



Machine learning applications in predictive maintenance: Data balancing and feature selection analysis

Ali Narin^{1*}, Nuriye Bektaşoğlu¹, Uğur İleri²

¹Department of Electrical and Electronics Engineering, Faculty of Engineering, Zonguldak Bülent Ecevit University, 67100, Zonguldak, Türkiye

²Department of Electrical-Electronics and Computer Engineering, Institute of Science, Duzce University, 81620, Düzce, Türkiye

Highlights:

- Comparison of SMOTE and Random Selection methods for data imbalance problem
- Z-score normalization performed better than Min-max normalization
- Feature selection with backward elimination method

Keywords:

- Predictive Maintenance
- Feature Selection
- Machine Learning
- Data Balancing

Article Info:

Research Article

Received: 14.04.2025

Accepted: 11.11.2025

DOI:

10.17341/gazimmfd.1676005

Correspondence:

Author: Ali Narin

e-mail: alinarin@beun.edu.tr

phone: +90 553 502 7125

Graphical/Tabular Abstract

Accurate prediction of machine failures plays a critical role in ensuring operational continuity, reducing costs, and increasing the efficiency of industrial systems. In predictive maintenance (PdM), it is crucial to determine critical parameters affecting equipment health status. However, high-dimensional sensor data can lead to increased computational load and decreased classification performance. In the study using AI4I 2020 dataset, a comprehensive method is proposed to overcome this challenge. The proposed approach includes data preprocessing steps, Synthetic Minority Oversampling Technique (SMOTE) to eliminate class imbalance, feature selection based on Backward Elimination method, and finally classification processes using Decision Tree (DT), Support Vector Machines (SVM), and k-Nearest Neighbors (k-NN) algorithms. The general flow of the methodology applied in the study is presented in Figure A.

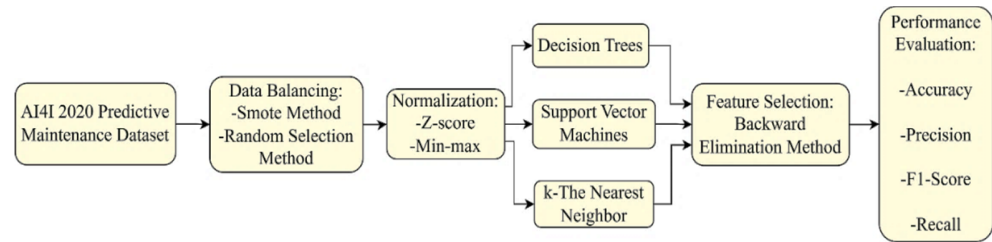


Figure A. Flowchart of the Proposed PdM Approach

Purpose: The main objective of this study is to improve the performance of predictive maintenance using machine learning techniques supported by feature selection. The research aims to optimize system reliability and minimize unexpected equipment failures by accurately predicting machine health status through data-driven methods.

Theory and Methods: In this study, AI4I 2020 dataset was used. In the first stage, data normalization was applied to ensure that all features were comparable and then the class imbalance was eliminated with the SMOTE method. The failure modes were divided into binary classes and normalization and feature selection processes were performed in each mode. Backward Elimination method was used for feature selection. Finally, the classification performance was evaluated with DT, SVM and k-NN algorithms and model performance was measured with metrics such as accuracy, precision, sensitivity and F1 score.

Results: In the analysis performed on the original unbalanced data set, the highest classification performance was obtained with 98.0% in the DT algorithm. However, in the data set balanced with SMOTE, the highest accuracy was achieved with 99.2% in the k-NN algorithm. In the evaluations specific to the failure modes, each failure mode was divided into binary classes and the optimal features were determined with Backward Elimination, and SVM and k-NN reached 100.0% accuracy in the PWF-OSF failure class. Similar high performances were obtained in other failure modes.

Conclusion: In this study, the analysis performed on the AI4I 2020 dataset showed that the data balanced with SMOTE increased the classification performance. The critical features determined by the Backward Elimination method significantly improved the performance of the classification algorithms, especially in error modes. The classifications made with DT, SVM and k-NN algorithms proved their effectiveness for PdM applications with high success criteria.



Kestirimci bakımda makine öğrenimi uygulamaları: Veri dengeleme ve öznelik seçimi analizi

Ali Narin^{1*}, Nuriye Bektaşoğlu¹, Uğur İleri²

¹Zonguldak Bülent Ecevit Üniversitesi, Mühendislik Fakültesi, Elektrik-Elektronik Mühendisliği Bölümü, 67100, Zonguldak, Türkiye

²Düzce Üniversitesi, Lisansüstü Eğitim Enstitüsü, Elektrik-Elektronik ve Bilgisayar Mühendisliği Bölümü, 81620, Düzce, Türkiye

ÖNEÇIKANLAR

- SMOTE, veri dengesizliğini gidermede rastgele seçime kıyasla daha yüksek başarı sağlamıştır
- Z-skor normalizasyonu, makine arızası sınıflandırmasında min-max normalizasyondan daha etkilidir
- Geriye doğru eleme yöntemi, ikili arıza sınıflarında gereksiz öznelikleri ayıklayarak doğruluğu artırmıştır

Makale Bilgileri

Araştırma Makalesi

Geliş: 14.04.2025

Kabul: 11.11.2025

DOI:

10.17341/gazimmfd.1676005

Anahtar Kelimeler:

Kestirimci bakım,
arıza sınıflandırması,
veri dengeleme,
geriye doğru eleme yöntemi

ÖZ

Makine öğrenimi ve nesnelerin internetindeki gelişmeler, kestirimci bakımı sanayi sistemleri için vazgeçilmez bir bileşen hâline getirmiştir. Ancak bu yaklaşımların başarımı, veri işleme süreçlerinin etkinliğine doğrudan bağlıdır. Özellikle veri dengesizliği, öznelik seçimi ve normalizasyon teknikleri, model performansını belirleyen temel unsurlar arasındadır. Bu çalışmada, AI4I 2020 veri seti kullanılarak makine arızalarının ve dört farklı arıza modunun sınıflandırılması hedeflenmiştir. Karar Ağaçları (DT), Destek Vektör Makineleri (SVM) ve k-En Yakın Komşu (k-NN) gibi sığ makine öğrenmesi algoritmalarının performansları kapsamlı biçimde analiz edilmiştir. Veri dengesizliğini gidermek amacıyla rastgele seçim ve SMOTE yöntemleri uygulanmış, SMOTE'un daha başarılı sonuçlar sunduğu görülmüştür. Ayrıca min-max ve z-skor normalizasyon teknikleri karşılaştırılmış ve z-skor normalizasyonunun sınıflandırma başarımını artırdığı belirlenmiştir. Arıza modları ikili sınıflara ayrılarak en uygun öznelik seti oluşturulmuş, ardından geriye doğru eleme yöntemiyle beş öznelik kademeli olarak çıkarılarak performansa etkileri incelenmiştir. Sonuçlar, önerilen yaklaşımın sınıflandırma performansını anlamlı biçimde iyileştirdiğini göstermektedir. Özellikle k-NN algoritması, SMOTE ile dengelenmiş veri setinde %99,2 doğruluk oranı elde ederek en yüksek başarımı sağlamıştır. Bu çalışma, kestirimci bakım uygulamalarının güvenilirliğini artırmaya yönelik özgün bir katkı sunmaktadır.

Machine learning applications in predictive maintenance: Data balancing and feature selection analysis

HIGHLIGHTS

- SMOTE has achieved higher success in eliminating data imbalance compared to random selection
- Z-score normalization is more effective than min-max normalization in machine failure classification
- The backward elimination has improved accuracy by removing unnecessary features in binary failure classes

Article Info

Research Article

Received: 14.04.2025

Accepted: 11.11.2025

DOI:

10.17341/gazimmfd.1676005

Keywords:

Predictive maintenance,
fault classification,
data balancing,
backward elimination
method

ABSTRACT

Advances in machine learning and the Internet of Things have made predictive maintenance an indispensable component of industrial systems. However, the effectiveness of these approaches is directly dependent on the efficiency of data preprocessing techniques. In particular, data imbalance, feature selection, and normalization methods are among the key factors that determine model performance. In this study, the classification of machine failures and four different failure modes is investigated using the AI4I 2020 dataset. The performance of shallow machine learning algorithms, including Decision Trees (DT), Support Vector Machines (SVM), and k-Nearest Neighbors (k-NN), is comprehensively analyzed. To address data imbalance, random sampling and the Synthetic Minority Over-sampling Technique (SMOTE) are applied, with SMOTE demonstrating superior performance. In addition, min-max and z-score normalization techniques are compared, and z-score normalization is found to enhance classification performance. Failure modes are further divided into binary classes to determine the most suitable feature set, after which five features are sequentially eliminated using the backward elimination method and their effects on classification performance are examined. The results indicate that the proposed approach significantly improves classification performance. Notably, the k-NN algorithm achieves the highest accuracy of 99.2% on the SMOTE-balanced dataset. This study provides an original contribution toward improving the reliability of machine learning-based predictive maintenance applications.

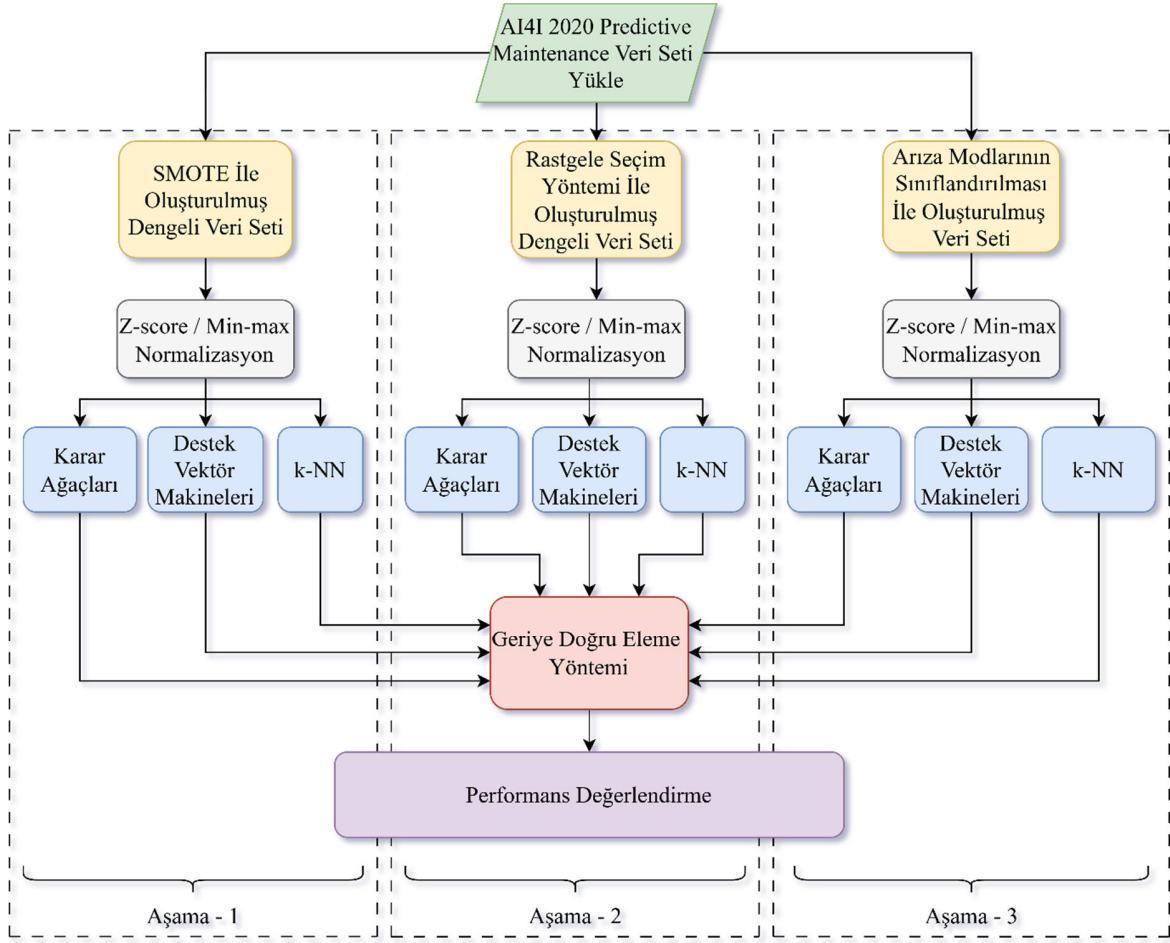
1. Giriş (Introduction)

Son yıllarda dijital teknolojiler ve akıllı sistemlerin sanayiye entegrasyonu, şirketlerin yapay zeka ve makine öğrenimi kullanarak yenilikçi çözümler geliştirmesine ve verimlilik artışları sağlamasına olanak tanımaktadır [1]. Özellikle endüstriyel nesnelerin interneti (Industrial Internet of Things-IIoT) ve makine öğrenimi teknolojilerinin gelişimi, ekipman performansını izleme, bakım ihtiyaçlarını öngörme ve üretim süreçlerini optimize etme gibi pek çok fayda sunmaktadır. IIoT, internet tabanlı nesnelerin endüstriyel verileri ürettiği, paylaştığı ve analiz ettiği ortak bir ağ platformu olarak tanımlanabilmektedir [2]. Bu teknoloji, Endüstri 4.0'ın üretim tesislerine entegre ettiği ileri teknolojiler sayesinde daha akıllı fabrikalar kurmayı, karmaşık ürünleri daha kısa sürede ve yüksek kalitede üretmeyi mümkün kılmaktadır. Üretim süreçlerini dijitalleştirerek ve otomasyonu artırarak daha verimli, sürdürülebilir ve akıllı üretim ortamları sunmaktadır [3]. Bu dönüşüm, ürünlerin ömür devri yönetiminde devrim niteliğinde değişikliklere yol açmıştır. Bu dijital dönüşümün önemli bir çıktısı olarak ömür devri yönetimi, ürünlerin tasarım aşamasından kullanım dışı kalışına kadar olan tüm süreçlerini kapsayan bir yaklaşım haline gelmiştir. İyi bir ömür devri yönetimi, kaynakların etkin kullanılmasını, operasyonel verimliliğin artırılmasını ve maliyetlerin optimize edilmesini sağlamaktadır. Bu yaklaşım ile özellikle kritik bileşenlerin sağlığını izlemek ve arıza risklerini minimize etmek ön plana çıkmaktadır. Ömür devri yönetiminin başarılı bir şekilde uygulanması, özellikle ürünlerin ve ekipmanların sağlık durumunun izlenmesi ve potansiyel arıza risklerinin önceden tespit edilmesiyle doğrudan ilişkilidir. Bu noktada kestirimci bakım planlaması (Predictive Maintenance - PdM), ömür devri yönetiminin önemli bir bileşeni olarak karşımıza çıkmaktadır. Kestirimci bakım planlaması, makine ve ekipmanların arıza risklerini minimize etmek için düzenli aralıklarla yapılan bakım işlemleridir [4]. PdM, makinelerin durumunu ve performansını sürekli olarak izleyerek potansiyel arızaları öngörmeyi hedeflemektedir. Bu modern yöntem, geleneksel reaktif bakım uygulamalarına kıyasla daha etkin bir çözüm sunmaktadır. Geleneksel yaklaşımlar, beklenmeyen duraklamalara, yüksek bakım maliyetlerine ve düşük verimliliğe yol açarken, PdM yalnızca maliyetleri düşürmekle kalmaz, aynı zamanda üretkenliği artırır, bakım süreçlerini optimize eder ve ekipman ömrünü uzatır. PdM'nin başlıca avantajları arasında makine performansındaki anormalliklerin tespiti, arıza modlarının tahmini, makine kontrolü ve yüksek ürün kalitesi sağlanması yer almaktadır. Bununla birlikte, makine ömrünün uzatılması, bakım sürelerinin optimize edilmesi ve yedek parça kullanımının iyileştirilmesi gibi ek katkılar da sunmaktadır. PdM, sanayi uygulamalarında genellikle taşınabilir cihazlar ve makinelere yerleştirilen sensörler aracılığıyla elde edilen verilerin işlenmesiyle gerçekleştirilmektedir [5]. Bu sistemler, ekipmanların durumunu izler ve sıcaklık, tork, dönme hızı, akım ve basınç gibi kritik parametreleri gerçek zamanlı olarak ölçer. Bu veriler ekipmanın mevcut durumu hakkında bilgi verir ve olası arızalar önceden tahmin edilerek bakım süreçlerini optimize eder. Bu sayede sanayi kuruluşları, makinelerde meydana gelebilecek problemleri beklenmeyen arızalara dönüşmeden çözme fırsatı bulmaktadır.

Son yıllarda, kestirimci bakım uygulamalarının etkinliğini artırmak amacıyla, sensörlerden elde edilen verilerin işlenmesi ve sınıflandırılması üzerine çeşitli makine öğrenimi yöntemleri geliştirilmiştir. Literatürde, dengesiz veri kümeleri için Bayes optimizasyonuna dayalı makine öğrenimi modelleri sunulmuş, bu yöntem ile model hiperparametrelerinin iyileştirilerek sınıflandırma performansının artırıldığı gösterilmiştir [6]. Başka bir çalışmada, frekans alanı titreşim sinyalleri kullanılarak endüksiyon motoru arızalarını tespit etmek için bir Destek Vektör Makinesi (SVM) algoritması kullanılmıştır [7].

Yatak arızaları üzerine yapılan başka bir çalışmada, zaman-frekans alanı analiz aracı olan dalgacık dönüşümü, titreşim sinyalleri ve öznelite çıkarma yöntemleriyle birleştirilerek, makine öğrenimi algoritmalarının yardımıyla aşım seviyelerinin doğru bir şekilde sınıflandırılması sağlanmıştır [8]. Bunun yanında, IoT ve IIoT tabanlı durum izleme sistemleri de PdM alanında önemli bir yer edinmiştir. Çakır ve arkadaşları tarafından yapılan bir çalışmada, Endüstri 4.0 uyumlu ve düşük maliyetli bir IIoT tabanlı durum izleme sistemi geliştirilmiş, rulman arızalarının tespitinde titreşim verilerinin yanı sıra akım, sıcaklık ve dönme hızı gibi farklı sensör verileri de kullanılmıştır. Toplanan veriler SVM, lineer ayraç analizi (LDA), Rastgele Orman (RF), Karar ağaçları (DT), ve k-En Yakın Komşular (k-NN) algoritmaları ile sınıflandırılmış, %99'un üzerinde doğruluk ve F1-skor değeri elde edilmiştir. Modeller arasında DT, pratik uygulanabilirliği, hızlı ve kabul edilebilir sınıflandırma oranları verebilme yeteneği ile en uygun yöntem olarak öne çıkmıştır [9]. Başka bir çalışmada, ORS 6203 tipi rulman arızalarının tespiti için IoT tabanlı bir durum izleme uygulaması geliştirilmiş ve titreşim, sıcaklık, akustik emisyon, akım ve devir ölçümleri sensörler aracılığıyla toplanmıştır. Veriler mikrodenetleyici tabanlı sistem ile Android cihazlara kablosuz olarak aktarılmış ve eşik değerlerinin aşılması durumunda SMS/e-posta bildirimleri sağlanmıştır. Toplanan veriler LDA, RF, SVM, k-NN ve DT algoritmaları ile sınıflandırılmış, tüm modellerde %96'nın üzerinde doğruluk oranı elde edilmiştir [10]. Shivaramu ve arkadaşlarının çalışmasında, endüstriyel operasyonlarda planlanmamış duruşlar ile yetersiz bakımın yol açtığı üretim kayıpları ele alınmıştır. Bu amaçla AI4I 2020 PdM veri seti kullanılarak RF, SVM ve Yapay Sinir Ağları (ANN) algoritmalarıyla ekipman arızalarının tahmini gerçekleştirilmiş; ayrıca tanımlama, ölçme, analiz etme, iyileştirme ve kontrol etme adımlarını içeren metodoloji kapsamında bakım faaliyetlerinin optimizasyonu hedeflenmiştir. Elde edilen sonuçlar, arıza oranının %3,39'dan %2,00'a gerilediğini ve bununla birlikte genel ekipman etkinliği ile süreç kararlılığında kayda değer bir artış sağlandığını ortaya koymuştur [11]. TechRxiv'de yayımlanan bir çalışmada, arıza eğilimleri ve bunların operasyonel maliyetlere etkileri incelenmiş, ısı dağılımı arızalarının bakım maliyetlerinin %74'ünü oluşturduğu tespit edilmiştir [12]. Bu durum, arızaların azaltılması için gelişmiş soğutma sistemleri, yük optimizasyonu ve güç dengeleyiciler gibi yenilikçi çözümleri gerekli kılmaktadır. Sengupta vd. zırlı araçların kestirimci bakımında makine öğrenimi yaklaşımlarını ele alan çalışmasında, sensörlerden elde edilen verilere dayalı olarak çeşitli algoritmalar uygulanmış ve sistemin performansı k-katlı çapraz doğrulama ve TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) analizi ile değerlendirilmiştir [13]. Assagaf ve arkadaşları tarafından yapılan bir çalışmada, SVM algoritmalarının, makine arızalarının tahmin edilmesi ve bakım süreçlerinin düzenlenmesinde önemli bir araç olduğu belirtilmiştir [14]. Başka bir çalışmada, İleri ve arkadaşları, otomatik arıza sınıflandırması için tek boyutlu derin öğrenme modelleri (1D-LeNet, 1D-AlexNet, 1D-VGG16) ile sığ makine öğrenimi algoritmalarını (DT, SVM, k-NN) karşılaştırmıştır [15]. Liao ve arkadaşlarının çalışması, geliştirilen topluluk ağaç modelinin, çeşitli performans metriklerinde %99'a yaklaşan başarıya ulaştığını; ayrıca mevcut makine öğrenimi yöntemlerine kıyasla daha yüksek doğruluk ve daha kısa eğitim süresi sunduğunu göstermiştir. [16]. Bu araştırmalar, makine öğrenimi ve derin öğrenme tabanlı yöntemlerin PdM alanında sağladığı ilerlemeleri ve AI4I 2020-PdM veri seti üzerinde gösterdiği performansı vurgulamaktadır. Yukarıda bahsedilen çalışmalar, kestirimci bakım uygulamalarındaki veri dengesizliği, sınırlı ölçeklenebilirlik ve sığ makine öğrenimi yöntemlerine olan bağımlılık, hata sınıflandırma modellerinin performansını sınırlamaktadır.

Bu çalışmada, Şekil 1'de şematik olarak sunulan kestirimci bakım sürecini geliştirmek amacıyla üç temel araştırma adımı sistematik biçimde gerçekleştirilmiştir.



Şekil 1. PdM kapsamında gerçekleştirilecek çalışmanın akış şeması (Flowchart of the work to be carried out within the scope of PdM.)

İlk olarak, veri dengesizliği sorununa yönelik çözümler geliştirilmiş; ardından farklı normalizasyon tekniklerinin sınıflandırma performansı üzerindeki etkileri incelenmiş; son olarak ise en uygun öznelik kümesini belirlemek için geriye doğru eleme yöntemi uygulanmıştır.

Analizler, AI4I 2020-PdM kestirimci bakım veri seti kullanılarak üç aşamalı bir yapı çerçevesinde yürütülmüştür. Şekil 1, metodolojik süreci bütüncül bir bakış açısıyla sunmakta ve aşamalar arasındaki ilişkileri görselleştirmektedir. Özellikle veri ön işleme, sınıflandırma ve performans değerlendirme adımlarının nasıl birbirine bağlandığını açıkça ortaya koyarak metodolojinin daha anlaşılır bir biçimde aktarılmasına katkı sağlamaktadır.

Şekil 1'de özetlenen bu aşamalar, ayrıntılı olarak şu şekilde açıklanmıştır:

Aşama-1: Sentetik Azınlık Aşırı Örneklemme Tekniği (Synthetic Minority Over-sampling Technique -SMOTE) Yöntemi ile Dengeli Veri Seti Oluşturulması

Veri setindeki sınıf dengesizliğini gidermek amacıyla SMOTE yöntemi uygulanmıştır. Bu yöntem, azınlık sınıfına ait sentetik örnekler üreterek veri dağılımını dengeli hale getirmektedir. Elde edilen dengeli veri kümesine z-skor normalizasyonu uygulanmış, ardından DT, SVM ve k-NN algoritmaları ile sınıflandırma işlemi gerçekleştirilmiştir. Modelin doğruluğunu artırmak ve gereksiz

öznelikleri elemek amacıyla geriye doğru eleme yöntemi uygulanmıştır. Son olarak, SMOTE yöntemi ile oluşturulan dengelenmiş veri setinde, modellerin başarısı doğruluk, kesinlik, duyarlılık ve F1 skoru gibi performans metrikleri ile değerlendirilmiştir.

Aşama-2: Rastgele Seçim Yöntemi ile Dengeli Veri Seti Oluşturulması

Alternatif bir veri dengeleme yöntemi olarak rastgele seçim yaklaşımı uygulanmıştır. Bu yöntemde, azınlık sınıfına ait örnekler belirli bir oranla rastgele seçilerek çoğaltılmıştır. Dengelenmiş veri seti, z-skor normalizasyonu ile ölçeklendirilmiş ve DT, SVM ve k-NN algoritmaları ile sınıflandırma işlemi gerçekleştirilmiştir. Ayrıca, modelin doğruluğunu artırmak ve gereksiz öznelikleri elemek amacıyla geriye doğru eleme yöntemi uygulanmıştır. Son olarak, bu aşamada da modellerin başarısı doğruluk, kesinlik, duyarlılık ve F1 skoru gibi performans metrikleriyle değerlendirilmiştir.

Aşama-3: Arıza Modlarının Sınıflandırılması ile Veri Seti Oluşturulması

Arıza modlarının daha ayrıntılı analiz edilmesini sağlamak amacıyla, arıza modları kendi içinde ikili sınıflara ayrılarak bir veri seti oluşturulmuştur. Bu veri setine z-skor normalizasyonu uygulanmış ve DT, SVM ve k-NN algoritmaları kullanılarak sınıflandırma işlemi gerçekleştirilmiştir. Modelin doğruluğunu artırmak ve gereksiz

öznitelikleri elemek için geriye doğru eleme yöntemi kullanılarak, istatistiksel olarak anlamlı olmayan öznitelikler aşamalı şekilde çıkarılmıştır. Elde edilen modellerin performansı, doğruluk, kesinlik, duyarlılık ve F1 skoru gibi değerlendirme metrikleri üzerinden analiz edilerek, arıza modlarının sınıflandırılmasına yönelik en etkili yöntem belirlenmiştir.

Gerçekleştirilen her bir aşamada, en yüksek başarımlı oranına sahip veri işleme yöntemi ve makine öğrenmesi modeli belirlenmiştir. Bu araştırmalar, ilgili alanda kullanılan veri işleme ve makine öğrenimi yöntemlerine katkı sağlayacak özgün yaklaşımların geliştirilmesine olanak tanımıştır. Bu bağlamda çalışmamızın özgün katkıları şu şekilde özetlenebilir:

- SMOTE ve rastgele seçim yöntemleri birlikte değerlendirilerek dengesizlik giderme stratejilerinin kestirimci bakım performansına etkileri ayrıntılı biçimde analiz edilmiştir;
- Geriye doğru eleme yöntemiyle kestirimci bakım için en kritik öznitelikler belirlenmiş, böylece model performansı artırılmıştır;
- Arıza modları ikili sınıflara ayrılarak veri seti yeniden yapılandırılmış ve daha hassas hata tespiti sağlanmıştır;
- Farklı normalizasyon yöntemleri (z-skor ve min-max) sistematik olarak karşılaştırılmış, z-skor'un daha yüksek doğruluk sağladığı ortaya konmuştur;
- DT, SVM ve k-NN algoritmaları farklı senaryolarda karşılaştırmalı olarak değerlendirilmiş ve kestirimci bakım süreçleri için en uygun modeller tartışılmıştır;
- Elde edilen bulgular, özellikle savunma sanayii ve ağır sanayi gibi kritik sektörlerde kestirimci bakım süreçlerinin nasıl daha verimli hale getirilebileceğine dair uygulamaya yönelik önerilerle desteklenmiştir.

Bu yönleriyle çalışma, AI4I 2020-PdM veri seti üzerine yapılmış önceki araştırmalardan ayrışmakta ve literatüre yenilikçi katkılar sunmaktadır. Bu çalışmada kullanılan AI4I 2020 veri seti, her ne kadar doğrudan bir endüstriyel üretim hattından elde edilmemiş olsa da, gerçek süreçlerde karşılaşılan tipik arıza modlarını ve operasyonel değişkenleri temsil etmesi bakımından kestirimci bakım araştırmaları için geçerli bir temel sağlamaktadır. Makine öğrenmesi ve veri dengeleme yöntemlerinin bu veri seti üzerinde uygulanması, özellikle arıza türlerinin dengesiz dağılımına rağmen yüksek doğruluk oranları elde edilmesine olanak tanımıştır. Bu bulgular, endüstriyel sistemlerde sıkça karşılaşılan veri dengesizliği sorununa çözüm getirmekte ve kestirimci bakım modellerinin güvenilirliğini artırmaktadır. Çalışmada önerilen yöntemler, makinelerde arıza tahminlerinin daha hassas yapılmasını sağlayarak, erken uyarı sistemlerinin geliştirilmesi, beklenmedik duruşların azaltılması ve bakım maliyetlerinin düşürülmesi gibi kestirimci bakımın temel hedeflerine doğrudan katkı sunmaktadır. Dolayısıyla, çalışma yalnızca makine öğrenmesi ve veri dengeleme perspektifinden değil, aynı zamanda kestirimci bakımın pratik uygulamalarına yönelik somut kazanımlar ortaya koyması bakımından da önem arz etmektedir.

Makalenin geri kalan bölümlerinde sırasıyla kullanılan veri seti, verilerin analiz için hazırlanması süreci, uygulanan sınıflandırma algoritmaları ve değerlendirilen performans metrikleri ele alınmakta; son bölümde ise elde edilen bulgular sunulup tartışılmaktadır.

2. Deneysel Metot (Experimental Method)

2.1. PdM Veri Seti (PdM Data Set)

Bu çalışmada, kestirimci bakım analizleri için halka açık bir şekilde sunulan AI4I 2020-PdM veri seti kullanılmıştır. Veri seti bir üretim

tesisinin endüstriyel üretim hattından toplanan sensör verilerini içermektedir [17]. Veri seti, Tablo 1'de gösterilen beş farklı arıza türünü içermekte olup, her bir arıza türünün veri setindeki yaklaşık yüzdesi aynı tabloda sunulmaktadır.

Örneğin Tablo 1'de sunulan TWF arıza modu, "alet (takım) aşınması" tanımı genel olarak aletlerin, özellikle de işleme sırasında kullanılan parçalardaki aşınmayı ifade etmektedir. Veri seti detaylı incelendiğinde; 200 ile 240 dakika arasında rastgele seçilen bir alet aşınma süresi ile ilişkilendirilmiş arızalar içerir. Bu süre içinde, takımların 69 kez değiştirilmesi ve 51 kez arıza vermesi (rastgele bir dağılım ile) gözlemlenmiştir. Bu, aletin kullanım süresinin sonunda, belirli bir noktada işlevselliğini yitirerek işlemin başarısız olmasına yol açan bir durumdur. Veri seti, yalnızca aletin aşınmaya uğradığı bir süreçten söz etmektedir. Diğer arıza modları ilgili kaynaktan detaylı bir şekilde sunulmaktadır [17]. Toplam 10.000 veriden oluşan veri setinin %3,4'ü arıza verisi içerirken %96,6'sı arıza verisi içermemektedir. Veri setinin öznitelikleri sırasıyla şu şekildedir:

- Hava Sıcaklığı (Air Temperature [K]): Makinenin bulunduğu ortam sıcaklığını temsil eder. Gerçek makinelerden toplanan sıcaklık sensör verilerini temsil etmektedir. 300 K etrafında bir random walk süreci ile oluşturulmuştur. Çevresel koşulların makine performansına etkisini modellemek için dikkate alınmıştır.
- Proses Sıcaklığı (Process Temperature [K]): Üretim sürecinin sıcaklığını ölçer. Yüksek veya düzensiz değerler mekanik parçaların aşınmasına veya arızasına yol açabileceği için kestirimci bakım açısından önemlidir.
- Dönme Hızı (Rotational Speed [rpm]): Motor veya ekipmanın dönme hızını gösterir; rulman, dişli ve şaft gibi bileşenlerde oluşabilecek mekanik arızaların öngörülmesine katkı sağlamaktadır.
- Tork (Torque [Nm]): Makinenin dönme hareketindeki yükü gösterir. Aşırı veya dengesiz torklar, mekanik bileşenlerde zorlanma ve arızaya neden olabilir.
- Takım Aşınması (Tool Wear [min]): Kesici veya hareketli parçaların aşınma seviyesini belirler ve makine güvenilirliği ile üretim kalitesi açısından kritik bir parametredir.

2.2. Normalizasyon (Normalization)

Çalışmada verilerin ölçüm farklarından etkilenmemesi için normalizasyon yöntemlerine başvurulmuştur. Bu nedenle, özniteliklerin ortak bir ölçekte ifade edilmesi, modelin daha hızlı ve kararlı bir şekilde öğrenmesini sağlamaktadır. Veri standardizasyonunda en sık kullanılan yöntemler, min-max ve z-skor normalizasyon yaklaşımlarıdır.

Min-max normalizasyonunda, bir veriyi min-max normalizasyon ile 0 ile 1 yeni veri aralığına indirmek için Eş. 1'deki ifade kullanılmaktadır [18].

$$y = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (1)$$

Eş. 1'de y normalize edilmiş veriyi, x_i girdi değerini, x_{min} girdi seti içerisinde yer alan en küçük sayıyı, x_{max} girdi seti içerisinde yer alan en büyük sayıyı göstermektedir.

Z-skor normalizasyonu, her değişkene karşı gelen sütunların aritmetik ortalaması (Eş. 2) ve standart sapması (Eş. 3) bulunarak hesaplanmaktadır [19].

$$\mu = \frac{1}{N} \sum_{i=1}^N (x_i) \quad (2)$$

Tablo 1. Arıza modları, açıklamaları ve yüzdelik dağılımları (Failure modes, descriptions and percentage distributions.)

Arıza Modu	Açıklama	Yüzdelik (%)
Tool Wear Failure (TWF)	Döner makinedeki takım, dişli gibi aletlerin aşımından kaynaklanan arızalar	13
Heat Dissipation Failure (HDF)	Motor veya ekipman ısı dağılımı değişimlerinden kaynaklanan arızalar	33
Power Failure (PWF)	Güç kaybı ve elektrik akımından kaynaklanan arızalar	26
Overstrain Failure (OSF)	Aşırı yüklenme veya tork değişimlerinden kaynaklanan arızalar	23
Random Failures (RNF)	Beklenmedik ve rastgele meydana gelen arızalar	5

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2} \quad (3)$$

$$z = \frac{x - \mu}{\sigma} \quad (4)$$

Eş. 4'de verilen z , normalize edilmiş değeri, μ ve σ ise sırasıyla ortalama ve standart sapmayı ifade etmektedir.

2.3. Veri Dengeleme (Data Balancing)

Veri dengesizliği, sınıflar arasında gözlem sayılarının eşit olmaması durumunda, özellikle azınlık sınıfına ait örneklerin sınıflandırma doğruluğunu olumsuz etkileyen önemli bir problemdir. Bu durum, makine öğrenmesi modellerinin genellikle çoğunluk sınıfına ağırlık vererek azınlık sınıfındaki örnekleri doğru bir şekilde tahmin edememesine yol açmaktadır. Bu soruna çözüm bulmak amacıyla iki farklı veri dengeleme yöntemi kullanılmıştır. Az sayıda gözlem içeren veri kümelerinde, çoğunluk sınıfındaki örneklerin sayısını azınlık sınıfındaki örneklerin sayısı ile dengelemek için rastgele seçim (random sampling) yöntemi tercih edilmiştir. Bu yöntem, çoğunluk sınıfına ait mevcut örneklerin rastgele seçilmesi yoluyla veri setindeki sınıf dengesizliğini azaltmayı hedeflemektedir. Bu sayede, çoğunluk sınıfındaki örneklerin çeşitliliği korunmuş ve dengeli bir veri kümesi elde edilmiştir.

Daha büyük veri setlerinde ise azınlık sınıfına yönelik olarak SMOTE yöntemi kullanılmıştır. SMOTE, azınlık sınıfındaki verileri artırmak için kullanılan bir aşırı örnekleme (oversampling) yöntemidir [20]. Bu teknik, azınlık sınıfına ait sentetik örnekler oluşturarak sınıf dengesizliğini azaltmayı hedeflemektedir. Sentetik örnekler, mevcut azınlık sınıfı örnekleri arasındaki ara değerlerden türetilir ve bu sayede modelin azınlık sınıfını daha iyi öğrenmesi sağlanır.

2.4. Sınıflandırma Algoritmaları (Classification Algorithms)

Sınıflandırma tabanlı makine öğrenmesi yaklaşımlarında modeller, veri kümesindeki her bir sınıfı temsil eden bir yapı oluşturmak için kategorik etiketleri tahmin etmektedir. Bu tür bir modelleme, verilen etiketli verilerle öğrenme yaparak, yeni verileri doğru sınıfa yerleştirmeyi amaçlamaktadır. Bu çalışmada, makine arızası ve hata türü etiketleri arasındaki tahminleri yapmak için makine öğrenmesi sınıflandırma algoritmalarından yararlanılmıştır.

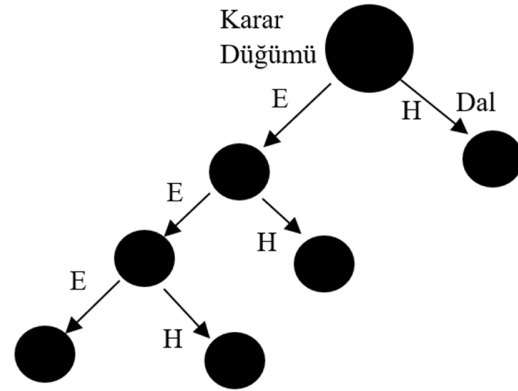
Bu algoritmalar, veri setindeki arızaların tanımlanması ve doğru sınıflandırılması için etkin bir yöntem sunmaktadır. Aşağıdaki bölümlerde, DT, SVM ve k-NN gibi yaygın kullanılan makine öğrenmesi algoritmalarının temel özellikleri ve çalışma prensipleri ele alınacaktır.

2.5. Karar Ağaçları (Decision Trees)

DT, verileri özelliklerine göre sıralayarak sınıflandıran ağaç yapılarıdır. Her bir düğüm, bir özelliği temsil ederken, dallar bu

özelliğin alabileceği farklı değerleri göstermektedir. Sınıflandırma işlemi, kök düğümünden başlanarak, her düğümdeki özellik değerlerine göre örneklerin uygun sınıflara yönlendirilmesiyle gerçekleşmektedir. DT, özellikle veri madenciliği ve makine öğrenmesi alanlarında, gözlemlerle hedef değerler arasında ilişki kurarak tahmin yapmak için kullanılan güçlü bir modelleme aracıdır. Bu modeller, genellikle sınıflandırma ve regresyon problemleri için kullanılmaktadır. DT algoritmasının doğru performans sergileyebilmesi için post-pruning (sonrasında budama) yöntemleri uygulanmaktadır.

Bu yöntem, doğrulama seti ile modelin performansını değerlendirir ve gereksiz düğümleri budayarak, her düğümdeki örneklerin en yaygın sınıfına göre yeniden sınıflandırma yapar [21]. DT modelinin şeması Şekil 2'de verilmiştir.



Şekil 2. Karar ağacı şeması (Decision tree diagram.)

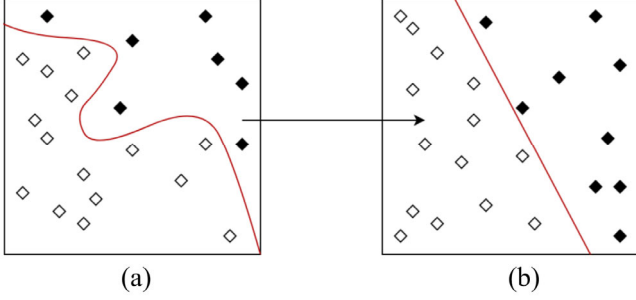
2.6. Destek Vektör Makineleri (Support Vector Machines)

SVM, Vapnik tarafından geliştirilmiş, sınıflandırma, regresyon ve aykırı değer tespiti gibi problemlerde kullanılan güçlü bir makine öğrenimi algoritmasıdır [22]. SVM, verileri iki sınıfa ayıran bir hiper düzlem oluşturur. Bu düzlem, her iki sınıfın veri noktaları arasındaki en geniş mesafeyi sağlayacak şekilde belirlenir, böylece modelin genelleme hatası azalır. SVM, doğrusal olarak ayrılamayan veriler için kernel fonksiyonları kullanarak daha yüksek boyutlu bir uzaya taşıyarak doğrusal ayrım yapmaktadır. Bu özellik, SVM'yi karmaşık veri setlerinde de etkili kılmaktadır. En temel SVM matematiksel ifadesi Eş. 5'de verilmiştir.

$$f(x) = w^T x + b \quad (5)$$

Bu eşitlikte, x , sınıflandırıcının girişi öznelik vektörünü, w , hiperdüzlemin ağırlık vektörünü, T , transpozunu ve son olarak b , hiperdüzlem konumu olarak ifade edilmektedir. Şekil 3'de SVM algoritmasının yapısı verilmiştir. Şekil 3a'da, iki farklı sınıfa ait veri noktalarının doğrusal bir ayırıcı ile birbirinden ayrılmadığı durum sunulmaktadır. Bu tür doğrusal olarak ayrılamayan veri

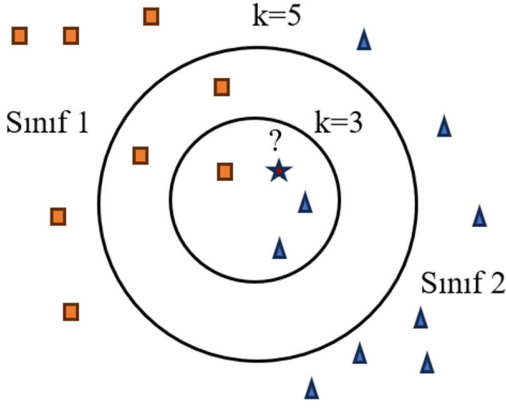
dağılımlarında, SVM algoritması çekirdek (kernel) fonksiyonları kullanarak veriyi daha yüksek boyutlu bir uzaya dönüştürmekte ve bu sayede sınıflar arasında doğrusal bir ayırım elde edilebilmektedir. Şekil 3b'de ise dönüştürülmüş uzayda elde edilen en uygun hiperdüzlem verilmiştir. Bu hiperdüzlem, sınıflar arası marjini maksimize ederek sınıflandırma başarısını artırmakta ve modelin genelleme yeteneğini güçlendirmektedir.



Şekil 3. SVM modeli (SVM model)

2.7. k-En Yakın Komşu (k-Nearest Neighbor)

k-NN kuralı, bir veri noktasının sınıfını en yakın komşusuna dayanarak belirler ve desen tanıma, metin sınıflandırma, nesne tanıma gibi alanlarda kullanılmaktadır. k-NN'ler eğitim süreci hızlı, basitliği, gürültülü eğitim verisine karşı dayanıklılığı ve öğrenme sürecindeki yüksek performansı nedeniyle araştırmacılara avantaj sağlamaktadır [23]. Cover ve Hart, bir veri örneğinin sınıfını belirlemek için belirli sayıda en yakın komşuyu (k) dikkate alan ve bu komşular arasından en sık görülen sınıfı seçerek test örneğinin sınıfını belirleyen k-NN algoritmasını önermiştir [24]. Daha sonra Bailey ve Jain, komşuların uzaklığa göre ağırlıklandırıldığı bir yöntemle algoritmayı geliştirmiştir. Şekil 4'te k-NN algoritmasının yapısı verilmiştir.



Şekil 4. k-En yakın komşular modeli (k-nearest neighbors model)

X_1 ve X_2 'nin özellik vektörleriyle temsil edildiğini varsayılırsa:

$$X_1 = \{ x_{11}, x_{12}, \dots, x_{1m} \} \quad (6)$$

$$X_2 = \{ x_{21}, x_{22}, \dots, x_{2m} \} \quad (7)$$

Eş. 6 ve Eş. 7'de verilen m, öznitelik uzayının boyutunu ifade etmektedir. Komşuluk ilişkisi mesafe ölçütleriyle belirlenmekte olup, en yaygın kullanılan ölçüt ise Eş. 8'de verilen Öklid mesafesidir:

$$\text{dist}(X_1, X_2) = \sqrt{\sum_{i=1}^m (x_{1i} - x_{2i})^2} \quad (8)$$

Burada x_{1i} ve x_{2i} , sırasıyla X_1 ve X_2 veri noktalarının i-inci öznitelikleridir.

2.8. Geriye Doğru Eleme Yöntemi (Backward Elimination Method)

Geriye doğru eleme, regresyon analizlerinde açıklayıcı değişkenlerin seçiminde kullanılan bir yöntemdir [25]. Bu yöntemde, modelde yer alan tüm değişkenler sırasıyla istatistiksel testlerle değerlendirilir ve en az anlamlı olanlar adım adım çıkarılır. Böylece, modelde yalnızca istatistiksel olarak anlamlı değişkenlerin kalması sağlanırken, anlamsız değişkenlerin çıkarılması modelin daha hızlı çalışmasına olanak tanımaktadır. Çünkü her bir eklenen değişken, modelin hesaplama süresini ve karmaşıklığını artırmaktadır. Anlamsız değişkenler ortadan kaldırıldığında, model daha basitleşir ve işlem süresi kısalmır. Bu durum, özellikle büyük veri setlerinde ve çok sayıda değişkenin bulunduğu durumlarda önemli bir fark yaratmaktadır. Böylece, hem modelin doğruluğu artırılır hem de daha verimli ve hızlı bir analiz sağlanır. Ayrıca, çok sayıda değişkenin bulunduğu modellerde aşırı uyum (overfitting) riski bulunmaktadır; geriye doğru eleme yönteminde ise gereksiz değişkenler çıkarılarak modelin genellebilirliği artırılmaktadır. Sonuç olarak, yalnızca anlamlı değişkenlerin modele dahil edilmesi, modelin daha doğru ve açıklayıcı olmasını sağlamaktadır.

2.9. Performans Metrikleri (Performance Metrics)

Modelin başarısını değerlendirmek için çeşitli metrikler kullanılmaktadır. Bu metrikler, modelin doğruluğunu, güvenilirliğini ve genel başarısını ölçmek amacıyla hesaplanmaktadır. Bu hesaplamalarda, hata matrisi (confusion matrix) temel alınmaktadır. Şekil 5'te verilen örnekte, iki sınıflı bir hata matrisi gösterilmektedir. Modelin sınıflandırma performansını gösteren karmaşıklık matrisi dört temel bileşenden oluşur [26]:

- Doğru Pozitifler (DP): Modelin pozitif sınıfı doğru şekilde tahmin ettiği örnek sayısı.
- Yanlış Pozitifler (YP): Modelin negatif sınıfı yanlış şekilde pozitif olarak tahmin ettiği örnek sayısı.
- Doğru Negatifler (DN): Modelin negatif sınıfı doğru şekilde tahmin ettiği örnek sayısı.
- Yanlış Negatifler (YN): Modelin pozitif sınıfı yanlış şekilde negatif olarak tahmin ettiği örnek sayısı.

		TAHMİN	
		Negatif	Pozitif
GERÇEK	Negatif	DN	YP
	Pozitif	YN	DP

Şekil 5. Hata Matrisi (confusion matrix)

Hata matrisi, modelin hangi sınıflarda güçlü olduğunu ve hangi sınıflarda hata yaptığını görmek için önemli bir araçtır [27]. Çalışmada kullanılan başlıca performans metrikleri ise şu şekildedir:

Doğruluk (Accuracy-Acc): Modelin doğru sınıflandırdığı örneklerin toplam örnek sayısına oranıdır. Yüksek doğruluk, modelin genel olarak doğru tahminlerde bulunduğunu göstermektedir (Eş. 9).

$$\text{Doğruluk}(Acc) = \frac{DP+DN}{DP+DN+YP+YN} \quad (9)$$

Kesinlik (Precision-Pre): Modelin pozitif sınıf olarak tahmin ettiği örneklerin ne kadarının doğru olduğunu ölçmektedir. Özellikle yanlış pozitiflerin önemli olduğu durumlarda kesinlik ön planda tutulmaktadır (Eş. 10).

$$\text{Kesinlik}(Pre) = \frac{DP}{DP+YP} \quad (10)$$

Duyarlılık (Recall-Rec): Gerçek pozitiflerin model tarafından ne kadarının doğru şekilde tahmin edildiğini ölçmektedir (Eş. 11).

$$\text{Duyarlılık}(Rec) = \frac{DP}{DP+YN} \quad (11)$$

F1-Skoru (F1): Kesinlik ve duyarlılığın harmonik ortalamasıdır. Bu metrik, her iki değeri dengeleyerek modelin genel başarısını daha objektif bir şekilde ölçmektedir. F1 skoru, kesinlik ve duyarlılık arasında bir denge kurmak gerektiğinde kullanılmaktadır (Eş. 12).

$$F1 \text{ Skor} (F1) = 2 \times \frac{\text{Kesinlik} \times \text{Duyarlılık}}{\text{Kesinlik} + \text{Duyarlılık}} \quad (12)$$

Performans sonuçlarının güvenilirliği, eğitim ve test veri setlerinin ayırımına doğrudan bağlıdır. Literatürde bu ayırım, genellikle k-parçalı çapraz doğrulama veya “birini dışarıda bırak” (leave-one-out) yöntemleriyle sağlanmaktadır. Bu çalışmada, modelin genellenabilirliğini artırmak ve aşırı öğrenme riskini en aza indirmek amacıyla 5-parçalı çapraz doğrulama yöntemi kullanılmış ve elde edilen sonuçların hem eğitim hem de test kümeleri üzerinde tutarlı olmasına özen gösterilmiştir.

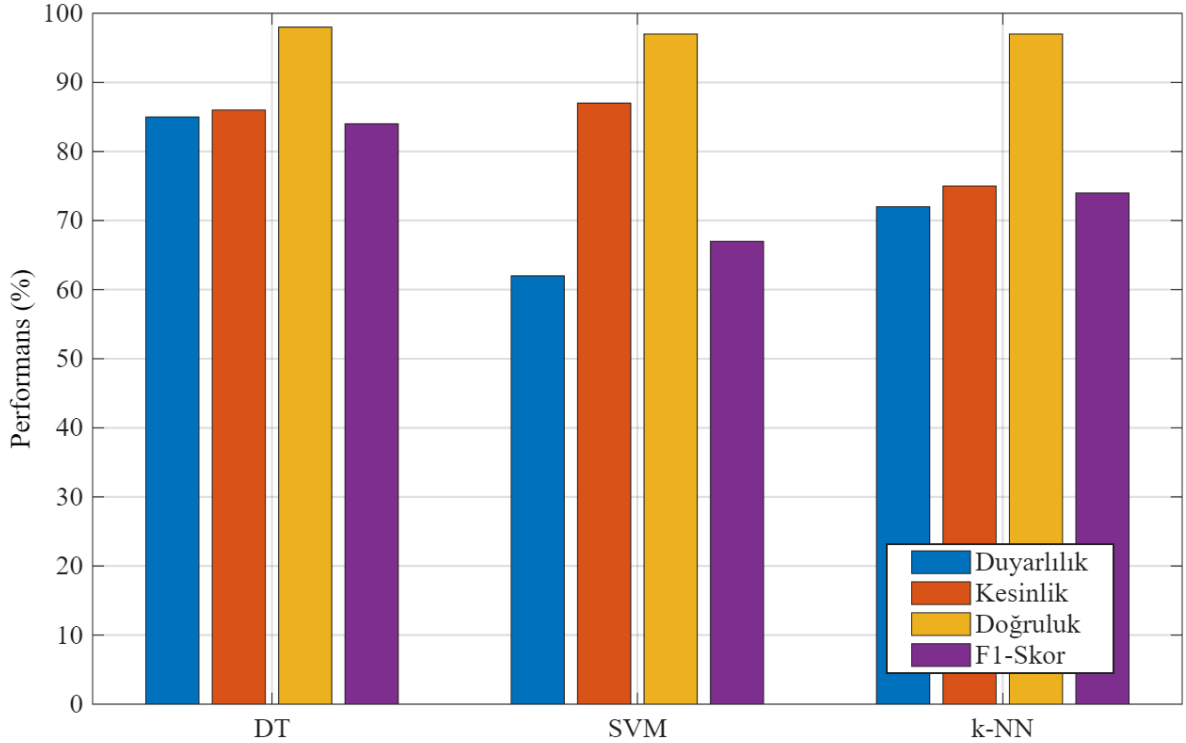
Bu yöntemde, her yinelemede verinin bir parçası test için, kalan dört parçası ise eğitim için kullanılmaktadır. Süreç, tüm parçaların bir kez test verisi olarak kullanılmasyla tamamlanmakta ve elde edilen

sonuçların ortalaması alınarak nihai performans değeri elde edilmektedir.

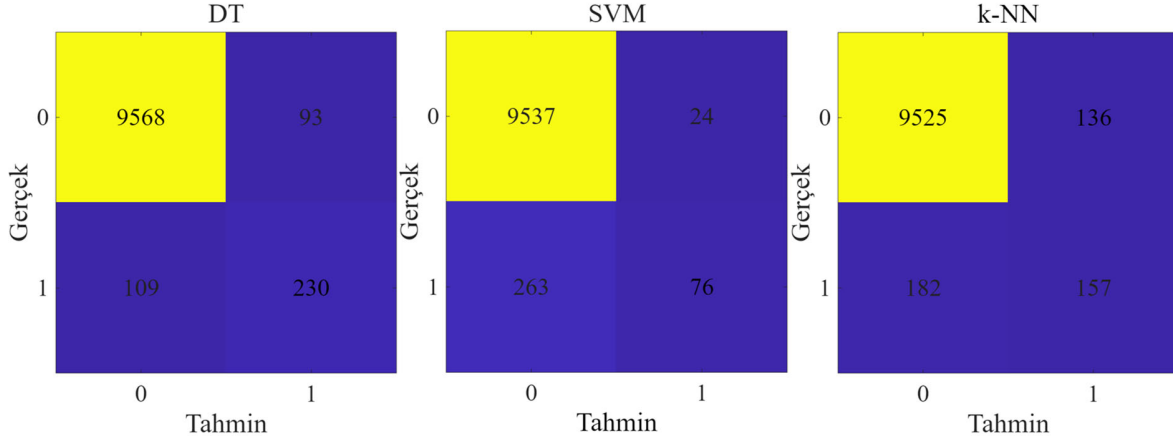
Ayrıca çalışmada ROC (Receiver Operating Characteristic) eğrisi ve (AUC – Area Under the Curve) değeri kullanılmıştır. ROC eğrisi, farklı eşiklerdeki duyarlılık ve 1-özellik değerlerini grafikleştirerek modelin pozitif ve negatif sınıfları ne kadar iyi ayırdığını gösterir. AUC değeri ise modelin genel performansını tek bir sayı ile özetler; 1'e yaklaştıkça modelin performansı yüksek, 0,5'e yaklaştıkça ise rastgele tahmine yakındır.

3. Sonuçlar ve Tartışmalar (Results and Discussions)

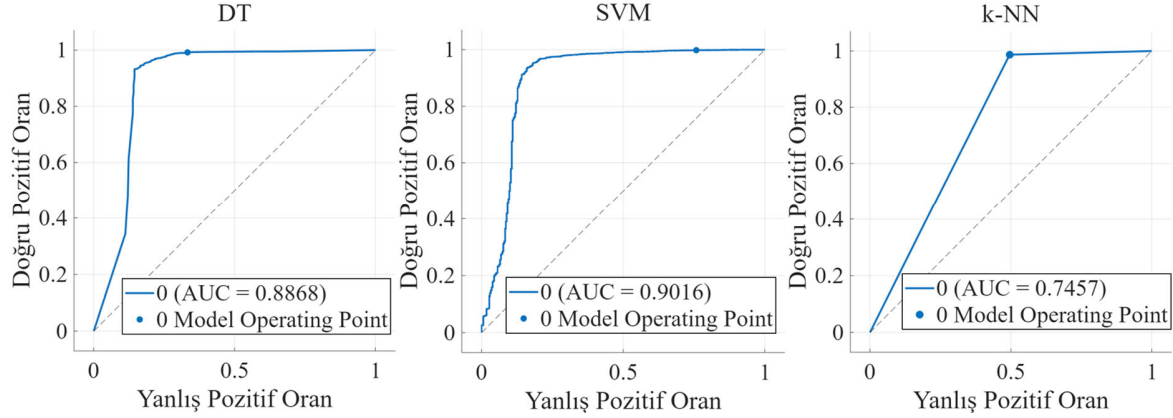
Bu çalışma kapsamında, makine arıza tipi ve dört farklı arıza modunun sınıflandırılmasında kullanılan DT, SVM ve k-NN algoritmalarının performansları detaylı bir şekilde değerlendirilmiştir. Dengesiz veri setleri üzerinde yapılan analizler sonucunda, rastgele seçim ve SMOTE yöntemleri kullanılarak veri dengesi sağlanmış ve SMOTE yönteminin daha başarılı sonuçlar verdiği görülmüştür. Daha sonra elde edilen dengeli veri setine, min-max ve z-skor normalizasyon işlemleri uygulanmış ve bu işlemler sonrasında algoritmaların performansları karşılaştırıldığında z-skor uygulanan verilerin daha iyi performans sergilediği tespit edilmiştir. Çalışmada performans değerlendirmesi için ilk olarak doğruluk oranı (accuracy) metriği esas alınmıştır. Burada doğruluk oranı, modelin toplam tahminlerinin ne kadarını doğru sınıflandırdığını göstermekte olup, makinedeki gerçek arıza oranına karşılık gelmemektedir. Dolayısıyla yüksek doğruluk oranı, modelin arıza ve normal durumları doğru ayırt etme başarısını ifade etmektedir. Ayrıca, TWF-HDF, TWF-PWF, TWF-OSF, HDF-PWF ve PWF-OSF arıza modları arasındaki sınıflandırma işlemleri gerçekleştirilmiş ve her bir mod için z-skor normalizasyonunun başarıyı artırdığı gözlemlenmiştir. Son aşamada, geriye doğru eleme yöntemi ile veri setindeki beş öznelik sırasıyla elenmiş ve en yüksek başarı oranları belirlenmiştir. Bu yöntemle, her arıza modu için kritik olmayan öznelikler tespit edilerek



Şekil 6. Dengesiz veri setinde algoritma performans karşılaştırması (Algorithm performance comparison on an imbalanced dataset.)



Şekil 7. Orijinal veri seti için algoritmaların hata matrisi (Confusion matrix of algorithms for the original dataset)



Şekil 8. Orijinal veri seti için algoritmaların ROC eğrileri ve AUC değerleri (ROC curves and AUC values of the algorithms for the original dataset)

sınıflandırma performansı artırılmıştır. Özellikle belirli özneliklerin çıkarılmasının performans metrikleri üzerinde olumlu etkiler sağladığı sonucuna ulaşılmıştır. Bu sonuçlar, önerilen yöntemlerin ve seçilen özneliklerin, makine arızası ve modlarının sınıflandırılmasında başarıyı önemli ölçüde arttırabileceğini göstermektedir. Gerçekleştirilen bu çalışmaların sonuçları alt başlıklarda detaylı bir şekilde ele alınmıştır.

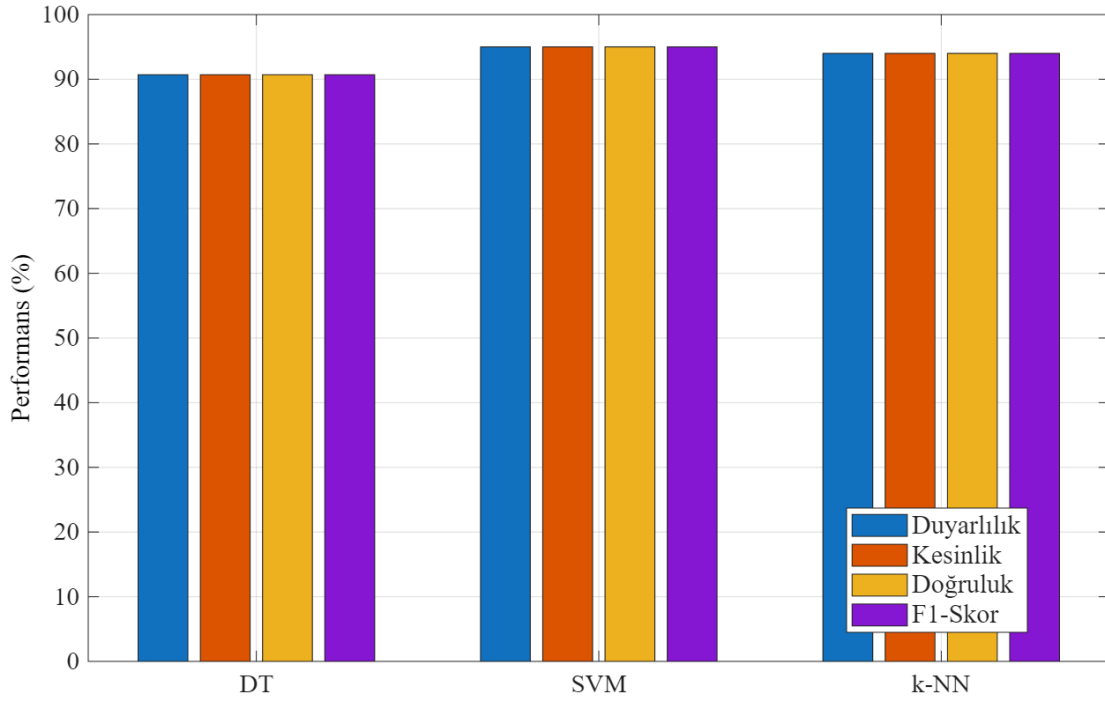
3.1. Orijinal Veri Seti Kullanılarak Elde Edilen Sonuçlar (The Results Obtained Using the Original Data Set)

Bu çalışmada kullanılan orijinal veri seti, toplamda 10.000 veri içermekte olup, 339 makine arıza durumu (MF) ve 9.661 makine arızasız (N-MF) durumu barındırmaktadır. Bu durum, veri setinin ciddi bir sınıf dengesizliğine sahip olduğunu göstermektedir. Bu bölümde, veri setine herhangi bir dengeleme yöntemi (SMOTE veya rastgele örnekleme) uygulanmadan veri seti üzerinde yapılan analizlerde, DT algoritması %98,0, SVM algoritması %97,1 ve k-NN algoritması %96,8 doğruluk oranı elde etmiştir. SVM algoritmasında %86,7 kesinlik değeri ile en yüksek sonuç elde edilmiştir. F1 değerleri ise sırasıyla DT %84,2, SVM %66,6 ve k-NN %74,0 şeklinde elde edilmiştir. Orijinal dengesiz veri setinde geriye doğru eleme yöntemi uygulandığında ise sırasıyla tüm öznelikler çıkarıldığında, model başarımında bir düşüş gözlemlenmiştir. Başlangıçta, tüm özneliklerin kullanılmasıyla elde edilen performans metrikleri Şekil 6'da sunulmuştur. Ek olarak, DT, SVM ve k-NN algoritmalarının hata matrisleri Şekil 7'de, ROC eğrileri ve AUC değerleri ise Şekil 8'de gösterilmiştir. Bu görselleştirmeler, her bir sınıflandırıcının doğru ve

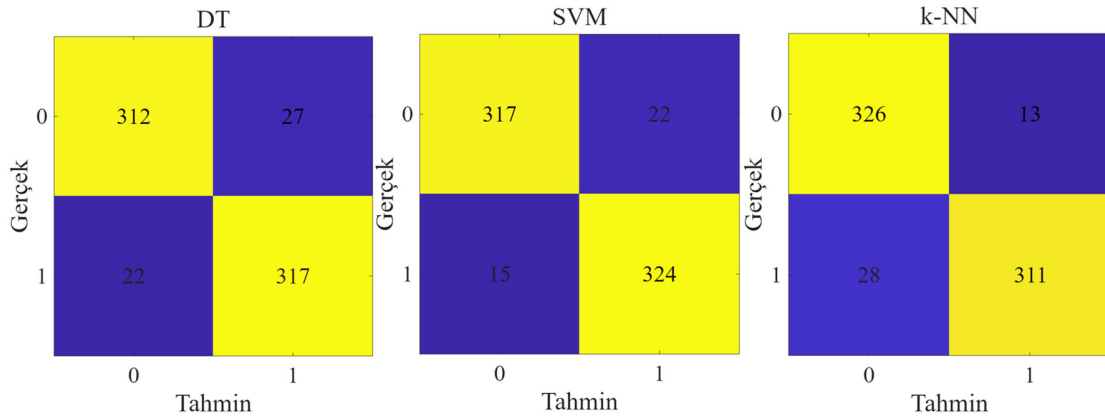
yanlış sınıflandırma dağılımlarını ve sınıflar arasındaki ayırt etme performansını net bir şekilde ortaya koymaktadır.

3.2. Rastgele Seçim Yöntemi Kullanılarak Oluşturulan Dengeli Veri Seti ve Elde Edilen Sonuçlar (The Balanced Data Set Created Using the Random Selection Method and the Results Obtained)

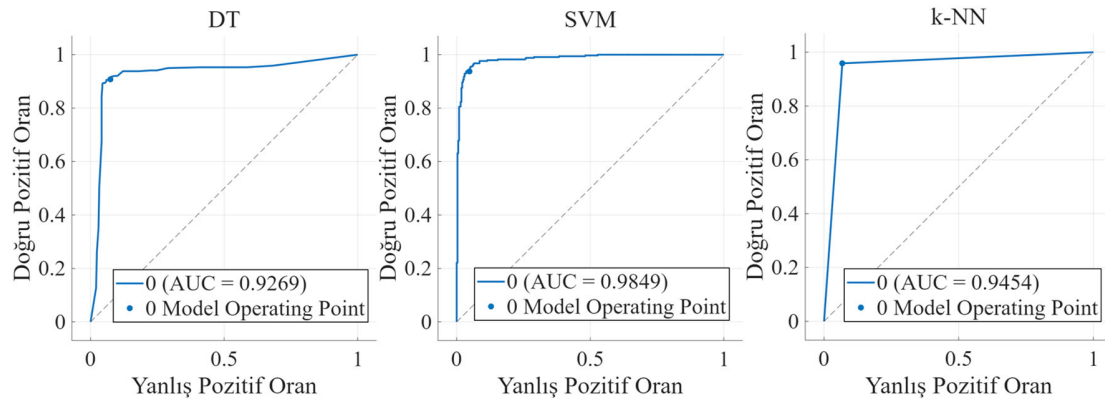
Bu çalışmada, dengesiz veri kümesindeki dengesizlik sorununu gidermek için rastgele seçim yöntemi kullanılarak dengeli bir veri seti oluşturulmuştur. Bu yöntem kullanılarak, 339 adet MF sınıfına karşı, 9.661 adet N-MF içinden rastgele 339 adet arızasız durum seçilmiş ve sonuçlar bu dengeli veri seti üzerinden elde edilmiştir. Şekil 9'da dengeli veri setinde farklı algoritmalar için performans ölçütlerinin karşılaştırılması verilmiştir. DT, SVM ve k-NN algoritmaları sırasıyla %90,7, %95,0 ve %94,0 doğruluk değeri elde etmiştir. En yüksek kesinlik değeri %95,0 ile SVM modeline aittir. Bu, modelin doğru pozitif sınıflandırmalarının sayısının yüksek olduğunu ve yanlış pozitiflerin minimumunda kaldığını göstermektedir. Yani, model, pozitif sınıfı doğru bir şekilde tanıma konusunda oldukça başarılıdır. Duyarlılık açısından ise yine en yüksek başarıyı SVM modeli %95,0 ile gösterirken, k-NN ve DT modelleri sırasıyla %94,0 ve %90,7 olarak tespit edilmiştir. F1 skorlarına da bakıldığında, SVM modelinin hem kesinlik hem de duyarlılık açısından en dengeli sonuçları verdiği görülmektedir. Bu, SVM modelinin doğruluk, kesinlik ve duyarlılık arasında güçlü bir denge kurarak her iki kriterde de yüksek katkı sağladığını göstermektedir.



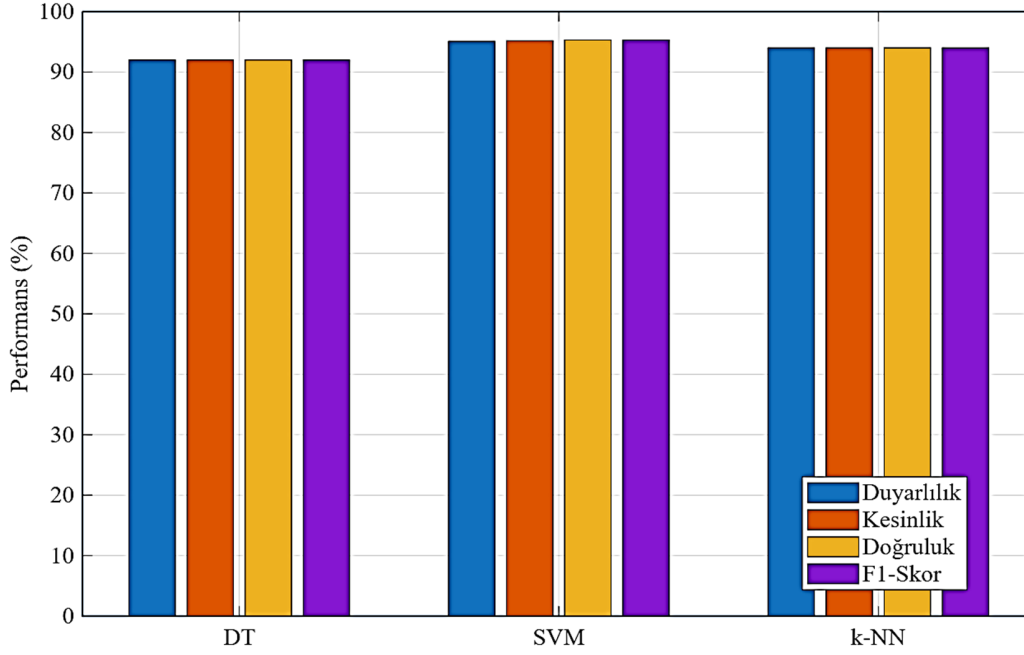
Şekil 9. Rastgele seçimle dengelenmiş veri setinde algoritma performans karşılaştırması
(Algorithm performance comparison on a balanced dataset with random selection.)



Şekil 10. Rastgele seçimle dengelenmiş veri seti için algoritmaların karmaşıklık matrisi
(Confusion matrix of algorithms for balanced dataset with random selection.)



Şekil 11. Rastgele seçimle dengelenmiş veri seti için algoritmaların ROC eğrileri ve AUC değerleri
(ROC curves and AUC values of the algorithms for the balanced data set with random selection.)



Şekil 12. Rastgele dengelenmiş veri setinde geriye doğru eleme ile performans karşılaştırması
(Performance comparison with backward elimination on a randomly balanced dataset)

Ayrıca, DT, SVM ve k-NN algoritmalarının dengeli veri seti üzerindeki hata matrisleri Şekil 10'da, ROC eğrileri ve AUC değerleri ise Şekil 11'de detaylı bir şekilde sunulmuştur. Hata matrisleri de, özellikle SVM modelinin sınıflar arasında daha yüksek ayırt edicilik sağladığını göstermektedir. ROC eğrileri de bu bulguları desteklemekte olup, SVM modelinin yalnızca temel performans ölçütlerinde değil, aynı zamanda sınıflar arasındaki ayırt edicilik açısından da diğer algoritmalarından daha başarılı olduğunu göstermektedir.

3.3. Rastgele Seçim Yöntemi ile Dengelenmiş Veri Setinde Geriye Doğru Eleme Yöntemiyle Elde Edilen Sonuçlar (The Results Obtained by the Backward Elimination Method in the Data Set Balanced by the Random Selection Method)

Geriye doğru eleme yöntemi uygulandıktan sonra model performansında önemli iyileşmeler gözlemlenmiştir. DT algoritmasının doğruluk oranı geriye doğru eleme sonrasında %92,0 başarı oranına yükselmiştir. SVM algoritmasında bu oran %95,3'e çıkmıştır. k-NN algoritmasında ise %94,0 doğruluk oranı ile aynı değerde kalmıştır. Sonuç olarak, geriye doğru eleme yöntemi kullanılarak elde edilen dengeli veri setinde DT ve k-NN modellerinde performansında artış sağlanmıştır. Özellikle SVM modelinin kesinlik ve duyarlılık değerleri sırasıyla %95,2 ve %95,1 seviyelerine yükselmiştir, bu da SVM modelinin öznelilik çıkarımı sonrasında daha doğru pozitif tahminler yaparak genel performansını artırdığını ortaya koymaktadır. Geriye doğru eleme sonucu dengeli veri setinde farklı algoritmalar için performans ölçütlerinin karşılaştırılması Şekil 12'de verilmiştir.

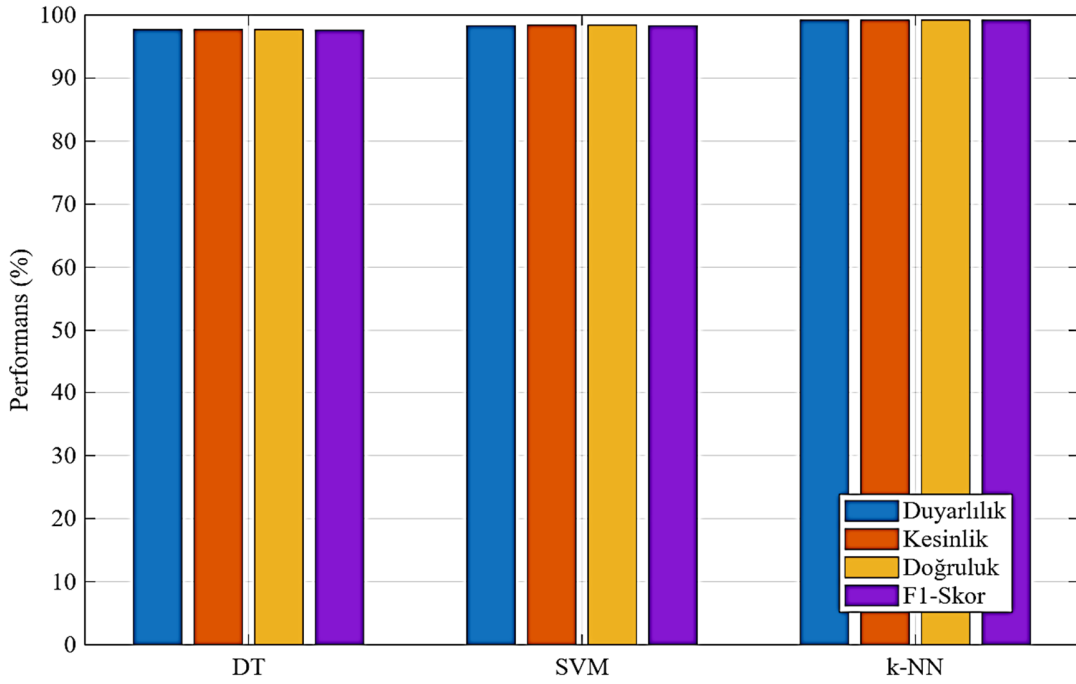
3.4. SMOTE Yöntemi Kullanılarak Oluşturulan Dengeli Veri Seti ile Elde Edilen Sonuçlar (The Results Obtained with the Balanced Data Set Created Using the SMOTE Method)

Veri setine SMOTE yöntemi uygulanarak azınlık sınıfı olan 339 adet MF, 9.492 örneğe çıkarılmış ve veri setindeki sınıf dengesizliği giderilmiştir. Şekil 13'te verilen grafiklere göre, DT, SVM ve k-NN

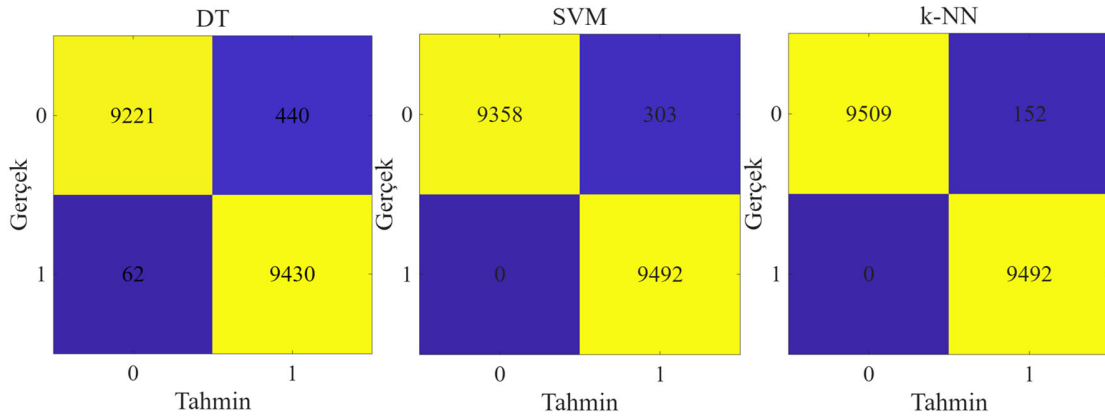
modelleri sırasıyla %97,7, %98,4 ve %99,2 doğruluk oranlarıyla oldukça başarılı sonuçlar elde etmiştir. Ancak doğruluk oranları, modelin yalnızca genel başarısını gösterdiğinden, daha ayrıntılı bir değerlendirme için diğer metrikler de göz önünde bulundurulmalıdır. Şekil 14'te sunulan hata matrisi sonuçları, k-NN modelinin en düşük hata oranı ile sınıfları daha başarılı şekilde ayırt ettiğini; Şekil 15'te verilen ROC eğrileri ve AUC değerleri ise tüm modellerin yüksek ayırt edicilik gücüne sahip olduğunu göstermektedir. SMOTE ile dengelenmiş veri setinde, k-NN modelinin %99,2 F1 puanı ile en yüksek başarıyı elde ettiği, ardından %98,3 başarı oranı ile SVM ve %97,5 ile DT modellerinin geldiği görülmektedir. Bu, k-NN modelinin geri çağırma ve kesinlik arasında güçlü bir denge sağladığını göstermektedir. Sonuç olarak, bu dengeli veri setinde, verilerin artırılması modellerin performansında olumlu değişikliklere neden olmuştur. k-NN modeli, daha fazla sınıfı ayırt ederek genel başarımda lider olmuştur.

3.5. SMOTE Yöntemi ile Dengelenmiş Veri Setinde Geriye Doğru Eleme Yöntemi ile Elde Edilen Sonuçlar (The Results Obtained by the Backward Elimination Method in the Data Set Balanced by the SMOTE Method)

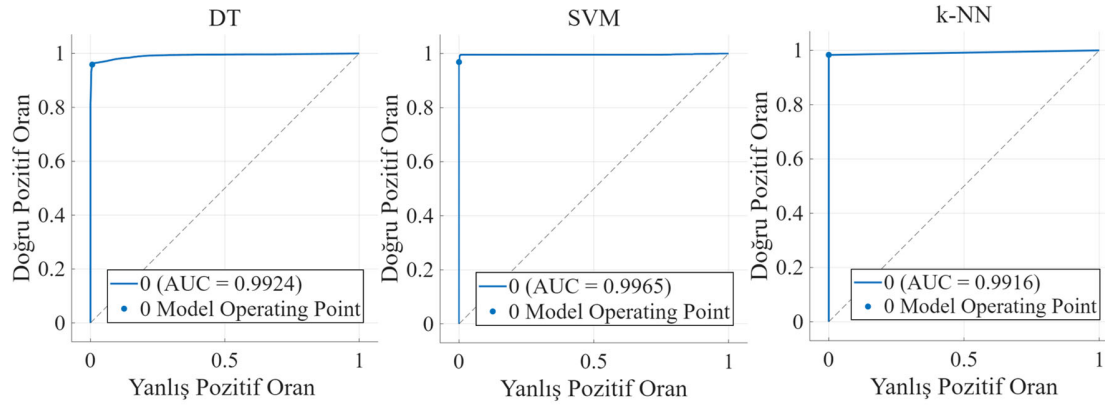
Bu bölümde, SMOTE yöntemi ile dengelenmiş veri setinde, geriye doğru eleme tekniği kullanılarak model başarımını optimize edilmiştir. Başlangıçta, tüm özneliliklerin kullanıldığı durumda DT, SVM ve k-NN modelleri sırasıyla %97,7, %98,4 ve %99,2 doğruluk oranlarıyla en yüksek başarıyı elde etmiştir. Bu, tüm özneliliklerin modelin sınıflandırma performansını en iyi şekilde desteklediğini göstermektedir. Ancak, geriye doğru eleme ile bazı öznelilikler çıkarıldığında, model başarımında bir düşüş gözlemlenmiştir (Şekil 16). Örneğin, hava sıcaklığı özelliği çıkarıldığında, doğruluk değerleri DT için %95,5, SVM için %96,1 ve k-NN için %98,9 olarak tespit edilmiştir. Geriye doğru eleme süreci, modelin daha sade hale gelmesine ve bazı özneliliklerin gereksiz olduğunu ortaya koymasına yardımcı olsa da, belirli özneliliklerin çıkarılması sınıflandırma başarımını olumsuz etkilemiştir. Bununla birlikte, k-NN modeli, hava sıcaklığı özelliği çıkarılsa bile, diğer modellere kıyasla yüksek doğruluk oranını koruyabilmiştir.



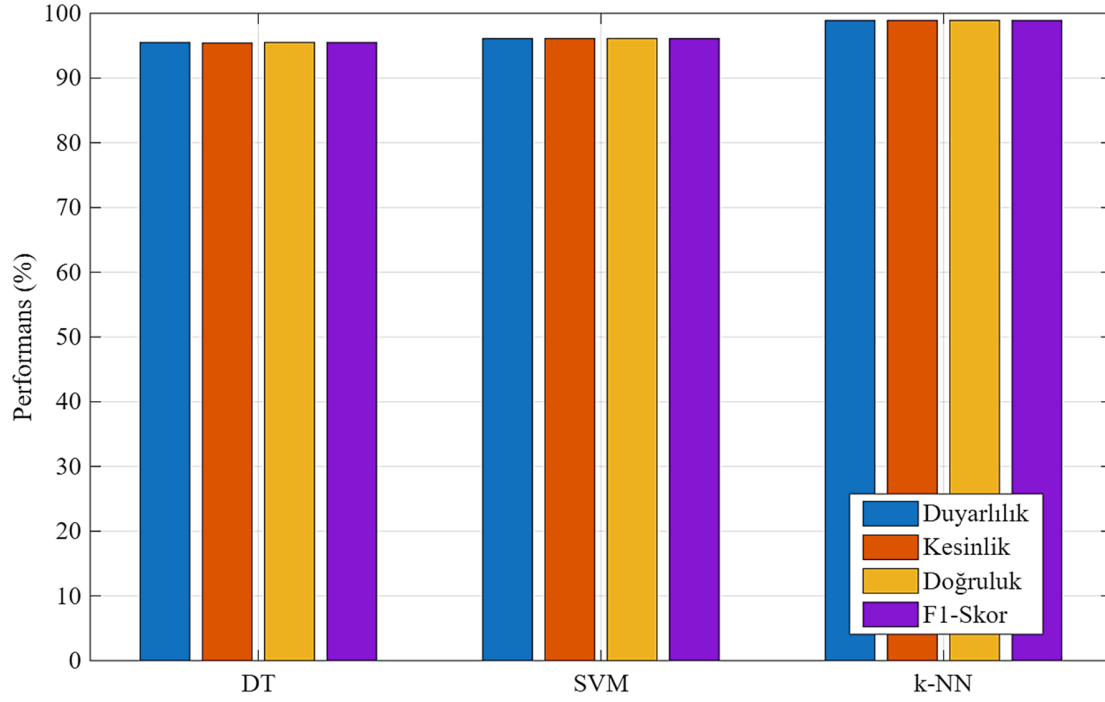
Şekil 13. SMOTE ile dengelenmiş veri setinde performans analizi (Performance analysis on the balanced dataset with SMOTE)



Şekil 14. SMOTE ile dengelenmiş veri seti için algoritmaların karmaşıklık matrisi (Confusion matrix of algorithms for the balanced dataset with SMOTE)



Şekil 15. SMOTE ile dengelenmiş veri seti için algoritmaların ROC eğrileri ve AUC değerleri (ROC curves and AUC values of the algorithms for the data set balanced with SMOTE)



Şekil 16. SMOTE ile dengelenmiş veri setinde geriye doğru eleme ile performans karşılaştırması (Performance comparison with backward elimination on a balanced dataset with SMOTE.)

Tablo 2. Geriye doğru eleme yöntemi öncesi performans metrikleri analizi (Analysis of performance metrics before the backward elimination method)

SINIF	DT (%)				SVM (%)				k-NN (%)			
	Acc	Pre	Rec	F1	Acc	Pre	Rec	F1	Acc	Pre	Rec	F1
TWF-HDF	98,8	98,5	98,5	98,5	90,1	91,8	81,1	84,5	97,5	96,7	94,1	95,3
TWF-PWF	97,1	96,7	96,7	96,7	92,1	94,8	87,8	90,3	95,7	95,1	95,1	95,1
TWF-OSF	97,1	97,2	96,0	96,6	75,4	86,8	60,5	59,7	93,5	93,9	90,8	92,1
HDF-PWF	95,1	95,1	95,1	95,1	97,1	97,0	97,1	97,0	95,6	96,1	95,9	96,0
HDF-OSF	95,0	92,9	93,2	93,0	93,0	94,1	94,4	94,0	95,5	96,5	96,5	96,5
PWF-OSF	98,8	98,8	98,8	98,8	98,2	98,2	98,2	98,2	98,2	98,1	97,0	98,0

3.6. Arıza Modlarının Performans Analizi: Geriye Doğru Eleme Yöntemi Öncesi (Performance Analysis of Failure Modes: Before the Backward Elimination Method)

Analizlerde, her bir arıza sınıfı için orijinal veri seti kullanılmış ve tüm öznelikler modele dahil edilmiştir. Veri seti sayıları şu şekildedir:

- TWF-HDF: 46 TWF, 115 HDF
- TWF-PWF: 45 TWF, 94 PWF
- TWF-OSF: 43 TWF, 95 OSF
- HDF-PWF: 112 HDF, 92 PWF
- HDF-OSF: 109 HDF, 92 OSF
- PWF-OSF: 83 PWF, 86 OSF

İlk aşamada tüm öznelikler kullanıldığında model başarı oranları TWF-HDF sınıfı için DT modeli %98,8, SVM %90,1 ve k-NN %97,5 şeklindedir. TWF-PWF sınıfı için bu oranlar DT için %97,1, SVM için %92,1 ve k-NN için %95,7 olarak hesaplanmıştır. TWF-OSF analizinde, SVM modelinin düşük bir performansla %75,4 doğruluk değeri elde ettiği görülmüştür. Ancak diğer modellerde DT modelinin

doğruluğunun %97,1 ve k-NN modelinin %93,5 olduğu gözlemlenmiştir. HDF-PWF sınıfında SVM %97,1 doğruluk oranı ile en yüksek başarı elde edilmiştir. PWF-OSF sınıfında ise DT modeli %98,8 ile en yüksek başarıyı göstermiştir. Bu sonuçlar, tüm özneliklerin dahil edilmesi durumunda modellerin genel olarak yüksek performans göstermesine rağmen, bazı modellerin belirli sınıflar için düşük başarı oranları sunduğunu ortaya koymuştur. Arıza modlarının geriye doğru eleme yöntemi öncesi performans analizi Tablo 2'de detaylı olarak verilmiştir.

3.7. Arıza Modlarının Performans Analizi: Geriye Doğru Eleme Yöntemi Sonrası (Performance Analysis of Failure Modes: After the Backward Elimination Method)

Geriye doğru eleme yöntemiyle öznelik sayısı azaltılarak model performansında belirgin iyileşmeler sağlanmıştır. Yapılan analiz sonucunda, geriye doğru eleme yöntemi ile seçilen öznelikler ve arıza modları Tablo 3'te verilmiştir.

TWF-HDF analizinde, hava sıcaklığı, dönme hızı ve takım aşınması özneliklerinin seçilmesiyle k-NN modeli en yüksek doğrulukla %99,4'e ulaşmış ve %99,1 kesinlik ile %99,3 F1 skor göstermiştir.

Tablo 3. Arıza modları ve dahil edilen öznelikler (Failure modes and selected features)

Arıza Modu	Seçilen Öznelikler
TWF-HDF	Hava Sıcaklığı, Dönme Hızı, Takım Aşınması
TWF-PWF	Dönme Hızı, Tork, Takım Aşınması
TWF-OSF	Dönme Hızı, Tork
HDF-PWF	Hava Sıcaklığı, Proses Sıcaklığı, Dönme Hızı, Tork
HDF-OSF	Hava Sıcaklığı, Dönme Hızı, Tork, Takım Aşınması
PWF-OSF	Dönme Hızı, Takım Aşınması

Tablo 4. Geriye doğru eleme yöntemi sonrası performans metrikleri analizi
(Analysis of performance metrics after the backward elimination method)

Sınıf	DT (%)				SVM (%)				k-NN (%)			
	Acc	Pre	Rec	F1	Acc	Pre	Rec	F1	Acc	Pre	Rec	F1
TWF-HDF	98,8	98,5	98,5	98,5	98,8	98,9	98,6	98,2	99,4	99,1	98,4	99,3
TWF-PWF	97,1	97,1	97,1	97,1	97,8	97,8	97,2	97,5	97,8	96,7	96,7	96,7
TWF-OSF	99,3	98,9	99,5	99,2	96,4	96,7	94,8	95,7	97,1	97,2	96,0	96,6
HDF-PWF	95,1	95,1	95,1	95,1	98,0	97,9	98,1	98,0	98,5	98,6	98,5	98,5
HDF-OSF	95,0	94,4	95,6	95,5	96,0	96,1	95,9	96,0	98,0	98,0	98,0	98,0
PWF-OSF	99,4	99,4	99,4	99,4	100	100	100	100	100	100	100	100

DT ve SVM modellerinin ise %98,8 doğruluğa ulaştığı gözlemlenmiştir. TWF-PWF sınıfında dönme hızı, tork ve takım aşınması öznelikleri seçildiğinde SVM ve k-NN modellerinin doğruluk oranları %97,8, DT modeli %97,1 olmuş ve SVM modeli %97,5 F1 skor ile en iyi performansı sunmuştur. TWF-OSF analizinde dönme hızı ve tork öznelikleri seçilerek DT modelinin doğruluğu %99,3, SVM modelinin %96,4 ve k-NN modelinin %97,1'e yükselmiştir. HDF-PWF analizinde hava sıcaklığı, proses sıcaklığı, dönme hızı ve tork öznelikleri seçildiğinde k-NN modeli %98,5 doğruluk ve F1 skor ile en yüksek performansı sergilemiştir. DT ve SVM modelleri ise sırasıyla %95,1 ve %98,0 olmuştur. HDF-OSF sınıfında hava sıcaklığı, dönme hızı, tork ve takım aşınması öznelikleri seçildiğinde DT modelinin doğruluğu %95,0'da sabit kalırken, SVM modeli %96,0 ve k-NN modeli %98,0 doğruluğa ulaşmıştır. PWF-OSF analizinde dönme hızı ve takım aşınması öznelikleri seçildiğinde SVM ve k-NN modelleri %100 doğruluk, kesinlik ve F1 skor ile maksimum başarı göstermiştir. DT modeli %99,4 doğruluk oranına yükselmiştir.

Tüm sınıflar için ortak olan dönme hızı özelliğinin seçilmesi, model performansında anlamlı iyileşmeler sağlamıştır. Bu sonuçlar, her arıza durumunda belirli özneliklerin seçilmesi veya çıkarılmasının model performansını olumlu yönde etkileyebileceğini göstermektedir. Arıza modlarının geriye doğru eleme yöntemi sonrası performans analizi Tablo 4'te detaylı olarak verilmiştir.

Bu çalışmada, kestirimci bakım planlaması kapsamında makine arızalarının ve arıza modlarının sınıflandırılmasında çeşitli makine öğrenimi algoritmalarının performansı değerlendirilmiştir. Özellikle, dengesiz veri setlerini dengelemek için kullanılan SMOTE yöntemi, hem rastgele seçim yöntemine kıyasla hem de literatürdeki diğer çalışmalarla karşılaştırıldığında daha yüksek sonuçlar sunmuştur. SMOTE yönteminin dengesiz veri kümelerinde sınıflandırma başarısını artırdığı literatürde birçok çalışmada vurgulanmıştır. Örneğin, Chawla ve arkadaşları SMOTE yönteminin azınlık sınıfların temsil gücünü artırarak model doğruluğunu yükselttiğini göstermiştir [28]. Benzer şekilde, Huang ve ekibi tarafından yapılan bir çalışmada, SMOTE yöntemi ile SVM algoritmasının birleştirilerek bir hibrit model oluşturulmuş ve diğer çeşitli tahmin modellerine kıyasla daha iyi sonuç verdiği gözlemlenmiştir [29]. Bu çalışmada da benzer şekilde SMOTE yöntemi, rastgele seçim yöntemine kıyasla daha başarılı sonuçlar sunmuştur.

Normalizasyon tekniklerine yönelik karşılaştırmalara bakıldığında, z-skor normalizasyonunun, min-max ölçeklemeye kıyasla daha kararlı sonuçlar sunduğu çeşitli çalışmalarda rapor edilmiştir. Özellikle, Herwanto ve ekibi, Java'dan elde edilen verilerin incelendiği ve z-skor normalleştirme yöntemiyle normalleştirilen harf özneliklerinin çıkarılmasının en iyi sonucu verdiği belirlenmiştir [30]. Ayrıca, literatürde sıkça kullanılan SVM, DT ve k-NN algoritmaları için farklı veri setleri üzerinde yapılan performans analizleri, belirli algoritmaların belirli arıza modlarında daha etkili olduğunu göstermektedir. Assagaf ve ekibi, SVM'nin karmaşık arıza modlarında yüksek başarı sunduğunu ortaya koyarken, k-NN algoritmasının küçük ve dengeli veri setlerinde oldukça etkili olduğunu belirlemiştir.

Geriye doğru eleme yöntemiyle gerçekleştirilen analizlerde ise model performansı artırılmış ve özellikle dönme hızı özelliğinin kritik bir belirleyici olduğu tespit edilmiştir. Literatürde öznelik seçiminin model başarısını üzerindeki etkisini inceleyen çalışmalar, gereksiz özneliklerin elimine edilmesinin sınıflandırma doğruluğunu artırabileceğini göstermektedir. Örneğin, Maulidina ve arkadaşları, geriye doğru eleme yöntemi ve SVM algoritması kullanılanlar PIMA Indians diyabet veri seti üzerinde sınıflandırma yapmıştır [31]. Sonuç olarak, SVM algoritmasının performansı artırılmış ve en yüksek doğruluk %85,71 olarak elde edilmiştir. Bu sonuçlar, literatürdeki birçok çalışmayla benzer eğilimler gösterse de, özellikle geriye doğru eleme yöntemiyle yapılan model sadeleştirilmesi literatürde daha sınırlı olarak ele alınmıştır. Bu bağlamda, çalışmanın PdM süreçlerinde veri ön işleme ve öznelik seçimi açısından literatüre katkı sunduğu düşünülmektedir.

Son olarak bu çalışmada elde edilen sonuçların literatürde yapılmış benzer çalışmalarla karşılaştırması Tablo 5'te sunulmuştur. Tabloya göre, önerilen yaklaşım özellikle k-NN algoritmasıyla %99,2 doğruluk elde ederek birçok literatür çalışmasından daha yüksek performans göstermiştir. Örneğin, Vandereycken ve Voorhaar [35] aşırı gradyan güçlendirme yöntemiyle %95,74 doğruluk elde ederken, Sharma vd. [33] rastgele orman algoritmasıyla %98,40 doğruluk raporlamıştır. Benzer şekilde, Harichandran vd. [32] hibrit gözetimli ve gözetimsiz makine öğrenmesi yaklaşımıyla %98,46 doğruluk oranına ulaşmıştır. Sengupta vd. [13] tarafından geliştirilen topluluk modeli %98,93 doğrulukla en yüksek sonuçlardan birini elde etmiştir. Ayrıca, İleri vd. [15] derin öğrenme tabanlı 1D-LeNet mimarisine

Tablo 5. AI4I 2020-PdM Veri Setinde Literatür Çalışmalarının Sınıflandırma Performansları
(Classification Performances of Literature Studies on AI4I 2020-PdM Dataset.)

Literatür	Sınıflandırma Algoritması	Sınıf	Acc (%)
Harichandran vd. [32]	Hibrit Gözetimsiz ve Gözetimli Makine Öğrenmesi	MF - nMF	98,46
Sharma vd. [33]	Rastgele Orman	MF - nMF	98,40
Sengupta vd. [13]	Hafif Gradyan Artırma Makinesi		98,47
	Karar Ağaçları		98,10
	Ekstra Ağaçlar Sınıflandırıcı		97,77
	k-En Yakın Komşular		97,41
	Ada Artırma Sınıflandırıcı		97,27
	Kudratik Ayırıcı Analizi		83,19
	Gradyan Artırma Sınıflandırıcı	MF - nMF	98,21
	Rastgele Orman Sınıflandırıcı		98,17
	Doğrusal Ayırıcı Analizi		96,87
	Lojistik Regresyon		96,80
Matzka [34]	Rastgele Sınıflandırıcı		96,76
	Naive Bayes		96,10
Matzka [34]	Önerilen Topluluk Modeli		98,93
	Bagged Trees		98,34
İleri vd. [15]	Karar Ağaçları		97,31
	Destek Vektör Makineleri		97,29
	k-En Yakın Komşular	MF - nMF	97,08
	1D-LeNet		98,50
	1D-AlexNet		98,09
Vandereycken ve Voorhaar [35]	1D-VGG16		97,95
	Aşırı Gradyan Güçlendirme	MF - nMF	95,74
Bu Çalışmada	Karar Ağaçları		97,70
	Destek Vektör Makineleri	MF - nMF	98,40
	k-En Yakın Komşular		99,20

%98,50 doğruluk oranı elde ederken, klasik makine öğrenmesi yöntemleri olan karar ağaçları, SVM ve k-NN modellerinde sırasıyla %97,31, %97,29 ve %97,08 doğruluk değerleri raporlamıştır. Bu karşılaştırmalar, önerilen yöntemin yalnızca geleneksel makine öğrenmesi algoritmalarıyla değil, aynı zamanda derin öğrenme tabanlı yaklaşımlarla da rekabet edebilecek düzeyde yüksek bir başarı sağladığını göstermektedir. Elde edilen bulgular, özellikle SMOTE yöntemiyle sınıf dengesizliğinin giderilmesi ve uygun öznelik seçimi adımlarının sınıflandırma başarımını anlamlı ölçüde artırdığını ortaya koymakta; böylece, önerilen yaklaşımın literatürde yer alan farklı yöntemlere kıyasla güçlü bir alternatif sunduğunu göstermektedir.

4. Sonuçlar (Conclusions)

Bu çalışmada, makine arızalarının ve arıza modlarının sınıflandırılmasında kullanılan DT, SVM ve k-NN algoritmalarının performansları detaylı bir şekilde değerlendirilmiştir. Elde edilen sonuçlar, veri dengesizliği probleminin, kestirimci bakım sistemlerinin başarımını olumsuz etkileyebileceğini göstermektedir. SMOTE yöntemi, rastgele seçime kıyasla daha yüksek başarı sağlamış ve veri dengesizliğini gidermede etkili olmuştur. Ayrıca, z-skor normalizasyonunun, min-max normalizasyona göre daha iyi performans sergileyerek sınıflandırma doğruluğunu artırdığı gözlemlenmiştir. Veri işleme adımlarının ardından yapılan değerlendirmelerde, geriye doğru eleme yöntemi, SMOTE ile dengelenmiş veri setinde başarıyı düşürürken, ikili sınıflara ayrılan arıza modlarından PWF-OSF sınıfında k-NN algoritması ile %100 doğruluk oranı elde edilerek en yüksek başarı sağlanmıştır. Bu bulgular, önerilen yöntemlerin ve seçilen özneliklerin makine arızası ve arıza modlarının sınıflandırılmasında başarıyı önemli ölçüde

artırabileceğini ve kestirimci bakım sistemlerinde makine öğrenimi yaklaşımlarının etkinliğini artırmada önemli bir katkı sunduğunu göstermektedir. Gelecekteki çalışmalarda, veri setinin çeşitlendirilmesi ve farklı algoritmalar, veri işleme yöntemleri ile öznelik seçim tekniklerinin kullanılmasıyla çalışmanın tekrarlanması, elde edilen bulguların kararlılığının değerlendirilmesine olanak sağlayacaktır.

Kaynaklar (References)

- Cioffi R., Travaglioni M., Piscitelli G., Petrillo A., De Felice F., Artificial intelligence and machine learning applications in smart production: Progress, trends, and directions, Sustainability, 12 (2), 492, 2020.
- Moura R., Ceotto L., Gonzalez A., Toledo R., Industrial Internet of Things (IIoT) platforms-an evaluation model, 2018 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas-A.B.D., 1002-1009, 13-15 December, 2018.
- Güzel H.Y., Basım Sanayisinde Endüstri 4.0 Teknolojilerinin Sürdürülebilir Üretime Etkileri, Yüksek Lisans Tezi, Marmara Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul, 2024.
- Fitouhi M.C., Nourelfath M., Integrating noncyclical preventive maintenance scheduling and production planning for a single machine, International Journal of Production Economics, 136 (2), 344-351, 2012.
- Hatipoğlu A., Güneri Y., Yılmaz E., A comparative predictive maintenance application based on machine and deep learning, Journal of the Faculty of Engineering and Architecture of Gazi University, 39 (2), 1037-1048, 2024.
- Ronzoni N., De Marco A., Ronchieri E., Predictive Maintenance Experiences on Imbalanced Data with Bayesian Optimization Approach, International Conference on Computational Science and Its Applications, Malaga-Spain, 120-137, July 4-7, 2022.

7. Baccarini L.M.R., Silva V.V.R., Menezes B.R., Caminhas W.M., SVM practical industrial application for mechanical faults diagnostic, *Expert Systems with Applications*, 38 (6), 6980-6984, 2011.
8. Zhang Z., Wang Y., Wang K., Fault diagnosis and prognosis using wavelet packet decomposition, Fourier transform and artificial neural network, *Journal of Intelligent Manufacturing*, 24, 1213-1227, 2013.
9. Cakir M., Guvenc M.A., Mistikoglu S., The experimental application of popular machine learning algorithms on predictive maintenance and the design of IIoT based condition monitoring system, *Computers & Industrial Engineering*, 151, 106948, 2021.
10. Çakir M., Güvenç M.A., Mistikoğlu S., IoT based condition monitoring system design for investigation of non-oil ball bearing in terms of vibration, temperature, acoustic emission, current and revolution parameters, 10th International Symposium on Intelligent Manufacturing and Service Systems, 1059-1068, September 2019.
11. Shivaramu P., Optimizing Manufacturing Processes with Predictive Maintenance Using Machine Learning and Lean Six Sigma, 2023.
12. Rayarao S.R., Rayarao S.R., Advanced Predictive Maintenance Strategies: Insights from the AI4I 2020 Dataset, *Authorea Preprints*, 2024.
13. Sengupta P., Mehta A., Rana P.S., Predictive Maintenance of Armoured Vehicles using Machine Learning Approaches, *arXiv preprint, arXiv:2307.14453*, 2023.
14. Assagaf I., Sukandi A., Abdillah A.A., Arifin S., Ga J.L., Machine predictive maintenance by using support vector machines, *Recent in Engineering Science and Technology*, 1 (01), 31-35, 2023.
15. Ileri U., Altun Y., Narin A., An Efficient Approach for Automatic Fault Classification Based on Data Balance and One-Dimensional Deep Learning, *Applied Sciences*, 14 (11), 4899, 2024.
16. Liao Y., Li M., Sun Q., Li P., Advanced stacking models for machine fault diagnosis with ensemble trees and SVM, *Applied Intelligence*, 55 (3), 1-12, 2025.
17. Meddaoui A., Hain M., Hachmoud A., The benefits of predictive maintenance in manufacturing excellence: a case study to establish reliable methods for predicting failures, *The International Journal of Advanced Manufacturing Technology*, 128 (7-8), 3685-3690, 2023.
18. Yavuz S., Deveci M., İstatiksel normalizasyon tekniklerinin yapay sinir ağı performansına etkisi, *Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 40, 167-187, 2012.
19. Schenatto K., De Souza E.G., Bazzi C.L., Gavioli A., Betzek N.M., Beneduzzi H.M., Normalization of data for delineating management zones, *Computers and Electronics in Agriculture*, 143, 238-248, 2017.
20. Blagus R., Lusa L., SMOTE for high-dimensional class-imbalanced data, *BMC Bioinformatics*, 14, 1-16, 2013.
21. Osisanwo F.Y., Akinsola J.E.T., Awodele O., Hinmikaiye J.O., Olakanmi O., Akinjobi J., Supervised machine learning algorithms: classification and comparison, *International Journal of Computer Trends and Technology (IJCTT)*, 48 (3), 128-138, 2017.
22. Panuganti J., Support Vector Machines, Medium, <https://jyothiffu.medium.com/support-vector-machines-65b4a04d177>, Yayın tarihi: 7 Şubat 2020, Erişim tarihi: 12 Haziran 2025.
23. Bhatia N., Survey of nearest neighbor techniques, *arXiv preprint, arXiv:1007.0085*, 2010.
24. Warfield S., Fast k-NN classification for multichannel image data, *Pattern Recognition Letters*, 17 (7), 713-721, 1996.
25. Fırat Ü.O., Regresyon analizinde bağımsız değişkenlerin belirlenmesini sağlayan algoritmalar, *Öneri Dergisi*, 1 (5), 49-54, 1996.
26. Uğurlu M., Doğru İ., Arslan R., Detection and classification of darknet traffic using machine learning methods, *Journal of the Faculty of Engineering and Architecture of Gazi University*, 38 (3), 1737-1746, 2023.
27. Bianchi G., Freddi F., Giuliani F., La Placa A., Implementation of an AI-based predictive structural health monitoring strategy for bonded insulated rail joints using digital twins under varied bolt conditions, *Railway Engineering Science*, 1-18, 2025.
28. Chawla N.V., Bowyer K.W., Hall L.O., Kegelmeyer W.P., SMOTE: sentetik azınlık aşırı örnekleme tekniği, *Journal of Artificial Intelligence Research*, 16, 321-357, 2002.
29. Huang X., Zhang C.Z., Yuan J., Predicting extreme financial risks on imbalanced dataset: A combined kernel FCM and kernel SMOTE based SVM classifier, *Computational Economics*, 56 (1), 187-216, 2020.
30. Herwanto H.W., Handayani A.N., Wibawa A.P., Chandrika K.L., Arai K., Comparison of min-max, z-score and decimal scaling normalization for zoning feature extraction on Javanese character recognition, 2021 7th International Conference on Electrical, Electronics and Information Engineering (ICEEIE), Jakarta-Indonesia, 1-3, 7-8 October, 2021.
31. Maulidina F., Rustam Z., Hartini S., Wibowo V.V.P., Wirasati I., Sadewo W., Feature optimization using Backward Elimination and Support Vector Machines (SVM) algorithm for diabetes classification, *Journal of Physics: Conference Series*, 012006, IOP Publishing, 1821 (1), 8-10 March, 2021.
32. Harichandran A., Raphael B., Mukherjee A., Equipment activity recognition and early fault detection in automated construction through a hybrid machine learning framework, *Computer-Aided Civil and Infrastructure Engineering*, 38 (2), 253-268, 2023.
33. Sharma N., Sidana T., Singhal S., Jindal S., Predictive maintenance: Comparative study of machine learning algorithms for fault diagnosis, *Proceedings of the International Conference on Innovative Computing & Communication (ICICC)*, June 2022.
34. Matzka S., Explainable artificial intelligence for predictive maintenance applications, 2020 Third International Conference on Artificial Intelligence for Industries (AI4I), 69-74, September 2020.
35. Vandereycken B., Voorhaar R., TTML: tensor trains for general supervised machine learning, *arXiv preprint, arXiv:2203.04352*, 2022.