

Araç Sensör Verilerinde Makine Öğrenimi Teknikleri ile Anomali Tespiti

Şevval Kaşan, Sümeyye Gülçeken

November 1, 2024

Abstract

Bu çalışmada, araç sensör verilerinde anomali tespiti için makine öğrenimi tabanlı bir yaklaşım sunulmaktadır. Çalışma, anormal durumları belirlemek amacıyla özellik mühendisliği ve topluluk modelleme tekniklerinin bir kombinasyonunu kullanmaktadır. Bu teknikler arasında Gradient Boosting, Random Forest ve XGBoost yer almaktadır. Sonuçlar orta düzeyde başarı göstermiş olup, gelecekte doğruluğu artırmak için sinir ağlarına dayalı yöntemlerin uygulanması planlanmaktadır.

1 Giriş

Araç sensör verileri, otomotiv sistemlerinin sağlık durumu ve performansı hakkında önemli bilgiler sağlar. Bu verilerdeki anomali tespiti, olası arızaların erken tespitine yardımcı olarak güvenliği artırır ve bakım maliyetlerini azaltır. Bu çalışmada, araç motor verilerindeki anomalileri tespit etmek için makine öğrenimi modellerinin bir kombinasyonu kullanılmıştır. Daha önceki araştırmalar anomali tespiti için istatistiksel ve makine öğrenimi yöntemleri kullanmıştır, ancak bu çalışmada tespit doğruluğunu artırmak amacıyla topluluk yaklaşımına odaklanılmıştır.

2 Veri Seti ve Ön İşleme

Bu çalışmada kullanılan veri seti, motor devri, yağ basıncı, yakıt basıncı, soğutucu basıncı ve sıcaklıklar gibi çeşitli motor parametrelerinden oluşmaktadır. Makine öğrenimi modelleri uygulanmadan önce veriler temizlenmiş ve eksik değerler ele alınmıştır. Sınıf dengesizliği, Sentetik Azınlık Aşırı Örnekleme Tekniği (SMOTE) kullanılarak giderilmiş ve veriler StandardScaler ile normalize edilmiştir.

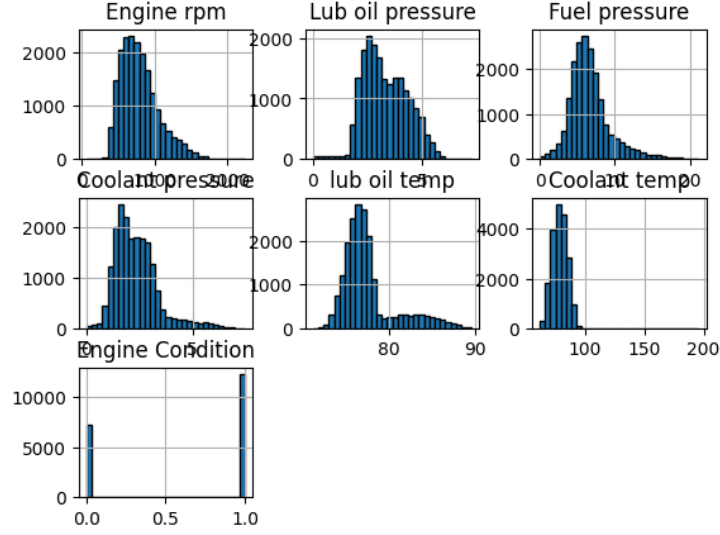


Figure 1: Veri setindeki her bir özelliğin dağılımını gösteren histogramlar. Grafiklerde "Engine rpm" (Motor Devri), "Lub oil pressure" (Yağ Basıncı), "Fuel pressure" (Yakıt Basıncı), "Coolant pressure" (Soğutucu Basıncı), "Lub oil temp" (Yağ Sıcaklığı), "Coolant temp" (Soğutucu Sıcaklığı) ve "Engine Condition" (Motor Durumu) değişkenlerinin dağılımları yer almaktadır. Her bir histogram, ilgili özelliğin veri setindeki dağılımını, yoğunluğunu ve olası aykırı değerleri anlamamıza yardımcı olmaktadır. Özellikle, "Engine Condition" değişkeni anomali tespiti için ikili bir dağılım sunmaktadır. Bu grafikler, modelin eğitiminde kullanılacak özelliklerin dağılımlarını analiz etmek ve aykırı değerleri gözlemlemek açısından önemlidir.

3 Özellik Mühendisliği

Model performansını artırmak için, mevcut veri setinden yeni ve anlamlı özellikler oluşturmak amacıyla özellik mühendisliği teknikleri uygulanmıştır. Bu özellikler şunlardır:

- **Yakıt ve Soğutucu Basıncı Oranı:** Yakıt basıncının soğutucu basıncına oranı.
- **Yağ ve Soğutucu Sıcaklık Oranı:** Yağ sıcaklığının soğutucu sıcaklığına oranı.
- **Motor Devri ve Soğutucu Basıncı Çarpımı:** Motor devrinin soğutucu basıncı ile çarpımı.
- **Basınç Toplamı:** Yağ basıncı ve yakıt basıncının toplamı.

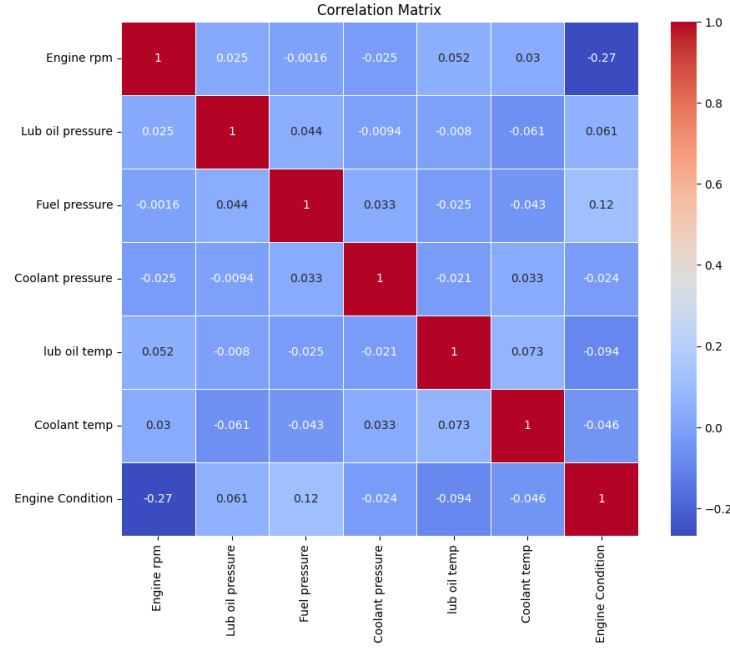


Figure 2: Veri setindeki özellikler arasındaki korelasyon matrisi. Bu grafik, yüksek derecede ilişkili özelliklerin belirlenmesine ve aşırı uyumun önlenmesine yardımcı olmaktadır. Örneğin, 'Fuel pressure' ve 'Pressure Sum' arasında yüksek bir korelasyon gözlemlenmiştir, bu da bu iki özelliğin benzer bilgi içerdiğini ve modelde birlikte kullanılmasının aşırı uyuma yol açabileceğini göstermektedir. Yüksek korelasyon, özelliklerin birlikte kullanıldığında modelin performansını nasıl etkileyebileceğini anlamamıza olanak sağlar. Düşük korelasyon seviyeleri ise modelin genelleme kapasitesini artırır.

4 Modelleme

Anomalileri tespit etmek için üç makine öğrenimi modeli kullanılmıştır:

- **Gradient Boosting Sınıflandırıcısı:** Zayıf öğrencilerin, genellikle karar ağaçlarının, bir topluluğunu oluşturarak doğruluğu artıran bir yükseltme tekniği.
- **Random Forest Sınıflandırıcısı:** Karar ağaçlarından oluşan bir topluluk modeli olup aşırı uyumu azaltır ve genelleme yeteneğini artırır.
- **XGBoost Sınıflandırıcısı:** Verimli ve ölçeklenebilir bir optimize edilmiş gradient boosting kütüphanesi.

Bu modeller ayrıca bir Voting Sınıflandırıcı kullanılarak birleştirilmiş ve her bir temel modelin güçlü yönlerinden yararlanmak amaçlanmıştır.

5 Sonuçların Değerlendirilmesi

Modellerin performansı doğruluk, kesinlik, geri çağırma ve F1-skora göre değerlendirilmiştir. Topluluk modeli, araç sensör verilerinde anomali tespiti için yaklaşık olarak

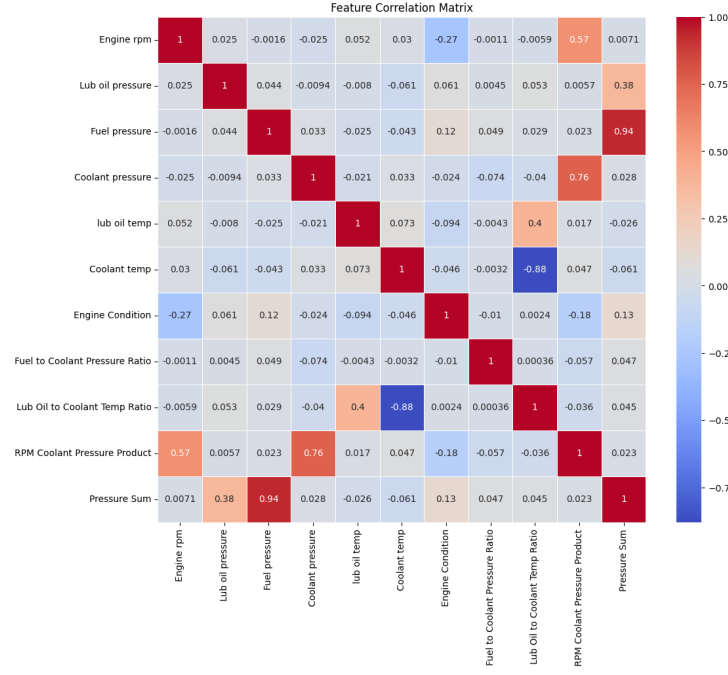


Figure 3: Özellik mühendisliği ile oluşturulan yeni özelliklerin mevcut özelliklerle olan ilişkisini gösteren korelasyon matrisi. Grafik, "Fuel to Coolant Pressure Ratio," "Lub Oil to Coolant Temp Ratio," "RPM Coolant Pressure Product," ve "Pressure Sum" gibi türetilmiş özelliklerin, orijinal özellikler ile korelasyonlarını göstermektedir. Özellikle "RPM Coolant Pressure Product" ve "Pressure Sum" özelliklerinin, diğer özelliklerle yüksek korelasyon gösterdiği gözlemlenmiştir. Bu yeni özelliklerin, model performansına etkisi ve aşırı uyuma yol açabilecek durumların değerlendirilmesi açısından önemli bir analiz sağlar.

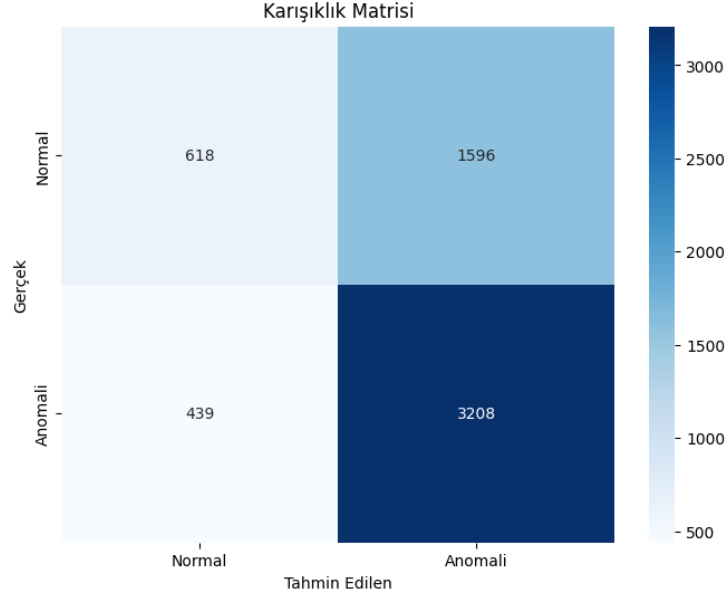


Figure 4: Topluluk Modeline Ait Karışıklık Matrisi. Matris, modelin tahmin performansını göstermektedir: Gerçek "Normal" sınıfında 618 doğru tahmin ve 1596 yanlış pozitif tahmin yer alırken, "Anomali" sınıfında 3208 doğru tahmin ve 439 yanlış negatif tahmin bulunmaktadır. Model, anomalileri daha yüksek doğrulukla tespit etmiş olup, yanlış negatiflerin düşük olması modelin güvenilirliğini artırmaktadır.

Sınıf	Kesinlik	Geri Çağırma	F1-skora
0 (Normal)	0.51	0.68	0.58
1 (Anomali)	0.75	0.60	0.66

Table 1: Topluluk Modeline Ait Sınıflandırma Raporu. Tablo, her sınıf için kesinlik, geri çağırma ve F1-skora değerlerini göstermektedir. Modelin anomali sınıfında daha yüksek kesinlik sağladığı gözlemlenmiştir, ancak geri çağırma oranı düşüktür. Bu durum, modelin anomalileri doğru tahmin etme eğiliminde olduğunu ancak bazı anormal durumları kaçırabileceğini göstermektedir.

6 Anomali Tespiti

Anomaliler, veri setindeki gözlemleri izole ederek aykırı değerleri belirleyen Isolation Forest modeli kullanılarak tespit edilmiştir. Model, veri noktalarının yaklaşık

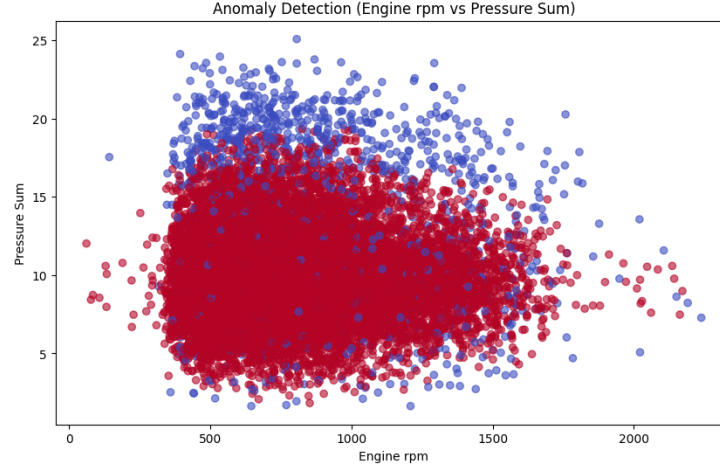


Figure 5: Anomali Tespiti Görselleştirme: Motor Devri (Engine rpm) ve Basınç Toplamı (Pressure Sum) arasındaki ilişki. Grafikte, normal veriler kırmızı renkte, anomali olarak işaretlenen veriler ise mavi renkte gösterilmiştir. Bu görselleştirme, motor devri ve basınç toplamı arasındaki ilişkiye dayanarak anomali tespiti yapılmasına olanak tanır. Mavi renkle işaretlenen anomaliler, veri setinde alışılmadık değerler olarak öne çıkmakta ve olası arıza veya olağandışı durumların göstergesi olarak değerlendirilmektedir. Bu tespitler, modelin araç sensör verilerinde anormal durumları belirlemesi açısından kritik bir rol oynamaktadır.

7 Sonuç ve Gelecek Çalışmalar

Sonuçlar, topluluk yaklaşımının araç sensör verilerinde anomali tespiti için etkili olduğunu ve orta düzeyde doğruluk elde ettiğini göstermektedir. Ancak, tespit oranlarını artırmak için daha fazla iyileştirme gerekmektedir. Gelecek çalışmalarda, TensorFlow kullanarak autoencoder ve LSTM ağları gibi derin öğrenme modelleri uygulanarak anomali tespiti yeteneklerinin artırılması hedeflenmektedir.

References

- [1] A. Himam, "Exploring Automotive Engine Health", <https://www.kaggle.com/code/goduguanilhimam/exploring-automotive-engine-health>
- [2] K. Kabil, "Machine Learning with Car Engine Health", <https://www.kaggle.com/code/kabil007/machine-learning-with-car-engine-health>

- [3] H. Sezen, "Automotive Vehicles Engine Error",<https://www.kaggle.com/code/huseyinsezen/automotive-vehicles-engine-error>
- [4] R. Yassminh, "Anomaly Detection Methods for Sensor Systems", Medium,<https://medium.com/@ryassminh/anomaly-detection-methods-for-sensor-systems-8955d7cca38>
- [5] G. Sönmez, "İzolasyon Ormanı (Isolation Forest)", Medium,<https://medium.com/@gokcenursonmez/i%CC%87zolasyon-orman%C4%B1-isolation-forest-1e1e24d70962>
- [6] "Comparing Anomaly Detection Algorithms for Outlier Detection on Toy Datasets in Scikit-Learn", GeeksforGeeks,
https://www.geeksforgeeks.org/comparing-anomaly-detection-algorithms-for-outlier-detection-on-toy-datasets-in-scikit-learn/?ref=oin_asr1
- [7] W. Dagang, "Demystifying Anomaly Detection with Autoencoder Neural Networks", Medium,
<https://medium.com/@weidagang/demystifying-anomaly-detection-with-autoencoder-neural-networks-1e235840d879>
- [8] IEEE, "Anomaly Detection in Automotive Systems",<https://ieeexplore.ieee.org/document/10419101>