Research Article/Araştırma Makalesi

Gazi AKADEMİK YAYINCILIK

GAZİ
JOURNAL OF ENGINEERING SCIENCES

DarkWEB Traffic Detection and Classification Using Machine Learning Method

Esen Gül İlgün[*a], Yusuf Sönmez[b], Murat Dener[c]

Submitted: 19.11.2023 Revised: 20.12.2023 Accepted: 20.12.2023 doi:10.30855/gmbd.0705S13

ABSTRACT

DarkWEB makes up 6% of DeepWEB, which contains data that search engines cannot index and is approximately 96% of all websites. DarkWEB is encrypted network traffic tunneled through special software such as TOR (The Onion Router) and provides a high level of anonymity with a series of anonymized connections that make the IP address untraceable. This makes it easier to carry out criminal activities such as media piracy, drug dealing, terrorism and child pornography. In this study, the statistical information of the packets was analyzed without decrypting this encrypted network traffic. Different data sets were obtained by applying categorical data coding, scaling, feature selection and data balancing pre-processes separately and together to the CIC-Darknet2020 data set used within the scope of the proposed methodology for high-accuracy detection and classification of DarkWEB traffic. Obtained data sets and Logistic Regression (LR), Gaussian Naive Bayes (GNB), Decision Tree (DT), K-Nearest Neighbor (KNN), Multi Layer Perceptron (MLP), Random Forest (RF), eXtreme Gradient Boosting (XGBoost), many DarkWEB traffic detection and classification models have been created using Light Gradient Boosting Machine (LightGBM), Category Boosting (CatBoost) machine learning algorithms. With the models created, Encryption (Encrypted, Standard), Category (Tor, Non-Tor, Non-VPN, VPN), Subcategory (Audio-Stream, Browsing, Chat, E-mail, P2P, Transfer, Video-Stream, VOIP) classes 2, 4 and 8 classifications were made. Accuracy rates of 99.9% were achieved in 2-fold and 4-fold classification, and 94% accuracy rate were achieved in 8-fold classification.

Keywords: DeepWEB, DarkWEB, encrypted network traffic, machine learning, classification

[a,*] Gazi University,
Graduate School of Natural And Applied Sciences,
Dept. of Information Security Engineering

06560 - Ankara, Türkiye
Orcid: 0000-0002-1719-5727
e mail: egul.ilgun@gazi.edu.tr

[b] Gazi University,
Technology Faculty,
Dept. of Computer Engineering
06560 - Ankara, Türkiye
Orcid: 0000-0002-9775-9835

[c] Gazi University,
Graduate School of Natural And Applied Sciences,
Dept. of Information Security Engineering
06261 – ANKARA, Türkiye
Orcid: 0000-0001-5746-6141

[*]Corresponding author:
egul.ilgun@gazi.edu.tr

Makine Öğrenme Yöntemi Kullanılarak DarkWEB Trafiği Tespiti ve Sınıflandırması

ÖZ

DarkWEB, arama motorlarının indeksleyemediği verileri içeren ve tüm web sitelerinin yaklaşık %96'sı olan DeepWEB'in %6'sını oluşturur. DarkWEB, TOR (The Onion Router) gibi özel yazılımlar ile tünellenen şifreli ağ trafiğidir ve IP adresini izlenemez hale getiren anonimleştirilmiş bir dizi bağlantı ile yüksek düzeyde anonimlik sağlar. Bu durum medya korsanlığı, uyuşturucu satıcılığı, terörizm, çocuk pornografisi gibi suç faaliyetlerinin gerçekleştirilmesini kolaylaştırır. Bu çalışmada, bu şifreli ağ trafiğinde deşifreleme işlemi yapılmadan, paketlerin istatistiki bilgileri analiz edilmiştir. DarkWEB trafiğinin yüksek doğrulukta tespiti ve sınıflandırılması için önerilen metodoloji kapsamında kullanılan CIC-Darknet2020 veri setine kategorik veri kodlama, ölçeklendirme, öznitelik seçimi ve veri dengeleme ön işlemleri ayrı ayrı ve de birlikte uygulanarak farklı veri setleri elde edilmiştir. Elde edilen veri setleri ve Logistic Regression (LR), Gaussian Naive Bayes (GNB), Decision Tree (DT), K-Nearest Neighbor (KNN), Multi Layer Perceptron (MLP), Random Forest (RF), eXtreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), Category Boosting (CatBoost) makine öğrenme algoritmaları kullanılarak çok sayıda DarkWEB trafiği tespit ve sınıflandırma modeli oluşturulmuştur. Oluşturulan modeller ile Encryption (Şifreli, Standart), Category (Tor, Non-Tor, Non-VPN, VPN), Subcategory (Audio-Stream, Browsing, Chat, E-mail, P2P, Transfer, Video-Stream, VOIP) sınıfları olmak üzere 2'li, 4'lü, 8'li sınıflandırmalar yapılmıştır. 2'li ve 4'lü sınıflandırmada %99.9, 8'li sınıflandırmada ise %94 doğruluk oranına ulaşılmıştır.

Anahtar Kelimeler: DeepWEB, DarkWEB, şifreli ağ trafiği, makine öğrenme, sınıflandırma

To cite this article: E. G. İlgun, Y. Sonmez 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

1. Giriş (Introduction)

Herkesin erişebileceği ve dizine eklenmiş adres alanı, clearnet olarak bilinen ağın dışında olan, standart arama motorları tarafından indekslenmeyen, internetin yer altı olarak tabir edilen Deep WEB'in, clearnet'in 500 katından fazlası olduğu tahmin edilmektedir [1]. DeepWEB'in asıl amacı kullanıcı anonimliğidir ve kullanımı yasaldır. DeepWEB içindeki ayrılmamış adres alanından oluşan ağlar ise DarkWEB olarak bilinir. DarkWEB [2], internet içinde şifrelenmiş, yalnızca belirli bir yapılandırma ve yetkilendirme ile erişilebilen bir yer paylaşımını ifade eder yani DeepWEB'in gizli kısmıdır. Kullanım amacına göre DarkWEB yasal ya da yasadışı olabilir fakat yaklaşık %95 oranında yasadışı faaliyetler için kullanılmaktadır.

DarkWEB'e özel bir yazılım ile kullanıcı yetkilendirmesi veya standart dışı iletişim protokolleri gerektiren bir yer paylaşımlı ağ ile ulaşılır [3]. Bu özel yazılımlardan en bilinenleri Tor ve FreeNet'tir. Tor 2002 yılında ABD Donanma Araştırma Laboratuarı ile Free Haven Projesi arasında ortak bir proje olarak ortaya çıkmıştır. Projenin amacı anonim, şifrelenmiş bir ağ oluşturmaktı. FreeNet ise "Dağıtılmış, Merkezi Olmayan Bilgi Depolama ve Alma Sistemi" adında bir tezden üretilmiş yazılımdır. Bu yazılımlar ile giriş yapılan web adresleri, anlamsız numaralar ile harflerden oluşmaktadır ve bu adresler düzenli olarak değiştirilmektedir. Bu ağlar, bir IP adresini izlenemez hale getiren anonimleştirilmiş bir dizi bağlantı, proxy ağları aracılığıyla yüksek düzeyde anonimlik sağlar. Bu durum bilgisayar korsanlığı, medya korsanlığı, uyuşturucu satıcılığı, terörizm, insan kaçakçılığı ve çocuk pornografisi gibi suç faaliyetlerinin gerçekleştirilmesini kolaylaştırır. Birçok yasa dışı ürün ve hizmetin satışa sunulduğu DarkWEB pazarlarında, 2020 yılı itibariyle [4] 500.000'den fazla kullanıcı işlem gerçekleştirdi, 2400'den fazla aktif satıcı olduğu tespit edildi, 320.000'den fazla alım-satım işlemi yapıldı, 4.650'den fazla bitcoin ve 12.800'den fazla monero kripto para el değiştirdi. Bu rakamların günümüze dek katlanarak arttığı tahmin edilmektedir. DarkWEB ortamında gerçekleştirilen suç faaliyetlerini engellemek için DarkWEB trafiğinin makine öğrenme ve derin öğrenme teknikleri kullanılarak tespiti ile sınıflandırılması, önemli bir çalışma alanıdır.

Bu çalışmanın amacı, Lashkari vd. [5] tarafından üretilen ve açık kaynak olan CICDarknet2020 veri setine çeşitli ön işlemler uygulayarak, kurulan makine öğrenme modellerinin yüksek doğrulukta DarkWEB trafiği tespiti ile sınıflandırması yapmasını sağlamaktır.

Bu makalenin ana katkı noktaları aşağıdaki gibi olmuştur:

- XGBoost gömülü öznitelik seçimi yöntemi kullanılarak modellerin tespit ve sınıflandırma başarısını olumsuz etkileyecek öznitelikler veri setinden çıkarılmıştır.
- Rastgele Yeniden Örneklendirme (Random Over Sampling, ROS) yöntemi kullanılarak modelin objektif tespit ve sınıflandırma yapmasını engelleyebilecek, veri seti içerisindeki sınıf dengesizliği sorunu giderilmiştir.
- Oluşturulan modeller ile CICDarknet2020 veri setindeki katmanlı yaklaşım baz alınarak Encryption (Şifreli, Standart), Category (Tor, Non-Tor, Non-VPN, VPN), Subcategory (Audio-Stream, Browsing, Chat, E-mail, P2P, Transfer, Video-Stream, VOIP), sınıfları olmak üzere 2'li, 4'lü, 8'li sınıflandırmalar yapılmıştır.
- CICDarknet2020 veri setine, kategorik veri kodlama, ölçeklendirme, öznitelik seçimi, veri dengeleme ön işlemlerinin ayrı ayrı ve birlikte uygulandığı dört farklı senaryo denenmiş ve senaryoların uygulanması ile elde edilen sonuçlar karşılaştırılmıştır.
- Yapılan çalışmada, CICDarknet2020 veri setinin kullanıldığı literatür ile karşılaştırıldığında daha yüksek sınıflandırma doğruluğuna ulaşılmıştır.

Çalışmanın 2. bölümünde, ilgili çalışmalar özetlenmiştir. 3. bölümde, önerilen metodoloji detaylı olarak anlatılmıştır. 4. bölümde, önerilen metodolojinin uygulaması ve alınan sonuçlar verilmiştir. 5. bölümde ise sonuçların analizi ve değerlendirilmesi yapılmıştır.

2. İlgili Çalışmalar (Related Works)

Literatürde DarkWEB trafiğinin tespiti ve sınıflandırılması konusunda birçok yöntem önerilmiş, bu yöntemlerin uygulanması ile başarılı sonuçlar alınmış ve bu konuda gelecekte yapılması planlanan çalışmalara değinilmiştir. Bu bölümde yapılan bu çalışmalardan bahsedilmektedir:
Muhammad Bılal Sarwar vd. [6] DarkWEB ağ trafiği tespiti ve sınıflandırması için veri setinden en uygun özellikleri seçerken, Principal component analysis (PCA), Decision Tree Classifier, XGBClassifier kullanmış, geleneksel makine öğrenme algoritmaları ve ön işlenmiş veri seti ile modeller oluşturmuştur. Ardından, ağ trafiğini daha doğru bir şekilde tanımak için değiştirilmiş CNN-LSTM ve CNN-GRU derin öğrenme tekniklerini uygulamıştır. Önerilen yaklaşım, XGB özellik seçme yöntemi ve CNN-LSTM kullanılarak DarkWEB trafiği tespitinde %96, sınıflandırmasında ise %89 doğruluk elde etmiştir. L. A. Iliadis ve T. Kaifas [7] CICDarknet2020 veri seti üzerinde, ikili ve çok sınıflı sınıflandırma görevinde Random Forest algoritması ile ortalama %98'in üzerinde bir tahmin doğruluğu elde etmiş ve özellik önem analizi ile birlikte bir ROC analizi yapmıştır. S. Sridhar and S. Sanagavarapu [8] CICDarknet2020 veri seti üzerinde, öznitelik seçimi için Ki-Kare yöntemi, sınıf dengesizliğini gidermek için Conditional Generative Adversarial aşırı örnekleme yöntemi uygulayarak çok sınıflı sınıflandırmada Random Forest ile 97,87 F1 Puanı elde etmiştir. Y. Li, Y. Lu ve S. Li [9] CIC-Darknet2020 veri setini kullanarak sıfır gün saldırılarını sınıflandırmak için CNN ve K-Means'ı birleştiren bir Şifreli Sıfır Gün Uygulamaları Sınıflandırması (EZAC) yöntemini önermiştir. İlk önce akışları sınıflandırmak için CNN, sıfır gün uygulamaları olabilecek akışları kategorize etmek için ise K-Means'ı kullanmıştır. Daha sonra bu akışlar manuel olarak etiketlenmiştir. EZAC yöntemi ile %97,4 doğru tespit oranı elde edilmiştir. M. Uğurlu vd. [10], CICDarknet2020 veri setindeki 82 adet öznitelik içerisinden ağırlıklandırma işlemi yaparak 30 adet öznitelik seçmiş, ROSE tekniği ile veri dengeleme yapmış, 10 faklı makine öğrenme algoritması ile model oluşturmuş ve hiper-parametre ayarı yapmıştır. Çalışmada önerilen yöntem ile Karar Ağacı algoritması ile kurulan model de %93,32 doğru tespit oranına ulaşılmıştır. Nhien Rust-Nguyen vd. [11] CIC-Darknet2020 veri setini kullanarak, modellerde hiper-parametre ayarı yapmış, kayıtları dört trafik sınıfına ve sekiz uygulama sınıfına göre sınıflandırmıştır. Sınıf dengesizliği için ise SMOTE seviyelerini denemiş, CNN ve AC-GAN için trafik özelliklerinin iki boyutlu temsillerini araştırmıştır. Ayrıca en iyi performans gösteren sınıflandırıcının kafasını karıştırmak için bir saldırganın bakış açısıyla yaklaşarak, uygulama sınıfı trafik özelliklerini etkili bir şekilde gizlemiştir. Random Forest ile trafik sınıflandırması için %99,8 F1 puanı ve uygulama sınıflandırması için %92,2 F1 puanı elde etmiştir. Ammar Almomani [12], DarkWEB trafiğini tespit etmek için iki aşamalı bir yöntem önermiştir. İlk aşamada ANN, RF ve SVM algoritmaları ile yapılan tahmin performanslarını, 2. Aşamada lojistik regresyon meta sınıflandırıcısına dayanan bir teknik ile son tahminleri oluşturmak için birleştirmiştir. Kümeleme topluluk öğrenimi denilen bu yöntem ile eğitim veri setinde %99'dan ve test veri setinde %97'den fazla hassasiyet ve doğruluk değerlerine ulaşılmıştır. Hardhik Mohanty vd. [13], RF, KNN, DT olmak üzere üç temel öğrenicinin tahminlerini en verimli şekilde birleştirmek için bir İstifleme Topluluğu (SE) modeli önermiştir. Önerilen model, Hızlı Gradyan İşaret Yöntemi (FGSM), Temel Yinelemeli Yöntem (BIM), DeepFool ve sınır saldırısı olmak üzere 4 saldırı çeşidi ile test edilmiş ve %98,89 doğru tespit oranı, %0,43 FPR elde edilmiştir. Qasem Abu Al-Haija vd. [14], Torbalama Karar Ağacı Toplulukları (BAG-DT), AdaBoost Karar Ağacı Toplulukları (ADA-DT), RUSBoosted Karar Ağacı Toplulukları (RUS-DT), Optimize Edilebilir Karar Ağacı (O- DT), Optimize Edilebilir k en yakın komşu (O-KNN) ve Optimize Edilebilir Discriminant (O-DSC) olmak üzere altı denetimli makine öğrenme tekniği ile DarkWEB trafik algılama modeli oluşturmuştur. Modelin CICDarknet2020 veri seti üzerindeki test performansı, torbalama topluluğu tekniği (BAG-DT) ile kurulan modelde 9,09 saniyede %99,50 sınıflandırma doğruluğu olmuştur. Yan Li ve Yifei Lu [15], şifreli uygulamaları sınıflandırmak için hem protokolleri hem de uygulamaları tanımlayabilen CNN (ETCC) yöntemini kullanan iki aşamalı, iki etiketli sınıflandırma önermiştir. İlk aşama, şifreli trafiği kullanılan protokolü sınıflandırır. İkinci aşama, uygulamaları trafik tarafından kullanılan protokole göre sınıflandırır. Önerilen ETCC yöntemi ile %97 sınıflandırma doğruluğuna ulaşılmıştır. Mahmoud Alimoradi vd. [16], DarkWEB trafiğini Tor, Tor olmayan, VPN ve VPN olmayan olmak üzere dört sınıfa ayırmış, 79 giriş nöronlu ve 6 gizli katmanlı derin sinir ağı mimarisi ile ham DarkWEB trafiğinden karmaşık, doğrusal olmayan ilişkileri keşfederek DIDarknet veri seti üzerinde %96 doğruluk oranına ulaşmıştır. Gerard Draper-Gil vd. [17], VPN trafiğini algılamak ve şifreli trafiği tarama, akış vb. trafik türüne göre farklı kategorilerde karakterize etmek için akışa dayalı zamanla ilgili özniteliklerin etkinliğini incelemiş, özniteliklerin doğruluğunu test etmek için makine öğrenme teknikleri (C4.5 ve KNN) kullanmıştır.
Literatür incelendiğinde DarkWEB trafiği tespiti ve sınıflandırması konusunda yeterince çalışma ve de veri seti olmadığı görülmüştür. Tablo 1'de ilgili çalışmaların bir özeti yer almaktadır.

Tablo 1. İlgili çalışmalar (Related works)

Yıl	Yazarlar	Model	Veri Seti	En iyi sonuçlar
2016	G. Draper-Gil vd.	C4.5, KNN	Real time dataset	Doğruluk:%80
2021	Y. Li, Y. Lu	CNN	CICDarknet2020	Doğruluk:%97
2021	M. B. Sarwar vd.	DT, GB, RFR, XGB, CNN-LSTM, CNN-GRU	CICDarknet2020	Doğruluk: %96, %89
2021	L. A. Iliadis, T. Kaifas	KNN, MLP, RF, DT, GB	CICDarknet2020	Doğruluk: 98.71 Hassasiyet: 0.9871 Kesinlik: 0.9870 F1-Ölçütü: 0.9870
2021	S. Sridhar, S. Sanagavarapu	RF	CICDarknet2020	F1- Ölçütü: 97,87
2022	H. Mohanty vd.	Stacking Ensemble (SE) model (RF-KNN-DT)	CICDarknet2020	Doğruluk:%98,89
2022	Q. A. Al-Haija vd.	BAG-DT, ADA-DT, RUS-DT, O-DT, O-KNN, O-DSC	CICDarknet2020	Doğruluk:%99,50
2022	M. Alimoradi vd.	DNN	CICDarknet2020	Doğruluk:%96
2023	M. Uğurlu vd.	KNN, LR, RF SVM, DT, GNB, LDA, GB, ET, XGBoost	CICDarknet2020	Doğruluk : %93,32
2023	N. Rust-Nguyen vd.	GBDT, XGBoost, KNN, MLP, SVM, RF, CNN, AC-GAN	CICDarknet2020	F1- Ölçütü: %99,8, %92,2
2023	A. Almomani	NN LG KNN SVM stacking ensemble learning (RF SVM ANN)	CICDarknet2020	Doğruluk:%99.4, %96.74
2023	E. G. İlgün vd.	LR, GNB, DT, KNN, MLP, RF, XGBoost, LightGBM, CatBoost	CICDarknet2020	Doğruluk: %99.9, %94

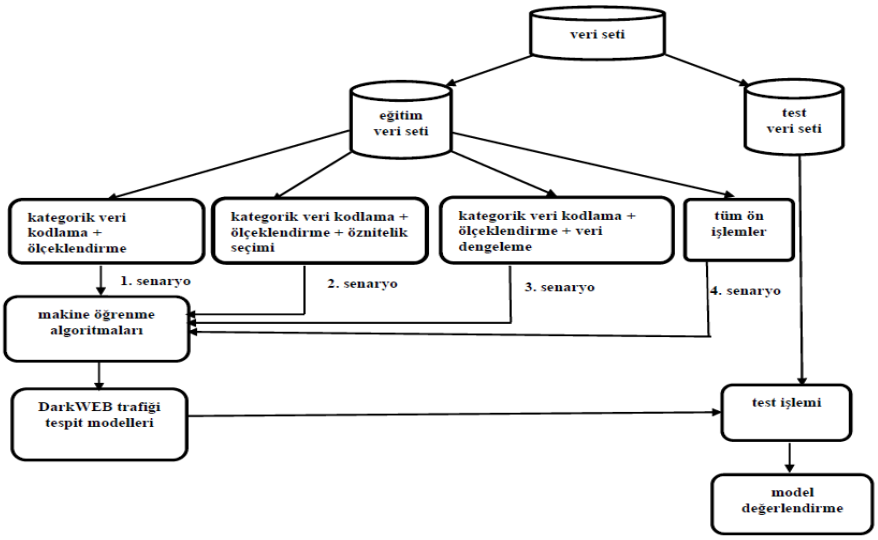
Tablo 1'de yer alan çalışmalar incelendiğinde, CICDarknet2020 veri seti ile yapılan DarkWEB trafiği tespit ve sınıflandırılması çalışmalarında, çoğunlukla geleneksel makine öğrenme yöntemlerinin kullanıldığı, %89-%99.8 doğruluk oranında sınıflandırma başarısına ulaşıldığı görülmektedir.

2. Önerilen Metodoloji (Recommended Methodology)

Bu çalışmada, DarkWEB trafiğinin yüksek doğrulukta tespiti ve sınıflandırması için üç aşamalı bir metodoloji önerilmiştir:

1) Veri setinin ön işlenmesi;
2) Ön işlenmiş veri setleri ve makine öğrenme algoritmaları ile saldırı tespit ve sınıflandırma modelleri oluşturulması;
3) Modellerin değerlendirilmesi;

Çalışmada önerilen metodoloji, Şekil 1'de gösterildiği gibi dört farklı senaryo ile uygulanmıştır.

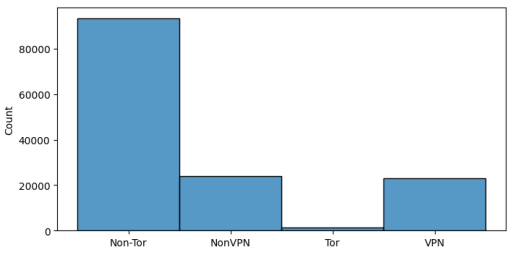


Şekil 1. Önerilen metodolojinin akış diyagramı
(Flowchart of the proposed methodology)

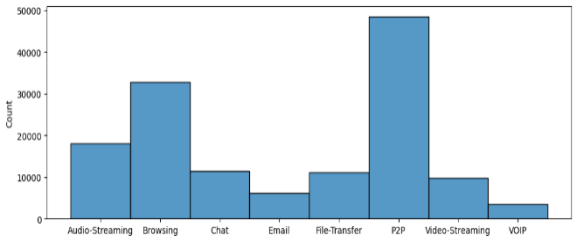
İlk senaryoda CIC-Darknet2020 veri setindeki kategorik veriler, label encoder ile kodlanmış ve veri setine min-max ölçeklendirme ön işlemi uygulanmıştır. 2. senaryoda 1. senaryonun uygulanması ile elde edilen veri setine, XGBoost gömülü öznitelik seçim yöntemi ile öznitelik seçimi ön işlemi uygulanmıştır. 3.senaryoda 1. senaryonun uygulanması ile elde edilen veri setine, ROS yöntemi ile veri dengeleme işlemi uygulanmıştır. 4. senaryoda 1. senaryonun uygulanması ile elde edilen veri setine, veri dengeleme işleminin ardından öznitelik seçimi ön işlemi uygulanmıştır. 4 farklı senaryonun uygulanması ile elde edilen veri setleri kullanılarak LR, GNB, KNN, MLP, DT, RF, XGBoost, LightGBM ve CatBoost makine öğrenme algoritmaları eğitilmiş ve çok sayıda saldırı tespit ve sınıflandırma modeli oluşturulmuştur, son olarak modellerin performansları test veri seti kullanılarak karşılaştırılmıştır.

3.1. CIC-Darknet2020 veri seti (CIC-Darknet2020 dataset)

Bu çalışmada kullanılan veri seti, ISCXTor2016 ve ISCXVPN2016 veri setlerinin bir karışımı olan CIC-Darknet2020 veri setidir [18]. Veri seti toplam 141530 örnek, 83 öznitelik ve 2 sınıf içermektedir. Üst sınıfta Tor, Non-Tor, Non-VPN, VPN olmak üzere 4 farklı trafik türü vardır. Alt sınıf ise üst sınıftaki trafik türlerinin elde edildiği uygulama türlerini içermektedir. Bunlar: Audio-Stream, Browsing, Chat, E-mail, P2P, Transfer, Video-Stream, VOIP'tir. Şekil 2, trafik türlerinin; Şekil 3 ise uygulama türlerinin veri setindeki dağılımını göstermektedir.



Şekil 2. Trafik türlerinin veri setindeki dağılımı
(Distribution of traffic types in the dataset)



Şekil 3. Uygulama türlerinin veri setindeki dağılımı
(Distribution of application types in the dataset)

Şekiller incelendiğinde trafik ve uygulama türlerinin dengesiz dağıldığı, özellikle trafik türlerinden Non-Tor sınıfının baskın, Tor sınıfının ise çok az sayıda olduğu görülmektedir.

3.2. Ön işlemler (pre-treatments)

Veri setlerindeki eksik değerli veriler, sınıfların dengesiz dağılması, kategorik değerli veriler, farklı aralıklardaki veri değerleri, veri setini temsil etmede yetersiz öznitelikler, makine öğrenme algoritmalarının sınıflandırma performansını düşürebilir [19]. Bu çalışmada, kullanılan makine öğrenme algoritmalarından en üst düzeyde performans elde edebilmek için CIC-Darknet2020 veri setine veri temizleme ve düzenleme, kategorik veri kodlama, ölçeklendirme, öznitelik seçimi ve veri dengeleme ön işlemleri uygulanmıştır.

Veri temizleme ve düzenleme: CIC-Darknet2020 veri seti, üst sınıfta 4 trafik türünü (Tor, Non-Tor, Non-VPN, VPN), alt sınıfta bu trafik türlerinin elde edildiği 8 uygulama türünü (Audio-Stream, Browsing, Chat, E-mail, P2P, Transfer, Video-Stream, VOIP), 83 özniteliği ve 141530 örneği içermektedir. Eksik değer içeren veriler, tüm verilerin yaklaşık %0.0004 olduğu ve bu verilerin kaldırılmasının eğitim süreci üzerinde kayda değer bir etkisi olmayacağı düşünüldüğü için veri setinden kaldırılmıştır. Yine her örnek için farklı değerlere sahip ve sınıflandırma performansına herhangi bir etkisi olmadığı düşünülen Flow ID özniteliği; her örnekte aynı formatta olmadığı için ön işlenemeyen Timestamp özniteliği veri setinden kaldırılmıştır. Veri setinden mümkün olduğunca fazla bilgi elde etmek için Source IP, Destination IP öznitelikleri ise octet'lerine ayrılmıştır. Ayrıca ikili sınıflandırma için veri setine Non-Tor ve Non-VPN trafik türlerinin 0 (standart), Tor ve VPN trafik türlerinin ise 1 (şifreli) olarak kodlandığı Encryption sınıfı eklenmiştir. Şekil 4 ve Şekil 5, şifreli ve standart trafiğe ait uygulamaların veri setindeki dağılımını göstermektedir. CIC-Darknet2020 veri setine uygulanan bu düzenlemeler sonunda 2, 4 ve 8 değişkene sahip üç sınıftan (Category, Subcategory, Encryption), 87 öznitelik ve 141481 örnekten oluşan bir veri seti elde edilmiştir.

Kategorik veri kodlama: Çoğu makine öğrenme algoritması kategorik değere sahip verileri, aralarında matematiksel ya da mantıksal bir ilişki bulunmadığı için işleyemez. Bu nedenle kategorik değerler sayısal değerlere dönüştürülmelidir. Bu çalışmada CIC-Darknet2020 veri setindeki kategorik değerli veriler, az sayıda değişkene sahip oldukları için en klasik yöntem olan label encoding ile sayısal formata dönüştürülmüştür.

Ölçeklendirme: Çalışmada, CIC-Darknet2020 veri setindeki farklı aralıklardaki veri değerlerinin ortak bir ölçekte eşleştirilerek söz konusu verilerin daha objektif karşılaştırılabilmesi için en küçük değere sahip veri 0, en büyük değere sahip veri 1 olacak şekilde, diğer bütün veri değerlerinin bu 0-1 aralığına yayıldığı min-max ölçeklendirme yöntemi kullanılmıştır.

Öznitelik seçimi: Ağ trafiğinde kullanılan protokolü ve uygulamaları ayırt etmek, trafik türlerini tanımlamak ya da sınıflandırmak için paket boyutu ve paket varış aralığı gibi öznitelikler belirleyicidir. Sınıflandırmaya en fazla katkısı olan özniteliklerin seçimi ile daha hızlı, daha doğru trafik türü tespiti ve sınıflandırması yapılabilir. Çalışmada, veri setini temsil etmede yetersiz, algoritmanın sınıflandırma performansını düşürecek öznitelikleri veri setinden çıkarmak için gömülü öznitelik seçimi yöntemi olan XGBoost kullanılmıştır. Yöntemin seçilmesinin nedeni, gömülü yöntemlerin aşırı öğrenmeye, sarmal ve filtreleme yöntemlerine kıyasla daha az eğilimli olması ve XGBoost ile hesaplama maliyetinin düşük olması, topluluk öğrenme algoritmaları içerisinde ve çevrim içi veri bilimcileri ile makine öğrenme uygulayıcıları topluluğu olan kaggle'ın düzenlediği yarışmalarda açık ara başarılı olmasıdır [20]. XGBoost gömülü öznitelik seçim yöntemi, öznitelikleri önemlerine göre sıralamakta ve önem değerleri için bir eşik değeri belirlenerek bu eşik değerinin üzerinde önem derecesine sahip öz nitelikler seçilebilmektedir. Fakat eşik değerinin ne olacağı bilgisi açık olmadığı için çalışmada, en iyi başarı oranlarını veren eşik değerleri, deneme yanılma yoluyla belirlenmiştir [21].

Veri dengeleme: Dengesiz sınıflandırma, veri setindeki örneklerin sınıflar arasındaki dağılımının eşit olmadığı bir sınıflandırma problemidir. Sınıf dağılımı dengesiz bir veri seti ile sınıflandırma modeli oluşturmak, modelde yanlılığa neden olarak modelin azınlık sınıfının aleyhine tahmin performansı ile sonuçlanır ki çoğu sınıflandırma probleminde ana amaç azınlık sınıfın doğru sınıflandırılmasıdır. Bu çalışmada da kullanılan veri setinde sınıflar arasında net bir dengesizlik söz konusudur ve bu sorunun çözümü için %70-%30 eğitim ve test veri seti olarak ayrılan CIC-Darknet2020 veri setinin eğitim verilerine, ROS veri dengeleme yöntemi uygulanmıştır. ROS yöntemi ile azınlık sınıfından örnekler rastgele kopyalanır ve eğitim veri setine eklenir, bu şekilde azınlık sınıfından örnekler eğitim veri setine birden çok kez eklenebilir. Bu yöntem ile bir sınıflandırıcı, görünüşte doğru olan ancak aslında çoğaltılmış örnekleri kapsayan kurallar oluşturabilir [22]. Tüm bunlar kurulan modelin aşırı öğrenme olasılığını artırabilir. Bu nedenle, yöntemin etkisi hakkında bilgi edinmek için, yüksek hızda örneklemeden sonra hem eğitim hem de test veri setlerindeki performansı izlemek ve sonuçları orijinal veri setindeki aynı algoritmayla karşılaştırmak önerilmiştir [23].

Çalışmada ROS yöntemi kullanılırken bu öneriler dikkate alınmıştır. Ayrıca ROS yönteminde "sampling strategy" parametresi ile ikili sınıflandırma problemleri için azınlık sınıfın, çoğunluk sınıfının ne kadar fazlasının örneklenmesinin yapılacağını belirlemek mümkündür. Fakat bu çalışmada 2'li sınıflandırma ile birlikte 4'lü ve 8'li sınıflandırmalar yapıldığı için bu öneri uygulanamamıştır. Eğitim veri setinde veri dengeleme ön işleminden önce ve sonra Encryption, Category, Subcategory sınıflarına ait değişkenlerin örnek sayısı Tablo 2 'de gösterilmiştir.

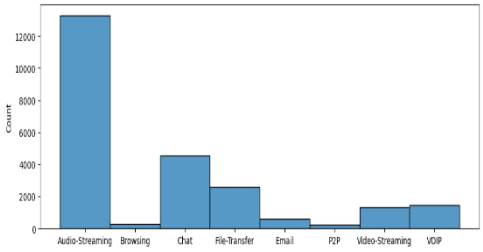
Tablo 2. ROS öncesi ve sonrası Encryption, Category, Subcategory sınıflarına ait değişkenlerin örnek sayısı
(Number of examples of variables belonging to Encryption, Category, Subcategory classes before and after ROS)

Encryption sınıfı	**ROS öncesi örnek sayısı**	**ROS sonrası örnek sayısı**
Şifreli (0)	82018	82018
Standart (1)	17018	82018

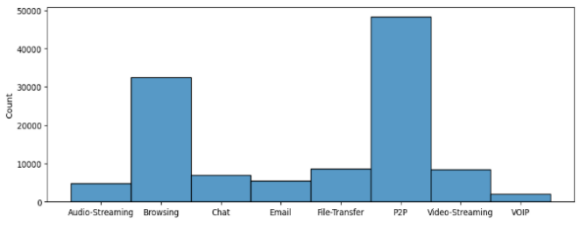
Category sınıfı	**ROS öncesi örnek sayısı**	**ROS sonrası örnek sayısı**
Non-Tor (0)	93309	65180
NonVPN (1)	23861	65180
Tor (2)	1392	65180
VPN (3)	22919	65180

Subcategory sınıfı	**ROS öncesi örnek sayısı**	**ROS sonrası örnek sayısı**
Audio-Streaming (0)	18050	33903
Browsing (1)	32808	33903
Chat (2)	11473	33903
Email (3)	6143	33903
File-Transfer (4)	11173	33903
P2P (5)	48520	33903
VOIP (6)	3566	33903
Video-Streaming (7)	9748	33903

Şekil 4 ve Şekil 5'de CIC-Darknet2020 veri setindeki trafik türlerinin üretildiği uygulama türlerinin şifreli ve standart trafiğe ait dağılımları gösterilmektedir.



Şekil 4. Şifreli trafiğe ait uygulama türlerinin dağılımı
(Distribution of application types for standard traffic)



Şekil 5. Standart trafiğe ait uygulama türlerinin dağılımı
(Distribution of application types for encrypted traffic)

Şekil 4 ve 5'e göre veri setinde şifreli ağ trafiğini en çok içeren uygulama türü Audio-Streaming, en az içeren uygulama türü Browsing ve P2P'tir. Standart trafikte ise durum nerdeyse tam tersidir. Bu durum uygulama türlerinin şifreli ya da standart trafiğe göre değişiklik gösterdiğini ve uygulama türüne göre sınıflandırma yapılmasının anlamlı olabileceğini göstermektedir.

3.3. DarkWEB trafiği tespit ve sınıflandırma modellerinin oluşturulması (creation of DarkWEB traffic detection and classification models)

Önerilen metodolojinin 2. aşamasında LR, GNB, KNN, MLP, DT, RF, XGBoost, LightGBM ve CatBoost makine öğrenme algoritmaları, metodolojinin 1. aşamasında uygulanan ön işlemler sonucu elde edilen veri setleri ile eğitilerek çok sayıda DarkWEB trafiği tespit ve sınıflandırma modeli oluşturulmuştur. Oluşturulan modeller ile **Encryption** (Şifreli, Standart), **Category** (Tor, Non-Tor, Non-VPN, VPN)**, Subcategory** (Audio-Stream, Browsing, Chat, E-mail, P2P, Transfer, Video-Stream, VOIP) sınıfları olmak üzere 2'li, 4'lü, 8'li sınıflandırmalar yapılmıştır.

3.4. Modellerin değerlendirilmesi (evaluation of models)

Önerilen metodolojinin 3. aşamasında, ön işlenmiş veri setleri ile makine öğrenme algoritmaları kullanılarak oluşturulan modellerin performansları, test veri seti ile Tablo 3'te gösterilen karışıklık matrisinden elde edilen değerlendirme metrikleri ölçüt alınarak değerlendirilmiştir.

Tablo 3. Karışıklık matrisi (Confusion matrix)

Karışıklık matrisi	**Tahmin edilen pozitif (saldırı)**	**Tahmin edilen negatif (normal)**
Gerçek Pozitif	Doğru Pozitif (DP)	Yanlış Negatif (YN)
Gerçek Negatif	Yanlış Pozitif (YP)	Doğru Negatif (DN)

Doğruluk= (DP + DN) \ (DP + DN + YP + YN)
Kesinlik= DP \ (DP + YP)
Duyarlılık= DP \ (DP + YN)
F1-Ölçütü= 2 * ((kesinlik * duyarlılık) \ (kesinlik + duyarlılık))

4. Uygulama Sonuçları ve Tartışmalar (Implementation Results and Discussions)

Yapılan çalışmada önerilen metodoloji dört farklı senaryo ile uygulanmış, her senaryonun uygulanması sonucunda DarkWEB trafiği tespit ve sınıflandırma modelleri oluşturulmuş, test veri seti kullanılarak modellerin başarı durumları, doğruluk, kesinlik, duyarlılık ve F1-Ölçütü metrikleriyle analiz edilmiştir.

4.1. 1.Senaryo: Makine öğrenme algoritmalarının label encoding ve min-max ölçeklendirme ön işlemleri yapılan veri setleri üzerinde çalıştırılması (scenario 1: running machine learning algorithms on datasets with label encoding and min-max scaling pre-processing)

CIC-Darknet2020 veri setine label encoding ve min-max ölçeklendirme ön işlemleri uygulanarak LR, GNB, KNN, MLP, DT, RF, XGBoost, LightGBM ve CatBoost algoritmaları ile modeller oluşturulmuş ve test veri seti ile test edilmiştir. Birinci senaryonun uygulanması ile alınan sonuçlar (bkz. Tablo 4) değerlendirildiğinde; 2'li (şifreli-standart) ve 4'lü (trafik türü) sınıflandırmada KNN, MLP, DT, RF, XGBoost, LightGBM ve CatBoost ile oluşturulan modellerde dört metriğin değerinin de birbirine yakın ve oldukça yüksek olduğu, LR ve GNB ile kurulan modellerde ise performansın diğer yedi algoritma ile kurulan modellere göre daha düşük olduğu görülmektedir.

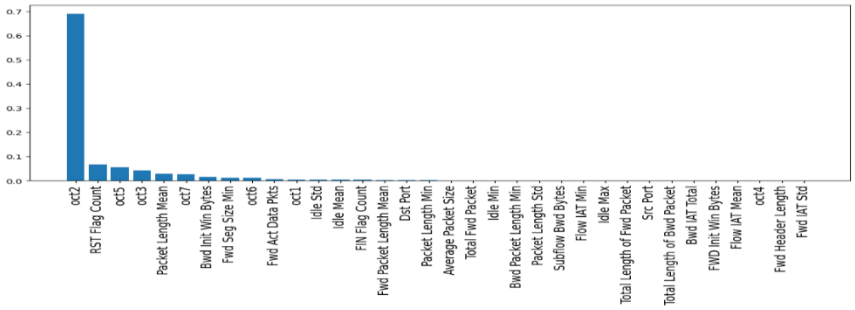
8'li (uygulama türü) sınıflandırmada ise beklenildiği gibi 2'li ve 4'lü sınıflandırmaya kıyasla performansın düştüğü, %94 ile en yüksek doğruluk değerinin XGBoost ile kurulan modelde, %32.6 ile en düşük doğruluk değerinin ise GNB ile kurulan modelde olduğu görülmüştür.

Algoritmalar	Doğruluk			Kesinlik			Duyarlılık			F1 -Ölçütü		
	Sınıf sayısı			Sınıf sayısı			Sınıf sayısı			Sınıf sayısı		
	2'li	4'lü	8'li	2'li	4'lü	8'li	2'li	4'lü	8'li	2'li	4'lü	8'li
XGBoost	0.999	0.999	0.940	0.999	0.999	0.940	0.999	0.999	0.940	0.999	0.999	0.940
LGBM	0.999	0.999	0.939	0.999	0.999	0.939	0.999	0.999	0.939	0.999	0.999	0.938
RF	0.999	0.999	0.931	0.999	0.999	0.931	0.999	0.999	0.931	0.999	0.999	0.931
CatBoost	0.999	0.999	0.930	0.999	0.999	0.930	0.999	0.999	0.930	0.999	0.999	0.929
DT	0.999	0.998	0.928	0.999	0.998	0.928	0.999	0.998	0.928	0.999	0.998	0.928
MLP	0.998	0.998	0.871	0.998	0.998	0.872	0.998	0.998	0.871	0.998	0.998	0.868
KNN	0.997	0.996	0.878	0.997	0.995	0.876	0.997	0.996	0.878	0.997	0.995	0.876
GNB	0.859	0.729	0.326	0.903	0.766	0.452	0.859	0.729	0.326	0.871	0.699	0.269
LR	0.986	0.973	0.725	0.986	0.973	0.709	0.986	0.973	0.725	0.986	0.973	0.709

Tablo 4. Senaryo 1: Label encoding+min-max ölçeklendirme ön işlemleri uygulanarak alınan sonuçlar
(Scenario 1: Results by applying label encoding+min-max scaling preprocesses)

4.2. 2. Senaryo: Makine öğrenme algoritmalarının label encoding, min-max ölçeklendirme ve öznitelik seçimi ön işlemleri yapılan veri setleri üzerinde çalıştırılması (scenario 2: running machine learning algorithms on datasets with label encoding, min-max scaling and feature selection pre-processing)

2. senaryoda 1. senaryonun uygulanması ile elde edilen veri setinden, XGBoost gömülü öznitelik seçim yöntemi ile öznitelikler seçilmiştir. Seçilen özniteliklerin, önem sıralaması Şekil 6'daki gibidir.



Şekil 6. XGBoost gömülü öznitelik seçim yönteminin öznitelik önem sıralaması
(Attribute importance ranking of XGBoost embedded feature selection method)

Önem katsayısı en yüksek öznitelikleri seçmek için deneme yanılma yoluna gidilerek, eşik değer 0.001 olarak belirlenmiş ve 87 öznitelik içerisinden 34'ü seçilmiştir. Seçilen öznitelikler Tablo 5'te yer almaktadır.

Tablo 5. XGBoost gömülü öznitelik seçim yöntemi ile seçilen öznitelikler
(Attributes selected with the XGBoost embedded feature selection method)

Seçilen öznitelikler
'Src Port', 'Dst Port', 'Total Fwd Packet', 'Total Length of Fwd Packet', 'Total Length of Bwd Packet', 'Fwd Packet Length Mean', 'Bwd Packet Length Min', 'Flow IAT Mean', 'Flow IAT Min', 'Fwd IAT Std', 'Bwd IAT Total', 'Fwd Header Length', 'Packet Length Min', 'Packet Length Mean', 'Packet Length Std', 'FIN Flag Count', 'RST Flag Count', 'Average Packet Size', 'Subflow Bwd Bytes', 'FWD Init Win Bytes', 'Bwd Init Win Bytes', 'Fwd Act Data Pkts', 'Fwd Seg Size Min', 'Idle Mean', 'Idle Std', 'Idle Max', 'Idle Min', 'oct1', 'oct2', 'oct3', 'oct4', 'oct5', 'oct6', 'oct7'

Label encoding, min-max ölçeklendirme ve öznitelik seçimi ön işlemleri uygulanan veri seti ve LR, GNB, KNN, MLP, DT, RF, XGBoost, LightGBM ve CatBoost algoritmaları ile oluşturulan modellerin test veri seti üzerindeki performansları Tablo 6'da yer almaktadır.

Tablo 6. Senaryo 2: Label encoding+min-max ölçeklendirme+öznitelik seçimi ön işlemleri uygulanarak alınan sonuçlar
(Scenario 2: Results obtained by applying label encoding + min-max scaling + feature selection pre-processes)

Algoritmalar	Doğruluk			Kesinlik			Duyarlılık			F1-Ölçütü		
	Sınıf sayısı			Sınıf sayısı			Sınıf sayısı			Sınıf sayısı		
	2'li	4'lü	8'li	2'li	4'lü	8'li	2'li	4'lü	8'li	2'li	4'lü	8'li
XGBoost	0.999	0.999	0.940	0.999	0.999	0.939	0.999	0.999	0.940	0.999	0.999	0.939
LGBM	0.999	0.999	0.937	0.999	0.999	0.937	0.999	0.999	0.937	0.999	0.999	0.936
RF	0.999	0.999	0.940	0.999	0.999	0.939	0.999	0.999	0.940	0.999	0.999	0.939
CatBoost	0.999	0.999	0.929	0.999	0.999	0.929	0.999	0.999	0.929	0.999	0.999	0.928
DT	0.999	0.999	0.930	0.999	0.999	0.930	0.999	0.999	0.930	0.999	0.999	0.930
MLP	0.997	0.998	0.860	0.997	0.998	0.861	0.997	0.998	0.860	0.997	0.998	0.857
KNN	0.997	0.996	0.878	0.997	0.996	0.876	0.997	0.996	0.878	0.997	0.996	0.876
GNB	0.968	0.907	0.389	0.968	0.927	0.520	0.968	0.907	0.389	0.968	0.914	0.310
LR	0.980	0.964	0.698	0.980	0.964	0.683	0.980	0.964	0.698	0.980	0.964	0.679

Tablo 6'daki sonuçlar değerlendirildiğinde 2'li sınıflandırmada KNN, MLP, DT, RF, XGBoost, LightGBM ve CatBoost ile oluşturulan modellerde dört metriğin değerinde 1.senaryoya göre pek bir değişiklik olmamıştır. GNB ile kurulan modelde ise doğruluk değeri %85.9'dan %96.8 'e yükselmiştir. 4'lü sınıflandırmada KNN, MLP, DT, RF, XGBoost, LightGBM ve CatBoost ile oluşturulan modellerde dört metriğin değerinde 1.senaryoya göre pek bir değişiklik olmamıştır. GNB ile kurulan modelde ise doğruluk değeri %72.9'dan %90.7 'ye yükselmiş, LR ile kurulan modelde ise %97.3 'ten %96.4'e küçük bir düşüş yaşanmıştır. 8'li sınıflandırmada KNN, DT, XGBoost, LightGBM ve CatBoost ile oluşturulan modellerde dört metriğin değerinin de 1.senaryoya göre pek bir değişiklik olmamıştır. RF ile kurulan modelde doğruluk değerinde %93.1'den %94'e bir yükseliş olmuştur. MLP ile kurulan modelde doğruluk değerinde %87.1'den %86'ya bir düşüş olmuştur. GNB ile kurulan modelde ise doğruluk değeri %32.6'dan %38.9 'a yükselmiş, LR ile kurulan modelde ise %72.5 'ten %69.8'e düşmüştür. Öznitelik seçimi ön işlemi, modellerin sınıflandırma performanslarında küçük çaplı yükseliş ve azalışlara neden olmakla beraber beklenen performans artışı sağlanamamıştır. Bu artışı gözlemleyene dek farklı öznitelik seçim yöntemleri ve farklı eşik değerleri denenebilir.

4.3. 3.Senaryo: Makine öğrenme algoritmalarının label encoding, min-max ölçeklendirme ve veri dengeleme ön işlemleri yapılan veri setleri üzerinde çalıştırılması (scenario 3: running machine learning algorithms on datasets with label encoding, min-max scaling and data balancing pre-processing)

3. senaryoda CIC-Darknet2020 veri setine label encoding, min-max ölçeklendirme ve veri dengeleme ön işlemleri uygulanarak LR, GNB, KNN, MLP, DT, RF, XGBoost, LightGBM ve CatBoost algoritmaları ile modeller oluşturulmuş ve test veri seti ile test edilmiştir. 3.senaryonun uygulanması ile alınan sonuçlar (bkz. Tablo 7) değerlendirildiğinde; 2'li sınıflandırmada KNN, MLP, DT, RF, XGBoost, LightGBM ve CatBoost ile oluşturulan modellerde dört metriğin değerinde 1.senaryoya göre pek bir değişiklik olmamıştır. GNB ile kurulan modelde ise doğruluk değerinde %1.8'lik, LR ile kurulan modelde ise %0.4'lük bir düşüş görülmüştür. 4'lü sınıflandırmada KNN, MLP, DT, RF, XGBoost, LightGBM ve CatBoost ile oluşturulan modellerde dört metriğin değerinde 1.senaryoya göre pek bir değişiklik olmamıştır. GNB ile kurulan modelde ise doğruluk değerinde %1.8'lik bir yükseliş, LR ile kurulan modelde ise %1.7'lik bir düşüş görülmüştür.

Tablo 7. Senaryo3: Label encoding+min-max ölçeklendirme+veri dengeleme ön işlemleri uygulanarak alınan sonuçlar

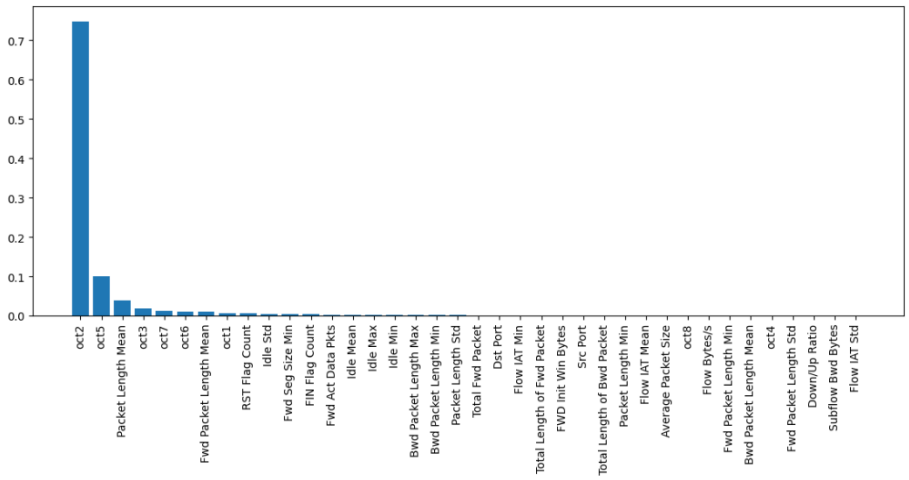
Algoritmalar	Doğruluk			Kesinlik			Duyarlılık			F1 -Ölçütü		
	Sınıf sayısı			Sınıf sayısı			Sınıf sayısı			Sınıf sayısı		
	2'li	4'lü	8'li	2'li	4'lü	8'li	2'li	4'lü	8'li	2'li	4'lü	8'li
XGBoost	0.999	0.999	0.936	0.999	0.999	0.942	0.999	0.999	0.936	0.999	0.999	0.937
LGBM	0.999	0.999	0.933	0.999	0.999	0.940	0.999	0.999	0.933	0.999	0.999	0.934
RF	0.999	0.998	0.934	0.999	0.998	0.936	0.999	0.998	0.934	0.999	0.998	0.934
CatBoost	0.999	0.999	0.928	0.999	0.999	0.935	0.999	0.999	0.928	0.999	0.999	0.929
DT	0.999	0.998	0.928	0.999	0.998	0.928	0.999	0.998	0.928	0.999	0.998	0.928
MLP	0.996	0.995	0.862	0.995	0.996	0.877	0.995	0.995	0.862	0.995	0.995	0.865
KNN	0.995	0.996	0.889	0.997	0.996	0.894	0.997	0.996	0.889	0.997	0.996	0.891
GNB	0.841	0.747	0.325	0.895	0.796	0.463	0.841	0.747	0.325	0.855	0.736	0.269
LR	0.982	0.956	0.687	0.983	0.961	0.730	0.982	0.956	0.687	0.982	0.958	0.697

(Scenario 3: Results obtained by applying label encoding + min-max scaling + data balancing pre-processes)

8'li sınıflandırmada XGBoost ile kurulan modelin doğruluğunda %0.4'lük bir düşüş, LGBM'de %0.6'lık bir düşüş, CatBoost'ta %0.2'lik bir düşüş, MLP'de %0.9'luk bir düşüş, RF'de %0.3'lük bir yükseliş, LR'de %3.8'lik bir düşüş, KNN'de %1.1'lik bir yükseliş görülmüştür. DT ve GNB'de pek bir değişiklik gözlemlenmemiştir. Özetle 3. senaryonun uygulanması, özellikle veri dengeleme işlemi, modellerin 8'li sınıflandırma performanslarında önemli bir değişikliğe neden olmamıştır. 3.senaryonun uygulanması ile modellerin performanslarında genel anlamda küçük çaplı düşüşler olduğu görülmüştür.

4.4. 4.Senaryo: Makine öğrenme algoritmalarının label encoding, min-max ölçeklendirme, veri dengeleme ve öznitelik seçimi ön işlemleri yapılan veri setleri üzerinde çalıştırılması (scenario 4: running machine learning algorithms on data sets with label encoding, min-max scaling, data balancing and feature selection pre-processing.)

4. senaryoda label encoding, min-max ölçeklendirme, veri dengeleme ön işlemleri uygulanan veri setinden XGBoost gömülü öznitelik seçim yöntemi ile öznitelik seçimi yapılmıştır. Seçilen öznitelikler Şekil 7'de görülmektedir.



Şekil 7. Veri dengeleme ön işlemi yapılan veri setinde XGBoost yönteminin öznitelik önem sıralaması
(Attribute importance ranking of the XGBoost method in the data set with data balancing pre-processing)

Önem katsayısı en yüksek öznitelikleri seçmek için deneme yanılma yoluna gidilerek eşik değer 0.001 olarak belirlenmiş ve 87 öznitelik içerisinden 38'i seçilmiştir. Seçilen öznitelikler Tablo 8'de yer almaktadır.

Tablo 8. Label encoding min-max ölçeklendirme veri dengeleme ön işlemleri uygulanan veri setinden XGBoost gömülü öznitelik seçim yöntemi ile seçilen öznitelikler (Attributes selected with the XGBoost embedded feature selection method from the dataset with label encoding min-max scaling data balancing pre-processing applied.)

Seçilen öznitelikler
'Src Port', 'Dst Port', 'Total Fwd Packet', 'Total Length of Fwd Packet', 'Total Length of Bwd Packet', 'Fwd Packet Length Min', 'Fwd Packet Length Mean', 'Fwd Packet Length Std', 'Bwd Packet Length Max', 'Bwd Packet Length Min', 'Bwd Packet Length Mean', 'Flow Bytes/s','Flow IAT Mean', 'Flow IAT Std', 'Flow IAT Min', 'Packet Length Min', 'Packet Length Mean', 'Packet Length Std', 'FIN Flag Count', 'RST Flag Count', 'Down/Up Ratio', 'Average Packet Size', 'Subflow Bwd Bytes', 'FWD Init Win Bytes', 'Fwd Act Data Pkts', 'Fwd Seg Size Min', 'Idle Mean', 'Idle Std', 'Idle Max', 'Idle Min', 'oct1', 'oct2', 'oct3', 'oct4', 'oct5', 'oct6', 'oct7', 'oct8'

4.senaryonun uygulanması ile elde edilen modellerin test veri seti üzerindeki performansları Tablo 9'da yer almaktadır.

Tablo 9. Senaryo 4: Tüm ön işlemlerin uygulanması ile alınan sonuçlar
(Scenario 4: Results obtained by applying all pre-treatments)

Algoritmalar	Doğruluk			Kesinlik			Duyarlılık			F1 -Ölçütü		
	Sınıf sayısı			Sınıf sayısı			Sınıf sayısı			Sınıf sayısı		
	2'li	4'lü	8'li	2'li	4'lü	8'li	2'li	4'lü	8'li	2'li	4'lü	8'li
XGBoost	0.999	0.999	0.934	0.999	0.999	0.939	0.999	0.999	0.934	0.999	0.999	0.935
LGBM	0.999	0.999	0.931	0.999	0.999	0.939	0.999	0.999	0.931	0.999	0.999	0.932
RF	0.999	0.999	0.940	0.999	0.999	0.942	0.999	0.999	0.940	0.999	0.999	0.941
CatBoost	0.999	0.999	0.925	0.999	0.999	0.933	0.999	0.999	0.925	0.999	0.999	0.926
DT	0.999	0.998	0.930	0.999	0.998	0.930	0.999	0.998	0.930	0.999	0.998	0.930
MLP	0.997	0.993	0.861	0.997	0.995	0.879	0.997	0.993	0.861	0.997	0.993	0.861
KNN	0.996	0.997	0.887	0.996	0.997	0.892	0.996	0.997	0.887	0.996	0.997	0.888
GNB	0.954	0.879	0.467	0.957	0.913	0.553	0.954	0.879	0.467	0.955	0.892	0.420
LR	0.975	0.940	0.654	0.977	0.950	0.703	0.975	0.940	0.654	0.976	0.944	0.667

4.senaryonun uygulanması ile alınan sonuçlar (bkz. Tablo 9) değerlendirildiğinde; 2'li sınıflandırmada KNN, MLP, DT, RF, XGBoost, LightGBM ve CatBoost ile oluşturulan modellerde dört metriğin değerinde 1.senaryoya göre pek bir değişiklik olmamıştır. GNB ile kurulan modelde ise doğruluk değerinde %9.5'lik bir yükseliş, LR ile kurulan modelde ise %1.1'lik bir düşüş görülmüştür. 4'lü sınıflandırmada KNN, DT, RF, XGBoost, LightGBM ve CatBoost ile oluşturulan modellerde dört metriğin değerinde 1.senaryoya göre pek bir değişiklik olmamıştır. MLP'de %0.5'lik, LR ile kurulan modelde ise %3.3'lük bir düşüş görülmüştür. GNB ile kurulan modelde ise doğruluk değerinde %15'lik bir yükseliş, görülmüştür. 8'li sınıflandırmada XGBoost ile kurulan modelin doğruluk değerinde %0.6'lık, LGBM'de %0.8'lik, CatBoost'ta %0.5'lik, MLP'de %1'luk, LR'de %7.1'lik düşüşler gözlemlenirken; RF'de %0.9'luk, DT'de %0.2'lik, KNN'de %0.9'luk bir yükseliş, GNB'de % 14.1'lik yükseliş gözlemlenmiştir.

Dört senaryonun uygulanması sonucunda, modellerin 2'li ve 4'lü sınıflandırmada doğruluk oranlarında pek bir değişiklik olmamıştır ve senaryolar boyunca doğruluk oranı KNN, MLP, DT, RF, XGBoost, LightGBM ve CatBoost ile oluşturulan modellerde %99.9-%99.3 arasında bir değer almıştır. Fakat 8'li sınıflandırmada senaryolar boyunca modellerin doğruluk oranlarında değişkenlik gözlemlenmiştir. Senaryolar boyunca 8'li sınıflandırmada modellerin doğruluk metriklerindeki değişimin görüldüğü Tablo 10'daki sonuçlar değerlendirildiğinde: XGBoost'un %94 doğruluk oranı ile en başarılı olduğu senaryolar 1. ve 2. senaryolardır. LGBM'in %93.9 doğruluk oranı ile en başarılı olduğu senaryo 1.senaryodur. RF'ın %94 doğruluk oranı ile en başarılı olduğu senaryolar 2. ve 4. senaryolardır. CatBoost'un %93 doğruluk oranı ile en başarılı olduğu senaryo 1. senaryodur. DT'nin %93 doğruluk oranı ile en başarılı olduğu senaryo 2. ve 4. senaryolardır. MLP'nin %87.1 doğruluk oranı ile en başarılı olduğu senaryo 1.senaryodur. KNN'nin %88.9 doğruluk oranı ile en başarılı olduğu senaryo 3.senaryodur. GNB'nin %46.7 doğruluk oranı ile en başarılı olduğu senaryo 4.senaryodur. LR'nin %72.5 doğruluk oranı ile en başarılı olduğu senaryo 1.senaryodur.

Tablo 10. 8'li sınıflandırmada senaryolar boyunca modellerin sınıflandırma doğruluğu değişimleri
(Classification accuracy changes of models across scenarios in 8-class classification)

Algoritmalar	1.senaryo	2.senaryo	3.senaryo	4.senaryo
XGBoost	**%94**	**%94**	%93.6	%93.4
LGBM	**%93.9**	%93.7	%93.3	%93.1
RF	%93.1	**%94**	%93.4	**%94**
CatBoost	**%93**	%92.9	%92.8	%92.5
DT	%92.8	**%93**	%92.8	**%93**
MLP	**%87.1**	%86	%86.2	%86.1
KNN	%87.8	%87.8	**%88.9**	%88.7
GNB	%32.6	%38.9	%32.5	**%46.7**
LR	**%72.5**	%69.8	%68.7	%65.4

Tablo 10 ile ilgili genel bir çıkarım yapılırsa, 8'li sınıflandırmada modellerin sınıflandırma doğruluğu metriğinde, en yüksek değerlere 1. ve 4. senaryoların uygulanması sonucunda ulaşılmıştır.

Senaryolar boyunca 8'li sınıflandırmada modellerin, dengesiz veri setleri ile yapılan çalışmalarda dikkate alınması gereken bir metrik olan F1-Ölçütü'ndeki değişimin görüldüğü Tablo 11'deki sonuçlar değerlendirildiğinde ise: XGBoost'un %94 doğruluk oranı ile en başarılı olduğu senaryo 1. senaryodur.

LGBM'in %93.8 doğruluk oranı ile en başarılı olduğu senaryo da 1.senaryodur. RF'ın %94.1 doğruluk oranı ile en başarılı olduğu senaryo 4. senaryodur. CatBoost'un %92.9 doğruluk oranı ile en başarılı olduğu senaryo 1. senaryodur. DT'nin %93 doğruluk oranı ile en başarılı olduğu senaryo 2. ve 4. senaryolardır. MLP'nin %86.8 doğruluk oranı ile en başarılı olduğu senaryo 1.senaryodur. KNN'nin %89.1 doğruluk oranı ile en başarılı olduğu senaryo 3.senaryodur. GNB'nin %42 doğruluk oranı ile en başarılı olduğu senaryo 4.senaryodur. LR'nin %70.9 doğruluk oranı ile en başarılı olduğu senaryo 1.senaryodur.

Tablo 11. 8'li sınıflandırmada senaryolar boyunca modellerin F1-Ölçütü değişimleri
(F1-Criteria changes of models across scenarios in 8-class classification)

Algoritmalar	1.senaryo	2.senaryo	3.senaryo	4.senaryo
XGBoost	**0.940**	0.939	0.937	0.935
LGBM	**0.938**	0.936	0.934	0.932
RF	0.931	0.939	0.934	**0.941**
CatBoost	**0.929**	0.928	**0.929**	0.926
DT	0.928	**0.930**	0.928	**0.930**
MLP	**0.868**	0.857	0.865	0.861
KNN	0.876	0.876	**0.891**	0.888
GNB	0.269	0.310	0.269	**0.420**
LR	**0.709**	0.679	0.697	0.667

Tablo 11 ile ilgili genel bir çıkarım yapılırsa, 8'li sınıflandırmada modellerin F1-Ölçütü metriğinde, en yüksek değerlere 1. ve 4. senaryoların uygulanması sonucunda ulaşılmıştır. Özetle senaryolara göre modellerin 8'li sınıflandırmada F1-Ölçütü metriğindeki değişimleri ile doğruluk metriğindeki değişimleri benzerdir ve en başarılı sonuçlar, kategorik veri kodlama ve ölçeklendirme ön işlemlerinin uygulandığı 1.senaryo ile kategorik veri kodlama, ölçeklendirme, veri dengeleme ön işlemlerinin uygulandığı veri setinden XGBoost gömülü yöntemle öznitelik seçiminin yapıldığı, özetle tüm ön işlemlerin yapıldığı 4.senaryonun uygulanması ile elde edilmiştir.

5. Sonuçlar ve Öneriler (Conclusions and Recommendations)

Bu çalışmada, DarkWEB trafiğinin yüksek doğrulukta tespiti ve sınıflandırılması için veri setinin ön işlenmesi, ön işlenmiş veri setleri ve makine öğrenme algoritmaları ile DarkWEB trafiği tespit ve sınıflandırma modelleri oluşturulması, modellerin değerlendirilmesi olmak üzere üç aşamalı bir metodoloji önerilmiştir. Çalışmada önerilen metodoloji dört farklı senaryo ile uygulanmıştır. Metodolojinin 1.aşamasında, veri setine kategorik veri kodlama, ölçeklendirme, öznitelik seçimi ve veri dengeleme ön işlemleri ayrı ayrı ve birlikte uygulanarak 4 farklı senaryo için 4 farklı veri seti oluşturulmuştur. Metodolojinin 2.aşamasında, 1.aşamada oluşturulan veri setleri ve LR, GNB, KNN, MLP, DT, RF, XGBoost, LightGBM ve CatBoost makine öğrenme algoritmaları kullanılarak DarkWEB trafiği tespit ve sınıflandırma modelleri oluşturulmuştur. Oluşturulan modeller ile **Encryption** (Şifreli, Standart), **Category** (Tor, Non-Tor, Non-VPN, VPN)**, Subcategory** (Audio-Stream, Browsing, Chat, E-mail, P2P, Transfer, Video-Stream, VOIP) sınıfları olmak üzere 2'li, 4'lü, 8'li sınıflandırmalar yapılmıştır.
Metodolojinin 3.aşamasında oluşturulan modellerin performansları doğruluk, kesinlik, duyarlılık ve F1-Ölçütü metrikleri ile değerlendirilmiştir. 2'li ve 4'lü sınıflandırmada doğruluk oranları birbirine yakın ve KNN, MLP, DT, RF, XGBoost, LightGBM ve CatBoost ile oluşturulan modellerde %99.9-%99.3 arasında bir değer almıştır. 2'li sınıflandırmada GNB'de %96.8 ile en yüksek doğruluk oranı 2.senaryo ile elde edilmiştir. LR'de ise %98.6 ile en yüksek doğruluk oranı 1.senaryo ile elde edilmiştir. 4'lü sınıflandırmada, GNB'de %90.7 ile en yüksek doğruluk oranı 2.senaryo ile elde edilmiştir. LR'de ise %97.3 ile en yüksek doğruluk oranı, 1.senaryo ile elde edilmiştir. 8'li sınıflandırma ise en başarılı modeller, label encoding yöntemi ile kategorik veri kodlama, min-max ölçeklendirme ve XGBoost öznitelik seçimi ön işlemlerinin yapıldığı 2. senaryonun uygulanması sonucunda elde edilen %94 doğruluk oranıyla XGBoost ve RF ile kurulan modeller olmuştur. Veri dengelemenin öne çıktığı 3. senaryonun ve öznitelik seçimin öne çıktığı 2.senaryonun uygulanması ile beklenen performans artışı sağlanamamıştır. Bu sonuçların 3.senaryo için nedeninin, veri dengeleme için kullanılan ROS yönteminde, azınlık sınıfı örneklerinin tam kopyaları oluşturulduğu için kullanılan sınıflandırıcının, görünüşte doğru olan ancak aslında çoğaltılmış örnekleri kapsayan kurallar oluşturmuş ve sonuçta fazla uydurma olasılığını arttırmış olması sonucuna varılmıştır. Veri dengelemenin neden olduğu

aşırı öğrenmeyi azaltmak için GAN (Generative Adversarial Networks) gibi daha gelişmiş yöntemlerin kullanılması hedeflenmektedir. 2. senaryonun uygulanması ile beklenen performans artışını sağlamak için ise daha farklı öznitelik seçim yöntemleri ve eşik değerlerinin kullanılması önerilmektedir.

Gelecekte yapılacak çalışmalarda, DarkWEB trafiği tespit modellerinin performanslarınıcddö arttırmak için veri setine uygulanan ön işlemlerin kullanılacak her bir algoritmaya ve veri setine uygun seçilmesi, en doğru hiper-parametrelerin belirlenebilmesi için çok sayıda parametre kombinasyonunun denenmesine olanak sağlayacak kaynakların temin edilmesi, oluşturulacak modellerinin başarısının, güncel ve farklı veri setleri ile test edilerek güvenilirliği yüksek DarkWEB trafiği tespit ve sınıflandırma modelleri geliştirilmesi, ayrıca son yıllarda oldukça başarılı sonuçlar alınan derin öğrenme modellerinin kullanımı da hedeflenmektedir.

Çıkar Çatışması Beyanı (Conflict of Interest Statement)

Yazarlar tarafından herhangi bir çıkar çatışması bildirilmemiştir.

Kaynaklar (References)

[1] G. Weımann, "Going Darker? The Challenge of Dark Net Terrorism", wilsoncenter.org, [Online]. Available: https://www.wilsoncenter.org/sites/default/files/media/documents/publication/going_darker_challenge_of_dark_net_terrorism.pdf. [Accessed: Jun. 6, 2023].

[2] R. Badhwar, *The CISO's Next Frontier: Dark Web & Dark Net*, Springer Nature Switzerland AG 2021.

[3] K. Demertzis, K. Tsiknas, D. Takezis, C. Skianis and L. Iliadis, "Darknet traffic bigdata analysis and network management for real-time automating of the malicious intent detection process by a weight agnostic neural networks framework", *Electronics*, vo.10, no.7, pp.781, 2021. doi: 10.3390/electronics10070781

[4] A. Bracci, M.Nadini, M. Aliapoulios, D. McCoy, I. Gray, A. Teytelboym, A. Gallo and A. Baronchelli, "Dark Web Marketplaces and COVID-19: before the vaccine," *EPJ Data Sci*, vol.10, no. 6, 2021. doi: 10.1140/epjds/s13688-021-00259-w

[5] A.H. Lashkari, G. Kaur and A. Rahali, "DIDarknet: A Contemporary Approach to Detect and Characterize the Darknet Traffic using Deep Image Learning," *10th International Conference on Communication and Network Security*, *2020, Tokyo*, pp. 1-13, November, 2020.

[6] M. B. Sarwar, M. K. Hanif, R. Talib, M. Younas and M. U. Sarwar, "DarkDetect: Darknet Traffic Detection and Categorization Using Modified Convolution-Long Short-Term Memory," *in IEEE Access*, vol. 9, pp. 113705-113713, 2021, doi: 10.1109/ACCESS.2021.3105000.

[7] L. A. Iliadis and T. Kaifas, "Darknet Traffic Classification using Machine Learning Techniques," *2021 10th International Conference on Modern Circuits and Systems Technologies (MOCAST), Thessaloniki, July 2021, Greece* [Online]. Available: IEEE Xplore, https://ieeexplore.ieee.org/document/9493386. [Accessed: 10 Sept. 2023].

[8] S. Sridhar and S. Sanagavarapu, "DarkNet Traffic Classification Pipeline with Feature Selection and Conditional GAN-based Class Balancing," *2021 IEEE 20th International Symposium on Network Computing and Applications (NCA)*, Boston, MA, USA, 2021, [Online]. Available: IEEE Xplore, https://ieeexplore.ieee.org/document/9685743. [Accessed: 20 May. 2023].

[9] Y. Li, Y. Lu and S. Li, "EZAC: Encrypted Zero-day Applications Classification using CNN and K-Means," *2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, Dalian, China, 2021, [Online]. Available: IEEE Xplore, https://ieeexplore.ieee.org/document/9437716. [Accessed: 12 Feb. 2023].

[10] M. Ugurlu, İ. Dogru, ve R. S. Arslan, "Karanlık ağ trafiğinin makine öğrenmesi yöntemleri kullanılarak tespiti ve sınıflandırılması," *GUMMFD*, vol. 38, no. 3, pp. 1737–1746, 2023, doi: 10.17341/gazimmfd.1023147.

[11] N. Rust-Nguyen, S. Sharma, and M. Stamp, "Darknet traffic classification and adversarial attacks using machine learning," *Comput. Secur,* vol. 127, pp.16, 2023. doi: 10.1016/j.cose.2023.103098

[12] A. Almomani, "Darknet traffic analysis, and classification system based on modified stacking ensemble learning algorithms," *Inf Syst E-Bus Manage*, 2023. doi: 10.1007/s10257-023-00626-2

[13] H. Mohanty, A. H. Roudsari, and A. Habibi Lashkari, "Robust stacking ensemble model for darknet traffic classification under adversarial settings," *Comput. Secur,* vol.120, Sep. 2022. doi: 10.1016/j.cose.2022.102830

[14] Q. A. Al-Haija, M. Krichen and W. A. Elhaija, "Machine-Learning-Based Darknet Traffic Detection System for IoT Applications," *Electronics,* vol. 11, no.4, pp.556, 2022. doi:11. 556. 10.3390/electronics11040556.

[15] Y. Li and Y. Lu, " ETCC: Encrypted Two-Label Classification Using CNN," *Sec. and Commun. Netw.* vol.2021, pp.11, 2021. doi:10.1155/2021/6633250

[16] M. Alimoradi, M. Zabihimayvan, A. Daliri, R. Sledzik and R. Sadeghi, "Deep Neural Classification of Darknet Traffic," In book:

Artificial Intelligence Research and Development, Edition: printChapter: 356, Publisher: IOS Press, 2022, pp.105-114

[17] A. H. Lashkari, G. Draper Gil, M. Mamun and A. Ghorbani, "Characterization of Encrypted and VPN Traffic Using Time-Related Features," *The International Conference on Information Systems Security and Privacy (ICISSP),* Feb 2016, Italy, [Online]. Available: IEEE Xplore, https://doi.org/10.5220/0005740704070414. [Accessed: 10 Apr. 2023].

[18] A. H. Lashkari, G. Kaur and A. Rahali, "DIDarknet: A Contemporary Approach to Detect and Characterize the Darknet Traffic using Deep Image Learning," *10th International Conference on Communication and Network Security, November 2020, Tokyo, Japan*, [Online]. Available: https://doi.org/10.1145/3442520.3442521. [Accessed: 20 May. 2023].

[19] E. G. İlgün ve R. Samet, "Veri setine uygulanan ön işlemler ile makine öğrenimi yöntemi kullanılarak geliştirilen saldırı tespit modellerinin performanslarının arttırılması," GUMMFD, vol. 39, no. 2, pp. 679–692, 2023, doi: 10.17341/gazimmfd.1122021.

[20] E. G. İlgün, "Veri setine uygulanan ön işlemlerin anomali tabanlı saldırı tespit modellerinin performansları üzerindeki etkisinin incelenmesi," Yüksek Lisans Tezi, Ankara Üniversitesi, Ankara, Türkiye, 2022.

[21] O. Kaynar, H. Arslan, Y. Görmez ve Y. E. Işık, "Makine Öğrenmesi ve Öznitelik Seçim Yöntemleriyle Saldırı Tespiti," *Bilişim Teknolojileri Dergisi*, 11 (2), pp.175-185, 2018. doi: 10.17671/gazibtd.368583

[22] A. Fernandez, S. Garcia, M. Galar, R.C. Prati, B. Krawczyk and F. Herrera, "Learning from Imbalanced Data Sets," *Cambridge International Law Journal,* pp. 83, 2018. doi:10.1007/978-3-319-98074-4

[23] J. Brownlee, "Random Oversampling and Undersampling for Imbalanced Classification," *machinelearningmastery.com*, Jan. 15, 2020. [Online]. Available: https://machinelearningmastery.com/random-oversampling-and-undersampling-for-imbalanced-classification/on. [Accessed: 12 Apr. 2023].

** This paper was presented at the 5th International Conference on Artificial Intelligence and Applied Mathematics in Engineering (ICAIAME 2023) and the abstract was published as an e-book.*