International Journal of Multidisciplinary Studies and Innovative Technologies

e-ISSN: 2602-4888
dergipark.org.tr/en/pub/ijmsit
Research Article

2026, 10 (1), 72-78
DOI: 10.36287/ijmsit.10.1.8
Received: July 10, 2026; Accepted: Jul7 28, 2026

Ağ Trafiği Anomali Tespitinde MLOps Tabanlı Makine Öğrenmesi Yaklaşımlarının Karşılaştırmalı Analizi

Yunus Emre Demirdağ[1*], Betül Ay [2]

[1*]*Fırat Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, Elazığ, Türkiye (demirdag.emre.y@gmail.com) (ORCID: 0009-0004-4700-6538)*
[2] *Fırat Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, Elazığ, Türkiye (betulay@firat.edu.tr) (ORCID: 0000-0002-3060-0432)*

Türkçe Özet – Ağ trafiği anomali tespiti, siber güvenlik sistemlerinde kritik bir rol oynamaktadır. Bu çalışmada ağ trafiği anomali tespiti problemi için farklı makine öğrenmesi modüllerinin performansı, uçtan uca bir MLOps platformu üzerinde karşılaştırmalı olarak değerlendirilmiştir. Çalışma kapsamında TensorFlow Decision Forests (TFDF), LightGBM, System AutoML ve H2O AutoML modülleri aynı veri seti ve deney protokolü kullanılarak analiz edilmiştir. Modellerin performansı Accuracy, AUC, Precision, Recall, F1, Kappa ve MCC metrikleri ile değerlendirilmiştir. Deneysel sonuçlar, TFDF Gradient Boosted Trees modelinin %92.50 doğruluk skoru ile en dengeli performansı sağladığını göstermektedir. Çalışma, model seçiminin yalnızca doğruluk metriğine göre yapılmasının yeterli olmadığını ve özellikle güvenlik uygulamalarında Recall ve F1 dengesinin kritik önem taşıdığını ortaya koymaktadır. Ayrıca MLOps platformu üzerinde çalışan farklı modüllerin deneylerin tekrarlanabilirliğini ve karşılaştırılabilirliğini artırdığı gösterilmiştir.

Anahtar Kelimeler *– Ağ Trafiği Anomali Tespiti, MLOps, Makine Öğrenmesi, AutoML, Gradyan Artırma, Siber Güvenlik*

Atıf: Demirdağ, Y., Ay, B. (2026). Ağ Trafiği Anomali Tespitinde MLOps Tabanlı Makine Öğrenmesi Yaklaşımlarının Karşılaştırılmalı Analizi. International Journal of Multidisciplinary Studies and Innovative Technologies, 10(1): 72-78.

A Comparative Study of MLOps-Driven Machine Learning Approaches for Network Traffic Anomaly Detection

Extended Abstract

Research Problem/Questions – *Network traffic anomaly detection plays a critical role in cybersecurity systems. This study investigates how the performance of different machine learning modules varies in the context of network traffic anomaly detection and aims to determine which model provides the most balanced performance.*

Short Literature Review – *In the literature, various machine learning and AutoML approaches have been proposed for network traffic anomaly detection. Gradient boosting-based methods and AutoML solutions stand out due to their high performance. However, existing studies often use different datasets and experimental setups, limiting the direct comparability of models. This highlights the importance of conducting comparative analyses within a standardized MLOps environment.*

Methodology – *In this study, TensorFlow Decision Forests (TFDF), LightGBM, System AutoML, and H2O AutoML modules were evaluated on an end-to-end MLOps platform using the same dataset and experimental protocol. Model performance was assessed using multiple evaluation metrics, including Accuracy, AUC, Precision, Recall, F1-score, Kappa, and MCC.*

Results and Conclusions – *Experimental results show that the TFDF Gradient Boosted Trees model achieved the most balanced performance with an accuracy score of 92.50%. The findings indicate that relying solely on accuracy is insufficient for model selection, especially in cybersecurity applications, where Recall and F1-score are of critical importance. Furthermore, the use of an MLOps platform significantly improves the reproducibility and comparability of experiments.*

Keywords *– Network Traffic Anomaly Detection, MLOps, Machine Learning, AutoML, Cybersecurity, Gradient Boosting*

Citation: Demirdağ, Y., Ay, B. (2026). A Comparative Study of MLOps-Driven Machine Learning Approaches for Network Traffic Anomaly Detection. International Journal of Multidisciplinary Studies and Innovative Technologies, 10(1): 72-78.

I. GİRİŞ

Günümüz iletişim altyapılarında ağ trafiğinin hızla artması, siber güvenlik tehditlerinin tespitini kritik bir araştırma alanı haline getirmektedir. Geleneksel kural tabanlı saldırı tespit sistemleri, bilinmeyen veya değişen saldırı örüntülerine karşı yetersiz kalmakta; bu durum makine öğrenmesi tabanlı yaklaşımların önemini artırmaktadır [1].

Makine öğrenmesi algoritmalarının ağ anomali tespitindeki etkinliği pek çok çalışmada gösterilmiş olsa da farklı yaklaşımların aynı koşullar altında sistematik biçimde karşılaştırıldığı çalışmalar sınırlı sayıda kalmaktadır [2]. Bu boşluğu gidermek amacıyla, mevcut çalışmada uçtan uca bir MLOps platformu üzerinde TFDF, LightGBM, System AutoML ve H2O AutoML yaklaşımları aynı veri seti ve optimizasyon protokolü kullanılarak karşılaştırılmıştır.

MLOps (Machine Learning Operations), makine öğrenmesi modellerinin geliştirilmesi, dağıtılması ve izlenmesine yönelik süreçleri otomatikleştiren bir yaklaşımdır [3]. Çalışmada kullanılan MLOps platformu; veri yönetimi, model eğitimi, hiperparametre optimizasyonu ve deployment aşamalarını tek bir ortamda bütünleştirmektedir. Bu sayede deney tekrarlanabilirliği ve karşılaştırma tutarlılığı sağlanmaktadır.

Çalışmanın temel katkıları şu şekilde özetlenebilir: (i) Dört farklı yaklaşımın aynı MLOps altyapısı üzerinde karşılaştırılması, (ii) Optuna tabanlı otomatik hiperparametre optimizasyonu ile adil karşılaştırma koşullarının sağlanması, (iii) TFDF Gradient Boosted Trees'in ağ anomali tespitindeki üstünlüğünün gösterilmesi.

II. İLGİLİ ÇALIŞMALAR

Ağ trafiği anomali tespiti, siber güvenlik alanında uzun süredir aktif olarak araştırılan bir konudur. Bu problem kapsamında hem istatistiksel yöntemler hem de makine öğrenmesi temelli yaklaşımlar literatürde geniş biçimde incelenmiştir. Chandola ve ark. [1], anomali tespiti alanındaki yöntemleri kapsamlı bir biçimde ele alarak istatistiksel modeller, yakınlık (proximity) tabanlı yöntemler, yoğunluk tabanlı yaklaşımlar ve sınıflandırma temelli yöntemler gibi farklı kategorileri karşılaştırmıştır. Bu çalışma, anomali tespitinin farklı veri türlerinde uygulanabilirliğini ortaya koyan temel referans çalışmalarından biri olarak kabul edilmektedir.

Makine öğrenmesi yöntemlerinin gelişmesiyle birlikte ağ saldırı tespiti problemlerinde denetimli öğrenme algoritmalarının kullanımı yaygınlaşmıştır. Kaur ve Singh [2], ağ saldırı tespiti için karar ağacı ve rastgele orman (Random Forest) algoritmalarını incelemiş ve ensemble tabanlı yöntemlerin tekil sınıflandırıcılara kıyasla daha yüksek doğruluk ve genelleme başarısı sağlayabildiğini göstermiştir. Benzer şekilde birçok çalışma, özellikle tablo (tabular) veri üzerinde karar ağacı tabanlı yöntemlerin güçlü performans sergilediğini ortaya koymaktadır [4]. Zoppi ve ark. [5], karar ağaçlarına dayalı sınıflandırıcıların derin öğrenme modellerine kıyasla daha yüksek performans sağladığını, daha hızlı eğitildiği ve daha az sayıda, anlaşılması ve ayarlanması kolay hiperparametreye sahip olduğunu vurgulamıştır. Ayrıca çalışmada [5], ağaç tabanlı modeller ile derin öğrenme modellerinin birlikte kullanıldığı ensemble yaklaşımlarının performansı artırmadığı gösterilmiştir.

Son yıllarda gradient boosting algoritmaları, özellikle yapılandırılmış veri üzerinde yüksek performans sağlamaları nedeniyle yaygın olarak kullanılmaktadır [6, 7]. Chen ve Guestrin [9] tarafından önerilen XGBoost algoritması, ölçeklenebilirliği ve paralel hesaplama yetenekleri sayesinde veri madenciliği ve makine öğrenmesi yarışmalarında önemli başarılar elde etmiştir. Benzer şekilde Prokhorenkova ve ark. [8] tarafından geliştirilen CatBoost algoritması, kategorik değişkenleri daha etkin şekilde işleyebilmesi ve aşırı uyumu (overfitting) azaltan teknikleri sayesinde güçlü bir alternatif sunmaktadır.

Karar ağacı tabanlı modellerin modern makine öğrenmesi altyapılarıyla entegrasyonu da son yıllarda önem kazanmıştır [10]. TensorFlow Decision Forests (TFDF), karar ağaçları ve ensemble yöntemlerini TensorFlow ekosistemi ile birleştiren bir modül olarak geliştirilmiştir. TFDF; Random Forest, Gradient Boosted Trees ve CART gibi modelleri yüksek performanslı bir şekilde eğitmeye olanak sağlamaktadır ve özellikle büyük veri kümelerinde ölçeklenebilirliği ile dikkat çekmektedir [11].

Makine öğrenmesi modellerinin geliştirilmesi ve üretim ortamına taşınması sürecinde ise MLOps (Machine Learning Operations) yaklaşımı önemli bir rol oynamaktadır. Sculley ve ark. [3], makine öğrenmesi sistemlerinde ortaya çıkan "teknik borç" kavramını inceleyerek veri bağımlılıkları, pipeline karmaşıklığı ve model yönetimi gibi sorunların üretim sistemlerinde ciddi zorluklar oluşturabileceğini göstermiştir. Bu nedenle model geliştirme, deney yönetimi, veri versiyonlama ve model dağıtım süreçlerinin otomatikleştirilmesi büyük önem taşımaktadır.

Son yıllarda otomatik makine öğrenmesi (AutoML) yaklaşımları da model geliştirme sürecini hızlandırmak amacıyla yaygın biçimde kullanılmaktadır [12]. H2O AutoML [13] gibi açık kaynak platformlar, farklı algoritmaları ve hiperparametre konfigürasyonlarını otomatik olarak deneyerek en iyi modeli belirlemeyi hedeflemektedir. Bu yaklaşım, özellikle farklı makine öğrenmesi algoritmalarının karşılaştırmalı analizini gerçekleştirmek isteyen araştırmacılar için önemli avantajlar sağlamaktadır.

Bu çalışmada ise literatürdeki çalışmalardan farklı olarak TFDF, LightGBM [14], System AutoML ve H2O AutoML modülleri aynı veri seti ve aynı hiperparametre optimizasyon protokolü kullanılarak uçtan uca bir MLOps platformu üzerinde karşılaştırmalı olarak değerlendirilmiştir. Böylece farklı makine öğrenmesi modüllerinin performansı tutarlı deney koşulları altında analiz edilmiştir.

III.YÖNTEM

A. *Veri Problemi ve Tanımı*

Bu çalışmada veri kaynağı olarak Kaggle üzerinde yayımlanan Network Traffic Anomaly Detection Dataset kullanılmıştır [15]. Deneylerde kullanılan tabloda her satır bir ağ trafiği gözlemini (paket/akış temsili) ifade etmektedir. Tahmin hedefi, label değişkeni üzerinden ikili sınıflandırma olarak tanımlanmıştır: 0 normal trafik, 1 anomali (zararlı) trafik.

Özellik uzayı, ağ davranışını farklı boyutlardan temsil eden öznitelikleri içermektedir: paket boyutu ve zamansal ölçüler (ör. packet_size, inter_arrival_time), bağlantı/port bilgileri (src_port, dst_port), kısa zaman penceresinde trafik yoğunluğu (packet_count_5s), istatistiksel/sinyal tabanlı göstergeler (spectral_entropy, frequency_band_energy) ve protokol tabanlı ikili göstergeler (ör. protocol_type_TCP, protocol_type_UDP). Bu yapı, hem klasik ağaç tabanlı

modellerin hem de AutoML sistemlerinin anomali ayrımında faydalanabileceği heterojen bir temsil sağlamaktadır.

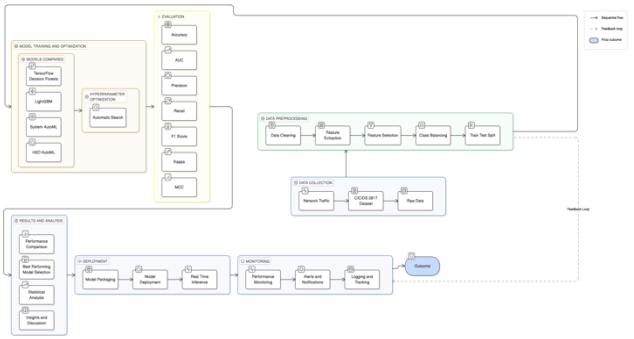
Veri ön işleme aşamasında eksik kayıtlar temizlenmiştir. Ek olarak, src_ip_* ve dst_ip_* sütunları ile mean_packet_size değişkeni veri setinden çıkarılmıştır. Bu değişkenlerin sabit/aynı değer taşıdığı ve ayırt edici bilgi üretmediği gözlemlendiğinden, model eğitimine anlamlı katkı sağlamayacağı varsayımıyla temizlenmiştir. Sonrasında eğitim-doğrulama-test ayrımı stratified (etiket oranını koruyan) biçimde yapılmış ve MLOps hattında tekrar üretilebilirlik için aynı dönüşümler tüm modüllere aynen uygulanmıştır. Sınıf dağılımındaki dengesizlik (normal sınıfın baskınlığı) nedeniyle değerlendirmede yalnızca doğruluk metriğine bağlı kalınmamış; ROC-AUC ve F1 skorları birincil karşılaştırma ölçütleri olarak kullanılmıştır.

B. *MLOps Platformu*

Deneyler, uçtan uca bir MLOps platformu üzerinde gerçekleştirilmiştir. Platform; veri setinin yüklenmesi, ön işleme, model eğitimi, hiperparametre optimizasyonu, sonuç karşılaştırması ve model dağıtımı aşamalarını tek bir çatı altında yönetmektedir. Altyapı düzeyinde gerçekleştirilen optimizasyonlar sayesinde kısa sürede çok sayıda deney tamamlanabilmektedir.

Platform, hiperparametre optimizasyonu için Optuna kütüphanesini [16] entegre biçimde kullanmaktadır. Her deneyde 100 deneme (trial) ile accuracy-maximize hedefi doğrultusunda otomatik arama yapılmış; bu yaklaşım farklı modüller arasında tutarlı ve adil bir karşılaştırma zemini sağlamıştır. Eğitim tamamlandıktan sonra modeller platform bünyesindeki deployment ortamına aktarılarak servis edilebilir hale getirilmektedir.

Bu çalışmada kullanılan MLOps tabanlı makine öğrenmesi pipeline'ı Şekil 1'de gösterilmektedir. Süreç; veri toplama, veri ön işleme, model eğitimi ve optimizasyonu, değerlendirme, model dağıtımı ve izleme aşamalarından oluşmaktadır.



Şekil 1. MLOps tabanlı ağ trafiği anomali tespiti pipeline'ının genel mimarisi

Şekil 1 incelendiğinde, veri toplama aşamasından başlayarak model eğitimi, hiperparametre optimizasyonu ve performans değerlendirme süreçlerinin entegre bir yapı içerisinde gerçekleştirildiği görülmektedir. Ayrıca model dağıtımı ve izleme aşamaları ile sistemin üretim ortamına uygunluğu sağlanmaktadır.

C. *Kullanılan Yöntemler*

Bu çalışmada ağ trafiği anomali tespiti problemi kapsamında farklı makine öğrenmesi yaklaşımlarının performansını karşılaştırmak amacıyla dört farklı modül değerlendirilmiştir. Seçilen modüller, hem klasik makine öğrenmesi algoritmalarını hem de otomatik makine öğrenmesi (AutoML) yaklaşımlarını temsil edecek şekilde belirlenmiştir. Böylece farklı model geliştirme paradigmalarının aynı veri ve deney protokolü altında karşılaştırılması mümkün olmuştur.

TFDF (TensorFlow Decision Forests): TensorFlow Decision Forests (TFDF), Google tarafından geliştirilen ve karar ağacı tabanlı modelleri TensorFlow ekosistemi içerisinde sunan açık kaynaklı bir makine öğrenmesi kütüphanesidir. TFDF; Random Forest, Gradient Boosted Trees ve CART gibi ağaç tabanlı algoritmaların yüksek performanslı ve ölçeklenebilir şekilde eğitilmesini sağlamaktadır. Modül, özellikle yapılandırılmış (tabular) veri üzerinde güçlü performans göstermesi ile bilinmektedir. Bu çalışmada TFDF modülüü kapsamında Gradient Boosted Trees ve Random Forest modelleri eğitilmiş ve performansları karşılaştırılmıştır.

LightGBM: LightGBM, Microsoft tarafından geliştirilen ve gradient boosting tabanlı karar ağacı algoritmalarını yüksek verimlilikle uygulayan bir makine öğrenmesi modülüdür. Geleneksel boosting yöntemlerinden farklı olarak yaprak bazlı büyüme stratejisi (leaf-wise tree growth) kullanması sayesinde daha hızlı eğitim süresi ve daha yüksek doğruluk sağlayabilmektedir. Ayrıca histogram tabanlı öğrenme mekanizması ve paralel eğitim desteği ile büyük veri setlerinde ölçeklenebilir bir yapı sunmaktadır. Bu çalışma kapsamında LightGBM'in farklı booster türleri değerlendirilmiştir. Deneylerde gbdt (Gradient Boosting Decision Tree), dart (Dropouts meet Multiple Additive Regression Trees), goss (Gradient-based One-Side Sampling) ve rf (Random Forest) konfigürasyonları test edilmiştir.

System AutoML: System AutoML, çalışmada kullanılan MLOps platformu içerisinde yer alan entegre otomatik makine öğrenmesi modülüdür. Bu modül, farklı makine öğrenmesi algoritmalarını otomatik olarak eğitmekte ve hiperparametre optimizasyonunu gerçekleştirmektedir. System AutoML, scikit-learn tabanlı algoritmaları kullanarak model seçimi ve hiperparametre arama süreçlerini otomatikleştirir. Bu çalışmada Decision Tree, Random Forest, AdaBoost ve Logistic Regression gibi algoritmalar değerlendirilmiştir. Hiperparametre optimizasyonu sürecinde Optuna kütüphanesi kullanılmış ve her model için belirlenen arama uzayında otomatik denemeler gerçekleştirilmiştir.

H2O AutoML: H2O AutoML, H2O.ai tarafından geliştirilen açık kaynaklı bir otomatik makine öğrenmesi platformudur. Bu modül, veri seti üzerinde farklı algoritmaları otomatik olarak eğiterek en iyi performansı veren modeli belirlemeyi amaçlamaktadır. H2O AutoML; Generalized Linear Models (GLM), Gradient Boosting Machines (GBM), Random Forest ve Stacked Ensemble gibi çeşitli algoritmaları paralel olarak eğitmekte ve model performanslarını karşılaştırmaktadır. Ayrıca otomatik özellik işleme, model sıralama ve ensemble oluşturma gibi özellikler sunarak model geliştirme sürecini hızlandırmaktadır. Bu çalışmada H2O AutoML modülü kullanılarak veri seti üzerinde otomatik model araması gerçekleştirilmiş ve elde edilen en iyi modeller diğer modüller ile karşılaştırılmıştır.

Bu dört modülün aynı veri seti ve hiperparametre optimizasyon protokolü altında değerlendirilmesi, farklı makine öğrenmesi yaklaşımlarının performansını adil ve tekrarlanabilir bir biçimde karşılaştırmayı mümkün kılmıştır.

IV. DENEYSEL SONUÇLAR

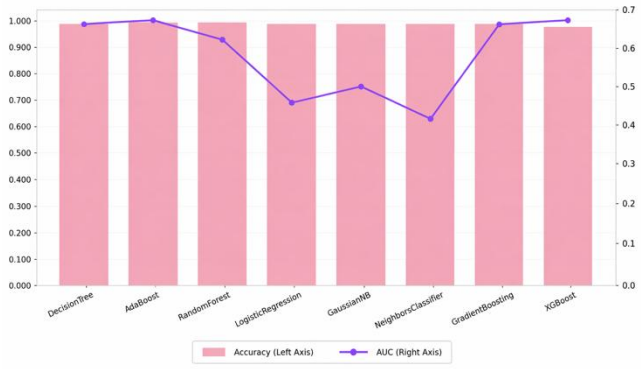
Bu çalışmada farklı makine öğrenmesi modüllerinin ağ trafiği anomali tespiti problemindeki performansını değerlendirmek amacıyla çeşitli deneyler gerçekleştirilmiştir. Deneyler sırasında tüm modeller aynı veri seti ve aynı ön işleme adımları kullanılarak eğitilmiş ve sonuçlar ortak bir MLOps platformu üzerinden karşılaştırılmıştır. Böylece farklı modüllerin performansı tutarlı ve tekrarlanabilir koşullar altında analiz edilmiştir.

Model performanslarını değerlendirmek amacıyla Accuracy, AUC, F1 Score, Cohen's Kappa, Matthews Correlation Coefficient (MCC), Precision ve Recall metrikleri kullanılmıştır. Accuracy metriği modelin genel sınıflandırma başarısını ölçerken, Precision ve Recall değerleri özellikle anomali tespitinde yanlış pozitif ve yanlış negatif oranlarının analiz edilmesini sağlamaktadır. F1 skoru, precision ve recall değerlerinin harmonik ortalaması olup dengesiz veri setlerinde daha güvenilir bir performans göstergesi olarak kabul edilmektedir. AUC (Area Under ROC Curve) metriği modelin farklı eşik değerleri altında sınıfları ayırt edebilme kapasitesini ölçmektedir. Cohen's Kappa ve MCC metrikleri ise özellikle sınıf dengesizliği bulunan veri setlerinde model performansının daha güvenilir biçimde değerlendirilmesini sağlamaktadır.

A. System AutoML Sonuçları

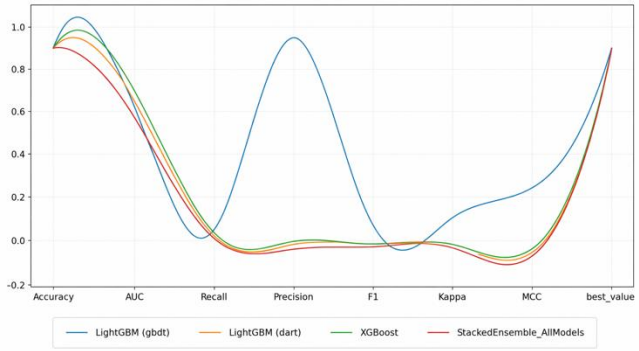
System AutoML modülü kapsamında Decision Tree, AdaBoost, Random Forest, Logistic Regression ve Gaussian Naive Bayes algoritmaları değerlendirilmiştir.

AutoML modellerinin doğruluk ve AUC performanslarının karşılaştırılması Şekil 2'de gösterilmektedir.



Şekil 2. System AutoML modellerinin Accuracy ve AUC metriklerine göre karşılaştırılması

Şekil 2 incelendiğinde Decision Tree ve AdaBoost modellerinin yaklaşık %92 doğruluk ile en yüksek doğruluk değerine ulaştığı görülmektedir. Modellerin sınıf bazlı performansını incelemek için diğer metriklerin de değerlendirilmesi gerekmektedir. AutoML kapsamında kullanılan modellerin tüm performans metriklerine göre karşılaştırılması Şekil 3'te sunulmaktadır.



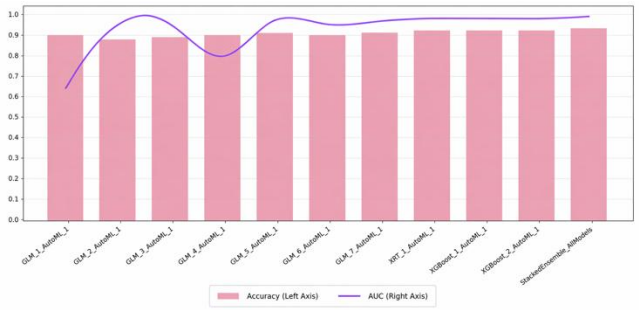
Şekil 3. System AutoML modellerinin Accuracy, AUC, Recall, Precision, F1, Kappa ve MCC metriklerine göre performans karşılaştırması

Şekil 3 incelendiğinde Decision Tree ve AdaBoost modellerinin yüksek doğruluk değerlerine rağmen F1 skorunun 0.1111 seviyesinde kaldığı görülmektedir. Bu durum modellerin anomali sınıfını doğru tespit etme performansının sınırlı olduğunu göstermektedir.

B. H2O AutoML Sonuçları

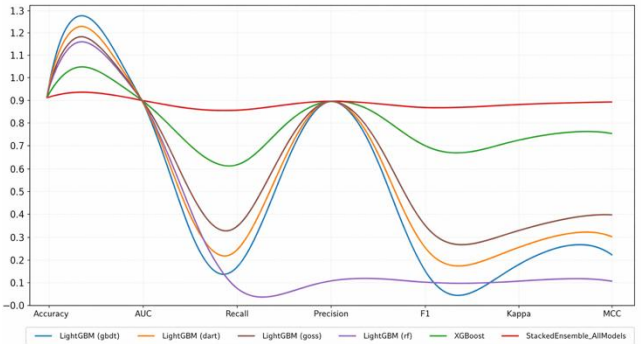
H2O AutoML modülü kapsamında GLM, GBM ve Stacked Ensemble modelleri otomatik olarak eğitilmiştir.

Bu modellerin doğruluk ve AUC performanslarının karşılaştırılması Şekil 4'te gösterilmektedir.



Şekil 4. H2O AutoML modellerinin Accuracy ve AUC metriklerine göre karşılaştırılması

Şekil 4'te görüldüğü üzere bazı modeller oldukça yüksek AUC değerleri üretmiştir. Özellikle GBM tabanlı modellerin ayrıştırma performansının yüksek olduğu gözlemlenmektedir. H2O AutoML kapsamında elde edilen tüm metriklerin karşılaştırılması ise Şekil 5'te verilmiştir.



Şekil 5. H2O AutoML modellerinin Accuracy, AUC, Recall, Precision, F1, Kappa ve MCC metriklerine göre performans karşılaştırması

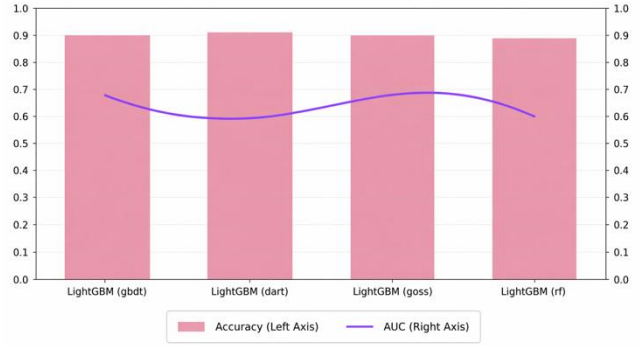
Şekil 5 incelendiğinde GBM_3 modelinin 0.9948 AUC değeri ile en yüksek ayrıştırma performansını gösterdiği görülmektedir. Ancak modelin F1 skorunun görece düşük

kalması, anomali sınıfının doğru tespit edilme oranının sınırlı olduğunu göstermektedir.

C. LightGBM Sonuçları

LightGBM modülü kapsamında gbdt, dart, goss ve rf booster türleri değerlendirilmiştir.

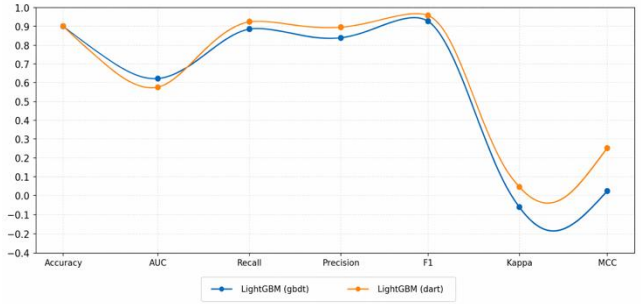
Bu modellerin doğruluk ve AUC performanslarının karşılaştırılması Şekil 6'da sunulmaktadır.



Şekil 6. LightGBM modellerinin Accuracy ve AUC metriklerine göre karşılaştırılması

Şekil 6 incelendiğinde LightGBM modellerinin doğruluk değerlerinin birbirine oldukça yakın olduğu görülmektedir.

LightGBM modellerinin tüm metrikler açısından performans karşılaştırması ise Şekil 7'de gösterilmektedir.



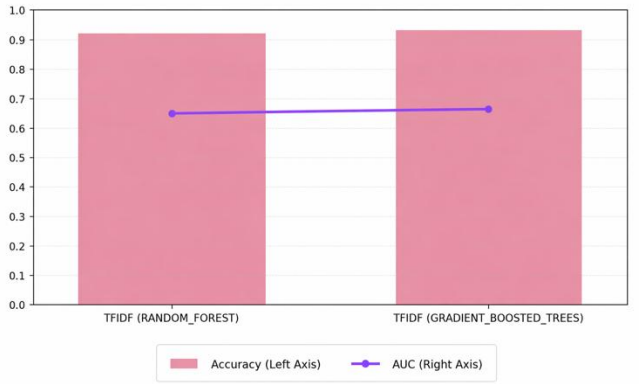
Şekil 7. LightGBM modellerinin Accuracy, AUC, Recall, Precision, F1, Kappa ve MCC metriklerine göre performans karşılaştırması

Şekil 7 incelendiğinde LightGBM (dart) modelinin 0.8861 F1 skoru ile LightGBM modelleri arasında en iyi performansı gösterdiği görülmektedir. Ayrıca modelin precision değerinin de oldukça yüksek olduğu dikkat çekmektedir.

D. TFDF Sonuçları

TensorFlow Decision Forests modülü kapsamında Random Forest ve Gradient Boosted Trees modelleri değerlendirilmiştir.

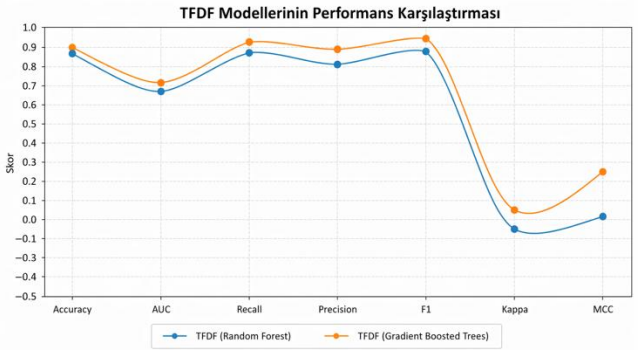
Bu modellerin doğruluk ve AUC performanslarının karşılaştırılması Şekil 8'de verilmiştir.



Şekil 8. TFDF modellerinin Accuracy ve AUC metriklerine göre karşılaştırılması

Şekil 8 incelendiğinde Gradient Boosted Trees modelinin Random Forest modeline kıyasla daha yüksek doğruluk performansı gösterdiği görülmektedir.

TFDF modellerinin tüm metriklere göre performans karşılaştırması ise Şekil 9'da sunulmaktadır. Şekil 9 incelendiğinde TFDF Gradient Boosted Trees modelinin %92.50 doğruluk, 0.8969 F1 skoru ve 0.3297 MCC değeri ile en iyi genel performansı elde ettiği görülmektedir.



Şekil 9. TFDF modellerinin Accuracy, AUC, Recall, Precision, F1, Kappa ve MCC metriklerine göre performans karşılaştırması

Farklı makine öğrenmesi modüllerinin performansını daha kompakt ve karşılaştırılabilir bir şekilde sunmak amacıyla, her yaklaşımın en iyi performans gösteren modeli seçilerek Tablo 1'de özetlenmiştir. Bu tablo, modellerin yalnızca doğruluk değil, aynı zamanda AUC, F1 skoru ve MCC gibi kritik metrikler açısından genel performanslarını karşılaştırmalı olarak değerlendirmeye olanak sağlamaktadır.

Tablo 1. Farklı makine öğrenmesi yaklaşımlarının en iyi modellerine ait performans karşılaştırması

Model	Accuracy	AUC	F1 Score	MCC
System AutoML (Best)	~0.92	~0.67	0.1111	Low
H2O AutoML (GBM_3)	~0.92	0.9948	Low	Medium
LightGM (DART)	~0.91	~0.58	0.8861	Medium
TFDF (Gradient Boosted)	0.9250	~0.70	0.8969	0.3297

Tablo 1 incelendiğinde, TFDF Gradient Boosted Trees modelinin doğruluk, F1 skoru ve MCC metrikleri açısından en dengeli performansı sağladığı görülmektedir. H2O AutoML kapsamında elde edilen GBM modeli, 0.9948 AUC değeri ile en yüksek ayrıştırma performansını sunmasına rağmen, F1 ve recall metriklerinde sınırlı kalmıştır. LightGBM DART modeli, özellikle F1 skoru açısından güçlü bir performans sergilemiştir. Buna karşın System AutoML modelleri yüksek doğruluk değerlerine ulaşmasına rağmen, düşük F1 skorları nedeniyle anomali tespitinde yetersiz kalmıştır. Bu sonuçlar, model seçiminde yalnızca doğruluk metriğine odaklanmanın yeterli olmadığını ve özellikle dengesiz veri setlerinde F1 ve MCC gibi metriklerin dikkate alınması gerektiğini göstermektedir.

E. Açıklanabilirlik ve Güvenilirlik Analizi

Bu çalışmada makine öğrenmesi modellerinin yalnızca performans metrikleri değil, aynı zamanda açıklanabilirlik ve güvenilirlik boyutları açısından da değerlendirilmesi hedeflenmiştir. Bu kapsamda, özellikle ağaç tabanlı modeller ve AutoML yaklaşımlarının karar mekanizmalarını analiz etmek amacıyla SHAP (Shapley Additive Explanations) yöntemi kullanılmıştır.

SHAP analizi sonucunda, model kararlarını en fazla etkileyen özelliklerin başında trafik yoğunluğu ve zaman temelli değişkenlerin (packet_count_5s ve inter_arrival_time) geldiği gözlemlenmiştir. Bu durum, anomali davranışlarının çoğunlukla kısa zaman aralıklarında oluşan trafik dalgalanmaları ile ilişkili olduğunu göstermektedir.

TFDF ve LightGBM gibi ağaç tabanlı modellerin, özellik katkılarını daha açık ve izlenebilir biçimde sunabilmesi nedeniyle daha yüksek yorumlanabilirlik sağladığı görülmüştür. Buna karşın H2O AutoML gibi otomatik modelleme yaklaşımlarının, özellikle ensemble yapılar içermesi nedeniyle karar süreçlerinin daha karmaşık hale geldiği ve yorumlanabilirliğin sınırlı olduğu tespit edilmiştir.

Güvenilirlik açısından değerlendirildiğinde, yalnızca yüksek doğruluk veya AUC değerine sahip modellerin güvenlik uygulamaları için yeterli olmadığı; özellikle düşük recall değerine sahip modellerin kritik saldırıları kaçırma riski taşıdığı görülmüştür. Bu nedenle model seçimi sürecinde performans metrikleri ile birlikte açıklanabilirlik ve hata türlerinin de dikkate alınması gerektiği ortaya konulmuştur.

V. SONUÇ

Bu çalışmada ağ trafiği anomali tespiti problemi kapsamında TFDF, LightGBM, System AutoML ve H2O AutoML modülleri hem model performansı hem de operasyonel uygulanabilirlik açısından karşılaştırmalı olarak değerlendirilmiştir. Deneysel sonuçlar, farklı makine öğrenmesi modüllerinin aynı veri seti üzerinde farklı performans sonuçları sergilediğini ve bu nedenle tek bir modelin tüm senaryolar için en uygun çözüm olmayabileceğini göstermektedir. Bu bulgu, model seçiminde problem bağlamının ve uygulama gereksinimlerinin dikkate alınması gerektiğini ortaya koymaktadır.

Özellikle siber güvenlik uygulamalarında yanlış negatiflerin (false negative) yüksek maliyet oluşturabildiği durumlarda, yalnızca genel doğruluk (accuracy) metriğine odaklanmanın yeterli olmadığı görülmüştür. Bu tür senaryolarda Recall ve F1 skoru gibi metriklerin dengeli biçimde değerlendirilmesi, anomali tespit sistemlerinin etkinliği açısından daha kritik bir rol oynamaktadır. Deneysel sonuçlar, özellikle gradient boosting tabanlı yöntemlerin yapılandırılmış ağ trafiği verisi üzerinde daha dengeli bir performans sergilediğini göstermektedir.

Çalışmanın temel katkısı, model seçimini yalnızca istatistiksel performans metrikleri üzerinden değil; aynı zamanda eğitim süresi, çıkarım gecikmesi, kaynak tüketimi ve bakım kolaylığı gibi MLOps kriterleriyle birlikte ele alan bütüncül bir değerlendirme yaklaşımı sunmasıdır. Bu yaklaşım, makine öğrenmesi modellerinin araştırma ortamından üretim sistemlerine aktarılabilirliğini artırmakta ve daha sürdürülebilir bir model yönetimi sürecinin oluşturulmasına katkı sağlamaktadır.

Gelecek çalışmalarda MLOps süreçlerine açıklanabilir yapay zeka bileşenlerinin daha sistematik biçimde entegre edilmesi hedeflenmektedir. Bu yaklaşım, yalnızca yüksek performanslı değil aynı zamanda şeffaf, denetlenebilir ve güvenilir yapay zeka sistemlerinin geliştirilmesine katkı sağlayacaktır. Ayrıca değişken ağ ortamlarına uyum sağlayabilen çevrim içi öğrenme (online learning) mekanizmalarının entegrasyonu planlanmaktadır. Ek olarak veri dağılımındaki değişimleri tespit edebilmek amacıyla concept drift algılama ve otomatik model güncelleme boru hatlarının geliştirilmesi hedeflenmektedir. Model seçiminin daha etkin bir şekilde yapılabilmesi için başarı, gecikme ve hesaplama maliyeti gibi farklı kriterlerin birlikte ele alındığı çok amaçlı optimizasyon (multi-objective optimization) yaklaşımlarının da değerlendirilmesi planlanmaktadır.

REFERENCES

[1] V. Chandola, A. Banerjee, ve V. Kumar, "Anomaly detection: A survey," ACM Computing Surveys, c. 41, s. 3, ss. 1–58, 2009.

[2] S. Kaur ve M. Singh, "Hybrid intrusion detection and signature generation using deep recurrent neural networks," Neural Computing and Applications, c. 32, ss. 7859–7877, 2020.

[3] D. Sculley ve ark., "Hidden technical debt in machine learning systems," Advances in Neural Information Processing Systems, c. 28, 2015.

[4] Kumar, LK Suresh, et al. "Anomaly-based intrusion detection on benchmark datasets for network security: a comprehensive evaluation." Scientific Reports 16.1 (2026): 8507.

[5] Zoppi, Tommaso, Stefano Gazzini, and Andrea Ceccarelli. "Anomaly-based error and intrusion detection in tabular data: No DNN outperforms tree-based classifiers." Future Generation Computer Systems 160 (2024): 951-965.

[6] Shwartz-Ziv, Ravid, and Amitai Armon. "Tabular data: Deep learning is not all you need." Information fusion 81 (2022): 84-90.

[7] McElfresh, Duncan, et al. "When do neural nets outperform boosted trees on tabular data?." Advances in Neural Information Processing Systems 36 (2023): 76336-76369.

[8] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, ve A. Gulin, "CatBoost: Unbiased boosting with categorical features," NeurIPS, 2018.

[9] T. Chen ve C. Guestrin, "XGBoost: A scalable tree boosting system," KDD, ss. 785–794, 2016.

[10] Guillame-Bert, Mathieu, et al. "Yggdrasil decision forests: A fast and extensible decision forests library." Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2023.

[11] M. Passos, "TensorFlow Decision Forests," Google Research, 2021. [Çevrimiçi].

[12] Hutter, Frank, Lars Kotthoff, and Joaquin Vanschoren. Automated machine learning: methods, systems, challenges. Springer, 2019.

[13] LeDell, Erin, and Sebastien Poirier. "H2o automl: Scalable automatic machine learning." Proceedings of the AutoML Workshop at ICML. Vol. 2020.

[14] Ke, Guolin, et al. "Lightgbm: A highly efficient gradient boosting decision tree." Advances in neural information processing systems 30 (2017).
[15] Network Traffic Anomaly Detection Dataset, Kaggle, 2024. Erişim: https://www.kaggle.com/datasets/ziya07/network-traffic-anomaly-detection-dataset
[16] T. Akiba, S. Sano, T. Yanase, T. Ohta, ve M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," KDD, ss. 2623–2631, 2019.