Research Article/Araştırma Makalesi

GAZİ
JOURNAL OF ENGINEERING SCIENCES

Classification of Encrypted Traffic with Machine Learning Algorithms

Nur Betül Demirel[*a], Aydın Erden [b]

Submitted: 04.06.2024 Revised: 09.12.2024 Accepted: 23.01.2025 doi:10.30855/gmbd.070525N04

ABSTRACT

In this study, machine learning and ensemble learning methods have been developed for the classification of encrypted network traffic. Using the ISCXVPN2016 dataset, the performances of Light Gradient Boosting Machine (LGBM), k-Nearest Neighbors (KNN), Support Vector Machines (SVM), and Logistic Regression (LR) algorithms were compared. The data underwent various processes, such as feature selection, addressing class imbalances via Synthetic Minority Over-Sampling Technique (SMOTE), and hyperparameter optimization. The LGBM algorithm stood out due to its high accuracy and fast performance, providing an effective model for accurate and rapid classification of encrypted traffic. This work contributes to the development of effective solutions to detect and prevent cybersecurity threats.

Keywords: Machine Learning, Ensemble Learning, Encrypted Network Traffic, Network Traffic Classification, Cybersecurity

[a,*] Marmara University, Institute for Graduate Studies in Pure Applied Sciences, Dept. of Computer Engineering 34722 - İstanbul, Türkiye
Orcid: 0009-0001-9782-2507
e mail: nur.okur@tubitak.gov.tr

[b] Marmara University, Faculty of Business Administration, Dept. of Management Information Systems 34854 - İstanbul, Türkiye
Orcid: 0000-0002-5124-8335

*Corresponding author: nur.okur@tubitak.gov.tr

Makine Öğrenmesi Algoritmaları ile Şifreli Trafiğin Sınıflandırılması

ÖZ

Bu çalışmada, şifreli ağ trafiğinin sınıflandırılmasına yönelik olarak makine öğrenmesi ve topluluk öğrenmesi yöntemleri geliştirilmiştir. Çalışmada, ISCXVPN2016 veri seti kullanılarak Light Gradient Boosting Machine (LGBM), k-Nearest Neighbours (KNN), Support Vector Machines (SVM) ve Logistic Regression (LR) algoritmalarının performansları karşılaştırılmıştır. Veriler özellik seçimi, Synthetic Minority Over-Sampling Technique (SMOTE) ile sınıf dengesizliklerinin giderilmesi ve hiperparametre optimizasyonu gibi süreçlerden geçirilerek analiz edilmiştir. LGBM algoritması, yüksek doğruluk ve hızlı performansıyla öne çıkmış, şifreli trafiğin doğru ve hızlı sınıflandırılması için etkin bir model sunulmuştur. Çalışma, siber güvenlik tehditlerini tespit etmek ve önlemek için etkili çözümler geliştirilmesine katkıda bulunmaktadır.

Anahtar Kelimeler: Makine Öğrenmesi, Topluluk Öğrenmesi, Şifreli Ağ Trafiği, Ağ Trafiği Sınıflandırma, Siber Güvenlik

To cite this article: N. B. Demirel, A. Erden, "Classification of Encrypted Traffic with Machine Learning Algorithms", *Gazi Journal of Engineering Sciences,* vol.xx, no.x, pp. xx-xx, 2025. doi:10.30855/gmbd.070525N04

1. Giriş (Introduction)

Günümüzde internet, bilgiye erişim ve iletişimin sağlanmasında temel bir araç haline gelmiştir. Dünya genelinde milyarlarca insana, geniş bir bilgi yelpazesine anında erişebilme imkanı sunmaktadır [1]. İnternetin yaygın olarak kullanılması; sağlık, güvenlik, sanat, eğlence, teknoloji ve ticaret gibi birçok alanda önemli faydalar sağlamaktadır. Ancak, internet üzerinden kişiler ve kurumlar tarafından kişisel, finansal, askeri ve diğer birçok hassas verinin iletilmesi, sağlanan bu faydalarla birlikte güvenlik endişelerini de beraberinde getirmektedir [2]. Bu nedenle, internet üzerinden iletilen verilerin güvenliğinin sağlanması büyük önem taşımaktadır. Verilerin güvenliği sağlanamadığında, siber saldırılara ve veri ihlallerine maruz kalma riski artmaktadır [3].

Şifreleme, hassas verilerin izinsiz erişime karşı korunması amacıyla uygulanan temel güvenlik önlemlerinden biridir. Böylece iletilen verilerin gizliliği sağlanmaya çalışılır. Ancak, ağa sızmış olan saldırganlar da şifreli trafiği kullanarak saldırılar gerçekleştirebilmekte ve izlerini gizleyebilmektedirler. Bu tür saldırılara karşı önlem alabilmek için çeşitli çözümler geliştirilmiş olsa da mevcut ticari güvenlik çözümleri, şifrelenmiş trafiği deşifre etmeden analiz yapamadıkları için şifreli trafiğin sınıflandırılmasında yetersiz kalabilmektedir [4,5].

Makine öğrenmesi algoritmaları, şifreli trafiği deşifre etmeden, başka bir deyişle veri paketinin içeriğini görmeden paket sayısı, uzunluğu, boyutu, iletişim süreleri ve hedef IP adresleri gibi özellikleri kullanarak şifreli trafikteki veri paketlerini kategorize edebilmekte ve şifreli trafiği sınıflandırabilmektedir. Böylece, siber tehditlerin daha doğru ve hızlı bir şekilde tespit edilmesinde ve saldırıların önlenmesinde etkili bir araç haline gelmektedir [6].

Bu çalışmada, ağ trafiği deşifre edilmeden, ağ üzerinden gerçekleşen veri akışına ait paket sayısı, uzunluğu, boyutu, iletişim süreleri ve hedef IP adresleri gibi değişkenler makine öğrenmesi teknikleri ile analiz edilmiş ve sınıflandırılmıştır. Şifreli trafiğin sınıflandırılmasında makine öğrenmesi tekniklerinin etkinliğinin test edilebilmesi amacıyla denetimli makine öğrenmesi algoritmaları ve makine öğrenmesinin bir alt dalı olan topluluk öğrenmesi algoritmaları kullanılmıştır. Çalışmada, yüksek başarı oranına ve hızlı performansa sahip algoritmalar ile bu algoritmalar için en iyi parametrelerin tespit edilmesi hedeflenmiştir. Analizlerde, Kanada Siber Güvenlik Enstitüsü tarafından yayımlanan şifreli trafik verilerini içeren ISCXVPN2016 veri seti [7] kullanılmıştır.

Çalışmanın ana katkıları aşağıdaki gibidir:

- ISCXVPN2016 veri seti kullanılarak yapılan bu çalışmada, daha önce bir arada ele alınmamış Light Gradient-Boosting Machine (LGBM) topluluk öğrenmesi algoritması ile k-Nearest Neighbours (KNN), Support Vector Machines (SVM) ve Logistic Regression (LR) denetimli makine öğrenmesi algoritmaları, hem başarı oranı hem de hız açısından karşılaştırılmıştır.
- Önerilen modelin test edilebilmesi için, kullanılan veri setindeki tüm senaryolara ilişkin veriler analiz edilmiştir. Bu süreçte Synthetic Minority Over-Sampling Technique (SMOTE) yöntemiyle sınıf dengesizlikleri giderilmiş ve hiperparametre optimizasyonu amacıyla rastgele örneklenen hiperparametre kümeleri üzerinde çapraz doğrulama yapılarak en iyi kombinasyon belirlenmiştir.
- Hem doğruluk hem de hız bakımından diğer yöntemlere kıyasla daha yüksek performans gösteren bir LGBM topluluk öğrenmesi modeli önerilmiştir.

Makalenin devamında öncelikle ağ trafiğinin şifrelenmesi ve sınıflandırılmasına ilişkin yöntemler ele alınarak ilgili literatür incelenmiştir. Ardından kullanılan algoritmalar, üzerinde çalışılan veri seti, önerilen iki aşamalı model ve model performans metrikleri açıklanmıştır. Son bölümde elde edilen sonuçlar karşılaştırmalı analizlerle değerlendirilmiş, çalışmanın hem akademik hem de sektörel açıdan pratik katkıları ortaya konmuş ve gelecekteki çalışmalara yönelik öneriler sunularak makale sonuca bağlanmıştır.

2. Ağ Trafiğinin Sınıflandırılması ve İlgili Araştırmalar (Classification of Network Traffic and Related Research)

İnternetin yaygın kullanımı, bilgiye hızlı erişim imkanı sunmanın yanı sıra sağlık, güvenlik, sanat, eğlence, teknoloji ve ticaret gibi pek çok alanda önemli avantajlar sağlamaktadır. Ancak bu avantajlarla birlikte, internet üzerinden kişisel, finansal, askeri ve diğer hassas verilerin iletilmesi, söz konusu verilerin güvenliğine ilişkin endişeleri artırmaktadır.
Bu endişelerin giderilmesi ve olası tehditlerin önlenmesi amacıyla ağ trafiğinin şifrelenmesi gerekliliği doğmuştur. Ağ trafiğinin şifrelenmesi, internet üzerindeki hassas bilgilerin güvenli bir şekilde aktarılmasına olanak tanımaktadır [8].

Hypertext Transfer Protocol Secure (HTTPS) protokolü ve Virtual Private Network (VPN) teknolojisi, güvenli iletişim sağlamak için sıkça kullanılan yöntemler arasındadır [9]. HTTPS protokolünde, kullanıcı ile bağlandığı sunucu arasındaki veri trafiği şifrelenerek kullanıcının internet üzerinden sunulan hizmetlere güvenli biçimde erişmesi sağlanmaktadır [10]. VPN teknolojisinde ise iletişim şifreli bir tünel aracılığıyla gerçekleştirilmektedir. Veriler bu güvenli tünelden iletilirken kapsüllenerek farklı bir paket içerisine alınmakta ve şifrelenmektedir [11]. Böylece kullanıcılar, internet üzerinden güvenli bir bağlantı sağlayabilmektedir.

Şifreleme yöntemleri ve protokolleri, kullanıcıların güvenliğini artırırken, saldırganların ağ üzerindeki kötü niyetli faaliyetlerini gizlemesine de olanak tanıyabilmektedir. Bu durum, kötü niyetli etkinliklerin tespitini zorlaştırmaktadır. Ayrıca, şifreli iletişim kanallarının yaygınlaşması bilgi güvenliği açısından olumlu bir gelişme olsa da, siber güvenlik uzmanlarının trafiği analiz etme ve ağdaki güvenlik tehditlerini tanımlama süreçlerini karmaşıklaştırabilmektedir. Tüm bu etkenler, şifreli trafiğin sınıflandırılmasını ağ güvenliğinin sağlanmasında önemli bir konu haline getirmektedir [12].

Ağ trafiğinin sınıflandırılmasına yönelik olarak literatürde sıkça kullanılan üç temel yöntem bulunmaktadır. Bunlar; port tabanlı sınıflandırma, içerik tabanlı sınıflandırma ve makine öğrenmesi tabanlı sınıflandırma yaklaşımlarıdır [12,13].

2.1. Port tabanlı sınıflandırma (Port-based classification)

Sınıflandırma işlemi için kullanılan en eski yöntem, internet kaynaklarının tahsisi ve düzenlenmesinden sorumlu Internet Assigned Numbers Authority (IANA) tarafından yönetilen "Service Name and Transport Protocol Port Number Registry" veritabanındaki port numaralarını sorgulamaktır. Bu veritabanı, internet hizmetlerinin adları ile bu hizmetlere tahsis edilmiş port numaralarını içermekte ve siber güvenlik uzmanlarına saldırıların tanımlanmasında bir referans kaynağı sunmaktadır [14]. Ancak saldırganlar, port sahtekarlığı, port atlama ve yanıltıcı trafik yönlendirmesi gibi hileli taktikler kullanarak bu yöntemi atlatabilmektedir. Bu nedenle port tabanlı sınıflandırma yöntemi düşük bir güvenilirlik seviyesine sahip olarak değerlendirilmektedir [15].

2.2. İçerik tabanlı sınıflandırma (Content-based classification)

İçerik tabanlı sınıflandırma yönteminde ağ trafiğindeki veri paketlerinin içeriği ayrıntılı bir şekilde incelenerek sınıflandırma gerçekleştirilmektedir [16]. Bu yaklaşım genellikle Deep Packet Inspection (DPI) tekniklerini kullanmakta olup, trafiğin yüksek doğrulukla analiz edilmesini sağlamaktadır. Ancak DPI, imza tabanlı bir yaklaşım benimsediğinden, belirli örüntü veya imzaları tespit etmek amacıyla sürekli güncellenmesi gereken bir imza veritabanına ihtiyaç duymaktadır. Bu veritabanı güncel tutulmadığında, yeni ve değişen tehditlerin tespitinde yetersiz kalınabilmektedir. Bunun yanı sıra, içerik tabanlı sınıflandırmada derinlemesine inceleme yapılması gizlilik endişelerine neden olabilmektedir. En önemli zayıflığı ise şifrelenmiş paket içeriklerini tanımlayamamasıdır [12,17].

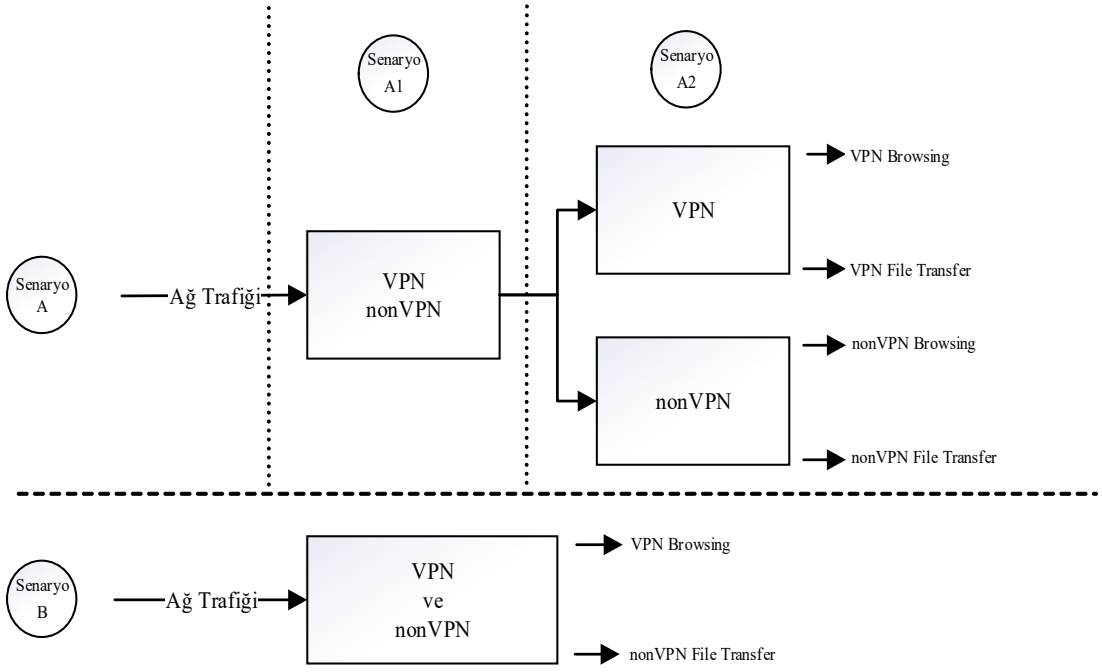
2.3. Makine öğrenmesi tabanlı sınıflandırma (Machine Learning-based classification)

Topluluk öğrenmesi ve derin öğrenme gibi çeşitli öğrenme tekniklerini de içeren bu yöntem ile, trafiğin şifreli olduğu durumlarda, ağ üzerindeki veri iletişiminine ait; paket sayısı, uzunluğu, boyutu, iletişim süresi ve hedef IP adresi gibi şifrelenmemiş bazı veriler kullanılarak sınıflandırma yapılabilmektedir. Bu sayede, siber güvenlik uzmanları ağda iletilen verilere ilişkin önemli bilgilere ulaşabilmektedir [18,19]. Diğer yöntemlere göre daha esnek ve duyarlı olan bu yaklaşım, karmaşık ve değişken ağ trafiği üzerinde etkin sınıflandırma olanağı sunmaktadır. Ayrıca, büyük hacimli veri trafiği üzerinde daha yüksek performans göstermektedir [16,20].

2.4. İlgili araştırmalar (Related Research)

Şifreli ağ trafiğinin sınıflandırılması, ağ güvenliği ve trafik yönetimi açısından kritik bir araştırma alanıdır. Yapılan çalışmada şifreli trafik verilerini içeren ISCXVPN2016 veri seti kullanılmıştır. ISCXVPN2016 veri seti temel alınarak gerçekleştirilen birçok çalışma farklı makine öğrenmesi, topluluk öğrenmesi ve derin öğrenme algoritmaları kullanarak yüksek doğruluk oranlarına ulaşmayı amaçlamaktadır. Draper-Gil vd [7], ISCXVPN2016 veri setini oluşturarak şifreli ağ trafiğinin sınıflandırılması için önemli bir temel oluşturmuşlardır. Bu veri seti, 14 farklı etiketten yararlanan akış tabanlı bir sınıflandırma yaklaşımına dayanmaktadır. Gerçek internet trafik verilerinden elde edilen veri setinde, ağ trafiğinde yaygın olarak görülen Browsing, Email, Chat, Streaming, File Transfer, Voice over Internet Protocol (VoIP) ve Peer-to-Peer (P2P) olmak üzere toplam 7 sınıf bulunmaktadır. Veri setinde, HTTPS protokolüyle şifrelenmiş trafik ve VPN teknolojisiyle şifrelenmiş trafik verileri yer almakta; bu trafik türleri veri seti içerisinde sırasıyla non-VPN ve VPN olarak 2 türde kategorize edilmektedir.

Veri setinde, Şekil 1'de gösterildiği üzere Senaryo A ve Senaryo B olmak üzere iki farklı senaryo bulunmaktadır. Senaryo A'da sınıf türü ve trafik türüne göre kategorizasyon iki adımda, Senaryo B'de ise tek adımda gerçekleştirilmiştir. Her iki senaryoda da veriler, zamanla ilgili özellikler doğrultusunda akış sürelerine göre 15, 30, 60 ve 120 saniyelik veri gruplarından oluşmaktadır.



Şekil 1. Veri setinde yer alan senaryolar
(Scenarios included in the dataset)

Draper-Gil vd. [7], yukarıda ayrıntıları verilen kendi oluşturdukları ISCXVPN2016 veri setini kullanarak,

C4.5 ve KNN algoritmalarıyla çalışmalar gerçekleştirmiş ve senaryoya bağlı olarak VPN ve non-VPN sınıflayıcılarında %80,9 ila %90,6 arasında değişen doğruluk oranlarına ulaşmışlardır. Bu veri seti ve çalışma, literatürde sonraki araştırmalara önemli bir rehber niteliği taşımaktadır.

Bu veri seti kullanılarak şifreli trafiğin makine öğrenmesi yöntemleriyle sınıflandırılmasına ilişkin yapılan çalışmalar incelendiğinde, senaryo bazlı karşılaştırmaların sınırlı sayıda olduğu görülmektedir. Karşılaştırılabilir sonuçlar elde etmek amacıyla, bu çalışmada genel başarı oranları yerine "senaryo" bazlı başarı oranlarını raporlayan çalışmalar [2,5,7,21,22] temel alınarak değerlendirme yapılmıştır.

Ayrıca, ilgili veri setinin yayımlanmasından bu yana ISCXVPN2016 veri seti kullanılarak şifreli trafiğin sınıflandırılmasına ilişkin yürütülen çalışmalar [2,5,7,8,15,19,21-34] incelenmiş ve bu alanda gerçekleştirilen az sayıda çalışmada [35], günümüzde hemen her alanda önemli bir kriter olan zaman unsuruna ilişkin test süresi metriğinin [36] ele alındığı görülmüştür. Test süresi, eğitilen algoritmanın test aşamasında harcadığı süreyi ifade etmektedir. Bu çalışmada kullanılan algoritmaların test süreleri de hesaplanarak karşılaştırılmıştır.

Aşağıda, kullanılan yöntemlere göre literatürde öne çıkan çalışmalar ve bulgular ele alınmıştır.

2.4.1. Makine öğrenmesi yöntemleri ile sınıflandırma (Classification with machine learning methods)

Birçok çalışma, ISCXVPN2016 veri setini kullanarak makine öğrenmesi algoritmalarının etkinliğini analiz etmiştir. Yamansavaşçılar vd. [23], farklı algoritmalar arasında en iyi sonucu %93,94 doğruluk oranıyla KNN algoritması ile elde etmişlerdir. Caicedo-Muñoz vd. [5], trafik sınıflandırması için özel bir Quality of Service (QoS) sınıflayıcı önermiş ve bu doğrultuda ISCXVPN2016 veri setinden QoS etiketli bir veri kümesi oluşturmuşlardır. Çalışmanın sonucunda, Bagging algoritması ile A2 senaryosunda %94,42, B senaryosunda ise %86,94 doğruluk oranları elde etmişlerdir. Zhang vd. [27], dengesiz şifreli trafiğin sınıflandırılması için Random Forest (RF) algoritmasını kullanarak bir şema önermişler ve bu yöntemle %93 doğruluk oranına ulaşmışlardır. Bagui vd. [24], RF ve Gradient Boosting Tree (GBT) algoritmaları ile %90 doğruluk oranı elde etmişlerdir. Zhou vd. [30] ise RF algoritması kullanarak %98 doğruluk oranına ulaşmıştır.

Literatürdeki bu çalışmalar, KNN, Bagging ve RF algoritmalarının ön planda olduğunu ve başarılı sonuçlar verdiğini göstermektedir. KNN algoritması %93,94, Bagging algoritması %94,42 ve RF algoritması %98 doğruluk oranına ulaşmıştır. Ayrıca GBT algoritması da etkili sonuçlar elde edilen bir diğer yöntem olarak öne çıkmaktadır.

2.4.2. Derin öğrenme yöntemleri ile sınıftandırma (Classification with deep learning methods)

Derin öğrenme yöntemleri, şifreli trafiğin sınıflandırılmasında kayda değer başarılar elde etmiştir. Örneğin, Guo vd. [28], Convolutional Auto Encoder (CAE) modelini kullanarak %99,87 doğruluk oranına ulaşmış; Ismailaj vd. [33] ise Convolutional Neural Network (CNN) tabanlı yaklaşımlarıyla %90 doğruluk oranı elde etmiştir. Obasi [19], Artificial Neural Network (ANN), CNN ve Long Short-Term Memory (LSTM) gibi derin öğrenme modellerini Decision Tree (DT) ve RF algoritmalarıyla karşılaştırmış ve RF ile %96 doğruluk oranına varmıştır. Bu sonuç, Bagui vd. [24] ile Zhou vd. [30] tarafından bildirilen bulgularla da örtüşerek, RF algoritmasının sınıflandırmadaki üstün başarısını ön plana çıkarmaktadır. Böylece RF, ANN, CNN ve LSTM gibi derin öğrenme modelleriyle kıyaslandığında, %96 doğruluk oranı ile yüksek bir performans sergilemektedir.

2.4.3. Hibrit ve topluluk öğrenmesi yöntemleri ile sınıflandırma (Classification with hybrid and ensemble learning methods)

Hibrit modeller ve topluluk öğrenmesi yöntemleri, birden fazla algoritmayı bir araya getirerek performansı artırmayı amaçlamaktadır. Huang vd. [8], önerdikleri hibrit bir yaklaşım ile %80 F1 skoru elde etmişlerdir. Bozkır [22] ise Extreme Gradient Boosting (XGBoost) algoritması ile %99 F1 skoru yakalayarak en yüksek sonuçlardan birini elde etmiştir. Uğurlu vd. [2], XGBoost algoritması kullanarak %93,04 doğruluk oranına ulaşmışlardır. Almomani [34], ANN, RF ve SVM algoritmalarını bir araya getirerek oluşturduğu topluluk öğrenmesi modeli ile %98 doğruluk oranı elde etmiştir.

Afuwape vd. [32], şifreli trafiği sınıflandırmak için AdaBoost, GBT, RF, Bagging Decision Tree (BDT), KNN, LR, Multi Layer Perceptron (MLP) ve Naive Bayes (NB) tekniklerini ISCXVPN2016 veri seti üzerinde kullanmışlardır. Bu çalışmada, topluluk algoritmalarının tekli algoritmalardan daha iyi performans gösterdiği sonucuna varılmış ve RF ile %93,8 doğruluk oranına ulaşılmıştır. Genel olarak, hibrit modeller daha yüksek doğruluk ve F1 skorları sağlarken, topluluk öğrenmesi yöntemleri birden fazla algoritmanın birleşimi sayesinde performansı önemli ölçüde artırmaktadır. Ayrıca, farklı tekniklerin uygulandığı çalışmalarda da topluluk algoritmalarının tekli algoritmalardan daha iyi sonuçlar verdiği gözlemlenmiştir.

2.4.4. Alternatif veri setleri ve karşılaştırmalar (Alternative datasets and comparisons)

ISCXVPN2016 verisetinin yanı sıra, bazı araştırmacılar kendi veri setlerini de kullanmışlardır. Khatouni ve Heywood [35], Gradient Descent, DT, NB, SVM ve RF algoritmalarını kendi veri setlerinde kullanarak %85 doğruluk oranına ulaşmışlardır. Ayrıca RF algoritmasının doğruluk açısından üstün olmasına rağmen, DT algoritmasının işlem hızı konusundaki avantajlarına dikkat çekmişlerdir. Majeed vd. [21], şifreli trafiğin sınıflandırılması için özellik tabanlı Horizontal Federated Learning yöntemini önermiş ve bu yöntemle %86 doğruluk oranı elde etmişlerdir.

Bunların yanı sıra, karanlık ağ (Dark Web) trafiğinin sınıflandırılması için makine öğrenmesi tabanlı yöntem kullanılarak yapılan çalışmalara da [37-40] rastlanmıştır.

Sonuç olarak literatürde yer alan araştırmalar KNN, Bagging, RF gibi makine öğrenmesi algoritmalarının yanı sıra, CNN ve CAE gibi derin öğrenme modellerinin yüksek doğruluk oranları sunduğunu göstermiştir. Hibrit ve topluluk modelleri, bu yöntemleri birleştirerek daha yüksek performans ve doğruluk sağlamıştır. Çeşitli veri setleri ve alternatif yöntemler kullanılarak yapılan çalışmalar, özellikle RF algoritmasının sınıflamada üstün başarı gösterdiğini ortaya koymuştur. Literatür taraması kapsamında değerlendirilen tüm önceki çalışmalar Tablo 1'de ayrıca özetlenmiştir.

Tablo 1. Literatür özeti (Literature Review)

Yıl	Kaynak	Veri Seti	Algoritma Türü	Algoritma Adı	Sonuç
2016	Draper-gil vd. [7]	ISCXVPN2016 (Veri setini kendileri hazırlamıştır.)	Makine öğrenmesi	C4.5 ve KNN	C4.5 ile Senaryo A1 %90,6 kesinlik Senaryo A2 %89 kesinlik Senaryo B %80,9 kesinlik
2017	Yamansavaşçılar vd. [25]	ISCXVPN2016	Makine öğrenmesi	J-48, RF, KNN ve Bayes Net	KNN ile %93,94 doğruluk
2017	Bagui vd. [24]	ISCXVPN2016	Makine öğrenmesi, Topluluk öğrenmesi	NB, SVM, KNN, LR, RF ve GBT	RF ve GBT ile %90 doğruluk
2018	Caicedo vd. [5]	ISCXVPN2016	Makine öğrenmesi, Topluluk öğrenmesi	C4.5, Bagging ve Boosting	Bagging ile Senaryo A2 %94,42 doğruluk Senaryo B %86,94 doğruluk
2020	Zhang vd. [27]	ISCXVPN2016	Makine öğrenmesi, Topluluk öğrenmesi	C4.5, KNN ve RF	RF ile 0,93 doğruluk
2020	Guo vd. [28]	ISCXVPN2016	Derin öğrenme, Makine öğrenmesi	CAE, CNN, KNN ve C4.5	CAE ile %99,87 genel doğruluk
2020	Obasi [19]	ISCXVPN2016 ve Solana Networks	Derin öğrenme, Makine öğrenmesi, Topluluk öğrenmesi	ANN, CNN, LSTM, CapsNet, DT ve RF	RF ile %96 doğruluk
2020	Zhou vd. [30]	ISCXVPN2016 ve ISCX-Tor/Non-Tor	Derin öğrenme, Makine öğrenmesi, Topluluk öğrenmesi	ANN, SVM, NB, LR ve RF	RF ile %98 doğruluk
2020	Majeed vd. [21]	ISCXVPN2016	Makine öğrenmesi	Horizontal Federated Learning yöntemi	Senaryo A %86, Senaryo B %81 doğruluk
2021	Afuwape vd. [32]	ISCXVPN2016	Makine öğrenmesi, Derin öğrenme, Topluluk öğrenmesi	KNN, LR, NB, MLP, AdaBoost, GBT, RF ve BDT	RF ile %93,80 doğruluk
2021	Uğurlu vd. [2]	ISCXVPN2016	Makine öğrenmesi, Topluluk öğrenmesi	DT, XGBoost ve RF	XGBoost ile Senaryo A1 %93,04 kesinlik Senaryo A2 %94,53 doğruluk Senaryo B %86,06 kesinlik
2021	Huang vd. [8]	ISCXVPN2016	Derin öğrenme, Makine öğrenmesi	Kendi önerdikleri hibrit yaklaşım	Önerilen yöntem ile %80 F1 Skoru
2021	Ismailaj vd. [33]	ISCXVPN2016	Derin öğrenme, Topluluk öğrenmesi	CNN ve LGBM	CNN ile %90 doğruluk
2022	Almomani [34]	ISCXVPN2016	Derin öğrenme, Makine öğrenmesi	ANN, SVM ve kendi önerdikleri yöntem	Önerilen yöntem ile %98 doğruluk
2023	Bozkır vd. [22]	ISCXVPN2016	Topluluk öğrenmesi	GBT, LGBM ve XGBoost	XGBoost ile Senaryo A ve Senaryo B %99 F1

3. Materyal ve Yöntem (Materials and Methods)

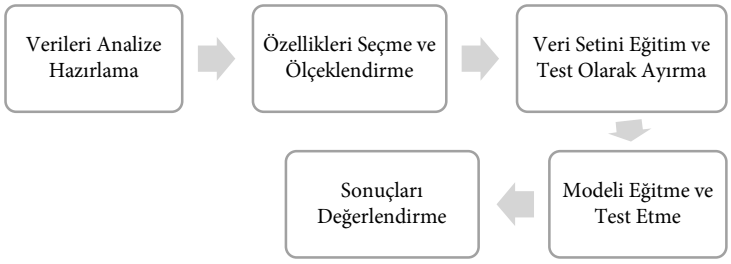
Yapılan çalışmada, Drapper-Gil vd. tarafından oluşturulan ve Kanada Siber Güvenlik Enstitüsü tarafından yayımlanan kapsamlı bir etiketli veri seti olan ISCXVPN2016 kullanılmıştır [7]. Bu veri seti için KNN, SVM ve LR denetimli makine öğrenmesi algoritmalarının yanı sıra LGBM topluluk öğrenmesi algoritması ile çalışılmıştır.

Çalışmada veriler öncelikle, eksik veri kontrolü ve sayısallaştırma yapılarak analize hazır hale getirilmiştir. Sonrasında önem derecesine göre özellik seçimi yapılmış ve her bir özellik için ölçeklendirme işlemi gerçekleştirilmiştir. Ölçeklendirme sonrası veriler, %70'i eğitim ve %30'u test verisi olacak şekilde rastgele bir şekilde bölünerek ayrılmıştır.

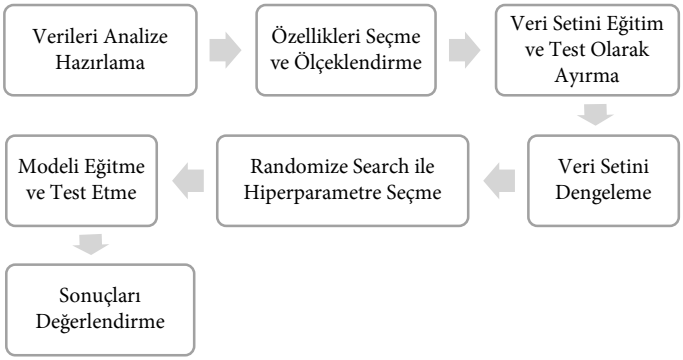
Çalışmanın birinci aşaması olan orijinal veri seti ile eğitim ve test, yukarıda yapılan işlemler sonrasında gerçekleştirilmiştir. İkinci aşamada ise dengesiz dağılıma sahip veri setleri SMOTE tekniği ile dengelenmiş, hiperparametrelerin rastgele örnekleri seçilerek çapraz doğrulama uygulanmış ve böylece en iyi hiperparametre kombinasyonu belirlenmiştir. Ardından bu optimize edilmiş hiperparametreler kullanılarak tekrar eğitim ve test adımları gerçekleştirilmiştir.

Çalışma kapsamında makine öğrenmesi algoritmaları ile şifreli trafiğin sınıflandırılmasında izlenen bu iki

aşamanın adımları sırasıyla Şekil 2 ve Şekil 3'te gösterilmektedir.



Şekil 2. Birinci aşama araştırma metodolojisinde izlenen adımlar
(The steps followed in the research methodology of the first stage)



Şekil 3. İkinci aşama araştırma metodolojisinde izlenen adımlar
(The steps followed in the research methodology of the second stage)

Çalışmada kullanılan ISCXVPN2016 veri seti etiketli verilerden oluşmaktadır [7]. Veri setinin analize uygun hale getirilmesi için eksik veriler kontrol edilmiş ve eksik veri bulunmadığı tespit edilmiştir. Ardından, veri setinde kategorik verilerin varlığı incelenmiş ve sınıf etiketlerinin kategorik olduğu belirlenmiştir. Bu kategorik sınıf etiketleri sayısal değerlere dönüştürülmüştür (Tablo 2).

Tablo 2. Sınıf etiketlerine atanan sayısal değerler (Numerical values assigned to class labels)

Trafik Sınıfı	Atanan Değer
Browsing	0
Chat	1
File Transfer	2
Email	3
P2P	4
Streaming	5
VoIP	6
VPN-Browsing	7
VPN-Chat	8
VPN-File Transfer	9
VPN-Email	10
VPN-P2P	11
VPN-Streaming	12
VPN-VoIP	13

Makine öğrenmesi modellerinin başarı oranlarını etkilemeden çalışma performansını artırmak, hesaplama maliyetini düşürmek ve gereksiz özelliklerin etkisini azaltmak amacıyla özellik seçimi yapılmaktadır [41]. Bu çalışmada, Decision Tree Classifier kullanılarak modelde yer alan özelliklerin önem dereceleri hesaplanmıştır. Bu önem dereceleri, toplamı 1 olacak şekilde ağırlıklandırma yapılmıştır. Tüm sınıfları içeren Senaryo B-15s veri setinde, ağırlığı 0,01'in üzerinde ağırlığa sahip Tablo 3'de yer alan ilk 15 özellik seçilerek kullanılmıştır.

Tablo 3. Senaryo B 15s'ye ait özelliklerin ağırlık değerleri (The weight values of the features for Scenario B 15s)

No	Özellik Adı	Ağırlığı
1	total_biat	0,143661
2	flowBytesPerSecond	0,117527
3	duration	0,114499
4	min_flowiat	0,087023
5	total_fiat	0,065198
6	std_flowiat	0,060411
7	max_fiat	0,059739
8	max_flowiat	0,056023
9	mean_biat	0,046961
10	min_biat	0,045398
11	max_biat	0,045074
12	mean_flowiat	0,040779
13	flowPktsPerSecond	0,031403
14	min_fiat	0,020947
15	mean_fiat	0,018952
16	std_idle	0,008859
17	mean_idle	0,007039
18	max_active	0,006434
19	min_idle	0,006154
20	std_active	0,005152
21	mean_active	0,004819
22	min_active	0,004677
23	max_idle	0,003259

Sonrasında veri seti analiz edilmiş ve değerlerin farklı ölçeklere sahip olduğu görülmüştür. Örnek olarak aşağıda bulunan Tablo 4'te Senaryo B 15s'ye ait betimsel istatistiki veriler paylaşılmıştır. Tablodaki veriler incelendiğinde, değişkenlerin aritmetik ortalamaları ile minimum ve maksimum değerleri arasında ciddi farklılıklar olduğu gözlemlenmiştir. Örneğin, bazı değişkenlerde maksimum değerlerin milyonlarla ifade edilmesi, diğer değişkenlerde ise oldukça düşük değerlere sahip olması, verilerde ölçek farklılıklarını ortaya koymaktadır. Bu tür dengesizlikler, özellikle makine öğrenimi modellerinde kullanılan metriklerin dengesiz çalışmasına ve model performansında sapmalara yol açabilmektedir.

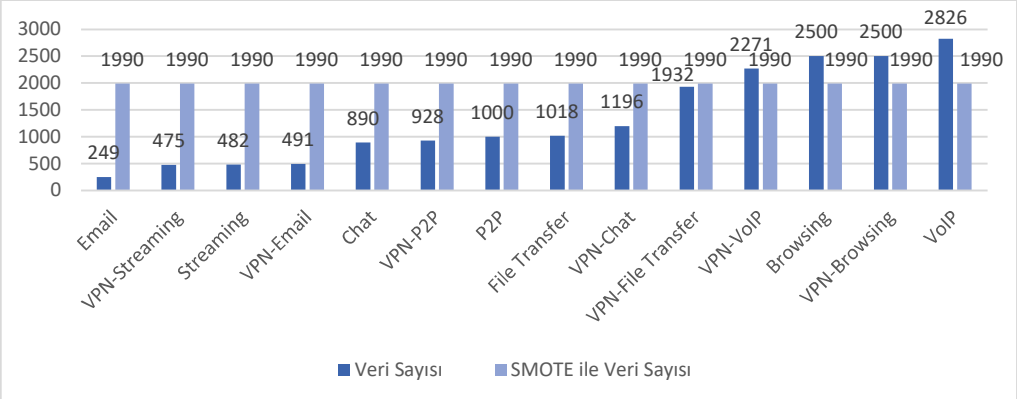
Tabloda görülen özelliklerden flowBytesPerSecond isimli olan byte/saniye, flowPktsPerSecond isimli olan paket/saniye, kalan diğer tüm veriler ise saniye cinsindendir. Hem özellikler arasındaki dengesizlikler hem de farklı cinslerde özelliklerin bir arada kullanımı verilerin normalleştirilmesi ihtiyacını ortaya çıkarmaktadır. Bu durumu dengelemek ve verileri aynı ölçek seviyesine getirmek amacıyla min-max normalizasyonu ile ölçeklendirme işlemi yapılmıştır [40].

Tablo 4. Senaryo B 15s'ye ait betimsel istatistiki veriler (Descriptive statistical data for Scenario B 15s)

	Veri Sayısı	Aritmetik Ortalama	Standart Sapma	Minimum Değer	Maksimum Değer
total_biat (s)	13.130	638.875	2.362.490	1	43.002.380
flowBytesPerSecond (byte/saniye)	13.130	454.718	8.982.233	0	565.000.000
Duration (s)	13.130	9.788.733	14.807.250	0	601.405.000
min_flowiat (s)	13.130	43.427	466.325	721	18.919.900
total_fiat (s)	13.130	619.785	2.274.082	1	37.680.790
std_flowiat (s)	13.130	1.032.781	3.595.225	0	136.000.000
max_fiat (s)	13.130	1.016.790	3.328.565	0	152.000.000
max_flowiat (s)	13.130	3.852.325	14.311.790	1	601.109.700
mean_biat (s)	13.130	604.968	2.390.375	0	98.000.000
min_biat (s)	13.130	2.871.231	10.275.070	1	600.109.700
max_biat (s)	13.130	928.233	2.504.501	0	43.000.000
mean_flowiat (s)	13.130	480.361	1.475.132	0	60.700.000
flowPktsPerSecond (paket/saniye)	13.130	1.909	15.646	0	666.666
min_fiat (s)	13.130	3.306.506	10.627.500	1	303.595.700
mean_fiat (s)	13.130	8.410.637	4.509.882	0	215.000.000

$$z = \frac{x - min\,(x)}{max(x) - min\,(x)} \tag{1}$$

Min-max normalizasyon ölçeklendirmesi, yukarıda yer alan formüldeki (1) gibi, bir veri setindeki bir değerin minimum değere olan uzaklığının, maksimum ve minimum değerler arasındaki farka bölünmesiyle gerçekleştirilir. Bu yöntem, veri setindeki değerleri 0 ile 1 aralığına indirger. Böylece, değerler arasındaki büyük farklılıklar ortadan kaldırılarak modelin daha doğru ve dengeli sonuçlar üretmesi sağlanmaktadır. Ölçeklendirme sonrası veriler, %70'i eğitim ve %30'u test verisi olacak şekilde rastgele bölünerek ayrılmış ve çalışmanın birinci aşaması olan orijinal veri seti ile eğitim ve test gerçekleştirilmiştir. İkinci aşamada, sınıf dengesizliği bulunan A2 ve B senaryolarındaki veri setleri için dengeleme işlemi uygulanmıştır. Bu işlemde, yeniden örnekleme yöntemlerinden SMOTE kullanılmıştır. SMOTE, az sayıdaki sınıflar için sentetik örnekler oluşturarak veri setindeki dengesizliği gidermeyi amaçlamaktadır. Tüm sınıfları içeren Senaryo B-15s veri setinde, dengeleme öncesi ve sonrası veri dağılımı Şekil 4'te gösterilmiştir.



Şekil 4. Senaryo B-15s için SMOTE ile verilerin dengelenmesi
(Balancing the data for Scenario B-15s using SMOTE)

Hiperparametreler, algoritmaların performansını doğrudan etkilemektedir [42]. Çalışmanın ikinci aşamasında, algoritmalar için en iyi parametre kombinasyonunu belirlemek amacıyla rastgele arama tekniği kullanılmıştır. Bu teknikte, belirlenen bir aralıktaki hiperparametre değerleri rastgele denenerek optimize edilir. Bu yaklaşım, hesaplama süresi açısından daha az zaman gerektirdiği için avantaj sağlamaktadır. Ayrıca, tüm parametrelere eşit kaynak ayırmak yerine, önemli parametrelere odaklandığından, verimli ve pratik bir

yöntemdir [43].

Kullanılan makine öğrenmesi algoritmalarının parametreleri, belirlenen aralıklar içinde rastgele arama yöntemine tabi tutulmuş ve sınıflandırma işlemi için en iyi sonuçları veren parametreler belirlenmiştir. Parametre seçimi için kullanılan aralıklar Tablo 5'te gösterilmiştir.

Tablo 5. Kullanılan hiperparametre aralıkları (The hyperparameter ranges used)

Algoritma Adı	Hiperparametre Aralıkları
KNN	'n_neighbors': [3, 5, 7, 9], 'weights': ['uniform', 'distance'], 'metric': ['euclidean', 'manhattan', 'minkowski'].
SVM	'C': [0.1, 1, 10, 100], 'gamma': ['scale', 0.1, 1, 10], 'kernel': ['linear', 'rbf', 'sigmoid'].
LR	'C': [0.0001, 0.001, 0.01, 0.1, 1, 10, 100, 1000, 10000], 'penalty': ['none', 'l1', 'l2'], 'solver': ['newton-cg', 'lbfgs', 'liblinear', 'sag', 'saga'], 'max_iter': [30, 50, 100, 200, 500, 1000, 2000, 5000].
LGBM	'num_leaves': sp_randint(6, 100), 'learning_rate': [0.01, 0.05, 0.1, 0.3, 0.5, 0.7], 'max_depth': sp_randint(3, 30), 'min_child_samples': sp_randint(100, 1000), 'subsample': [0.6, 0.7, 0.8, 0.9, 1.0], 'colsample_bytree': [0.5, 0.6, 0.7, 0.8, 0.9, 1.0], 'reg_alpha': [0, 0.1, 0.5, 1, 2, 5, 10], 'reg_lambda': [0, 0.1, 0.5, 1, 2, 5, 10], 'min_child_weight': [1e-3, 1e-2, 0.1, 0.5, 1, 5, 10], 'n_estimators': sp_randint(50, 1000).

Rastgele arama işlemi ile birlikte hiperparametre seçiminde çapraz doğrulama da yapılmıştır. Çapraz doğrulama işleminde veri setinin her bir parçası hem eğitim hem de test verisi olarak kullanılmaktadır. Böylece veri, daha verimli değerlendirilmekte ve daha güvenilir tahmin yapılmaktadır. Çapraz doğrulama veri setini k sayıda alt kümeye bölmekte ve her alt küme için modeli eğitmektedir. Bu işlem k kez tekrarlanmakta ve sonuçların ortalamaları alınarak performans değeri elde edilmektedir [44]. Bu çalışmada tüm algoritmalar için çapraz doğrulama işleminde kullanılan katman sayısı 3 olarak belirlenmiştir.

Çalışmada performans metriklerinden doğruluk, F1 skor ve test süresi dikkate alınmıştır. Performans metriklerinin hesaplanmasında, makine öğrenmesi modelinin doğru ve yanlış yaptığı sınıflandırma sayılarına ihtiyaç duyulmaktadır. Bu kapsamda gerçek ve tahmin değerlerini içeren karmaşıklık matrisi kullanılmaktadır. Karmaşıklık matrisi yapısı Tablo 6'da yer almaktadır.

Tablo 6. Karmaşıklık matrisi yapısı (The structure of the confusion matrix)

Karmaşıklık Matrisi		**Tahmin**	
		X_1 (Pozitif)	X_2 (Negatif)
Gerçek	X_1 (Pozitif)	DP	YN
	X_2 (Negatif)	YP	DN

Gerçekte pozitif olup pozitif sınıflandırılan Doğru Pozitif (DP), gerçekte negatif olan ama pozitif sınıflandırılan Yanlış Pozitif (YP), gerçekte negatif olup negatif sınıflandırılan Doğru Negatif (DN), gerçekte pozitif olan ama negatif sınıflandırılan ise Yanlış Negatif (YN) örnek sayısını vermektedir [45].

Doğruluk, doğru sınıflandırılan tahminlerin oranını bulmak için kullanılmakta ve aşağıda yer alan formülle (2) hesaplanmaktadır.

$$Doğruluk = \frac{DP + DN}{DP + DN + YP + YN} \quad (2)$$

Kesinlik, pozitif olarak tahmini yapılan örneklemlerin gerçekte ne kadarının pozitif olduğunu ölçmek için kullanılmaktadır. Aşağıda yer alan formülle (3) hesaplanmaktadır.

$$Kesinlik = \frac{DP}{DP + YP} \quad (3)$$

Duyarlılık, gerçek pozitif örneklerin ne kadarının doğru tahmin edildiğini ölçmek için kullanılmaktadır. Aşağıda yer alan formülle (4) hesaplanmaktadır.

$$Duyarlılık = \frac{DP}{DP + YN} \quad (4)$$

F1 skor (F1-score), hassasiyet ve duyarlılığın harmonik ortalamasını alarak modelin ne kadar dengeli bir ölçüm sağladığını ölçmek için kullanılmakta ve aşağıdaki formülle (5) hesaplanmaktadır.

$$F1\ skor = \frac{2 * Kesinlik * Duyarlılık}{Kesinlik + Duyarlılık} \quad (5)$$

Test süresi, bir sınıflandırma sürecinde algoritmanın sınıflandırma işlemini ne kadar sürede gerçekleştirdiğini ifade etmektedir. Bu süre bir sınıflandırma algoritmasının ne kadar hızlı veya ne kadar yavaş çalıştığının belirlenmesine yardımcı olmaktadır. Algoritmaların hızı, sınıflandırma işleminde yapılan tespitin daha hızlı olabilmesini sağlayacağından daha etkin bir sınıflandırma ve kaynakların daha verimli kullanılması açısından fayda sağlayacaktır [36].

4. Bulgular ve Yorumlar (Findings and Interpretations)

Senaryo A1'de yer alan 4 adet, Senaryo A2'de yer alan 8 adet ve Senaryo B'de yer alan 4 adet zaman bazlı veri setlerinin her biri için ayrı ayrı denetimli makine öğrenmesi ve topluluk öğrenmesi algoritmaları kullanılarak çalışma yapılmıştır.

4.1. Birinci aşamada elde edilen sonuçlar (The results obtained in the first stage)

Çalışmanın bu ilk aşamasında, veri dengeleme işlemi ve hiperparametre seçimi yapılmamıştır. LGBM, KNN, SVM ve LR algoritmaları kullanılarak orijinal veri seti ile elde edilen sonuçlar Tablo 7'de gösterilmiştir. Tablo 8'de ise algoritmalar için test süreleri raporlanmıştır.

Tablo 7. Orijinal veri seti ile başarı oranları (The success rates with the original dataset)

Algoritma	LGBM		KNN		SVM		LR	
Senaryolar	Doğruluk	F1 Skor	Doğruluk	F1 Skor	Doğruluk	F1 Skor	Doğruluk	F1 Skor
Senaryo A1 15s	**0,9087***	**0,9085***	**0,8211***	**0,8212***	0,6075	0,5924	**0,5721***	**0,5678***
Senaryo A1 30s	0,8924	0,8922	0,8144	0,8140	0,6078	0,5859	0,5612	0,5389
Senaryo A1 60s	0,8901	0,8912	0,7889	0,7891	0,6123	0,6108	0,5603	0,5231
Senaryo A1 120s	0,8927	0,8926	0,7935	0,7930	**0,6173***	**0,6125***	0,5518	0,5156
Senaryo A2 No-VPN 15s	**0,9468****	**0,9461****	0,8539	0,8521	**0,6874****	**0,6310****	0,5825	0,4790
Senaryo A2 VPN 15s	0,8999	0,8988	0,8043	0,8028	0,5752	0,5106	0,4847	0,3503
Senaryo A2 No-VPN 30s	0,9383	0,9373	**0,8613****	**0,8578****	0,6460	0,5791	0,5742	0,4801
Senaryo A2 VPN 30s	0,9130	0,9121	0,8311	0,8290	0,5890	0,5277	0,5050	0,3862
Senaryo A2 No-VPN 60s	0,9169	0,9166	0,8050	0,8013	0,5812	0,5262	0,5190	0,4460
Senaryo A2 VPN 60s	0,8813	0,8785	0,7967	0,7915	0,5320	0,4448	0,4786	0,3827
Senaryo A2 No-VPN 120s	0,9353	0,9341	0,8551	0,8480	0,6578	0,5767	**0,6255****	**0,5380****
Senaryo A2 VPN 120s	0,8728	0,8684	0,7834	0,7791	0,5811	0,4841	0,5053	0,3644
Senaryo B 15s	**0,8658***	**0,8657***	0,7166	0,7147	**0,4394***	**0,3987***	0,2601	0,1882
Senaryo B 30s	0,8576	0,8562	**0,7232***	**0,7286***	0,3938	0,3595	0,3219	0,2420
Senaryo B 60s	0,8335	0,8234	0,7187	0,7245	0,3767	0,3443	0,3289	0,2619
Senaryo B 120s	0,8250	0,8222	0,6754	0,6681	0,4049	0,3655	**0,3354***	**0,2634***

Orijinal veri seti ile

* Kalın yazılan değerler, ilgili senaryo ve algoritmada elde edilen en yüksek başarı oranlarıdır.

**Kalın ve kırmızı yazılan değerler, ilgili algoritmada tüm senaryolar içinde elde edilen en yüksek başarı oranlarıdır.

Tablo 8. Orijinal veri seti ile test süreleri (The test durations with the original dataset)

Algoritma	Test Edilen Girdi Sayısı (adet)	LGBM		KNN		SVM		LR	
Senaryolar		Toplam Test Süresi (s)	Birim Test Süresi (s/ad)	Toplam Test Süresi (s)	Birim Test Süresi (s/ad)	Toplam Test Süresi (s)	Birim Test Süresi (s/ad)	Toplam Test Süresi (s)	Birim Test Süresi (s/ad)
Senaryo A1 15s	5.628	0,01	0,000002	0,34	0,000060	2,73	0,000485	0,04	0,000007
Senaryo A1 30s	4.397	0,01	0,000002	0,43	0,000098	2,09	0,000475	0,01	0,000002
Senaryo A1 60s	4.655	0,03	0,000006	0,56	0,000120	1,80	0,000387	0,01	0,000002
Senaryo A1 120s	3.236	0,13	0,000040	0,61	0,000189	1,13	0,000349	0,01	0,000003
Senaryo A2 No-VPN 15s	2.690	**0,04**	**0,000015**	0,25	0,000093	0,63	0,000234	0,01	0,000004
Senaryo A2 VPN 15s	2.938	0,05	0,000017	0,16	0,000054	0,96	0,000327	0,01	0,000003
Senaryo A2 No-VPN 30s	2.076	0,05	0,000024	**0,09**	**0,000043**	0,41	0,000197	0,01	0,000005
Senaryo A2 VPN 30s	2.321	0,04	0,000017	0,13	0,000056	0,59	0,000254	0,01	0,000004
Senaryo A2 No-VPN 60s	2.574	0,04	0,000016	0,09	0,000035	0,69	0,000268	0,01	0,000004
Senaryo A2 VPN 60s	2.081	0,03	0,000014	0,22	0,000106	0,46	0,000221	0,01	0,000005
Senaryo A2 No-VPN 120s	1.546	0,02	0,000013	0,07	0,000045	0,18	0,000116	0,01	0,000006
Senaryo A2 VPN 120s	1.690	0,12	0,000071	0,39	0,000231	0,33	0,000195	0,01	0,000006
Senaryo B 15s	5.704	0,17	0,000030	0,72	0,000126	6,31	0,001106	0,01	0,000002
Senaryo B 30s	4.543	0,14	0,000031	0,49	0,000108	2,97	0,000654	0,01	0,000002
Senaryo B 60s	4.825	0,12	0,000025	0,45	0,000093	2,03	0,000421	0,01	0,000002
Senaryo B 120s	3.501	0,12	0,000034	0,39	0,000111	1,62	0,000463	0,01	0,000003
Ortalama Test Süreleri		0,07	0,000022	0,33	0,000098	1,56	0,000385	0,01	0,000004

Orijinal veri seti ile

* Kalın yazılan değerler, en yüksek başarı oranı elde edilen senaryo ve algoritmalardaki test süreleridir.

Tüm senaryolar göz önünde bulundurulduğunda, orijinal veri setiyle en yüksek başarı oranlarına denetimli makine öğrenmesi algoritmaları arasında KNN ile ulaşılmıştır. KNN algoritması ile elde edilen en yüksek başarı oranları, 0,8613 doğruluk ve 0,8578 F1 Skoru ile Senaryo A2 No-VPN 30s veri setinde kaydedilmiştir. Test süresi 2.076 girdi için 0,09 saniye olmuş, bu da girdi başına 0,000043 saniyede kategorizasyon yapılabildiği anlamına gelmektedir. Topluluk öğrenme algoritması LGBM'de ise ulaşılan en yüksek başarı oranları, 0,9468 doğruluk ve 0,9461 F1 Skoru ile Senaryo A2 No-VPN 15s veri setinde elde edilmiştir. Test süresi 2.690 girdi için 0,04 saniye olarak hesaplanmış ve bu, her bir girdinin 0,000015 saniyede kategorize edildiğini göstermektedir. Sonuç olarak, LGBM ile elde edilen başarı oranlarının, hem doğruluk ve F1 Skoru açısından hem de kategorizasyonun gerçekleştirilme süresinin kısalığı bakımından KNN'den daha yüksek performans gösterdiği görülmektedir.

Çalışmanın ilk aşamasında algoritmaların varsayılan parametre değerleri kullanılmıştır. KNN ve LGBM algoritmalarında kullanılan varsayılan parametre değerleri Tablo 9'da gösterilmiştir.

Tablo 9. KNN ve LGBM algoritmaları için varsayılan parametre değerleri
(The default parameter values for the KNN and LGBM algorithms)

Algoritma	Parametrenin Adı	Varsayılan Değer
KNN	n_neighbors	5
	weights	uniform
	metric	minkowski
LGBM	colsample_bytree	1.0
	learning_rate	0.1
	max_depth	-1
	min_child_samples	20
	min_child_weight	0.001
	n_estimators	100
	num_leaves	31
	reg_alpha	0.0
	reg_lambda	0.0
	subsample	1.0

Tablolar incelendiğinde hem test süreleri hem de performans metrikleri açısından öne çıkan modelin LGBM olduğu görülmektedir. LGBM, çoğu senaryoda diğer algoritmalardan daha yüksek doğruluk ve F1 skor değerlerine ulaşırken, test süresi ve birim test süresi bakımından da son derece verimli görünmektedir. KNN, performans açısından LGBM'nin gerisinde kalsa da özellikle A2 senaryolarında makul doğruluk ve F1 skor değerlerine ulaşabilmekte, ancak birim test süresi LGBM'ye kıyasla daha uzun kalmaktadır. SVM, bazı durumlarda orta düzeyde performans sunarken, test sürelerinin göreceli olarak daha yüksek olduğu ve birim test süresinin KNN ve LR'a kıyasla daha maliyetli olduğu dikkat çekmektedir. LR ise genellikle en düşük doğruluk ve F1 skoruna sahip olmakla beraber, test süresi ve birim test süresi açısından en hızlı sonuç veren yöntemdir. Sonuç olarak, yüksek performans ile düşük test süresini dengelemek isteyenler için LGBM öne çıkarken, çok hızlı ancak düşük doğruluk isteyen durumlar için LR, daha ılımlı bir dengenin gerektiği senaryolar için ise KNN veya SVM alternatif olarak değerlendirilebilir. Ayrıca, No-VPN koşullarının performansa genellikle olumlu etki yaptığı görülmektedir.

4.2. İkinci aşamada elde edilen sonuçlar (The results obtained in the second stage)

Çalışmanın ikinci aşamasında, Senaryo A1'de VPN ve No-VPN olarak yapılan ikili sınıflandırmada verilerin dengeli olduğu gözlemlenmiştir. Bu nedenle, Senaryo A1'deki zaman bazlı 4 adet veri seti için veri seti dengeleme işlemi yapılmamıştır. Ancak, Senaryo A2 ve Senaryo B'de yer alan toplam 12 adet veri setinde SMOTE kullanılarak dengeleme işlemi gerçekleştirilmiştir. Ayrıca hiperparametrelerin rastgele örnekleri seçilip çapraz doğrulama yapılarak en iyi kombinasyon belirlenmiş ve hiperparametre optimizasyonu

sağlanmıştır. Yapılan optimizasyon sonrasında veriler tekrar eğitim ve teste tabi tutulmuştur. LGBM, KNN, SVM ve LR algoritmaları kullanılarak yapılan optimizasyon sonrası elde edilen sonuçlar Tablo 10'da yer almaktadır. Dengeleme işlemi yapılmayan Senaryo A1'de yer alan 4 adet veri seti için optimizasyon sonrası en iyi hiperparametre seçimi yapıldıktan sonra elde edilen sonuçlara da aynı tabloda yer verilmiştir. Tablo 11'de ise optimizasyon sonrası algoritmalar için test süreleri raporlanmıştır.

Tablo 10. Veri setinde optimizasyon sonrası başarı oranları (The success rates after optimization on the dataset)

Algoritma	**LGBM**		**KNN**		**SVM**		**LR**	
Senaryolar	Doğruluk	F1 Skor	Doğruluk	F1 Skor	Doğruluk	F1 Skor	Doğruluk	F1 Skor
Senaryo A1 15s	**0,9135***	**0,9173***	**0,8277***	**0,8421***	0,6167	0,6113	**0,5963***	**0,5877***
Senaryo A1 30s	0,8994	0,9081	0,8279	0,8289	0,6445	0,6367	0,5956	0,5874
Senaryo A1 60s	0,8915	0,9005	0,8089	0,8191	0,6323	0,6308	0,5603	0,5231
Senaryo A1 120s	0,8870	0,9015	0,8127	0,8209	**0,6576***	**0,6388***	0,5904	0,5802
Senaryo A2 No-VPN 15s	**0,9653****	**0,9459****	**0,9126****	**0,8591****	0,7403	0,7633	**0,6646****	**0,7197****
Senaryo A2 VPN 15s	0,9267	0,8989	0,8695	0,8203	0,6511	0,6411	0,5692	0,5486
Senaryo A2 No-VPN 30s	0,9468	0,9302	0,9109	0,8505	0,6427	0,6866	0,6489	0,6746
Senaryo A2 VPN 30s	0,9299	0,9076	0,8989	0,8454	**0,7602****	**0,7176****	0,6278	0,5735
Senaryo A2 No-VPN 60s	0,9294	0,9169	0,8675	0,8186	0,6372	0,6374	0,5537	0,6009
Senaryo A2 VPN 60s	0,9228	0,8678	0,8739	0,7999	0,6866	0,6293	0,5430	0,5044
Senaryo A2 No-VPN 120s	0,9437	0,9271	0,8993	0,8433	0,6748	0,6711	0,6171	0,6293
Senaryo A2 VPN 120s	0,9149	0,8614	0,8725	0,7972	0,6969	0,6049	0,5518	0,4746
Senaryo B 15s	**0,9005***	**0,8614***	0,8219	0,7334	**0,5914***	**0,5696***	0,3608	0,4008
Senaryo B 30s	0,8989	0,8561	**0,8379***	**0,7357***	0,4245	0,4234	**0,3950***	**0,3778***
Senaryo B 60s	0,8335	0,8234	0,8187	0,7245	0,3767	0,3443	0,3289	0,2619
Senaryo B 120s	0,8869	0,8253	0,8214	0,7140	0,4128	0,3810	0,3705	0,3281

* Kalın yazılan değerler, ilgili senaryo ve algoritmada elde edilen en yüksek başarı oranlarıdır.

**Kalın ve kırmızı yazılan değerler, ilgili algoritmada tüm senaryolar içinde elde edilen en yüksek başarı oranlarıdır.

Tablo 11. Veri setinde optimizasyon sonrası test süreleri (The test durations after optimization on the dataset)

Algoritma		**LGBM**		**KNN**		**SVM**		**LR**	
Senaryo	**Test Edilen Girdi Sayısı (adet)**	Test Süresi (s)	Birim Test Süresi (s/ad)	Test Süresi (s)	Birim Test Süresi (s/ad)	Test Süresi (s)	Birim Test Süresi (s/ad)	Test Süresi (s)	Birim Test Süresi (s/ad)
Senaryo A1 15s	5.628	0,06	0,000011	0,43	0,000076	2,56	0,000455	0,01	0,000002
Senaryo A1 30s	4.397	0,10	0,000023	0,31	0,000071	1,82	0,000414	0,01	0,000002
Senaryo A1 60s	4.655	0,25	0,000054	0,44	0,000095	1,30	0,000279	0,01	0,000002
Senaryo A1 120s	3.236	0,34	0,000105	0,60	0,000185	0,96	0,000297	0,01	0,000003
Senaryo A2 No-VPN 15s	2.690	**0,27**	**0,000100**	**1,50**	**0,000558**	3,45	0,001283	0,01	0,000004
Senaryo A2 VPN 15s	2.938	0,93	0,000317	0,17	0,000058	1,79	0,000609	0,01	0,000003
Senaryo A2 No-VPN 30s	2.076	0,03	0,000014	0,10	0,000048	1,45	0,000698	0,01	0,000005
Senaryo A2 VPN 30s	2.321	0,04	0,000017	0,15	0,000065	1,31	0,000564	0,01	0,000004
Senaryo A2 No-VPN 60s	2.574	0,03	0,000012	0,11	0,000043	2,94	0,001142	0,01	0,000004
Senaryo A2 VPN 60s	2.081	0,03	0,000014	0,14	0,000067	1,31	0,000630	0,01	0,000005
Senaryo A2 No-VPN 120s	1.546	0,02	0,000013	0,06	0,000039	0,74	0,000479	0,01	0,000006
Senaryo A2 VPN 120s	1.690	0,11	0,000065	0,45	0,000266	1,10	0,000651	0,01	0,000006
Senaryo B 15s	5.704	0,16	0,000028	1,46	0,000256	12,75	0,002235	0,01	0,000002
Senaryo B 30s	4.543	0,64	0,000141	0,65	0,000143	7,70	0,001695	0,01	0,000002
Senaryo B 60s	4.825	0,23	0,000048	0,60	0,000124	6,50	0,001347	0,01	0,000002
Senaryo B 120s	3.501	0,79	0,000226	0,54	0,000154	6,00	0,001714	0,01	0,000003
Ortalama Test Süreleri		0,25	0,000074	0,48	0,000140	3,35	0,000906	0,01	0,000003

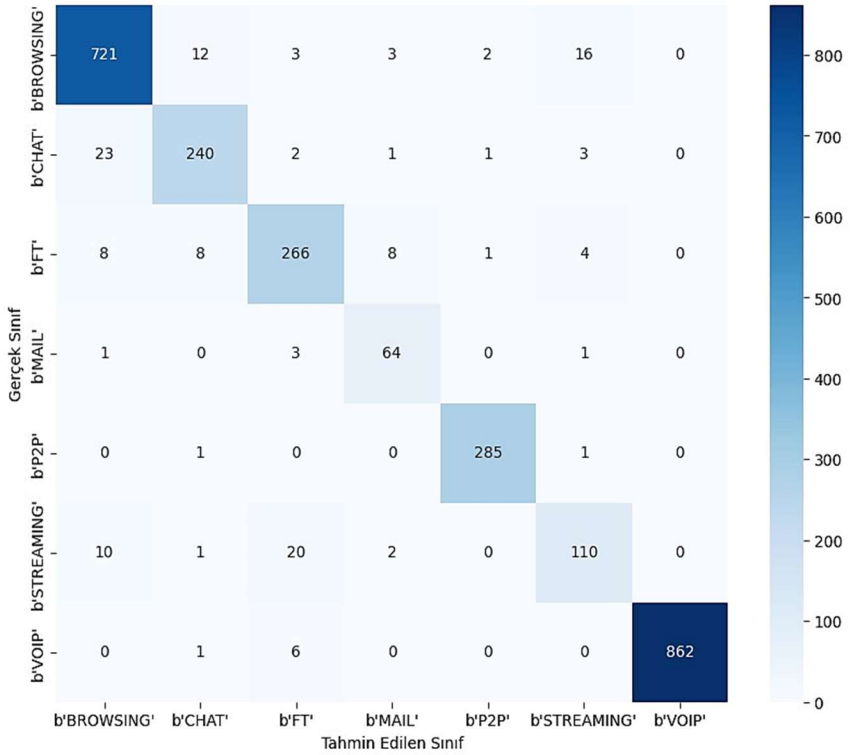
* Kalın yazılan değerler, en yüksek başarı oranı elde edilen senaryo ve algoritmalardaki test süreleridir.

Tüm senaryolarda optimizasyon sonrasında en yüksek başarı oranlarına denetimli makine öğrenmesi algoritmaları arasından KNN ile ulaşılmıştır. KNN algoritmasında elde edilen en yüksek başarı oranları, Senaryo A2 No-VPN 15s veri setinde 0,9126 doğruluk ve 0,8591 F1 skor olarak gerçekleşmiştir. Bu veri setinde 2.690 girdi için test süresi 1,5 saniye sürmüş, bu da her bir girdinin ortalama 0,000558 saniyede kategorize edilebildiğini göstermektedir.

Topluluk öğrenmesi algoritmalarından LGBM'de ise en yüksek başarı oranlarına, yine Senaryo A2 No-VPN 15s veri setinde 0,9653 doğruluk ve 0,9459 F1 skor ile ulaşılmıştır. Ayrıca LGBM, her bir girdiyi ortalama 0,0001 saniyede kategorize edebilmiştir.

Sonuç olarak LGBM'nin hem doğruluk ve F1 skor değerleri hem de kategorizasyon süresinin kısalığı bakımından KNN'den daha etkin olduğu görülmektedir.

Senaryo A2 No-VPN 15s veri seti ile LGBM algoritması kullanılarak elde edilen en başarılı sonuç için Şekil 5'deki karmaşıklık matrisine göre yapılan analizde, sınıf türlerini tahmin etmedeki en belirgin hatanın Chat ile Browsing sınıfları arasında olduğu görülmektedir. Bu durum, Chat sınıfına ait örneklerin Browsing olarak ya da Browsing sınıfına ait örneklerin Chat sınıfı olarak tahmin edildiğini göstermektedir.



Şekil 5. LGBM algoritması ile Senaryo A2 No-VPN 15s veri seti karmaşıklık matrisi
(The confusion matrix for the Scenario A2 No-VPN 15s dataset with the LGBM algorithm)

KNN ve LGBM algoritmalarına ait parametre optimizasyonu sonrasında elde edilen en iyi parametreler Tablo 12'de sunulmaktadır.

Tablo 12. KNN ve LGBM algoritmaları için en iyi parametre değerleri
(The best parameter values for the KNN and LGBM algorithms)

Algoritma	Parametrenin Adı	Hiperparametre Aralıkları	En İyi Değer
KNN	n_neighbors	[3, 5, 7, 9]	7
	weights	['uniform', 'distance']	distance
	metric	['euclidean', 'manhattan', 'minkowski']	manhattan
LGBM	colsample_bytree	[0.5, 0.6, 0.7, 0.8, 0.9, 1.0]	0.8
	learning_rate	[0.01, 0.05, 0.1, 0.3, 0.5, 0.7]	0.5

max_depth	sp_randint(3, 30)	13
min_child_samples	sp_randint(100, 1000)	257
min_child_weight	[1e-3, 1e-2, 0.1, 0.5, 1, 5, 10]	0.01
n_estimators	sp_randint(50, 1000)	604
num_leaves	sp_randint(6, 100)	18
reg_alpha	[0, 0.1, 0.5, 1, 2, 5, 10]	0
reg_lambda	[0, 0.1, 0.5, 1, 2, 5, 10]	1
subsample	[0.6, 0.7, 0.8, 0.9, 1.0]	0.9

Çalışmanın her iki aşamasında da algoritmaların test süreleri incelendiğinde, senaryolardaki en düşük toplam test sürelerinin genellikle LGBM ve LR algoritmalarında elde edildiği, buna karşılık KNN ve özellikle SVM'nin daha yüksek test sürelerine sahip olduğu görülmektedir. Birim test süresi değerlendirildiğinde LR, çoğunlukla her bir giriş başına en düşük süreyi sunarken (yaklaşık 2 ila 7 mikro saniye), LGBM de bu açıdan rekabetçi sonuçlar ortaya koymaktadır. KNN ve SVM ise hem toplam hem de birim başına test süresi anlamında diğer iki algoritmaya kıyasla daha uzun test süreleriyle dikkat çekmektedir. Özellikle SVM'nin birim test başına milisaniyeye yaklaşan veya geçen süreleri (0,000279 s/ad ve üzerine çıkan değerler) onu hız açısından dezavantajlı hale getirirken, LR ve LGBM test koşullarında en hızlı tepki verebilen modeller olarak öne çıkmıştır. Bu veriler genel olarak, test süresinin kritik olduğu hallerde LR ve LGBM'nin tercih edilebileceğine, KNN ve SVM'nin ise daha uzun işlem süresi gerektirdiğine işaret etmektedir.

5. Sonuçlar ve Tartışma (Results and Discussion)

Bu çalışmada, siber güvenlik uzmanlarının ağ üzerindeki tehditleri tespit etme ve saldırıları önleme yeteneklerinin artırılarak sistemlerin daha güvenli hale getirilmesi amaçlanmıştır. Bu sınıflandırma sayesinde, siber güvenlik uzmanları belirledikleri politikalar doğrultusunda, örneğin yetkisiz erişim olarak etiketlenmiş istenmeyen ağ trafiğini tespit edip engelleyebileceklerdir. Çalışma kapsamında, ağ trafiğinde yaygın olarak kullanılan sınıfları içeren ve gerçek internet trafiğinden elde edilen ISCXVPN2016 veri seti kullanılmıştır.

Bu veri seti için denetimli makine öğrenmesi algoritmalarından KNN, SVM ve LR ile topluluk öğrenmesi algoritmalarından LGBM algoritması kullanılmıştır. Kullanılan algoritmalar için öncelikle orijinal veri seti ile eğitim ve test gerçekleştirilmiştir. Daha sonra, dengesiz dağılımlı veri setleri SMOTE ile dengelenip, rastgele arama ve çapraz doğrulama yöntemleri ile parametre optimizasyonu yapılarak tekrar eğitim ve test gerçekleştirilmiştir. Optimizasyon sonucunda başarı oranlarının arttığı görülmüştür.

Bu çalışmada, optimizasyon sonrası LGBM algoritması ile elde edilen en başarılı sonuçlar ile literatürde yer alan diğer araştırmacıların geliştirdiği algoritmaların başarı oranlarının karşılaştırması Tablo 13'te sunulmuştur.

Tablo 13. Literatür karşılaştırması (Literature comparison)

Senaryo	Yazar(lar)	Algoritma	Doğruluk	F1 Skor	Kesinlik
A1	Draper-gil vd. [8]	C4.5	-	-	90,60%
	Majeed vd. [38]	Federated Learning	86%	-	-
	Uğurlu vd. [2]	XGBoost	93,04%	-	-
	Bozkır vd. [22]	XGBoost	-	99%	-
	Kendi çalışmamız	LGBM	91,35%	91,73%	91,99%
A2	Draper-gil vd. [8]	C4.5	-	-	89%
	Caicedo vd. [6]	Bagging	94,42%	-	-
	Majeed vd. [38]	Federated Learning	86%	-	-
	Uğurlu vd. [2]	XGBoost	94,53%	-	-
	Bozkır vd. [22]	XGBoost	-	99%	-
	Kendi çalışmamız	LGBM	96,53%	94,59%	94,86%
B	Draper-gil vd. [8]	C4.5	-	-	80,90%
	Caicedo vd. [6]	Bagging	86,94%	-	-

Majeed vd. [38]	Federated Learning	81%	-	-
Uğurlu vd. [2]	XGBoost	86,06%	-	-
Bozkır vd. [22]	XGBoost	-	99%	-
Kendi çalışmamız	LGBM	90,05%	86,14%	86,83%

Senaryo A1'de en yüksek başarı oranları 15s veri setinde elde edilmiş olup doğruluk %91,35, F1 Skor %91,73 ve kesinlik %91,99'dur. Draper-gil vd. [7] ve Majeed vd. [21] tarafından yapılan çalışmalara göre daha yüksek başarı oranlarına ulaşıldığı görülmüştür. Ancak Uğurlu vd. [2] ve Bozkır vd. [22] tarafından yapılan çalışmalardan daha az başarı elde edilmiştir. Senaryo A2'de en yüksek başarı oranları 15s veri setinde elde edilmiş olup doğruluk %96,53, F1 Skor %94,59 ve kesinlik %94,86'dır. Senaryo B'de en yüksek başarı oranları da yine 15s veri setinde elde edilmiş olup doğruluk %90,05, F1 Skor %86,14 ve kesinlik %86,83'dir. Her iki senaryoda da Draper-gil vd. [7], Caicedo vd. [5], Majeed vd. [21] ve Uğurlu vd. [2] tarafından yapılan çalışmalara göre daha yüksek başarı oranlarına ulaşıldığı görülmüştür. Ancak Bozkır vd. [22] tarafından yapılan çalışmadan daha az başarı elde edilmiştir.

Şifreli ağ trafiğinin sınıflandırılmasına ilişkin bu çalışmada, başarı oranlarının yanı sıra günümüzde hemen her alanda önemli bir kriter haline gelen test süresi de ölçülmüştür. Optimize edilmiş koşullarda, LGBM topluluk öğrenmesi algoritmasının ortalama girdi kategorizasyon süresi 0,000074 saniye olarak hesaplanırken, KNN makine öğrenmesi algoritmasının ortalama kategorizasyon süresi 0,00014 saniye olarak belirlenmiştir. Bu değerlere göre LGBM algoritması, KNN'ye kıyasla girdi kategorizasyonunda daha hızlı çalışmaktadır. Elde edilen bulgular, Khatouni ve Heywood [35] tarafından yapılan çalışmanın aksine, şifreli trafiğin sınıflandırılmasında topluluk öğrenmesi algoritmasının makine öğrenmesi algoritmasına oranla daha yüksek bir hız sağlayabileceğini ortaya koymaktadır.

Yapılan çalışmada hem doğruluk hem de hız açısından yüksek performanslı bir LGBM topluluk öğrenmesi modeli geliştirilmiştir. Şifreli trafiğin sınıflandırılmasına odaklanan bu yaklaşım, mevcut tehdit tespit ve engelleme sistemlerinin iyileştirilmesini ve güçlendirilmesini mümkün kılmaktadır. Önerilen model, şifreli trafiği daha doğru ve hızlı biçimde sınıflandırırken, yeni nesil güvenlik duvarları veya ağ izleme sistemleriyle entegre edilerek daha etkin ve uyarlanabilir bir güvenlik yapısı oluşturulabilir. Gelecekte, farklı veri kaynaklarının kullanımıyla daha kapsamlı ve güvenilir sınıflandırma modellerinin geliştirilmesi, modelin hızını artıracak yeni yöntemlerin tasarlanması ve veri bütünlüğünü korumak amacıyla zaman damgası entegrasyonunun sağlanması öngörülmektedir. Bu alanda sürdürülen çalışmalar ilerledikçe, modellere daha yüksek güvenilirlik düzeyleri kazandırılacak ve böylece sistemler daha güvenli bir hale getirilebilecektir.

Çıkar Çatışması Beyanı (Conflict of Interest Statement)

Yazarlar tarafından herhangi bir çıkar çatışması bildirilmemiştir.

Kaynaklar (References)

[1] B. M. Leiner, R. E. Kahn, J. Postel, and L. Kleinrock, "A brief history of the internet," *AC M SIGCO MM Computer Communication Review,* vol. 39, no. 5, pp. 22-31, Oct. 2009. doi: 10.1145/1629607.1629613

[2] M. Uğurlu, İ. A. Doğru, and R. S. Arslan, "A new classification method for encrypted internet traffic using machine learning," *Turkish Journal of Electrical Engineering and Computer Sciences,* vol. 25, no. 9, pp. 2450-2468, 2021. doi: 10.3906/ELK-2011-31

[3] E. Cengiz and M. Gök, "Reinforcement learning applications in cyber security: A Review," *Sakarya University Journal of Science*, vol. 27, no. 2, pp. 481-503, Apr. 2023. doi: 10.16984/saufenbilder.1237742

[4] M. K. Pehlivanoğlu, R. Atay, and D. E. Odabaş, "İki seviyeli hibrit makine öğrenmesi yöntemi ile saldırı tespiti," *Gazi Journal of Engineering Sciences,* vol. 5, no. 3, pp. 258-272, Dec. 2019. doi: 10.30855/gmbd.2019.03.07

[5] J. A. Caicedo-Muñoz, A. L. Espino, J. C. Corrales, and A. Rendón, "QoS-classifier for VPN and Non-VPN traffic based on time-related features," *Computer Networks,* vol. 144, pp. 271-279, Oct. 2018. doi: 10.1016/j.comnet.2018.08.008

[6] M. Uğurlu, "Şifrelenmiş internet trafiğinin makine öğrenmesi yaklaşımı ile sınıflandırılması," Yüksek Lisans Tezi, Gazi Üniversitesi, Fen Bilimleri Enstitüsü, Ankara, 2020

[7] G. Drapper-Gil, A. H. Lashkari, M. S. I. Mamun, and A. A. Ghorbani, "Characterization of encrypted and VPN traffic using time-related features," ICISSP 2016 - Proceedings of the 2nd International Conference on Information Systems Security and Privacy, 19-21 Şubat 2016, Roma, İtalya, SciTePress, pp. 407-414. doi: 10.5220/0005740704070414

[8] Y. F. Huang, C. B. Lin, C. M. Chung, and C. M. Chen, "Research on QoS classification of network encrypted traffic behavior based on machine learning," Electronics (Switzerland), vol. 10, no. 12, Jun. 2021. doi: 10.3390/electronics10121376

[9] S. Frankel (NIST), P. Hoffman (Virtual Private Network Consortium), A. Orebaugh (BAH), R. Park (BAH), "NIST SP 800-113 Guide to SSL VPNs." [Online]. Available: https://csrc.nist.gov/pubs/sp/800/113/final [Accessed: Mar. 24, 2024].

[10] N. H. Nguyen, *SSL/TLS interception challenge from the shadow to the light,* SANS Institute, 2019.

[11] S. Frankel, P. Hoffman, A. Orebaugh, and R. Park, NIST-"Guide to SSL VPNs," Recommendations of the National Institute of Standards and Technology. [Online]. Available: https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-113.pdf [Accessed: Mar. 21, 2024].

[12] Cisco, "Chapter: Traffic Classification," www.cisco.com. [Online]. Available: https://www.cisco.com/c/en/us/td/docs/nsite/enterprise/wan/wan_optimization/wan_opt_sg/chap05.html. [Accessed: Mar. 7, 2024].

[13] R. T. Elmaghraby, N. M. Abdel Aziem, M. A. Sobh, and A. M. Bahaa-Eldin, "Encrypted network traffic classification based on machine learning," Ain Shams Engineering Journal, vol. 15, no. 2, 2023. doi: 10.1016/j.asej.2023.102361

[14] Internet Assigned Numbers Authority, "Service name and transport protocol port number registry," www.iana.org. [Online]. Available: https://www.iana.org/assignments/service-names-port-numbers/service-names-port-numbers.xhtml. [Accessed: Mar. 19, 2024].

[15] D. T. Ergönül and O. Demir, "Real-time encrypted traffic classification with deep learning," *Sakarya University Journal of Science*, vol. 26, no. 2, pp. 313-332, Apr. 2022. doi: 10.16984/saufenbilder.1026502

[16] A. Azab, M. Khasawneh, S. Alrabaee, K. K. Choo, and M. Sarsour, "Network traffic classification: Techniques, Datasets, and Challenges," *Digital Communications and Networks. KeAi Communications Co.,* Sep. 2022. doi: 10.1016/j.dcan.2022.09.009

[17] V. A Muliukha, L. U Laboshin, and A. Lukashin, "Analysis and classification of encrypted network traffic using machine learning," XXIII International Conference on Soft Computing and Measurements, 27-29 May. 2020, Saint Petersburg, Russia, pp. 194-197. doi: 10.1109/SCM50615.2020.9198811

[18] A. Lichy, O. Bader, R. Dubin, A. Dvir, and C. Hajaj, "When an RF beats a CNN and GRU, together-a comparison of deep learning and classical machine learning approaches for encrypted malware traffic classification," Jun. 2022. [Online]. Available: http://arxiv.org/abs/2206.08004

[19] T. C. Obasi, "Encrypted network traffic classification using ensemble learning techniques," Ph.D. dissertation, Carleton University, Ottawa, Ontario, Canada, 2020. doi: 10.22215/etd/2020-14171

[20] G. Abbas, U. Farooq, P. Singh, S. S. Khurana, and P. Singh, "Feature engineering and ensemble learning-based classification of VPN and Non-VPN based network traffic over temporal features," *SN Comput Science*, vol. 4, Sep. 2023. doi: 10.1007/s42979-023-01944-5

[21] U. Majeed, L. U. Khan, and C. Seon Hong, "Cross-silo horizontal federated learning for flow-based time-related-features oriented traffic classification," APNOMS 2020 - 2020 21st Asia-Pacific Network Operations and Management Symposium: Towards Service and Networking Intelligence for Humanity, 23-25 Eylül 2020, Daegu, Güney Kore, pp. 389-392. doi: 10.23919/APNOMS50412.2020.9236971

[22] R. Bozkır, "Şifreli ağ trafiğinin içerik açısından sınıflandırılması," Yüksek Lisans Tezi, Bursa Uludağ Üniversitesi, Fen Bilimleri Enstitüsü, Bursa, 2022.

[23] B. Yamansavascilar, M. A. Guvensan, A. G. Yavuz, and M. E. Karsligil, "Application identification via network traffic classification," 2017 International Conference on Computing, Networking and Communications, 26-29 Ocak 2017, Santa Clara, CA, ABD, pp. 843-848. doi: 10.1109/ICCNC.2017.7876241.

[24] S. Bagui, X. Fang, E. Kalimantan, S. C. Bagui, and J. Sheehan, "Comparison of machine-learning algorithms for classification of VPN network traffic flow using time-related features," *Journal of Cyber Security Technology,* vol. 1, no. 2, pp. 108-126, Apr. 2017. doi: 10.1080/23742917.2017.1321891.

[25] M. Lotfollahi, M. Jafari Siavoshani, R. Shirali Hossein Zade, and M. Saberian, "Deep packet: A Novel Approach for Encrypted Traffic Classification using Deep Learning," *Soft Comput*, vol. 24, no. 3, pp. 1999-2012, Feb. 2019. doi: 10.1007/s00500-019-04030-2

[26] T. Shapira and Y. Shavitt, "Encrypted internet traffic classification is as easy as image recognition," IEEE INFOCOM 2019 - IEEE Conference on Computer Communications Workshops, 29 Nisan - 2 Mayıs 2019, Paris, France, pp. 680-687. doi: 10.1109/INFCOMW.2019.8845315.

[27] F Zhang, T Shang, and J Liu, "Imbalanced encrypted traffic classification scheme using random forest," 2020 International Conferences, Rhodes, Greece, pp. 837-842, 2020. doi: 10.1109/iThings-GreenCom-CPSCom-SmartData-Cybermatics50389.2020.00142

[28] L. Guo, Q. Wu, S. Liu, M. Duan, H. Li, and J. Sun, "Deep learning-based real-time VPN encrypted traffic identification methods," *Journal of Real-Time Image Processing 17*, Springer, pp. 103-114, Feb. 2020. doi: 10.1007/s11554-019-00930-6

[29] J. Cheng, R. He, E. Yuepeng, Y. Wu, J. You, and T. Li, "Real-time encrypted traffic classification via lightweight neural networks," 2020 IEEE Global Communications Conference, GLOBECOM 2020 - Proceedings, Institute of Electrical and Electronics Engineers Inc., 7-11 Aralık 2020, Taipei, Tayvan, pp. 1-6. doi: 10.1109/GLOBECOM42002.2020.9322309

[30] K. Zhou, W. Wang, C. Wu, and T. Hu, "Practical evaluation of encrypted traffic classification based on a combined method of entropy estimation and neural networks," *ETRI Journal*, vol. 42, no. 3, pp. 311-323, Jun. 2020. doi: 10.4218/etrij.2019-0190

[31] Z. Bu, B. Zhou, P. Cheng, K. Zhang, and Z. H. Ling, "Encrypted network traffic classification using deep and parallel network-in-network models," *IEEE Access*, vol. 8, pp. 132950-132959, 2020. doi: 10.1109/ACCESS.2020.3010637

[32] A. A. Afuwape, Y. Xu, J. H. Anajemba, and G. Srivastava, "Performance evaluation of secured network traffic classification using a machine learning approach," *Comput. Stand. Interfaces*, vol. 78, Oct. 2021. doi: 10.1016/j.csi.2021.103545

[33] K. Ismailaj, M. Camelo, and S. Latré, "When deep learning may not be the right tool for traffic classification," in Proc. of the 2021 IFIP/IEEE International Symposium on Integrated Network Management (IM), May 2021, Bordeaux, France, pp. 884-889. Available: IEEE Xplore, https://ieeexplore.ieee.org/abstract/document/9463927

[34] A. Almomani, "Classification of virtual private networks encrypted traffic using ensemble learning algorithms," *Egyptian Informatics Journal,* vol. 23, no. 4, pp. 57–68, Dec. 2022. doi: 10.1016/j.eij.2022.06.006

[35] A. S. Khatouni and N. Z. Heywood, " Integrating machine learning with off-the-shelf traffic flow features for HTTP/HTTPS traffic classification," 2019 IEEE Symposium on Computers and Communications (ISCC), 29 Haziran - 3 Temmuz 2019, Barc., Spain, pp. 1-7. doi: 10.1109/ISCC47284.2019.8969578

[36] A. Aydın, "Topluluk öğrenmesi kullanılarak zararlı alan adlarının sınıflandırılması," Yüksek Lisans Tezi, Marmara Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul, 2023.

[37] K. Eker, A. G. Eker, D. Mandal, M. Pehlivanoglu, and N. Duru, "Makine öğrenmesi yaklaşımları ile ağ trafik sınıflandırılması," 7th International Conference on Computer Science and Engineering, UBMK 2022, 14-16 Eylül 2022, Diyarbakır, Türkiye, pp. 393-397. doi: 10.1109/UBMK55850.2022.9919497

[38] Y. Alaca, "Siber güvenlikte CIC-Darknet2020 veri seti kullanarak VPN/NoVPN ve Tor/NoTor sınıflandırması: Basit ve Karmaşık Modellerin Kullanımı," *Fırat Üniversitesi Müh. Bilimleri Dergisi*, vol. 35, no. 2, pp. 569-579 Jul. 2023. doi: 10.35234/fumbd.1291388

[39] M. Uğurlu, İ. A. Doğru, and R. S. Arslan, "Detection and classification of darknet traffic using machine learning methods," *Journal of the Faculty of Engineering and Architecture of Gazi University*, vol. 38, no. 3, pp. 1737–1746, 2023. doi: 10.17341/gazimmfd.1023147

[40] E. G. İlgün, Y. Sönmez, and M. Dener, "DarkWEB traffic detection and classification using machine learning method," *Gazi Journal of Engineering Sciences*, vol. 9, no. 4, pp. 126-140, 2023. doi: 10.30855/gmbd.0705S13

[41] A. Zheng and A. Casari, "Feature engineering for machine learning: Principles and Techniques for Data Scientists," Chapter 2-Feature Selection, Almanya: O'Reilly Media. 2018.

[42] L. Yang and A. Shami, "On hyperparameter optimization of machine learning algorithms: Theory and Practice," *Neurocomputing,* vol. 415, pp. 295-316, Nov. 2020. doi: 10.1016/j.neucom.2020.07.061

[43] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," 2012. [Online]. Available: http://scikit-learn.sourceforge.net. [Accessed: March 05, 2024].

[44] C. Schaffer, "Selecting a classification method by cross-validation," Mach Learn 13, 135-143, 1993. doi: 10.1007/BF00993106

[45] X. Deng, Q. Liu, Y. Deng, and S. Mahadevan, "An improved method to construct basic probability assignment based on the confusion matrix for classification problem," *Inf Sci (N Y),* vol. 340–341, pp. 250-261, May 2016. doi: 10.1016/j.ins.2016.01.033