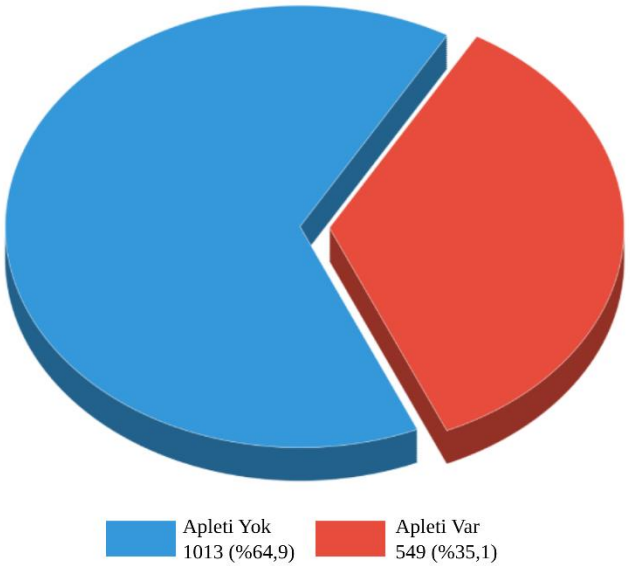
- Literatürde apleti arızalarına yönelik çalışmalar çoğunlukla sensör tabanlı tespit yaklaşımlarına odaklanırken, bu çalışmada gerçek işletme kayıtları, teker çapı değişimi, hat koşulları, yolcu yükü ve meteorolojik değişkenler birlikte kullanılarak apleti arızası erken risk tahmini problemi ele alınmıştır.
- Raylı sistem işletmeciliği açısından, apleti arızasının yalnızca oluştuktan sonra tespit edilmesi yerine, operasyonel ve çevresel koşullar altında arıza riskinin önceden tahmin edilmesine yönelik karar destek niteliğinde bir modelleme çerçevesi sunulmuştur.
- Gerçek saha verilerinde sık karşılaşılan sınıf dengesizliği problemi, veri sızıntısını önleyecek şekilde iç içe çapraz doğrulama yapısı içinde ele alınarak farklı örnekleme stratejilerinin model davranışı üzerindeki etkisi sistematik olarak karşılaştırılmıştır.
- SHAP tabanlı açıklanabilirlik analizi ile operasyonel ve meteorolojik değişkenlerin model kararına katkısı değerlendirilmiştir. Böylece model çıktılarının bakım uzmanları tarafından yorumlanabilir ve karar destek süreçlerine aktarılabilir hale getirilmesi sağlanmıştır.

2. Materyal ve Metot (Material and Method)

Bu çalışma, raylı sistem araç tekerleklerinde oluşan apleti arızalarının (wheel flat) erken aşamada tahmin edilmesini amaçlamaktadır. Apleti arızası, tekerleğin ray üzerinde kayarak dönmesi sonucu yuvarlanma dairesi üzerinde lokal düzleşme meydana gelmesiyle oluşan bir yüzey kusurudur. Bu düzleşme, tekerlek–ray temasında darbe etkisine neden olarak titreşim, gürültü artışı, konforsuz yolculuk ve ileri aşamalarda ray ile tekerlek profilinde ikincil hasarlara yol açabilmektedir. Şekil 1'de sahadan alınmış gerçek apleti örnekleri ve apletinin oluştuğu konum gösterilmektedir.

2.1. Apleti Veri Seti (Wheel Flat Dataset)

Apleti arızasının (wheel flat) tahminine yönelik olarak, Kayseri'de faaliyet gösteren Kayseri Ulaşım A.Ş. bünyesindeki raylı sistem araçlarına ait 2015–2024 yılları arasında toplanmış operasyonel veriler, aynı dönemlere ait meteorolojik verilerle entegre edilerek bütünleşik bir veri seti oluşturulmuştur. Oluşturulan veri seti toplam 1562 kayıttan oluşmakta olup, biri hedef değişken olmak üzere 19 özellik içermektedir. Hedef değişken, apleti arızasının var olma durumunu temsil eden ikili bir sınıf etiketidir. Şekil 2'de gösterildiği üzere kayıtların 549'unda (%35,1) apleti arızası gözlemlenmiştir.



Şekil 2. Veri setinde yer alan apleti hedef değişkeninin sınıf dağılımı
(Distribution of the wheel flat target class)

2.2. Veri Ön İşleme (Data Preprocessing)

Veri ön işleme aşamasında, veri setinde yer alan değişkenlerin analitik olarak kullanılabilir hâle getirilmesi amacıyla gerekli sayısal dönüşüm işlemleri gerçekleştirilmiştir. Bu kapsamda, veri setinde bulunan tüm değişkenler makine öğrenmesi modelleme sürecine uygun olacak şekilde sayısal formata dönüştürülmüştür. Veri setinde yer alan özelliklerin ayrıntılı tanımları ile eğitim öncesinde gerçekleştirilen veri dönüşüm süreci sonunda her bir değişkenin alabileceği değerler Tablo 1'de sunulmaktadır. Ayrıca veri setinin temel özelliklerini ortaya koymak ve değişkenlerin dağılım yapısını değerlendirmek amacıyla tanımlayıcı istatistiksel analizler gerçekleştirilmiştir. Bu analizler sonucunda elde edilen sayısal değişkenlere ilişkin istatistiksel bulgular Tablo 2'de, kategorik değişkenlere ait özet istatistikler ise Tablo 3'te sunulmuştur.

Tablo 1. Veri setinde kullanılan tüm özellikler, bu özelliklere ait açıklamalar ve uygulanan sayısal dönüşümler
(All features used in the dataset, their descriptions, and the applied numerical transformation)

Özellik Adı	Veri Tipi	Açıklama	Belirlenen Değerler
Boji Pozisyonu	Kategorik	Tekerleğin ait olduğu bojinin araç üzerindeki konumu	P = 0, MA =1, MB =2
Teker Çapı Değişimi	Sayısal	Tekerlek çapındaki aşınma veya tornalama sonrası oluşan değişim miktarı (mm)	Gerçek sayısal değer
Hat Kirlilik Durumu	Kategorik	Ray yüzeyindeki kir, yaprak, yağ vb. kayganlık artırıcı etkenlerin var olma durumu	Temiz = 0, Kirli = 1
Hat Yağlama Sistemi	Kategorik	Ray yağlama sisteminin aktif olup olmadığı	Aktif = 1, Pasif = 0
Ortalama Yolcu Sayısı	Sayısal	Sefer başına ortalama yolcu sayısı, yük etkisini temsil eder	Gerçek sayısal değer
Hava Durumu	Kategorik	Üç ardışık zaman adımındaki hava durumlarının birleşik kodu (1=Hava Açık, 2=Karlı, 3=Parçalı Bulutlu, 4=Hava Kapalı, 5=Yağmurlu)	111, 222, 233, 244, 253, 254, 333, 444, 533, 544, 555
Hava Sıcaklığı	Sayısal	Ölçüm zamanındaki ortam sıcaklığı (°C)	Gerçek sayısal değer
Hava Nem Seviyesi	Sayısal	Bağıl nem oranı (%)	0–100 arası sayısal değer
Kar Derinliği	Sayısal	Ray hattı çevresindeki kar kalınlığı (cm)	Gerçek sayısal değer
Çiğ Noktası	Sayısal	Havadaki nemin yoğunlaşmaya başladığı sıcaklık (°C)	Gerçek sayısal değer
Yıl	Sayısal	Kaydın ait olduğu takvim yılı	2015–2024 arası tam sayı
Ay	Kategorik	Yıl içindeki ay bilgisi	1–12
Yılın Haftası	Sayısal	Yıl içindeki hafta sırası	1–52
Yılın Günü	Sayısal	Yıl içindeki gün sırası	1–365/366
Yılın Çeyreği	Kategorik	Yılın hangi çeyreğinde olunduğu	Q1=1, Q2=2, Q3=3, Q4=4
Haftanın Günü	Kategorik	Haftanın günü bilgisi	Pazartesi=1… Pazar=7
Mevsim	Kategorik	Meteorolojik mevsim bilgisi	İlkbahar=1, Yaz=2, Sonbahar=3, Kış=4
Hafta Sonu Durumu	Kategorik	Günün hafta sonu olup olmadığı	Hafta sonu = 1, Hafta içi = 0
Apleti Arıza Durumu	Kategorik	Raylı sistem aracında apleti arızasının var olup olmadığını ifade eden hedef değişken	Var = 1, Yok = 0

Tablo 2. Veri setindeki sayısal özelliklere ait tanımlayıcı istatistikler (Descriptive statistics of the numerical features in the dataset)

Özellik Adı	Ortalama	Standart Sapma	En Küçük	En Büyük	Medyan
Teker Çapı Değişimi	5,13	3,26	0,22	30,70	4,24
Ortalama Yolcu Sayısı	2256,19	682,45	346,86	3514,76	2154,83
Hava Sıcaklığı	9,11	8,86	-14,80	28,00	7,60
Hava Nem Seviyesi	58,58	23,37	0,00	94,60	61,90
Kar Derinliği	3,09	6,94	0,00	42,30	0,00
Çiğ Noktası	7,65	19,70	-17,60	93,00	2,30
Yıl	2019,22	3,02	2015,00	2024,00	2019,00
Yılın Haftası	27,50	17,12	1,00	53,00	1,00
Yılın Günü	188,65	119,99	1,00	366,00	189,00

Tablo 3. Veri seti içerisindeki kategorik özelliklere ait tanımlayıcı istatistikler (Descriptive statistics of the categorical features in the dataset)

Özellik Adı	En Sık	Dağılım
Apleti Durumu	Yok (%64,9)	Yok (%64,9), Var (%35,1)
Boji Pozisyonu	MB (%35,08)	P (%32,65), MA (%32,27), MB (%35,08)
Hat Kirlilik Durumu	Temiz (%77,8)	Temiz (%77,8), Kirli (%22,2)
Hat Yağlama Sistemi	Pasif (%74,6)	Pasif (%74,6), Aktif (%25,4)
Hava Durumu	Parçalı Bulutlu (%38,7)	Parçalı Bulutlu (%38,7), Yağmurlu-Parçalı Bulutlu (%25,2), Hava Açık (%20,0), Karlı-Yağmurlu-Parçalı Bulutlu (%6,7), Karlı-Yağmurlu-Hava Kapalı (%5,1), Karlı-Parçalı Bulutlu (%2,0), Yağmurlu-Hava Kapalı (%1,3), Hava Kapalı (%0,4), Karlı-Hava Kapalı (%0,3), Yağmurlu (%0,3), Karlı (%0,1)
Ay	Aralık (%16,7)	Ocak (%13,7), Şubat (%7,5), Mart (%8,6), Nisan (%5,4), Mayıs (%7,2), Haziran (%6,0), Temmuz (%6,1), Ağustos (%5,1), Eylül (%8,8), Ekim (%6,3), Kasım (%8,6), Aralık (%16,7)
Yılın Çeyreği	Q4 (%31,6)	Q1 (%29,8), Q2 (%18,6), Q3 (%20,1), Q4(%31,6)
Haftanın Günü	Perşembe (%18,1)	Pazartesi (%12,7), Salı (%16,8), Çarşamba (%14,3), Perşembe (%18,1), Cuma (%14,6), Cumartesi (%11,3), Pazar (%12,1)
Mevsim	Kış (%37,9)	İlkbahar (%21,1), Yaz (%17,3), Sonbahar (%23,7), Kış (%37,9)
Hafta Sonu Durumu	Hafta İçi (%76,6)	Hafta İçi (%76,6), Hafta Sonu (%23,4)
Apleti Durumu	Yok (%64,9)	Yok (%64,9), Var (%35,1)
Boji Pozisyonu	MB (%35,08)	P (%32,65), MA (%32,27), MB (%35,08)

Veri setinde yer alan özelliklerin farklı ölçek ve değer aralıklarına sahip olması, özellikle mesafe temelli veya ağırlık katsayılarına duyarlı algoritmalarda model performansını doğrudan etkileyebilmektedir. Normalizasyon uygulanmadığı durumlarda, sayısal değeri yüksek olan değişkenler model tarafından daha baskın olarak değerlendirilebilmekte ve bu durum öğrenme sürecinde yanlılığa neden olabilmektedir [24].

Bu çalışmada, değişkenlerin farklı ölçeklerden kaynaklanan etkisini azaltmak amacıyla Min–Max Normalizasyonu kullanılmıştır. Bu yöntem, her bir değişkenin değerlerini kendi en küçük ve en büyük değerlerine göre yeniden ölçeklendirerek 0 ile 1 aralığına dönüştürmektedir. Böylece tüm değişkenler ortak bir aralıkta ifade edilmekte ve modellerin değişkenleri daha dengeli biçimde değerlendirmesi sağlanmaktadır [24]. Min–Max normalizasyonu Eş. 1'de gösterildiği şekilde hesaplanmaktadır.

$$X_N = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

Burada X_N normalizasyon sonucu elde edilen değeri, X veri değerini, X_{min} veri setindeki en küçük değeri ve X_{max} veri setindeki en büyük değeri ifade etmektedir.

2.3. Veri Dengeleme (Data Balancing)

Üretilen veri setinde hedef değişken olan apleti arıza durumu için sınıflar arası dengesizlik bulunmaktadır. Başlangıç veri seti toplam 1562 kayıttan oluşmakta olup, bu kayıtların 1013'ü "Apleti Yok" sınıfına, 549'u ise "Apleti Var" sınıfına aittir. Buna göre özgün veri setinde çoğunluk sınıfı olan "Apleti Yok" kayıtları toplam veri setinin %64,9'unu, azınlık sınıfı olan "Apleti Var" kayıtları ise %35,1'ini oluşturmaktadır. Bu durum, özellikle azınlık sınıfına ait örneklerin yetersiz temsili nedeniyle modelin çoğunluk sınıfına eğilim göstermesine ve genelleme performansının düşmesine yol açabilmektedir [24]. Bu nedenle çalışmada, sınıf dengesizliğinin model performansı üzerindeki etkisini incelemek amacıyla iki farklı veri dengeleme stratejisi uygulanmış ve sonuçları karşılaştırılmıştır.

İlk yaklaşımda, sınıf dengesizliğini gidermek amacıyla SMOTE (Sentetik Azınlık Aşırı Örnekleme Tekniği, Synthetic Minority Over-sampling Technique) [25] yöntemi kullanılmıştır. SMOTE, azınlık sınıfına (apleti var olma durumu) ait mevcut örnekler arasındaki komşuluk ilişkilerini dikkate alarak yeni ve sentetik örnekler üretmektedir. Bu yöntem, mevcut verileri çoğaltmakla kalmayıp, örnekler arasındaki doğrusal ara noktaları kullanarak veri çeşitliliğini artırmaktadır [26].

İkinci yaklaşımda iki aşamalı bir veri dengeleme işlemi uygulanmıştır. İlk aşamada, Tomek Bağları (Tomek Links) [27] yöntemi uygulanarak farklı sınıflara ait ve birbirine en yakın gözlem çiftleri belirlenmiştir. Bu çiftler, karar sınırında yer alan ve sınıflar arası örtüşmeye neden olabilecek örnekleri temsil etmektedir. Genellikle çoğunluk sınıfına ait olan bu gözlemler veri setinden çıkarılarak gürültü azaltılmış ve daha net bir karar sınırı oluşturulması amaçlanmıştır. İkinci aşamada ise Rastgele Aşağı Örnekleme (Random Under Sampling) [28] yöntemi uygulanmıştır. Bu yöntemde, çoğunluk sınıfına ait örnekler rastgele seçilerek veri setinden çıkarılmıştır. Böylece sınıflar arasındaki sayısal fark azaltılmış ve daha dengeli bir dağılım elde edilmiştir. Ancak bu yöntem, rastgele örnek çıkarmaktadır. Ayrıca, belirli düzeyde bilgi kaybı riski barındırmaktadır.

Veri sızıntısını önlemek amacıyla, veri dengeleme işlemleri iç içe çapraz doğrulama (nested cross-validation) yapısı içerisinde yalnızca ilgili eğitim katlarına uygulanmıştır. Bu kapsamda dış katlamada veri setinin %80'i eğitim, %20'si ise test amacıyla ayrılmıştır. Dış katlamadaki eğitim verisi, hiperparametre optimizasyonu için iç katlama (inner fold) sürecinde tekrar İç Eğitim (Inner Train) ve İç Doğrulama (Inner Validation) kümelerine bölünmüştür. SMOTE, Tomek Bağları ve Rastgele Aşağı Örnekleme işlemleri, yalnızca İç Eğitim kümesine uygulanarak İç Doğrulama ve dış test kümelerine herhangi bir örnekleme işlemi uygulanmamıştır. Böylece model, dengelenmiş iç eğitim kümesi üzerinde eğitilmiştir. İç katlama düzeyinde veri dengeleme işlemlerinin sınıf dağılımı üzerindeki etkisi Tablo 4 ve Şekil 3'te gösterilmektedir.

Tablo 4 incelendiğinde, örneklemesiz durumda İç Eğitim kümesinde sınıf oranının 1,85:1 olduğu görülmektedir. Yukarı örnekleme sonrasında azınlık sınıfına sentetik örnekler eklenerek iki sınıf 648'er örnekle dengelenmiş ve sınıf oranı 1:1'e getirilmiştir. Aşağı örnekleme sonrasında ise çoğunluk sınıfına ait örnek sayısı azaltılarak her iki sınıf 351'er örnekle temsil edilmiştir. Şekil 3, bu dengeleme işlemlerinin yalnızca İç Eğitim kümesiyle sınırlandırıldığını ve İç Doğrulama kümesinin özgün dağılımını koruduğunu göstermektedir. Bu yapı, modelin dengelenmiş veri üzerinde öğrenmesini sağlarken doğrulama performansının gerçekçi sınıf dağılımı üzerinde değerlendirilmesine olanak tanımaktadır.

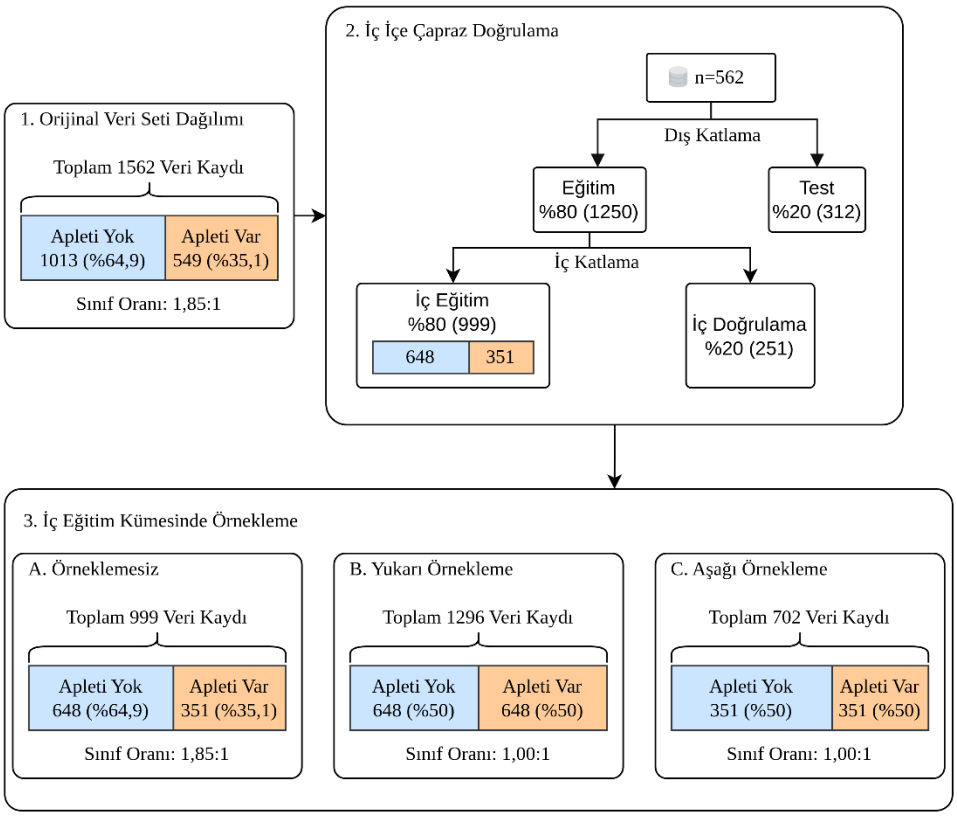
2.4. Makine Öğrenmesi Yöntemleri ile Apleti Arızasının Tahmini (Prediction of Wheel Flat Using Machine Learning)

Bu çalışmada, apleti arızasının tahmini amacıyla Lojistik Regresyon (Logistic Regression, LR) [29], Karar Ağacı (Decision Tree, DT) [30], Rastgele Orman (Random Forest, RF) [31], Destek Vektör Makineleri (Support Vector Machine, SVM) [32, 33], K-En Yakın Komşular (K-Nearest Neighbors, KNN) [34] ve Aşırı Gradyan Artırma (eXtreme Gradient Boosting, XGBoost) [35] olmak üzere altı farklı makine öğrenmesi algoritması kullanılmıştır. Bu algoritmalar doğrusal modeller, ağaç tabanlı yöntemler, örnek temelli yaklaşımlar ve gradyan artırma temelli yöntemler gibi farklı modelleme yaklaşımlarını temsil etmektedir. Böylece apleti arızasının tahmininde farklı yöntemlerin performansları karşılaştırmalı olarak değerlendirilmiştir.

Modellerin varsayılan parametreler ile sınırlı kalmaması ve her algoritmanın en uygun parametre yapısıyla değerlendirilmesi amacıyla hiperparametre optimizasyonu uygulanmıştır. Kullanılan algoritmaların hiperparametre arama uzayları sürekli, kesikli ve kategorik parametreler içerdiğinden, tüm olası kombinasyonların klasik ızgara aramasıyla denenmesi yüksek hesaplama maliyeti oluşturabilmektedir. Bu nedenle hiperparametre optimizasyonu, önceki denemelerden elde edilen performans bilgisini kullanarak daha başarılı parametre bölgelerine odaklanabilen Optuna kütüphanesi ve Ağaç Yapılı Parzen Tahmincisi (Tree-structured Parzen Estimator, TPE) tabanlı arama algoritması ile yürütülmüştür. Bu süreçte her bir

Tablo 4. İç katlama düzeyinde veri dengeleme işlemleri sonrasında sınıf dağılımındaki değişim
(Class distribution changes after data balancing at the inner-fold level)

Örnekleme Senaryosu	Toplam Eğitim Verisi	Apleti Yok Sınıfı	Apleti Var Sınıfı	Sınıf Oranı
Örneklemesiz	999	648 (%64,9)	351 (%35,1)	1,85:1
Yukarı Örnekleme	1296	648 (%50,0)	648 (%50,0)	1:1
Aşağı Örnekleme	702	351 (%50,0)	351 (%50,0)	1:1



Şekil 3. İç katlama düzeyinde veri dengeleme işlemlerinin uygulanma yapısı ve sınıf dağılımına etkisi
(Application structure of data balancing procedures and their effect on class distribution at the inner-fold level)

Tablo 5. Optuna ile hiperparametre optimizasyonunda kullanılan arama uzayları
(Hyperparameter search spaces used in Optuna-based hyperparameter optimization)

Algoritma	Hiperparametre Arama Uzayı
LR	Düzenlileştirme katsayısı (C) = 0.00001-1000; optimizasyon algoritması (solver)= “liblinear”, “lbfgs”, “saga”; maksimum iterasyon sayısı (max_iter) = 1000; ceza türü (penalty) = “l1”,“l2”, “elasticnet”, “none”; sınıf ağırlığı (class_weight) = “none”, “balanced”
RF	Ağaç sayısı (n_estimators) = 50-500; maksimum derinlik (max_depth) = “none” veya 3-30; bölünme için minimum örnek sayısı (min_samples_split) = 2-20; yaprak düğüm için minimum örnek sayısı (min_samples_leaf) = 1-20; özellik seçimi yöntemi (max_features) = “sqrt”, “log2”, “none”; sınıf ağırlığı (class_weight) = “none”, “balanced”, “balanced_subsample”; bölünme kriteri (criterion) = “gini”, “entropy”, “log_loss”
SVM	Düzenlileştirme parametresi (C) = 0.0001-1000.0; çekirdek fonksiyonu (kernel) = “linear”, “rbf”, “poly”, “sigmoid”; çekirdek katsayısı (gamma) = “scale”, “auto”, 0.0001-10; derece (degree) = 2-5; sınıf ağırlığı (class_weight) = “none”, “balanced”
KNN	Komşu sayısı (n_neighbors) = 1-30; ağırlıklandırma yöntemi (weights) = “uniform”, “distance”; uzaklık metriği (metric) = “minkowski”, “euclidean”, “manhattan”, “chebyshev”
DT	Maksimum derinlik (max_depth) = “none” veya 2-30; bölünme için minimum örnek sayısı (min_samples_split) = 2-20, yaprak düğüm için minimum örnek sayısı (min_samples_leaf) = 1-20; bölünme kriteri (criterion) = “gini”, “entropy”, “log_loss”; özellik seçimi yöntemi (max_features) = “sqrt”, “log2”, “none”; sınıf ağırlığı (class_weight) = “none”, “balanced”
XGBoost	Ağaç sayısı (n_estimators) = 50-600; maksimum derinlik (max_depth) = 2-12; öğrenme oranı (learning_rate) = 0.0001-0.4; alt örnekleme oranı (subsample) = 0.4-1.0; özellik alt örnekleme oranı (colsample_bytree) = 0.4-1.0; minimum çocuk ağırlığı (min_child_weight) = 1-20; L1 düzenlileştirme katsayısı (reg_alpha) = 0.00000001-10.0; pozitif sınıf ağırlığı (scale_pos_weight) = 0.5-5

algoritma için tanımlanan hiperparametre arama uzayı içerisinden en uygun parametre kombinasyonları, iç çapraz doğrulama performansına göre belirlenmiştir. Normalizasyon ve veri dengeleme işlemleri yalnızca ilgili eğitim katları üzerinde uygulanmıştır. Doğrulama ve test katlarında özgün veri dağılımı korunmuştur. Böylece veri sızıntısı önlenmiş ve modellerin performansları dış çapraz doğrulama katlarında elde edilen sonuçların ortalaması alınarak değerlendirilmiştir. İç içe çapraz doğrulama düzeninde en uygun hiperparametreler her dış kat (fold) için yeniden belirlenmektedir. Her bir algoritma için kullanılan hiperparametre arama uzayları Tablo 5’te ayrıntılı olarak sunulmuştur.

3. Sonuçlar ve Tartışmalar (Results and Discussions)

Bu çalışmada, raylı sistem araç tekerleklerinde meydana gelen apleti (wheel flat) arızalarının tahmin edilmesine yönelik olarak farklı makine öğrenmesi algoritmalarının performansları karşılaştırmalı biçimde değerlendirilmiştir. Bu amaç doğrultusunda literatürde yaygın olarak kullanılan LR, DT, RF, SVM, KNN ve XGBoost algoritmaları kullanılarak sınıflandırma modelleri oluşturulmuştur. Her bir algoritma, veri seti üzerinde aynı eğitim ve test koşulları altında uygulanarak modellerin tahmin performanslarının objektif bir şekilde karşılaştırılması hedeflenmiştir.

Geliştirilen modellerin performansları, sınıflandırma problemlerinde yaygın olarak kullanılan değerlendirme metrikleri aracılığıyla analiz edilmiştir. Bu kapsamda Doğruluk (Accuracy), Kesinlik (Precision), Duyarlılık (Recall), F1-Skor, Alıcı İşletim Karakteristiği Eğrisi Altında Kalan Alan (Receiver Operating Characteristic Area Under the Curve, ROC-AUC) [36], Kesinlik-Duyarlılık Eğrisi Altında Kalan Alan (Precision-Recall Area Under the Curve, PR-AUC) [37], Matthews Korelasyon Katsayısı (Matthews Correlation Coefficient, MCC) [38], Cohen Kappa Katsayısı (Cohen Kappa Coefficient) [39] ve karmaşıklık matrisi performans ölçütleri kullanılmıştır. Bu değerlendirme metrikleri kullanılarak farklı algoritmaların apleti arızalarını tahmin etme yetenekleri kapsamlı biçimde analiz edilmiş ve modellerin karşılaştırmalı performansları ortaya konulmuştur.

3.1. Veri Örnekleme Sonrası Model Performanslarının Değerlendirilmesi
(Evaluation of Model Performance After Data Sampling)

Veri setinin dengesiz sınıf dağılımına sahip olması nedeniyle, modellerin çoğunluk sınıfına yönelim eğilimini azaltmak amacıyla yukarı örnekleme (SMOTE) ve aşağı örnekleme (Tomek Bağları + Rastgele Aşağı Örnekleme) olmak üzere iki farklı veri dengeleme yaklaşımı uygulanmıştır. Tablo 6'da, farklı örnekleme senaryolarında dış çapraz doğrulama katlarından elde edilen ortalama sonuçları ve standart sapmaları sunulmaktadır. Ayrıca, algoritmaların doğruluk değerlerinin karşılaştırmaları Şekil 4'te grafiksel olarak gösterilmiştir.

Örnekleme uygulanmadan elde edilen bulgulara göre en yüksek doğruluk, kesinlik ve F1-Skor değerleri RF modelinde gözlenmiştir. RF modeli örneklemesiz durumda 0,8758 doğruluk, 0,8698 kesinlik, 0,7631 duyarlılık ve 0,8118 F1-Skor değerlerine ulaşarak tekil modeller arasında en dengeli genel performansı sağlamıştır. Örneklemesiz RF modelinin 0,9163 ROC-AUC, 0,8927 PR-AUC, 0,7242 MCC ve 0,7198 Cohen Kappa Katsayısı değerleriyle de öne çıktığı görülmektedir ile özellikle duyarlılık metriği açısından güçlü bir performans göstermiştir. Ayrıca XGBoost modelinin 0,9156 ROC-AUC ve 0,8934 PR-AUC değerlerine ulaşması, bu modelin pozitif sınıf olan "Apleti Var" durumunu ayırt etmede RF modeline oldukça yakın bir performans sunduğunu göstermektedir. Bu bulgular, örnekleme uygulanmadan da özellikle ağaç tabanlı topluluk modellerinin güçlü ve dengeli tahmin performansı sağlayabildiğini ortaya koymaktadır.

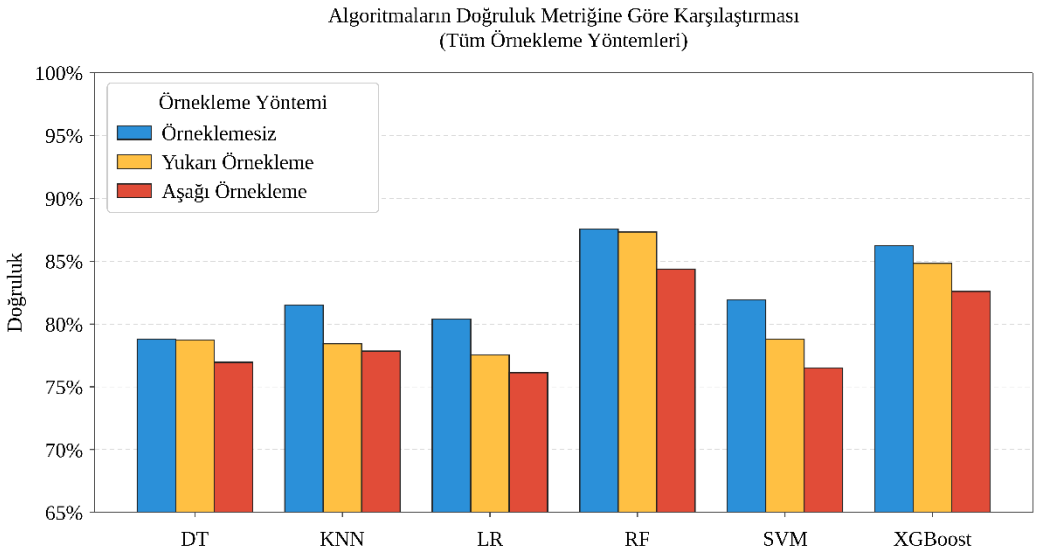
Yukarı örnekleme (SMOTE) uygulaması sonrasında özellikle LR, SVM, KNN ve DT modellerinde "Apleti Var" sınıfına ait örneklerin daha yüksek oranda doğru sınıflandırıldığı görülmüştür. Bu nedenle bu modellerde duyarlılık değerleri artmıştır. Ancak bu artış, modellerin genel performansına aynı ölçüde yansımamıştır. Çünkü duyarlılık yükselirken bazı modellerde kesinlik değerleri düşmüştür. Modeller apleti arızası olan örnekleri daha fazla doğru sınıflandırırken, apleti arızası olmayan bazı örnekleri de hatalı biçimde "Apleti Var" olarak tahmin etmiştir. Bu nedenle SMOTE, azınlık sınıfının öğrenilmesini kısmen desteklemiş olsa da doğruluk, kesinlik ve F1-skoru açısından genel ve tutarlı bir iyileşme sağlamamıştır. Özellikle LR, SVM ve KNN modellerinde duyarlılık artarken kesinlik değerlerinin görece düşmesi, modellerin apleti arızası bulunan örnekleri daha yüksek oranda doğru sınıflandırma eğilimi kazandığını, ancak bunun yanlış pozitif tahminleri artırabildiğini göstermektedir.

Tablo 6'ya göre SMOTE sonrası RF modeli 0,9178 ROC-AUC ve 0,8926 PR-AUC değerleri ile yüksek ayırt etme kapasitesini korumuştur. Bununla birlikte, RF modelinde MCC ve Cohen Kappa değerleri örneklemesiz duruma kıyasla sınırlı düzeyde azalmıştır. XGBoost modelinde ise SMOTE sonrası temel performans metriklerinin yanı sıra ROC-AUC, PR-AUC, MCC ve Cohen Kappa değerlerinde de düşüş gözlenmiştir. Bu nedenle SMOTE uygulamasının RF modeli için belirgin bir iyileşme sağlamadığı, XGBoost modelinde ise performans katkısı oluşturmadığı değerlendirilmiştir.

SMOTE uygulamasının genel performans metriklerinde tutarlı bir iyileşme sağlayamaması, apleti arızasının fiziksel oluşum süreciyle ilişkilendirilebilmektedir. SMOTE, azınlık sınıfına ait mevcut gözlemler arasında özellik uzayında doğrusal ara noktalar oluşturarak sentetik örnekler üretmektedir. Ancak apleti oluşumu yalnızca istatistiksel sınıf dağılımıyla açıklanabilecek bir durum değildir. Tekerlek–ray etkileşimi, frenleme davranışı, araç yükü, hat kirliliği ve yağlama durumu, bakım koşulları ile meteorolojik değişkenlerin birlikte etkisiyle ortaya çıkan karmaşık bir fiziksel süreçtir. Bu nedenle sentetik olarak üretilen bazı örneklerin gerçek saha koşullarında gözlenen fiziksel ilişkileri ve operasyonel korelasyonları tam olarak temsil etmesi garanti edilememektedir. Özellikle fiziksel olarak birlikte gerçekleşme olasılığı düşük değişken kombinasyonlarının sentetik örnekler içinde oluşması, modelin karar sınırını gerçek işletme koşullarından kısmen uzaklaştırabilmektedir. Bu durum, bu çalışmada SMOTE uygulamasının bazı modellerde duyarlılığı artırmasına rağmen doğruluk, kesinlik, F1-Skor, MCC ve Cohen Kappa gibi genel performans ölçütlerinde tutarlı bir iyileşme sağlayamamasının olası nedenlerinden biri olarak değerlendirilmektedir.

Aşağı örnekleme (Tomek Bağları + Rastgele Aşağı Örnekleme) uygulandığında LR, RF, SVM, KNN ve DT modellerinde duyarlılık değerlerinin arttığı görülmektedir. Bu sonuç, çoğunluk sınıfına ait örneklerin azaltılmasıyla modellerin azınlık sınıfına daha duyarlı hâle geldiğini göstermektedir. Bununla birlikte, aşağı örnekleme sonrasında genel eğilim doğruluk, kesinlik ve F1-Skor değerlerinde azalma yönündedir. Özellikle RF modelinde duyarlılık bakımından belirgin bir artış elde edilmesine rağmen kesinlik ve F1-Skor değerlerinin düşmesi, pozitif sınıfı tespit başarısındaki artışın yanlış pozitif tahminlerdeki artışla birlikte gerçekleştiğini göstermektedir. RF modelinin ROC-AUC değerinin 0,9016'ya, PR-AUC değerinin 0,8750'ye, MCC değerinin 0,6639'a ve Cohen Kappa değerinin 0,6628'e gerilediği görülmektedir. Benzer biçimde XGBoost modelinde de aşağı örnekleme örneklemesiz senaryonun gerisinde kalmıştır. Bu durum, çoğunluk sınıfından örnek çıkarılmasının karar sınırını temsil eden bazı bilgilerin kaybına yol açabileceğini ve modelin genelleme kapasitesini sınırlayabileceğini göstermektedir.

Genel olarak değerlendirildiğinde, veri dengeleme stratejilerinin etkisi modele bağlı olarak değişmiştir. SMOTE ve aşağı örnekleme yaklaşımları bazı modellerde duyarlılığı artırmış olsa da, bu artış çoğu durumda kesinlik, F1-Skor, MCC ve Cohen Kappa metriklerinde tutarlı bir iyileşme sağlamamıştır. En dengeli sonuçlar örneklemesiz RF modelinde elde edilirken, XGBoost modeli de duyarlılık ve PR-AUC bakımından güçlü bir alternatif olarak öne çıkmıştır. ROC-AUC ve PR-AUC değerleri, RF ve XGBoost modellerinin tüm örnekleme senaryolarında diğer algoritmalara kıyasla daha yüksek ayırt etme kapasitesine sahip olduğunu göstermektedir. MCC ve Cohen Kappa sonuçları ise özellikle örneklemesiz RF modelinin dengeli ve güvenilir bir sınıflandırma performansı sunduğunu desteklemektedir.

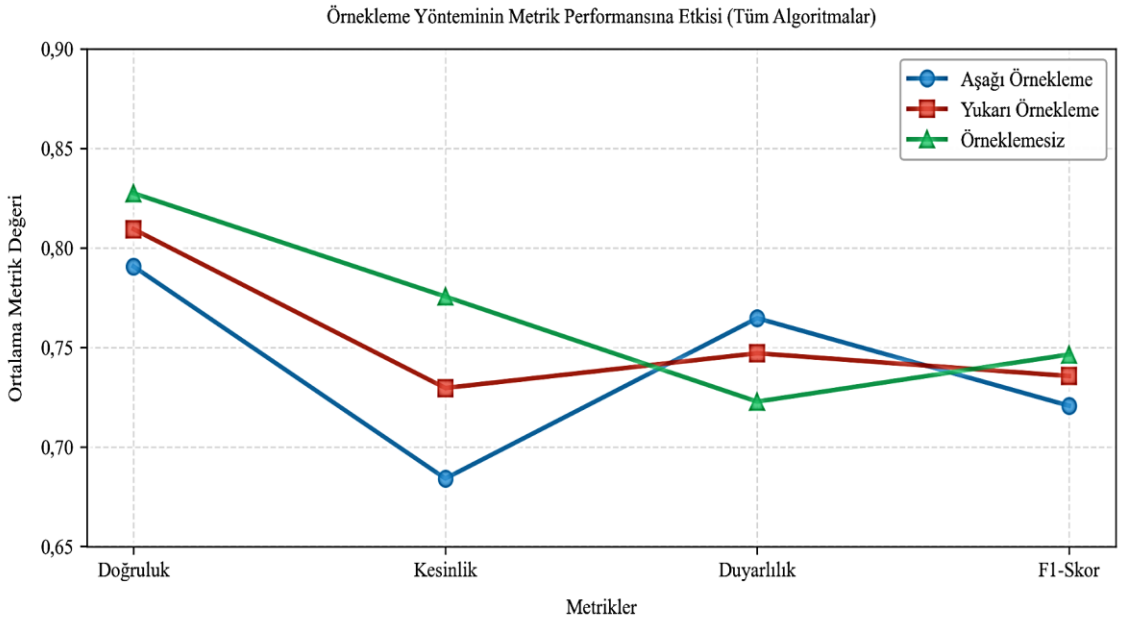


Şekil 4. Makine öğrenmesi algoritmalarının örnekleme öncesi ve sonrası doğruluk değerlerinin karşılaştırılması
(Comparison of the accuracy values of machine learning algorithms before and after sampling)

Tablo 6. Farklı örnekleme senaryolarında makine öğrenmesi algoritmalarının apleti arızası tahmin performansları
(Prediction performances of machine learning algorithms for wheel flat failure under different sampling scenarios)

Algoritma	Örnekleme Senaryosu	Doğruluk	Kesinlik	Duyarlılık	F1-Skor	ROC-AUC	PR-AUC	MCC	Cohen Kappa
LR	Örneklemesiz	0,8041 (0,0123)	0,7383 (0,0363)	0,6939 (0,0542)	0,7129 (0,0189)	0,8294 (0,0267)	0,7664 (0,0265)	0,5677 (0,0235)	0,5648 (0,0243)
	Yukarı Örnekleme	0,7753 (0,0270)	0,6595 (0,0357)	0,7467 (0,0412)	0,7002 (0,0360)	0,8248 (0,0287)	0,7577 (0,0328)	0,5244 (0,0576)	0,5216 (0,0572)
	Aşağı Örnekleme	0,7612 (0,0319)	0,6381 (0,0454)	0,7467 (0,0372)	0,6877 (0,0388)	0,8189 (0,0289)	0,7351 (0,0394)	0,5009 (0,0638)	0,4965 (0,0647)
RF	Örneklemesiz	0,8758 (0,0076)	0,8698 (0,0314)	0,7631 (0,0342)	0,8118 (0,0127)	0,9163 (0,0061)	0,8927 (0,0158)	0,7242 (0,0169)	0,7198 (0,0171)
	Yukarı Örnekleme	0,8732 (0,0148)	0,8637 (0,0451)	0,7631 (0,0294)	0,8091 (0,0186)	0,9178 (0,0069)	0,8926 (0,0147)	0,7189 (0,0321)	0,7147 (0,0304)
	Aşağı Örnekleme	0,8438 (0,0103)	0,7605 (0,0212	0,8124 (0,0075	0,7854 (0,0103)	0,9016 (0,0090)	0,8750 (0,0142	0,6639 (0,0183)	0,6628 (0,0194)
SVM	Örneklemesiz	0,8194 (0,0121)	0,7632 (0,0083)	0,7048 (0,0445)	0,7322 (0,0256)	0,8337 (0,0251)	0,7721 (0,0273	0,5981 (0,0306)	0,5964 (0,0321)
	Yukarı Örnekleme	0,7881 (0,0268)	0,6841 (0,0434)	0,7431 (0,0261)	0,7119 (0,0319)	0,8338 (0,0341)	0,7626 (0,0490	0,5465 (0,0529)	0,5449 (0,0539)
	Aşağı Örnekleme	0,7650 (0,0222)	0,6361 (0,0289)	0,7776 (0,0402)	0,6993 (0,0286)	0,8377 (0,0104)	0,7526 (0,0075	0,5173 (0,0456)	0,5097 (0,0456)
KNN	Örneklemesiz	0,8150 (0,0136)	0,7715 (0,0328)	0,6757 (0,0323)	0,7195 (0,0198)	0,8604 (0,0128)	0,8152 (0,0094)	0,5861 (0,0298)	0,5825 (0,0293)
	Yukarı Örnekleme	0,7842 (0,0109	0,6743 (0,0139)	0,7468 (0,0215)	0,7086 (0,0163)	0,8606 (0,0131)	0,8177 (0,0101)	0,5399 (0,0247)	0,5380 (0,0244)
	Aşağı Örnekleme	0,7785 (0,0279	0,6670 (0,0411)	0,7432 (0,0323)	0,7026 (0,0338)	0,8463 (0,0162)	0,7633 (0,0239)	0,5294 (0,0555)	0,5271 (0,0564)
DT	Örneklemesiz	0,7881(0,0175	0,6943 (0,0263)	0,7121 (0,0574)	0,7017 (0,0306)	0,8196 (0,0208)	0,7124 (0,0429)	0,5391 (0,0421)	0,5376 (0,0413)
	Yukarı Örnekleme	0,7874 (0,0244	0,6961 (0,0509)	0,7176 (0,0552)	0,7036 (0,0253)	0,8246 (0,0200)	0,7452 (0,0338)	0,5415 (0,0439)	0,5385 (0,0443)
	Aşağı Örnekleme	0,7695 (0,0086	0,6479 (0,0125)	0,7558 (0,0457)	0,6969 (0,0190)	0,8341 (0,0048)	0,7268 (0,0124)	0,5177 (0,0248)	0,5128 (0,0225)
XGBoost	Örneklemesiz	0,8623(0,0143	0,8161 (0,0271)	0,7868 (0,0316)	0,8006 (0,0209)	0,9156 (0,0112)	0,8934 (0,0154)	0,6964 (0,0309)	0,6956 (0,0314)
	Yukarı Örnekleme	0,8482 (0,0231	0,7997 (0,0553)	0,7649 (0,0408)	0,7802 (0,0301)	0,9073 (0,0144)	0,8831 (0,0155)	0,6666 (0,0491)	0,6645 (0,0482)
	Aşağı Örnekleme	0,8259(0,0119	0,7549 (0,0360)	0,7523 (0,0366)	0,7522 (0,0151)	0,8866 (0,0112)	0,8538 (0,0195)	0,6194 (0,0247)	0,6181 (0,0238)

Örnekleme yöntemlerinin ortalama metrikler üzerindeki etkisi Şekil 5'te ayrıca gösterilmektedir. Şekil 5 incelendiğinde, örneklemesiz yaklaşımın doğruluk, kesinlik ve F1-Skor metriklerinde genel olarak daha dengeli ve yüksek bir performans sunduğu görülmektedir. Buna karşılık, yukarı örnekleme ve aşağı örnekleme yaklaşımları özellikle duyarlılık metriğinde belirli bir artış oluşturmuştur. Bu durum veri



Şekil 5. Örnekleme yöntemlerinin ortalama performans metrikleri üzerindeki etkisi
(Effect of sampling methods on average performance metrics)

dengeleme yöntemlerinin, modelleri azınlık sınıfı olan "Apleti Var" durumuna karşı daha duyarlı hâle getirebildiğini göstermektedir. Ancak duyarlılıkta gözlenen bu artış, kesinlik, F1-Skor, MCC ile Cohen Kappa değerleriyle birlikte değerlendirildiğinde sınıflandırma başarısına tutarlı bir iyileşme olarak yansımamıştır. Dolayısıyla Şekil 5'teki ortalama metrik eğilimleri, Tablo 6'da model bazında gözlenen performans metriği sonuçları ile uyumludur.

Ayrıca kategorik değişkenlerin tekil kodlama (one-hot encoding) ile dönüştürülmesi sonucunda özellik uzayı genişlemektedir. Aşağı örnekleme sonrasında örnek sayısının azalması, bu genişleyen özellik uzayında bazı kategorik durumların daha az temsil edilmesine neden olabilmektedir. RF ve XGBoost gibi topluluk tabanlı yöntemler birden fazla ağacın birlikte çalışması ve kararları birleştirmesi nedeniyle bu bilgi kaybına karşı görece daha dayanıklı sonuçlar üretmiş olsa da, aşağı örnekleme senaryosunda bu modellerde de F1-Skor değerlerinin düştüğü görülmektedir. Bu nedenle Şekil 4 ve Tablo 6'da görülen performans farklılıkları, yalnızca örnekleme stratejisinden değil, aynı zamanda algoritmaların veri kaybına, özellik uzayına ve sınıf sınırlarının temsil gücüne olan duyarlılıklarından kaynaklanmaktadır.

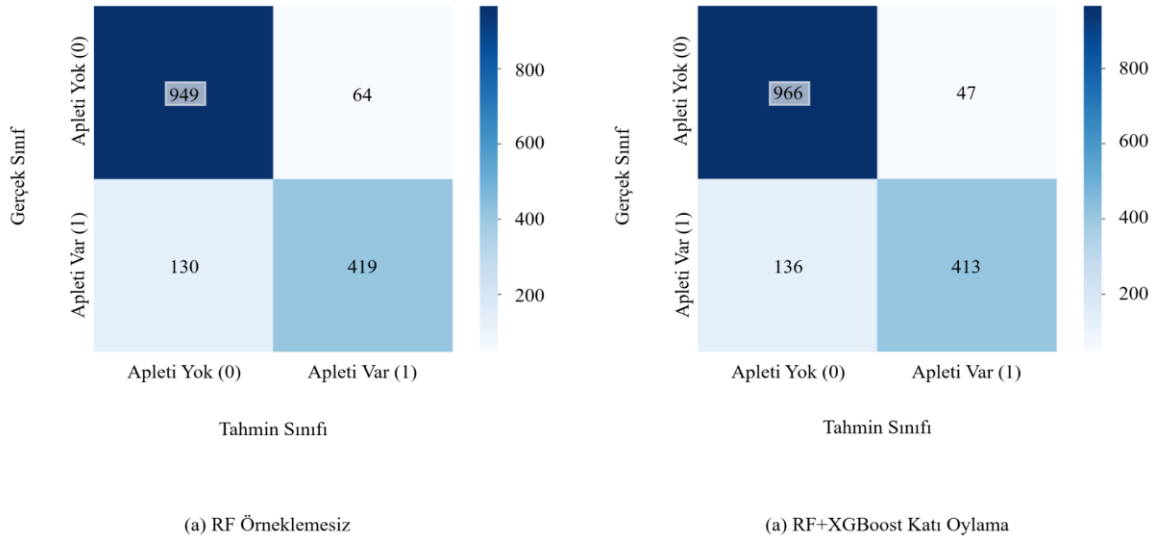
Sonuç olarak, iç içe çapraz doğrulama ve Optuna tabanlı hiperparametre optimizasyonu sonrasında elde edilen bulgular, örnekleme stratejilerinin model performansına katkısının sınırlı ve algoritmaya bağlı olduğunu göstermektedir. Yukarı örnekleme yaklaşımı LR, SVM, KNN ve DT modellerinde duyarlılığı artırmıştır. Ancak bu artış doğruluk, kesinlik ve F1-Skor metriklerinde genel ve tutarlı bir iyileşme oluşturmamıştır. Aşağı örnekleme yaklaşımı ise LR, RF, SVM, KNN ve DT modellerinde duyarlılığı artırmasına rağmen, çoğunluk sınıfına ait örneklerin azaltılmasına bağlı bilgi kaybı nedeniyle genel performansı sınırlamıştır. En dengeli tekil model performansı örneklemesiz RF modelinde elde edilmiştir. XGBoost modeli ise örneklemesiz durumda duyarlılık bakımından güçlü bir alternatif olarak öne çıkmıştır. Bu nedenle, çalışmada örnekleme stratejileri nihai performansı doğrudan artıran yöntemler olarak değil, sınıf dengesizliğinin model davranışı üzerindeki etkisini incelemek amacıyla karşılaştırmalı olarak değerlendirilen deneysel yaklaşımlar olarak ele alınmıştır.

3.2. Topluluk Öğrenme Tekniklerinin Performans Değerlendirmesi (Evaluation of Ensemble Learning Techniques)

Tekil modellerden elde edilen sonuçlar, özellikle RF ve XGBoost gibi ağaç tabanlı yaklaşımların dengeli bir performans sunduğunu göstermiştir. Bununla birlikte, farklı algoritmaların karar yapılarını bir araya getirerek daha kararlı ve genellenebilir tahminler üretip üretmediğini değerlendirmek amacıyla topluluk öğrenme teknikleri ayrıca incelenmiştir. Bu çalışmada, tekil modellerin performanslarının değerlendirilmesinin ardından farklı temel öğrenici kombinasyonlarından oluşan topluluk öğrenme yapıları karşılaştırmalı olarak analiz edilmiştir. Bu kapsamda LR, RF, SVM, KNN, DT ve XGBoost olmak üzere altı temel algoritma kullanılmıştır. Topluluk yapıları, en az iki ve en fazla altı temel modelin bir araya getirilmesiyle oluşturularak toplam 57 farklı model kombinasyonu elde edilmiştir. Her bir kombinasyon için katı oylama (hard voting) ve esnek oylama (soft voting) olmak üzere iki farklı oylama yaklaşımı değerlendirilmiştir. Ayrıca istifleme yaklaşımında üç farklı meta öğrenici kullanılmıştır. Buna ek olarak dört farklı temel algoritma için torbalama yapısı test edilmiştir. Böylece 114 oylama, 171 istifleme ve 4 torbalama deneyi olmak üzere toplam 289 farklı topluluk öğrenme deneyi gerçekleştirilmiştir. Gerçekleştirilen bu deneyler EK-1 Topluluk Öğrenmesi Deneyleri içerisinde verilmiştir.

Deneyler sonucunda, RF ve XGBoost modellerini içeren katı oylama yaklaşımının ortalama performans metrikleri açısından karşılaştırılan topluluk öğrenme modelleri arasında öne çıktığı görülmüştür. Bu model kombinasyonu ile doğruluk 0,8828, kesinlik 0,8984, duyarlılık 0,7522 ve F1-skoru 0,8182 değerlerine ulaşılmıştır. Elde edilen bulgular, RF+XGBoost tabanlı katı oylama yaklaşımının tekil modellere kıyasla ortalama doğruluk, kesinlik ve F1-Skor değerleri bakımından avantajlı bir sınıflandırma dengesi sunduğunu göstermektedir. En iyi tekil model olan RF ile karşılaştırıldığında, katı oylama yapısının doğruluk, kesinlik ve F1-Skor değerlerinde artış sağlaması, bu iki modelin karar yapılarının birbirini tamamladığını göstermektedir.

Tekil RF modeli ile en başarılı topluluk modeli olan RF+XGBoost katı oylama modelinin sınıf bazlı tahmin davranışlarını karşılaştırmak



(a) RF Örneklemesiz (a) RF+XGBoost Katı Oylama

Şekil 6. Tekil RF modeli ve RF+XGBoost katı oylama modeli için karmaşıklık matrisleri
(Confusion matrices for the single RF model and RF+XGBoost hard voting model)

Tablo 7. Tekil modeller için F1-Skor bazlı Friedman testi sonuçları (Friedman test results for individual models based on F1-Score)

Örnekleme Senayosu	Serbestlik Derecesi(df)	Ki-kare İstatistiği(x^2)	P Değeri
Örneklemesiz	5	19,7429	0,001396
Yukarı Örnekleme	5	17,5862	0,003512
Aşağı Örnekleme	5	18,8286	0,002069

amacıyla her iki modele ait karmaşıklık matrisleri Şekil 6'da birlikte sunulmuştur. Böylece topluluk öğrenme yaklaşımının yalnızca ortalama performans metrikleri üzerindeki etkisi değil, aynı zamanda yanlış pozitif ve yanlış negatif tahminler üzerindeki etkisi de karşılaştırmalı olarak değerlendirilmiştir.

Şekil 6 incelendiğinde, tekil RF modelinin "Apleti Yok" sınıfındaki örnekleri yüksek doğrulukla ayırt ettiği ve yanlış alarm üretme eğiliminin sınırlı kaldığı görülmektedir. Bununla birlikte, "Apleti Var" sınıfında bazı arızalı durumların "Apleti Yok" olarak sınıflandırılması, modelin belirli sayıdaki apleti arızasını kaçırabildiğini göstermektedir. RF+XGBoost katı oylama modeli ise tekil RF modeline kıyasla doğruluk, kesinlik ve F1-Skor değerlerinde sınırlı bir iyileşme sağlamıştır. Bu durum, topluluk öğrenme yaklaşımının farklı model kararlarını birleştirerek genel sınıflandırma performansını artırabildiğini göstermektedir.
Raylı sistem bakım süreçleri açısından yanlış negatif tahminler, apleti arızası bulunan durumların gözden kaçmasına neden olabileceğinden kritik öneme sahiptir. Yanlış pozitif tahminler ise arıza bulunmayan durumlarda aracın gereksiz kontrol veya bakıma gönderilmesine yol açabilir. Bu nedenle karmaşıklık matrisi sonuçları, RF+XGBoost katı oylama modelinin yalnızca ortalama metrikler açısından değil, bakım karar destek süreçlerinde oluşturabileceği hata türleri açısından da değerlendirilmesine olanak sağlamaktadır. Elde edilen sonuçlar, RF+XGBoost katı oylama modelinin riskli araç veya tekerleklerin önceliklendirilmesi amacıyla kullanılabilir bir kestirimci bakım karar destek performansı sunduğunu göstermektedir. Ancak yanlış negatif tahminlerin varlığı nedeniyle model çıktılarının nihai bakım kararı olarak değil, saha kontrollerini ve bakım planlamasını destekleyen erken uyarı bilgisi olarak değerlendirilmesi daha uygun olacaktır.
RF modelinin kararlı sınıflandırma yapısı ve yüksek kesinlik performansı ile XGBoost modelinin gradyan artırma temelli güçlü öğrenme kapasitesinin birlikte kullanılması, apleti arızası tahmininde daha dengeli bir topluluk modeli oluşturmuştur. Katı oylama mekanizması, temel modellerin sınıf tahminlerini çoğunluk kararına göre birleştirerek nihai sınıflandırmayı gerçekleştirmiştir. Bu yapı, özellikle yanlış pozitif tahminlerin sınırlanması ve genel sınıflandırma dengesinin desteklenmesi açısından avantaj sağlamıştır. Bununla birlikte, duyarlılık değerinin RF ve XGBoost tekil modellerine kıyasla daha yüksek olmaması, katı oylama yaklaşımının pozitif sınıfı tespit etme yönündeki eğiliminden ziyade genel karar güvenilirliği ve sınıflandırma dengesi açısından daha güçlü bir katkı sunduğunu göstermektedir.

Bununla birlikte, bu çalışmada kullanılan veri setinin 1562 gözlemden oluşması, değerlendirilen topluluk öğrenme kombinasyonlarının sayısı dikkate alındığında model genelleme performansı açısından dikkatli yorumlanması gereken bir sınırlılık oluşturmaktadır. Çok sayıda model kombinasyonunun sınırlı büyüklükteki bir veri seti üzerinde karşılaştırılması, bazı modellerin belirli çapraz doğrulama katlarına veya veri alt kümelerine daha iyi uyum sağlaması nedeniyle performans farklarının olduğundan daha belirgin görünmesine yol açabilmektedir. Bu nedenle topluluk öğrenme sonuçları yalnızca tek bir performans değerine dayalı olarak yorumlanmamıştır. Dış çapraz doğrulama katlarından elde edilen ortalama ve standart sapma değerleri, F1-Skor odaklı değerlendirme ve istatistiksel anlamlılık analizleri birlikte dikkate alınmıştır. Buna rağmen, önerilen modelin farklı işletme dönemleri, farklı hat koşulları veya daha geniş veri setleri üzerinde doğrulanması, modelin genelleme kapasitesinin daha güçlü biçimde ortaya konulması açısından gelecek çalışmalar için önemli bir gereklilik olarak değerlendirilmektedir.

3.3. İstatistiksel Anlamlılık Analizi (Statistical Significance Analysis)

Bu çalışmada elde edilen sınıflandırma sonuçlarının güvenilirliğini artırmak ve modeller arasındaki performans farklarının rastlantısal değişkenlikten kaynaklanıp kaynaklanmadığını değerlendirmek amacıyla istatistiksel anlamlılık analizleri gerçekleştirilmiştir. Veri setinin dengesiz sınıf yapısı dikkate alınarak istatistiksel karşılaştırmalarda ana değerlendirme metriği olarak F1-Skor

kullanılmıştır. Analizlerde, iç içe çapraz doğrulama sürecinin dış katlamalarında elde edilen kat (fold) bazlı F1-Skor değerleri dikkate alınmıştır.

İlk aşamada, altı farklı tekil makine öğrenmesi modelinin performansları örneklemesiz, yukarı örnekleme ve aşağı örnekleme senaryoları için ayrı ayrı değerlendirilmiştir. Bu kapsamda, LR, RF, SVM, KNN, DT ve XGBoost modelleri arasındaki genel performans farklılıklarını incelemek amacıyla parametrik olmayan Friedman testi [40] uygulanmıştır. Friedman testinin tercih edilmesinin nedeni, birden fazla modelin aynı çapraz doğrulama katları üzerinde karşılaştırılması ve performans değerlerinin normal dağılım varsayımını gerektirmeyen eşleştirilmiş bir test yapısına ihtiyaç duyulmasıdır.

Tekil modellere ait Friedman testi sonuçları Tablo 7'de sunulmaktadır. Tabloda serbestlik derecesi (df), karşılaştırılan model sayısına bağlı olarak hesaplanmıştır. Ki-kare istatistiği (χ^2), modellerin dış çapraz doğrulama katlarındaki F1-Skor sıralamaları arasındaki farklılığın büyüklüğünü göstermiştir. P-değeri ise modeller arasında F1-Skor açısından istatistiksel olarak anlamlı bir fark bulunup bulunmadığını değerlendirmek amacıyla kullanılmıştır.

Tablo 7 incelendiğinde, örnekleme senaryolarının tamamında p-değerlerinin 0,05 anlamlılık düzeyinin altında kaldığı görülmektedir. Bu sonuç, üç veri senaryosunda da LR, RF, SVM, KNN, DT ve XGBoost modellerinin F1-Skor performansları arasında istatistiksel olarak anlamlı farklılıklar bulunduğunu göstermektedir. En düşük p-değerinin örneklemesiz senaryoda elde edilmesi, modeller arasındaki performans ayrışmasının bu durumda daha belirgin olduğunu göstermektedir. Bununla birlikte, Friedman testi yalnızca modeller arasında genel bir fark olup olmadığını ortaya koymaktadır. Hangi model çiftleri arasında anlamlı fark bulunduğunu doğrudan göstermemektedir. Bu nedenle nihai karşılaştırmaya dahil edilen modeller için ek olarak post-hoc ikili karşılaştırmalar yapılmış ve önerilen RF+XGBoost katı oylama modeli diğer güçlü modellerle Wilcoxon işaretli sıra testi kullanılarak karşılaştırılmıştır.

Nihai modellerin karşılaştırılmasında, önerilen RF+XGBoost katı oylama modeli ile örneklemesiz RF, örneklemesiz XGBoost, en iyi istifleme modeli, en iyi esnek oylama modeli ve en iyi torbalama modeli dikkate alınmıştır. Bu modellerin dış çapraz doğrulama katlarında elde edilen F1-Skor değerleri kullanılarak Friedman testi uygulanmıştır. Friedman testi sonucunda modeller arasında istatistiksel olarak anlamlı bir genel fark olduğu belirlenmiştir (χ^2 = 11,0920; df = 5; p = 0,04959). Bu sonuç, modellerin F1-Skor performanslarının aynı düzeyde olmadığını göstermektedir.

Friedman testi sonrasında, en iyi RF+XGBoost katı oylama modeli referans alınarak diğer güçlü modellerle ikili karşılaştırmalar gerçekleştirilmiştir. İkili karşılaştırmalarda Wilcoxon işaretli sıra testi [41] kullanılmış ve çoklu karşılaştırmalardan kaynaklanabilecek Tip-I hata olasılığını azaltmak amacıyla p-değerleri Holm-Bonferroni yöntemiyle [42] düzeltilmiştir. Ayrıca modeller arasındaki ortalama F1-Skor farklarının yönünü ve büyüklüğünü göstermek amacıyla %95 güven aralıkları (GA) hesaplanmıştır. Elde edilen sonuçlar Tablo 8'de sunulmuştur

Tablo 8 incelendiğinde, RF+XGBoost katı oylama modelinin tüm nihai modellere göre pozitif ortalama F1-Skor farkı elde ettiği görülmektedir. Bu durum, önerilen modelin ortalama performans açısından diğer modellerin üzerinde yer aldığını göstermektedir. Bununla birlikte, Wilcoxon testi sonrasında uygulanan Holm-Bonferroni düzeltmesine göre ikili karşılaştırmaların hiçbiri 0,05 anlamlılık düzeyinde istatistiksel olarak anlamlı bulunmamıştır. Bu bulgu, önerilen modelin özellikle İstifleme, Esnek Oylama ve RF gibi güçlü modellerle arasındaki performans farkının sınırlı olduğunu göstermektedir. Buna karşın, önerilen RF+XGBoost katı oylama modeli nihai modeller arasında en iyi ortalama sıralama değerini elde etmiştir.

Genel olarak, istatistiksel anlamlılık analizi sonuçları iki temel bulguyu ortaya koymaktadır. İlk olarak, tekil modeller arasında tüm örnekleme senaryolarında F1-Skor bakımından istatistiksel olarak anlamlı farklılıklar bulunmaktadır. İkinci olarak, modeller arasında Friedman testi genel düzeyde anlamlı sonuç vermiştir. Ancak RF+XGBoost katı oylama modeli ile diğer güçlü modeller arasındaki ikili farklar çoklu karşılaştırma düzeltmesi sonrasında istatistiksel olarak anlamlı değildir.

3.4.Veri Setinde Yer Alan Özelliklerin Önem Derecelerinin Analizi (Analysis of Feature Importance in the Dataset)

Bu bölümde, topluluk öğrenme analizleri sonucunda ortalama performans ve model sadeliği açısından öne çıkan RF+XGBoost tabanlı katı oylama modeli için SHAP (Shapley Toplamsal Açıklamalar, SHapley Additive exPlanations) [21] yöntemi kullanılarak özelliklerin model kararına katkıları analiz edilmiştir. Analiz kapsamında hem küresel ölçekte özellik önem sıralaması hem de gözlem düzeyindeki SHAP etki dağılımı değerlendirilmiştir.

SHAP analizi, modelin yalnızca hangi değişkenleri önemli gördüğünü belirlemek amacıyla değil, aynı zamanda bu değişkenlerin apleti arızası tahminine hangi yönde katkı sunduğunu açıklamak amacıyla kullanılmıştır. SHAP değerleri yorumlanırken pozitif katkılar ilgili gözlemin "Apleti Var" sınıfına yaklaşmasına, negatif katkılar ise "Apleti Yok" sınıfına yaklaşmasına katkı sağlayan etkiler olarak değerlendirilmiştir. Böylece model karar mekanizması, yalnızca istatistiksel önem sıralaması üzerinden değil, operasyonel ve meteorolojik değişkenlerin mühendislik karşılıkları dikkate alınarak yorumlanmıştır.

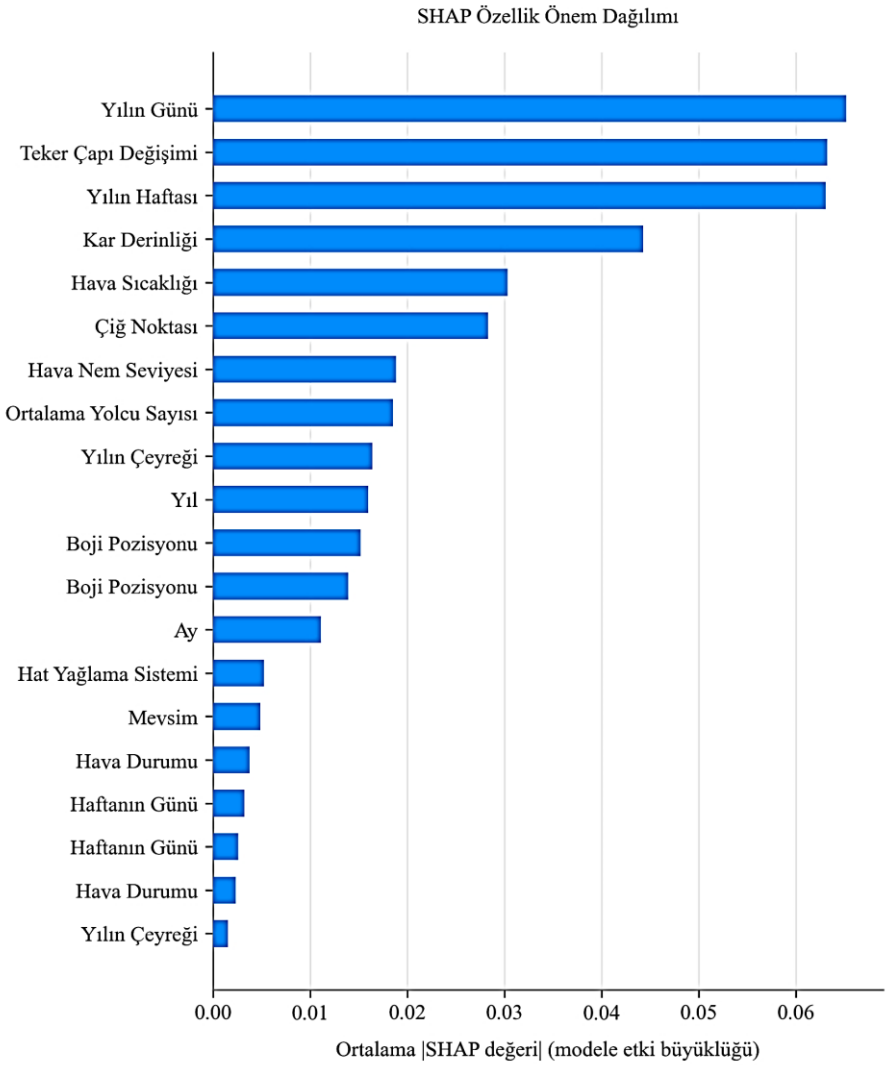
Şekil 7'de gösterilen SHAP özellik önem dağılımı grafiği incelendiğinde, model çıktısını en yüksek düzeyde etkileyen değişkenlerin sırasıyla Yılın Günü, Teker Çapı Değişimi ve Yılın Haftası olduğu görülmektedir. Bu üç değişkenin diğer özelliklerden belirgin biçimde ayrışması, apleti arızası tahmininde zamansal örüntüler ile tekerlek geometrisindeki değişimlerin model karar mekanizması üzerinde güçlü bir etkiye sahip olduğunu göstermektedir.

Yılın Günü ve Yılın Haftası değişkenlerinin üst sıralarda yer alması, apleti arızalarının yıl içindeki belirli dönemlerle ilişkili olabileceğine işaret etmektedir. Bu durum, takvimsel etkinin yanı sıra dönemsel işletme yoğunluğu, bakım çevrimleri, mevsim geçişleri, hat kullanım sıklığı, sıcaklık değişimleri ve yağış koşulları gibi birden fazla faktörün birlikte oluşturduğu operasyonel koşullar üzerinden değerlendirilmelidir. Dolayısıyla zamansal değişkenler, model açısından arıza oluşumuna zemin hazırlayan dönemsel risk koşullarını temsil eden önemli göstergeler olarak öne çıkmaktadır.

Teker Çapı Değişimi değişkeninin yüksek önem derecesine sahip olması ise tekerlek geometrisindeki değişimlerin apleti arızası tahmininde kritik bir operasyonel gösterge olduğunu ortaya koymaktadır. Tekerlek çapında meydana gelen değişimler, aşınma, tornalama işlemleri, profil bozulmaları veya tekerlek setleri arasındaki geometrik farklılıklarla ilişkili olabilmektedir. Bu tür değişimler ray-tekerlek temasındaki yük dağılımını, yuvarlanma davranışını ve frenleme anındaki temas koşullarını etkileyebilmektedir. Özellikle frenleme sırasında tekerleğin ray üzerinde kayma veya kızaklama eğiliminin artması, tekerlek yuvarlanma yüzeyinde lokal aşınma ve düzleşme oluşumunu destekleyen temel mekanizmalardan biridir. Bu

Tablo 8. RF+XGBoost katı oylama modeli ile son değerlendirmeye alınan modellerin ikili karşılaştırılması
(Pairwise comparison of the RF+XGBoost hard voting model and the models included in the final evaluation)

Karşılaştırma	Ortalama Fark (%95 GA)	Wilcoxon p-değeri	Holm-Bonferroni Düzeltilmiş p-değeri
RF+XGBoost Katı Oylama Modeli ile En İyi İstifleme Modeli (RF+DT+XGBoost)	+0,0012 [-0,0137; 0,0162]	0,6250	1,0000
RF+XGBoost Katı Oylama Modeli ile En İyi Esnek Oylama Modeli (RF+XGBoost)	+0,0055 [-0,0045; 0,0155]	0,3125	0,9375
RF+XGBoost Katı Oylama Modeli ile Rastgele Orman	+0,0064 [-0,0282; 0,0410]	0,6250	1,0000
RF+XGBoost Katı Oylama Modeli ile XGBoost	+0,0176 [-0,0069; 0,0420]	0,1875	0,7500
RF+XGBoost Katı Oylama Modeli ile En İyi Torbalama Modeli (DT)	+0,0487 [0,0136; 0,0838]	0,0625	0,3125



Şekil 7. SHAP özellik önem dağılımı grafiği (SHAP feature importance distribution plot)

nedenle modelin Teker Çapı Değişimi değişkenine yüksek önem atfetmesi, mühendislik açısından beklenen fiziksel süreçlerle uyumlu bir sonuç olarak yorumlanmaktadır.

SHAP önem sıralamasında Kar Derinliği, Hava Sıcaklığı, Çiğ Noktası ve Hava Nem Seviyesi gibi meteorolojik değişkenlerin de orta düzeyde etkili olduğu görülmektedir. Bu bulgu, çevresel koşulların apleti arızasını tek başına açıklayan bağımsız nedenler olmaktan çok, ray-tekerlek temas koşullarını etkileyen tamamlayıcı faktörler olarak model kararına katkı sunduğunu göstermektedir. Kar derinliği, buzlanma veya düşük aderanslı temas koşullarına işaret ederken, çiğ noktası ve nem seviyesi ise ray yüzeyinde yoğuşma veya ince su filmi



Şekil 8. SHAP özellik etki dağılımı grafiği (SHAP feature impact distribution plot)

oluşumuna işaret etmektedir. Bu koşullar frenleme veya kalkış sırasında tekerlek-ray temasındaki tutunmayı azaltarak kayma eğilimini artırabilmektedir.

Ortalama Yolcu Sayısı değişkeninin de orta-alt düzeyde katkı sağlaması, araç yükü ve operasyonel kullanım yoğunluğunun model tarafından dikkate alındığını göstermektedir. Yolcu yüküne bağlı olarak tekerlek-ray temas kuvvetleri ve frenleme davranışı değişebileceğinden, bu değişken arıza oluşumunu doğrudan açıklamaktan ziyade operasyonel yüklenme koşullarını temsil eden destekleyici bir gösterge olarak değerlendirilmektedir. Diğer tarafta Boji Pozisyonu, Ay, Hat Yağlama Sistemi, Mevsim, Hava Durumu, Haftanın Günü ve Yılın Çeyreği gibi değişkenlerin daha düşük SHAP önem değerlerine sahip olması, bu değişkenlerin model kararında daha sınırlı katkı sunduğunu göstermektedir.

Şekil 8'de yer alan SHAP özellik etki dağılımı grafiği, özelliklerin yalnızca önem düzeyini değil, aynı zamanda model çıktısını hangi yönde etkilediğini de göstermektedir. Grafikte her bir nokta bir gözlemi temsil etmekte, renkler ise ilgili değişkenin düşük veya yüksek değerlerini ifade etmektedir. Kırmızı renkler yüksek değişken değerlerini, mavi renkler ise düşük değişken değerlerini göstermektedir. Noktaların pozitif SHAP değerleri yönünde yer alması ilgili değişkenin "Apleti Var" tahminini artırıcı yönde katkı sunduğunu, negatif SHAP değerleri yönünde yer alması ise "Apleti Yok" tahminine katkı sağladığını göstermektedir.

Şekil 8 incelendiğinde, Yılın Günü ve Yılın Haftası değişkenlerinin belirli değer aralıklarında pozitif SHAP değerleri ürettiği görülmektedir. Bu durum, yılın belirli dönemlerinde apleti arızası riskinin model tarafından daha yüksek değerlendirildiğini göstermektedir. Ancak bu sonuç doğrudan takvimsel bir nedensellik olarak yorumlanmamalıdır. Yıl içindeki belirli dönemlerin işletme yoğunluğu, bakım planları, mevsimsel geçişler, yağış koşulları ve sıcaklık değişimleri gibi faktörlerin birleşik etkisini taşıdığı yönündedir.

Teker Çapı Değişimi değişkeninin SHAP etki dağılımı, bu değişkenin hem negatif hem de pozitif yönde belirgin etkiler oluşturabildiğini göstermektedir. Düşük çap değişimi değerleri bazı gözlemlerde "Apleti Yok" sınıfına katkı sağlarken, daha yüksek çap değişimi değerleri model çıktısını "Apleti Var" sınıfına yaklaştırabilmektedir. Bu eğilim, tekerlek geometrisindeki değişimlerin ray-tekerlek temas koşullarını bozarak apleti oluşum riskini artırabileceği yönündeki mühendislik beklentisiyle uyumludur. Bu nedenle Teker Çapı Değişimi, model açısından yalnızca istatistiksel olarak güçlü bir değişken değil, aynı zamanda fiziksel karşılığı olan kritik bir operasyonel gösterge olarak değerlendirilmiştir.

Meteorolojik değişkenler açısından Kar Derinliği, Hava Sıcaklığı, Çiğ Noktası ve Hava Nem Seviyesi değişkenlerinin SHAP dağılımları, çevresel koşulların model çıktısına yönlü katkı sağlayabildiğini göstermektedir. Özellikle Kar Derinliği değişkeninde yüksek değerlere karşılık gelen gözlemlerin pozitif SHAP bölgesinde yer alabilmesi, kar ve buzlanma koşullarının düşük aderanslı ray-tekerlek teması oluşturabileceğini düşündürmektedir. Bu koşullarda frenleme sırasında tekerlek kilitlenmesi veya kızaklama ihtimali artabilmektedir. Bu durum tekerlek yuvarlanma yüzeyinde lokal aşınma ve düzleşme oluşumuna sebep olabilmektedir.

Çiğ Noktası ve Hava Nem Seviyesi değişkenleri ise ray yüzeyinde yoğuşma, nemlenme veya ince su filmi oluşma olasılığıyla ilişkilendirilebilmektedir. Hava sıcaklığının çiğ noktasına yaklaşması veya bağıl nemin artması, temas yüzeyindeki aderans katsayısını düşürebilmektedir. Aderansın azalması, frenleme sırasında tekerlek ile ray arasındaki tutunmanın zayıflamasına ve mikro kayma oluşumuna neden olmaktadır. Bu nedenle bu değişkenler model açısından doğrudan arızayı oluşturan tekil nedenler olarak değil, kızaklama koşullarını kolaylaştıran çevresel göstergeler olarak yorumlanmaktadır.

Hava Sıcaklığı değişkeninin de SHAP dağılımında belirgin bir etkiye sahip olması, sıcaklık koşullarının ray-tekerlek temas davranışı üzerindeki dolaylı rolüne işaret etmektedir. Sıcaklık değişimleri, kar veya buz koşulları, yoğuşma, ray yüzeyi nemliliği ve frenleme dinamikleriyle birlikte değerlendirildiğinde apleti riskinin artmasına veya azalmasına sebep olabilmektedir. Bu nedenle Hava Sıcaklığı, tek başına belirleyici bir değişken olarak değil, diğer meteorolojik ve operasyonel değişkenlerle birlikte yorumlanması gereken tamamlayıcı bir çevresel faktördür.

Bununla birlikte, SHAP sonuçları doğrudan nedensel ilişki kanıtı olarak değerlendirilmemelidir. Bu analizler, modelin veri seti içerisinden öğrendiği ilişkileri açıklamakta ve hangi değişkenlerin tahmin sürecinde daha etkili olduğunu göstermektedir. Dolayısıyla Yılın Günü, Yılın Haftası, Teker Çapı Değişimi, Kar Derinliği, Çiğ Noktası, Hava Sıcaklığı ve Hava Nem Seviyesi gibi değişkenlerin etkisi, apleti arızasını tek başına tetikleyen bağımsız nedenler olarak değil, operasyonel dönemler, tekerlek geometrisi, ray-tekerlek aderansı, frenleme sırasında kayma/kızaklama eğilimi ve temas yüzeyinde lokal aşınma süreçleriyle ilişkili mühendislik göstergeleri olarak yorumlanmıştır.

Genel olarak SHAP analizi, apleti arızalarının tahmininde zamansal, operasyonel ve meteorolojik değişkenlerin birlikte değerlendirilmesi gerektiğini ortaya koymaktadır. Yılın Günü, Teker Çapı Değişimi ve Yılın Haftası değişkenleri model karar mekanizmasında en etkili göstergeler olarak öne çıkmaktadır. Bununla birlikte, Kar Derinliği, Hava Sıcaklığı, Çiğ Noktası ve Hava Nem Seviyesi gibi meteorolojik değişkenler ray-tekerlek temas koşullarını ve kayma eğilimini etkileyen tamamlayıcı çevresel faktörler olarak değerlendirilmiştir.

4. Sonuçlar (Conclusions)

Bu çalışma, raylı sistem araçlarında meydana gelen apleti (wheel flat) arızalarının tahmini için operasyonel ve meteorolojik değişkenleri bütünleşik biçimde ele alan veri odaklı bir makine öğrenmesi yaklaşımı önermektedir. Literatürde yaygın olarak yalnızca fiziksel sensör verilerine dayalı modeller geliştirilirken, bu çalışmada gerçek saha verileri kullanılarak çok boyutlu bir veri yapısı oluşturulmuş ve arıza oluşumunu etkileyen faktörler daha kapsamlı bir şekilde analiz edilmiştir.

Deneyler sonucunda, özellikle ağaç tabanlı topluluk modellerinin apleti arızası tahmininde güçlü ve dengeli sonuçlar sunduğu belirlenmiştir. Tekil modeller arasında RF modeli doğruluk, kesinlik ve F1-Skor açısından en başarılı performansı sergilemiştir. Duyarlılık bakımından en başarılı model olarak XGBoost öne çıkmıştır. Sınıf dengesizliğine yönelik uygulanan SMOTE tabanlı yukarı örnekleme ve Tomek Bağları + Rastgele Aşağı Örnekleme stratejileri, XGBoost haricindeki tüm modellerde duyarlılığı artırmış olsa da doğruluk, kesinlik ve F1-Skor metriklerinde genel ve tutarlı bir iyileşme sağlamamıştır. Bu çalışmada kullanılan veri seti üzerinde yapılan örnekleme işlemlerinin model bazında başarısını doğrudan artıran bir etkisi olmamıştır.

Farklı topluluk öğrenme yaklaşımlarının ve model kombinasyonlarının karşılaştırıldığı deneylerde, en yüksek başarıya RF+XGBoost modellerinin tahminlerinin dahil edildiği katı oylama yaklaşımı ile ulaşılmıştır. Bu model kombinasyonu ile 0,8828 doğruluk, 0,8984 kesinlik, 0,7522 duyarlılık ve 0,8182 F1-Skor değerlerine ulaşılmıştır. Elde edilen sonuçlar, RF modelinin kararlı sınıflandırma yapısı ile XGBoost modelinin güçlü öğrenme kapasitesinin birlikte kullanılmasının, tekil modellere kıyasla daha yüksek sınıflandırma başarısı sunduğunu göstermektedir. Özellikle yüksek kesinlik değeri, modelin apleti arızası olarak sınıflandırdığı durumlarda yanlış pozitif tahminleri sınırlama potansiyeline sahip olduğunu ortaya koymaktadır.

SHAP tabanlı açıklanabilirlik analizleri, model karar mekanizmasında en etkili değişkenlerin Yılın Günü, Teker Çapı Değişimi ve Yılın Haftası olduğunu ortaya koymuştur. Bu sonuç, apleti arızalarının yalnızca tekerlek geometrisindeki değişimlerle değil, aynı zamanda yıl içindeki dönemsel ve operasyonel koşullarla da ilişkili olduğunu göstermektedir. Teker Çapı Değişimi değişkeninin yüksek önem derecesine sahip olması, tekerlek geometrisindeki değişimlerin ray-tekerlek temas koşulları, yuvarlanma davranışı ve frenleme sırasında oluşabilecek kızaklama eğilimi açısından kritik bir gösterge olduğunu desteklemektedir. Meteorolojik değişkenlerin ise model kararına ikincil fakat dikkate değer katkılar sunduğu görülmüştür. Kar Derinliği, Hava Sıcaklığı, Çiğ Noktası ve Hava Nem Seviyesi gibi değişkenler, doğrudan arızayı oluşturan tekil nedenler olarak değil, ray-tekerlek temasındaki aderans koşullarını ve kayma eğilimini etkileyen tamamlayıcı çevresel göstergeler olarak yorumlanmıştır.

Apleti arızalarının erken tahmin edilmesi, teknik performansın artırılmasının yanı sıra işletme yönetimi süreçleri bakımından da stratejik bir potansiyel taşımaktadır. Çalışmada önerilen RF+XGBoost modellerinin tahminine dayanan katı oylama topluluk modeli, operasyonel ve meteorolojik veriler kullanılarak apleti oluşma riskinin önceden tahmin edilebileceğini ve bu tahminlerin bakım karar destek süreçlerinde kullanılabileceğini göstermektedir. Bu yaklaşım, özellikle plansız duruşların azaltılması, bakım faaliyetlerinin arıza riski oluşmadan önce planlanması ve bakım aralıklarının risk temelli olarak yeniden düzenlenmesi açısından önemli bir uygulama potansiyeli sunmaktadır. Ayrıca, arıza olasılığına dayalı çıktılar tekerlek seti ve ilgili yedek parçaların stok seviyelerinin talep odaklı biçimde planlanmasına katkı sağlayarak aşırı stok bulundurma veya kritik parça eksikliği risklerinin azaltılmasına imkân tanıyabilecektir. Gelecek çalışmalarda, önerilen modelin gerçek zamanlı veri akışları ile beslenen bakım yönetim sistemlerine ve kurumsal kaynak planlama (Enterprise Resource Planning, ERP) altyapılarına entegre edilmesi hedeflenmektedir.

Teşekkür (Acknowledgement)

Bu çalışmada kullanılan apleti verilerinin incelenmesine olanak sağlayan Kayseri Ulaşım A.Ş.'ye teşekkür ederiz. Bu çalışma, Erciyes Üniversitesi Bilimsel Araştırma Projeleri (BAP) Koordinasyon Birimi tarafından FDK-2023-13350 ve FDK-2026-15932 numaralı projeler kapsamında desteklenmiştir.

Kaynaklar (References)

1. Cui, Y., Kim, E., Yan, S., & Yu, Q., Wheel flat detection using long short-term memory and transformer models with a 1: 10 scale railway test rig, Advances in Mechanical Engineering, 17 (1), 16878132251314988, 2025.
2. Mosleh, A., Meixedo, A., Ribeiro, D., Montenegro, P., Calçada, R., Early wheel flat detection: an automatic data-driven wavelet-based approach for railways, Vehicle System Dynamics, 61 (6), 1644-1673,2023.
3. Iwnicki, S., Nielsen, J. C., Tao, G., Out-of-round railway wheels and polygonisation, Vehicle system dynamics, 61 (7), 1787-1830, 2023.
4. Fu, W., He, Q., Feng, Q., Li, J., Zheng, F., Zhang, B., Recent advances in wayside railway wheel flat detection techniques: A review, Sensors, 23 (8), 3916, 2023.
5. Xu, D., Wang, J., Yang, J., Wu, Y., Wen, X., Effect of Wheel Polygonalization on the Dynamic Characteristics of Gear-Transmission Systems of Urban Railway Vehicles, Machines, 13 (4), 323, 2025.
6. Salehi, M., Bagherzadeh, S. A., Fakhari, M., Experimental detection of train wheel defects using wayside vibration signal processing, Structural Health Monitoring, 22 (5), 3286-3301, 2023.
7. Mosleh, A., Ghiasi, R., Gordan, M., Ribeiro, D., Malekjafarian, A., 2026, Wayside monitoring for railway wheel flat identification: A multiclass supervised learning approach, Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, 240 (1), 112-129, 2026.
8. Ye, Y., Zhu, B., Huang, P., & Peng, B., OORNet: A deep learning model for on-board condition monitoring and fault diagnosis of out-of-round wheels of high-speed trains, Measurement, 199, 111268, 2022.
9. Shi, D., Ye, Y., Gillwald, M., & Hecht, M., Designing a lightweight 1D convolutional neural network with Bayesian optimization for wheel flat detection using carbody accelerations. International Journal of Rail Transportation, 9 (4), 311-341, 2021.
10. Bianchi, G., Fanelli, C., Freddi, F., Giuliani, F., & La Placa, A., Systematic review railway infrastructure monitoring: From classic techniques to predictive maintenance. Advances in Mechanical Engineering, 17 (1), 16878132241285631, 2025.
11. Wang, X., Liu, X., & Bian, Z., A machine learning based methodology for broken rail prediction on freight railroads: A case study in the United States, Construction and Building Materials, 346, 128353, 2022.
12. Shaikh, K., Hussain, I., & Chowdhry, B. S., Wheel defect detection using a hybrid deep learning approach, Sensors, 23 (14), 6248, 2023.
13. Bridgelall, R., & Tolliver, D. D., Railroad accident analysis using extreme gradient boosting, Accident Analysis & Prevention, 156, 106126, 2021.
14. Dong, Y., Han, C., Qu, S., Zhao, C., Madan, A., Fu, Y., Yang, Y., DAS-Accelerometer Data Fusion With Semi-Supervised Graph Variational Autoencoder for In-Service Train Wheel Flat Detection, IEEE Transactions on Intelligent Transportation Systems, 27 (1), 593-608, 2025.
15. Fu, W., He, Q., Liu, S., Feng, Q., Gao, R., Detection of Railway Wheel-Flat Based on CBAM-Enhanced ResNet for Imbalanced Data, IEEE Sensors Journal, 2025.
16. Kılıçarslan, S., Gögebakan, M., & Közkurt, C., Cervical cancer prediction using SMOTE algorithm and machine learning approaches, Journal of the Institute of Science and Technology, 13 (2), 747-759, 2023.
17. Ghofrani, F., Sun, H., & He, Q., Analyzing risk of service failures in heavy haul rail lines: A hybrid approach for imbalanced data, Risk analysis, 42 (8), 1852-1871, 2022.
18. Swana, E. F., Doorsamy, W., Bokoro, P., Tomek link and SMOTE approaches for machine fault classification with an imbalanced dataset, Sensors, 2 2(9), 3246, 2022.
19. Azizoglu, G., Toprak, A. N., MELD3: Integrating multi-task ensemble learning for driver distraction detection, IEEE Access, 12, 186022-186034, 2024.
20. Ou, J., Zhang, J., Li, H., Duan, B., Road damage prediction and intelligent maintenance methods based on stacking ensemble learning, Advanced Engineering Informatics, 66, 103466, 2025.
21. Lundberg, S. M., Lee, S. I., A unified approach to interpreting model predictions, Advances in neural information processing systems, 30, 2017.
22. Liu, D., Qin, Y., Zhao, Y., Yang, W., Hu, H., Yang, N., Liu, B., Fault early warning model for high-speed railway train based on feature contribution and causal inference, Sensors, 22 (23), 9184, 2022.
23. García-Méndez, S., de Arriba-Pérez, F., Leal, F., Veloso, B., Malheiro, B., Burguillo-Rial, J. C., An explainable machine learning framework for railway predictive maintenance using data streams from the metro operator of Portugal. Scientific Reports, 15 (1), 27495, 2025.
24. Zoralioğlu, Y., Gül, M. F., Azizoğlu, F., Azizoğlu, G., & Toprak, A. N., Predicting academic performance of students using machine learning techniques, In Innovations in Intelligent Systems and Applications Conference (ASYU), 1-6, 11-13 October, 2023.
25. Chawla, N. V., Bowyer, K. W., Hall, L. O., Kegelmeyer, W. P., SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research, 16, 321-357, 2002.
26. Alkan, A., Sevli, O., A classification study on predicting academic success using machine learning with oversampling and undersampling, Journal of the Faculty of Engineering and Architecture of Gazi University, 40 (4), 2191-2204, 2025.
27. Tomek, I., Two Modifications of CNN, IEEE Transactions on Systems Man and Communications SMC-6, 769-772, 1976.
28. He, H., Garcia, E. A., Learning from imbalanced data, IEEE Transactions on knowledge and data engineering, 21 (9), 1263-1284, 2009.
29. Cox, D. R., The regression analysis of binary sequences. Journal of the Royal Statistical Society Series B: Statistical Methodology, 20 (2), 215-232, 1958.
30. Breiman, L., Friedman, J., Olshen, R. A., Stone, C. J., Classification and regression trees, Chapman and Hall/CRC, 2017.
31. Breiman, L., Random forests, Machine learning, 45 (1), 5-32, 2001.
32. Cortes, C., Vapnik, V., Support-vector networks, Machine learning, 20 (3), 273-297, 1995.
33. Baş, S., Kiliçarslan, S., Elen, A., Közkurt, C., Adaptive arctan kernel: a generalized kernel for support vector machine, Sādhanā, 49 (2), 120, 2024.
34. Cover, T., Hart, P., Nearest neighbor pattern classification, IEEE transactions on information theory, 13 (1), 21-27, 1967.
35. Chen, T., Guestrin, C., Xgboost: A scalable tree boosting system, In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785-794, 13 - 17 August, 2016.
36. Fawcett, T., An introduction to ROC analysis, Pattern recognition letters, 27 (8), 861-874, 2006.
37. Davis, J., & Goadrich, M., The relationship between Precision-Recall and ROC curves, In Proceedings of the 23rd international conference on Machine learning, 233-240, 2006.
38. Matthews, B. W., Comparison of the predicted and observed secondary structure of T4 phage lysozyme, Biochimica et Biophysica Acta (BBA)-Protein Structure, 405 (2), 442-451, 1975.
39. Cohen, J., A coefficient of agreement for nominal scales, Educational and psychological measurement, 20 (1), 37-46, 1960.
40. Friedman, M., The use of ranks to avoid the assumption of normality implicit in the analysis of variance, Journal of the american statistical association, 32 (200), 675-701, 1937.
41. Wilcoxon, F., Individual comparisons by ranking methods, Biometrics bulletin, 1 (6), 80-83, 1945.
42. Holm, S., A simple sequentially rejective multiple test procedure, Scandinavian journal of statistics, 6 (2), 65-70, 1979.