Journal of Computer Science
https://dergipark.org.tr/en/pub/bbd

ISSN,e-ISSN: 2548-1304
Volume:IDAP-2022, pp:151-164, 2022
https://doi.org/ 10.53070/bbd.1173687
Research Paper

IoT Botnet Verisetlerinin Karşılaştırmalı Analizi

Comparative Analysis of IoT Botnet Datasets

Esin Gül ÖLMEZ[*1], Kenan İNCE[*1]

[1]Bilgisayar Mühendisliği Bölümü, İnönü Üniversitesi, Malatya, TÜRKİYE

(esinkarabacakoglu@gmail.com, kenanince@gmail.com)

Received:Sep.08,2022 Accepted:Sep.16,2022 Published:Oct.10,2022

Özetçe— Günümüzde IoT teknolojilerinin kullanımının yaygınlaşması birçok güvenlik sorunlarını da beraberinde getirmiştir. IoT cihazları çeşitli saldırıların hedefi haline gelmiştir. Bu saldırılarda en sık karşılaşılan tür botnet saldırılarıdır. IoT cihazlarda bu saldırıların sürekli çeşitlenerek gelişmesi ve donanımlarının kısıtlı olması sebebiyle geleneksel savunma yöntemlerinin uygulanamaması bu alanda yeni çalışmalara sebep olmuştur. Cihazlara yapılan saldırıların en kısa sürede tespit edilmesi, türlerine göre sınıflandırma yapılması güncel çalışmaların popüler konusu haline gelmiştir. Makine öğrenmesi yöntemleriyle sıfır gün saldırılarını tespit edip sınıflandırmak iyi bir yöntemdir. Yapılan bu çalışmada denetimli makine öğrenme yöntemlerinden Destek Vektör Makineleri (SVM) ile bir model oluşturulmuştur. Literatürde çokça kullanılan ve özellikle hem IoT botnet saldırı kayıtlarını hem de normal kayıt türlerini içeren verisetleri incelenmiştir. Bu veri setlerinden en uygun dört veriseti (Bot-IoT, CICIDS-2017, IoT-23 ve N-BaIoT) modelimiz üzerinde kullanılarak karşılaştırılmıştır. Yapılan değerlendirme sonucunda Bot-IoT veri seti için %99.94, CICIDS-2017 veri seti için %99.95, IoT-23 veri seti için %99.96 ve N-BaIoT veri seti için %99.92 oranında doğruluk değerlerine ulaşılmıştır. Bu sonuçlar değerlendirildiğinde makine öğrenme yöntemleri ile yapılan saldırı tespit ve sınıflandırma işlemlerinde seçmiş olduğumuz veri setlerinin kullanımının uygun olduğu görülmektedir.

Anahtar Kelimeler : *Nesnelerin İnterneti, Botnet Saldırıları, Saldırı Tespit Sistemleri, Destek Vektör Makineleri*

Abstract— Today, the widespread use of IoT technologies cause many security problems. IoT devices have became the target of various attacks. The most common type of these attacks are botnet attacks. The continuous diversification and development of these attacks on IoT devices and the inability to apply traditional defense methods due to the limited hardware have led to researchers to search for new solutions in this area. Detecting attacks on devices as soon as possible and classifying them according to their types has become a popular subject of current studies. It is a good method to detect and classify zero-day attacks using machine learning methods. In this study, a model was created with Support Vector Machines (SVM), one of the supervised machine learning methods. The datasets that are widely used in the literature and especially contain both IoT botnet attack records and benign record types have been examined. The four most appropriate data sets (Bot-IoT, CICIDS-2017, IoT-23 and N-BaIoT) from these data sets were compared using our model. As a result of the evaluation, accuracy values of 99.94% for Bot-IoT data set, 99.95% for CICIDS-2017 data set, 99.96% for IoT-23 data set and 99.92% for N-BaIoT data set were reached. When these results are evaluated, it is seen that the usage of the datasets we have chosen is appropriate for attack detection and classification processes carried out by machine learning methods.

Keywords : *Internet of Things, Botnet Attacks, Intrusion Detection Systems, Support Vector Machines*

1. Giriş

Nesnelerin İnterneti (Internet of Things - IoT), heterojen IoT cihazlarından elde edilen büyük miktarda veriyi algılayarak, işleyerek ve analiz ederek binlerce akıllı nesneyi/cihazı sorunsuz bir şekilde birbirine bağlamayı amaçlamaktadır (Ashton, 2009). Gelişen teknoloji ile IoT cihazları sağlık sektöründen, otomotiv sektörüne, akıllı ev sistemlerinden giyilebilir teknoloji cihazlarına kadar birçok alanda hayatımıza girmiştir. Teknolojinin her geçen gün gelişmesi sürekli yenilenen ağ güvenliğine tehdit olarak geliştirilen senaryoları da beraberinde getirmektedir. Saldırganlar devamlı olarak uygulamalardaki, sistemlerdeki ve cihazlardaki açıkları bulmaya çalışmaktadır. Bu saldırı türlerinden biri de IoT cihazlarına yönelik yapılan saldırılarıdır. Symantec, her iki dakikada bir bir IoT

cihazının saldırıya uğradığını bildirmiştir (Cisco, 2019). Ayrıca Kaspersky (Broadcom, 2020), 2018'de IoT cihazlarına saldıran 121.588 kötü amaçlı yazılım örneği topladığını belirtmiştir.

İnternet erişimine açık herhangi bir IoT cihazı birçok kötü amaçlı yazılım saldırısına açık bir hedef durumundadır. Bu saldırılara botnet saldırıları özellikle BASHLITE ve Mirai saldırıları örnek olarak verilebilir. Botnet saldırıları genellikle üç bileşenden oluşur (Nugraha vd., 2020). Bu bileşenler; ana saldırgan olan "bot yöneticisi (bot master)", siber saldırganın kontrolü altındaki enfekte olmuş "bot" adı verilen makineler ve son olarak komut ve kontrol (C&C) sunucusudur. Bot yöneticisi C&C sunucusu aracılığıyla ya da doğrudan her bir bot cihazına erişerek onları uzaktan kontrol eder ve saldırılarda bulunur. Bu durumda kullanıcıların kişisel veya kimlik bilgilerine erişimi ve Dağıtık Hizmet Reddi (DDos- Distrubuted Denial of Service Attack) saldırılarına açık hale gelinir. Botnet saldırıları, internete bağlı cihazlar arasında yayılan çok ciddi saldırılardır. IoT cihazlarını botnet saldırılarından korumak için uygun ve etkili çözümler bulmak önemlidir. Ancak eski teknolojilerle bu saldırılara karşı korunmada büyük boşluklar bulunmaktadır (Alkahtani ve Aldhyani, 2021).

Botnet saldırılarıyla başa çıkma çözümlerinden biri Saldırı Tespit Sistemleridir (IDS – Intrusion Detection Systems). IDS ile ağdaki kötü amaçlı hareketler tespit edilir ve loglanır. IDS'de tanımlama yöntemleri temel olarak iki türe ayrılır: imza tabanlı IDS'ler ve anomali tabanlı IDS'ler. İmza tabanlı yöntemlerde daha önceden tanımlanmış kural ve davranışlar ile saldırılar tespit edilirken, anomali tabanlı yöntemlerde ağdaki normal çalışma düzeninin bozulması durumunda anomali durumu olduğu alarmı verilmektedir. İmza tabanlı yöntemlerde belirtilmemiş bir durum olduğunda saldırı tespiti yapılmamaktadır. Bu durum yeni türdeki saldırılara (sıfır gün ataklarına) karşı sistemi savunmasız duruma düşürmektedir. Anomali tabanlı yöntemlerde her ne kadar yalancı pozitif ve negatif alarm ile karşılaşılsa da yeni tür saldırıları tespit etmede ve sınıflandırmada etkili bir yöntemdir. Bu tür yeni saldırı türü yakalama ve sınıflandırma işlemlerinde makine öğrenme yöntemleri kullanılmaktadır. Bu sayede saldırganlar tarafından sürekli güncellenerek yapılan saldırılar makine öğrenme yöntemleriyle farklı bir tür saldırı türü olarak sınıflandırılarak IoT'deki botnet yakalama sorununa çözüm olarak getirilmiştir.

Çalışmamızın amacı literatürde kullanılan IoT botnet veri setlerini karşılaştırarak güncel ve en çok kullanılan dört veri setini örnek bir uygulama üzerinden incelemektir. Uygulamamızda bir makine öğrenme tekniği olan Destek Vektör Makineleri (SVM) yöntemi kullanılmıştır. Amacımız tek bir yöntem ile içerik olarak oldukça zengin dört farklı veri setini analiz etmek, saldırı tiplerini başarılı bir şekilde sınıflandırmak ve bu veri setlerini kullanım kolaylığı açısından değerlendirmektir.

2. İlişkili Çalışmalar

Güncel çalışmalardaki birçok araştırmacı botnet saldırılarının önceden tespiti, IoT ortamlarındaki cihazların ve ağ trafiğinin güvenliğini korumak adına farklı yöntemler üzerinde çalışmaktadır. En etkili yöntemler arasında imza ve anomali tabanlı saldırı sistemlerinde çeşitli makine ve derin öğrenme yöntemlerinin kullanılması bulunmaktadır. Makine öğrenmesi yöntemleriyle yapılan bu tespit işlemleri oldukça etkili bir yöntemdir. Ancak yapılan çalışmaların çoğunda sadece bir çeşit veri setinden yani bir kaynaktan gelen veriler üzerinde test yapılmıştır. Bu yöntem ile yapılan sınıflandırmalar tüm IoT botnet saldırıları için genelleme yapıldığında doğruluk oranını değiştirebilir. Farklı veri kaynakları üzerinde tek bir yöntemin test edilmesi güncel veya gelecekte yapılacak çalışmalara daha doğru yorum yapabilme olanağı verebilir. Yeni saldırı guruplarının tespitine yardımcı olabilir.

Asadi'nin (Asadi, 2021) yaptığı çalışmada Uzun Kısa Süreli Bellek (Long Short Term Memory -LSTM), Otomatik Kodlayıcı (Auto Encoder - AE) ve SVM teknikleriyle botnet tespiti ve sınıflandırması yapılmıştır. Popoola ve meslektaşları tarafından yapılan başka bir çalışmada (Popoola vd., 2021) LSTM ve AE derin öğrenme yöntemlerini hibrit bir şekilde kullanarak IoT ağlarındaki botnet saldırılarının tespiti yapılmıştır. Apostol ve meslektaşlarının yaptığı çalışmada (Apostol vd., 2021), IoT botnet etkinliklerini belirlemek için denetimsiz derin öğrenme tekniklerini kullanan anormali tabanlı bir algılama çözümü önerilmektedir. Derin öğrenme tekniklerinden AE yöntemi seçilmiştir. Bu çalışmalarda veri seti olarak sadece BoT-IoT (Koroniotis vd., 2019) kullanılmıştır.

Basati ve Faghih'in (Basati ve Faghih, 2021) yapmış olduğu çalışmada APAE ismini verdikleri IoT saldırı tespit sistemlerinde otomatik kodlayıcı derin öğrenme yöntemini kullanmışlardır. Yaptıkları çalışmada UNSW-NB15 (Moustafa vd., 2015), CICIDS2017 (Sharafaldin vd., 2018) ve KDDCup99 (KDD-CUP99 Veriseti, 1999) veri setlerini kullanmışlardır. UNSW-NB15 veri seti 2015 yılında ve KDDCup99 veri seti 1999 yılında oluşturulmuş veri setleridir ve güncel saldırı türlerini içermemektedir. Kompougias ve meslektaşlarının yaptığı çalışmada (Kompougias vd., 2021) IoT ağında bulunan olağandışı davranışları tespit etmek için otomatik kodlayıcı yöntemiyle bir model oluşturulmuştur. Sadece CICIDS2017 veri seti kullanılmıştır.

Alkahtani ve Aldhyani (Alkahtani ve Aldhyani, 2021) tarafından yapılan çalışmada Evrişimsel Sinir Ağları (Convolutional Neural Network – CNN) ve LSTM derin öğrenme yöntemlerinin hibrit olarak çalıştırılmasıyla IoT botnet saldırılarının tespiti yapılmıştır. Önerilen sistemde, Mirai ve BASHLITE olmak üzere iki yaygın botnet saldırısının gerçek anlamda bulaştığı dokuz adet ticari IoT cihazından toplanan gerçek trafik verileri kullanıldığı

N-BaIoT (Meidan vd., 2018) veri seti kullanılmıştır. Song ve meslektaşlarının yaptığı çalışmada (Song vd., 2021) otomatik kodlayıcı tabanlı bir model kullanılarak NSL-KDD, IoTID20 ve N-BaIoT verisetleri üzerinde çalışılmıştır. NSL-KDD veri seti (2009), eski bir veri setidir. IoTID20 (2020) veri seti güncel bir veri setidir ancak çalışma kapsamımızda kullanılmamıştır. Hussain ve meslektaşları tarafından yapılan çalışmada (Hussain vd., 2021) endüstriyel IoT cihazlarına yapılabilecek botnet saldırılarına karşı LSTM ve CNN yöntemleri kullanılarak hibrit bir model önerilmiştir. Çalışmada N-BaIoT veri seti kullanılmıştır.

Sahu ve meslektaşları tarafından yapılan çalışmada (Sahu vd., 2021) CNN ve LSTM derin öğrenme teknikleri hibrit bir şeklide ile kullanılarak IoT cihazlarına yapılan kötü amaçlı saldırıları tespit edilmiştir. Abdalgawad ve meslektaşlarının yaptığı çalışmada (Abdalgawad vd., 2022), Çekişmeli Otomatik Kodlayıcı (Adversarial Autoencoders - AAE) ve Çift Yönlü Çekişmeli Üretici Ağ (Bidirectional Generative Adversarial Networks - BiGAN) derin öğrenme teknikleri kullanılarak IoT ağlarındaki bilinmeyen zararlı kayıtların tespiti yapılmıştır. Sahu (Sahu vd., 2021), Abdalgawad (Abdalgawad vd., 2022) ve meslektaşlarının yaptıkları bu çalışmalarda en güncel IoT verisetlerinden IoT-23 kullanılmıştır.

Bu çalışmaların yanında IoT cihazlarında makine öğrenmesi tekniklerinin kullanımını karşılaştıran çeşitli çalışmalar da bulunmaktadır. Ahmad ve meslektaşlarının yaptığı çalışmada (Ahmad vd., 2021) IDS'de kullanılan derin öğrenmeye dayalı olarak yapılmış çalışmalar karşılaştırılmıştır. Bu amaçla çeşitli tekil ve hibrit derin öğrenme sınıflandırıcıları kullanarak çeşitli veri kümelerini (eski, yeni, IoT olmayan ve IoT'ye özgü) analiz edilmiştir. Nugraha ve meslaktaşlarının yaptığı çalışmada (Nugraha vd., 2020) CNN, LSTM, hibrit CNN-LSTM ve Çok Katmanlı Algılayıcılar (Multi Layer Perception - MLP) derin öğrenme teknikleri kullanılarak bilinen ve bilinmeyen botnet trafiği tespiti yapılmıştır. Dört farklı tekniğin performans analizi CTU-13 (2011) veri seti kullanılarak yapılmıştır. CTU-13 veri seti içerik olarak IRC, P2P, HTTP saldırıları, SPAM mesajları, Click-Fraud (CF), port-scan (PS), DDoS saldırıları, C&C saldırıları gibi oldukça zengin saldırı türlerine sahiptir. Ancak bu veri seti eski olması ve dosya formatının ham verilerden oluşması nedeniyle çalışmamızda kullanılmamıştır.

Bu çalışmalara dayanarak Bot-IoT, CICIDS2017, IoT-23 ve N-BaIoT veri setlerinin çalışmamız kapsamında kullanımının uygun olacağı, saldırı türü içeriği açısından zengin olduğu ve güncel saldırı çeşitlerini içermesinden dolayı sınıflandırma işlemlerinde kullanılabilirliği görülmüştür.

3. Materyal ve Yöntem

Bu bölümde sınıflandırma işleminde kullanılan veri setleri, kullanılan model ve uygulama detayları anlatılacaktır.

3.1. Veri Setleri ve Veri Ön İşleme

Literatürde IoT botnet trafiğinin incelendiği birçok veriseti üzerinde çalışma yapılmıştır. Ancak kullanılan bu verisetlerinin bir kısmı güncelliğini kaybetmiştir, bir kısmı makine öğrenmesi yöntemleriyle işlenmeye uygun değildir. Karşılaştırma yaptığımız modelde bu verisetlerinden güncel ve işlenebilirliği en uygun verisetleri seçilmiştir. Bunlar Bot-IoT (Koroniotis vd., 2019), CICIDS2017 (Sharafaldin vd., 2018), IoT-23 (Garcia vd., 2020) ve N-BaIoT (Meidan vd., 2018) veri setleridir. Bu verisetlerinin temel özellikleri şu şekildedir:

3.1.1. Bot-IoT Veriseti

BoT-IoT veri seti (Koroniotis vd., 2019), UNSW Canberra Cyber Range Labaratuvarı'nda gerçekçi bir ağ ortamı tasarlanarak oluşturulmuştur. Tasarlanan veri seti temel olarak normal ve botnet verilerinin bulunduğu bir kombinasyondan oluşmaktadır ve kaynak dosyalar pcap, argus ve csv uzantılı şekilde kullanıma sunulmuştur. Dosyalar, etiketleme sürecine yardımcı olmak amacıyla saldırı kategorisine ve alt kategorilerine ayrılmıştır. Yakalanan pcap dosyalarının boyutu 69.3 GB'tır ve 72.000.000'den fazla kayıt bulunmaktadır. Çıkarılan akış trafiği, csv dosya formatında 16,7 GB boyutundadır.

Veri kümesinin işlenmesini kolaylaştırmak için, orijinal veri kümesinin %5'i çıkarılmıştır. Çıkarılan %5'lik veriseti, boyutu yaklaşık 1.07 GB olan 4 dosyadan ve 3668522 adet kayıttan oluşmaktadır. Tablo 1.'de verisetinde bulunan kayıtların tipleri ve verisetindeki kullanım sayıları bulunmaktadır.

3.1.2. CICIDS2017 Veriseti

CICIDS2017 veri seti (Sharafaldin vd., 2018) Kanada Siber Güvenlik Enstitüsü tarafından New Brunswick Üniversitesi'nde geliştirilmiştir. Gerçek gerçek dünya verilerine benzeyen, iyi huylu ve en güncel yaygın saldırıları içermektedir.

Tablo 1. Bot-IoT Veriseti

Bot-IoT		
	Tür	Verisetindeki Kullanım Sayısı
Attack tipine göre sınıflandırma		
	Attack	3668045
	Normal	477
Category tipine göre sınıflandırma		
	DDoS	1926624
	DoS	1650260
	Reconnaissance	91082
	Normal	477
	Theft	79
Subcategory tipine göre sınıflandırma		
	UDP	1981230
	TCP	1593180
	Service_Scan	73168
	OS_Fingerprint	17914
	HTTP	2474
	Normal	477
	Keylogging	73
	Data_Exfiltration	6

Tablo 2. CICIDS2017

CICIDS2017		
	Veri Türü	Verisetindeki Kullanım Sayısı
Monday-WorkingHours.pcap_ISCX.csv	BENIGN	529918
Tuesday-WorkingHours.pcap_ISCX.csv	BENIGN	432074
	FTP-Patator	7938
	SSH-Patator	5897
Wednesday-workingHours.pcap_ISCX.csv	DoS Hulk	231073
	BENIGN	440031
	DoS slowloris	5796
	DoS Slowhttptest	5499
	DoS GoldenEye	10293
	Heartbleed	11
Thursday-WorkingHours-Morning-WebAttacks.pcap_ISCX.csv	BENIGN	168186
	Web Attack - Brute Force	1507
	Web Attack - XSS	652
	Web Attack - Sql Injection	21
Thursday-WorkingHours-Afternoon-Infilteration.pcap_ISCX.csv	BENIGN	288566
	Infiltration	36
Friday-WorkingHours-Morning.pcap_ISCX.csv	BENIGN	189067
	Bot	1966
Friday-WorkingHours-Afternoon-PortScan.pcap_ISCX.csv	PortScan	158930
	BENIGN	127537
Friday-WorkingHours-Afternoon-DDos.pcap_ISCX.csv	DDoS	128027
	BENIGN	97718

Veri kümesi, beş günlük ağ trafiğini ve her güne ait veriler farklı kombinasyonda saldırı türlerini içermektedir. Pazartesi günkü trafik sadece normal verilerin bulunduğu trafikten oluşmaktadır. Diğer günlerde Brute Force FTP, Brute Force SSH, DoS, Heartbleed, Web Attack, Infiltrasyon, Botnet ve DDoS gibi çok çeşitli saldırı türü verileri bulunmaktadır. Beş güne ait veriler toplamda 79 adet sütunda tanımlanmış, ilk 78 sütun ağ trafik etiketlerini ve son sütun atak türü etiketini belirtmektedir. CICIDS2017 veri setine ait tür bilgileri ve kullanım sayıları Tablo 2.'de gösterilmektedir.

3.1.3. IoT-23 Veriseti

IoT-23 veri setine (Garcia vd., 2020) ait ağ trafiği, Çek Cumhuriyeti'ndeki CTU Üniversitesi'ne ait Stratosphere Laboratuvarı'nda elde edilmiştir. Veri seti, 3 adet normal IoT trafik ağına sahip cihazdan ve 20 adet kötü amaçlı yazılım bulaşmış cihazdan elde edilmiştir. Çalışmaları 23 adet farklı IoT ağ trafiği içeren senaryo içermektedir. Tüm (20 GB) ve küçük (8.7 GB) olmak üzere 2 farklı seçenekte sıkıştırılmış dosyalarda kullanıma sunulmuştur. Küçük veri setinin sıkıştırılmamış hali yaklaşık 43 GB veri içermektedir. Her bir senaryo "labeled" uzantılı dosyalara kaydedilmiş ve 20 adet veri özelliğine sahiptirler. Her bir veri "Benign" ve "Malicious" olarak ikili sınıf etiketine sahiptir. "Malicious" olarak etiketlendirilmiş verilerde saldırı türü detay bilgileri mevcuttur. Alt sınıf türleri; Benign, Attack, C&C, C&C-FileDownload, C&C-Torii, C&C-HeartBeat, C&C-HeartBeat-Attack, C&C-HeartBeat-FileDownload, C&C-PartOfAHorizontalPortScan, DDoS, FileDownload, Okiru, Okiru-Attack, PartOfAHorizontalPortScan, PartOfAHorizontalPortScan-Attack olarak etiketlendirilmiştir. IoT-23 Veriseti (2022) açıklamasına göre etiket atamasından sonraki veri oranın düzensizliğini açıkça görülebilir. Şöyle ki en yaygın üç kötü amaçlı etiketlendirilmiş veriler; PartOfAHorizontalPortScan (213852924 akış), Okiru (47381241 akış) ve DDoS (19538713 akış) iken, en az yaygın olan üç kötü niyetli etiketlendirilmiş veriler; C&C-Mirai (2 akış), PartOfAHorizontalPortScan-Attack (5 akış) ve C&C-HeartBeat-FileDownload (11 akış) verileridir.

3.1.4. N-BaIoT Veriseti

N-BaIoT veriseti (Meidan vd., 2018), gerçek IoT cihazlarından elde edilen verilerden oluşmaktadır. Dokuz adet ticari IoT cihazından Mirai ve BASHLITE (ya da Gafgyt) botnet türlerinde trafiği içerir. Bu veri kümesi, gerçek normal trafik (benign) ve 9 farklı IoT cihazının (kapı zilleri, termostat, bebek monitörü, güvenlik ve web kameraları) özelliklerinden oluşmaktadır. Verilerin toplanmasında kullanılan cihazlar; Danmini kapı zili, Ecobee temostat, Ennio kapı zili, Philips B120N10 bebek monitörü, Provision PT_737E güvenlik kamerası, Provision PT_838 güvenlik kamerası, Samsung SNH_1011_N web kamerası, SimpleHome XCS7_1002_WHT güvenlik kamerası, SimpleHome XCS7_1003_WHT güvenlik kamerasıdır. Her bir cihazdan elde edilen veriler "benign-traffic", "gafgyt_attacks" ve "mirai_attacks" şeklinde üç farklı dizinde csv dosya formatında sunulmuştur. Dosyalardaki veriler paket sayısı, gelen ve giden paketlerin boyutu ve paketlerin varışlar arası süreleri gibi trafik istatistiklerinin olduğu 23 özelliğinden oluşmaktadır (Song vd., 2021). Toplam etiket sayısı 115'tir, toplam atak türü 10 adettir.

Tablo 3. N-BaIoT Veriseti

N-BaIoT Veriseti (Danmini Kapı Zili)			
Veri Türü	Kategori	Orijinal Veri Sayısı	Çalışmamızda Kullanılan Miktar
Benign	Normal	49548	49000
Attack	Gafgyt_Combo	59718	5100
Attack	Gafgyt_Junk	29068	5100
Attack	Gafgyt_Scan	29849	5100
Attack	Gafgyt_TCP	92141	5100
Attack	Gafgyt_UDP	105874	5100
Attack	Mirai_ACK	102195	5100
Attack	Mirai_Scan	107685	5100
Attack	Mirai_Syn	122573	5100
Attack	Mirai_UDP	237665	5100
Attack	Mirai_UDP_Plain	81982	5100
	Toplam	**1018298**	**100000**

Veri setinin kapsamının çok büyük olması nedeniyle çalışmamız kapsamında bir adet cihazdan (Danmini kapı zili) elde edilen veriler kullanılmıştır. Danmini kapı ziline ait veri analizi Tablo3.'te verilmiştir.

3.1.5. Veri Ön İşleme

Sınıflandırma işlemlerinde veri setleri genellikle ilgili ve ilgisiz birçok veri içermektedir. Bu durumda alakasız özellikler sınıflandırmaya yardımcı olmaz ve geniş arama alanı nedeniyle sınıflandırma verimliliğini de azaltabilirler (Asadi, 2021). Bu nedenden dolayı ilerleyen bölümlerde anlatılacağı üzere veri setlerinden bazılarındaki ilgisiz sütunlar çıkarılmıştır. Bununla birlikte yapmış olduğumuz çalışmadaki tüm veriler programlama ile sınıflandırmaya uygun durumda değildi. Kodlamada hatalara neden olabilecek NaN, infinity vb. formundaki veriler güncellendi, boş satırlar temizlendi, CSV formatında oluşturulmamış veriler CSV formatına dönüştürüldü. Ayrıca kaynak limiti problemlerinden dolayı bazı veri setlerinin belirli kısımları test edilebilmiştir.

Bot-IoT veri setinde her bir kayıt 46 adet sütundan oluşacak şekilde etiketlendirilmiştir. Bu etiketlerden ilk 43 adeti ağ trafiğindeki bilgileri içeren etiketlerdir ve son üçü verilerin binary sınıflandırmasını (normal/attack), kategorisi ve alt kategori bilgilerini içermektedir. İlk 43 sütundaki bilgilerin tümü sınıflandırma işlemine katkıda bulunmamaktadır. Bu nedenle bu sütunlardan bazılarının elenmesi gerekmektedir. Mehdi Asadi'nin yaptığı çalışmaya göre (Asadi, 2021) her bir etiketin sınıflandırmadaki etkisi hesaplanmış ve etiketlere ilgilerine göre değerler verilmiştir. Yaptığımız çalışmada sınıflandırmada etkisi olan ilk sekiz etiket seçilmiştir bunlar; stime, pkts, ltime, spkts, sbytes, srate, TnBPSrcIP, Pkts_P_State_P_Protocol_P_SrcIP sütunları ve sınıflandırma verileri olan attack, category, subcategory sınıflarıdır. Bu etiketlere sahip All features kategorisindeki 4 csv dosyası SVM model ile sınıflandırmada kullanılmıştır.

CICIDS2017 veri setinde beş güne ait trafik akışını içeren dosyalar Tablo 2'de gösterildiği üzere 8 farklı csv dosyasında sunulmuştur. Bu veri setinin sınıflandırılmasında bazı problemler yaşanmıştır. Birincisi normal/atak türündeki ve atak alt türlerindeki veriler günlere göre homojen olarak bulunmamaktadır. Örneğin pazartesi gününe ait verilerin tümü "BENIGN" yani nomal tipte verilerdir. Bu durumda sınıflandırma işlemi yapılamamaktadır. Diğer bir sorun ise Çarşamba gününe ait verilerin 692703 adet olmasıdır ve modelimizin sınıflandırıcısı bu boyuttaki veriyi işleyememiştir. Bu durumda bu güne ait veriler iki kısımda işlenmiştir. Ayrıca csv dosyalarındaki verilerin bir kısmı boş satır, NaN ve Infinity değerler içermektedir. Sınıflandırma işleminden önce bu tür veriler temizlenmiştir.

N-BaIoT verisetinde 10 adet saldırı türü ve benign tipinde normal veriler bulunmaktadır. Tüm veriler 115 adet etikete sahiptir. Çalışmamızda Danmini kapı ziline ait veriler işlenmiştir. Farklı dosyalarda dizinlenen veriler tek bir csv dosyasında birleştirilmiştir. Orijinal etiketlere ek olarak binary (normal/atak) sınıflandırma için "Kategori", saldırı türü sınıflandırması için "Alt Kategori" etiketleri eklenmiştir. Danmini cihazına ait veri sayısı 1018298 adettir. Bu boyuttaki veri, kaynak yetersizliğinden dolayı işlenememiştir ve daha küçük boyuta indirgenmiştir. Test kapsamında 100000 veri işlenmiştir, bu verilerin 49000'i normal 51000'i ise atak türündedir.

Tablo 4. IoT-23 Veriseti (Sadeleştirilmiş)

IoT-23 Veriseti	
Kategori	**Çalışmamızda Kullanılan Miktar**
DDoS	14395
CCNormal	6720
Benign	5339
PartOfAHorizontalPortScan	157
OkiruNormal	18
CCTorii	16
CCFileDownload	14
CCHeartBeatFileDownload	9
Attack	4
FileDownload	3
CCHeartBeat	2
OkiruAttack	2
CCHeartBeatAttack	2
Toplam	**26681**

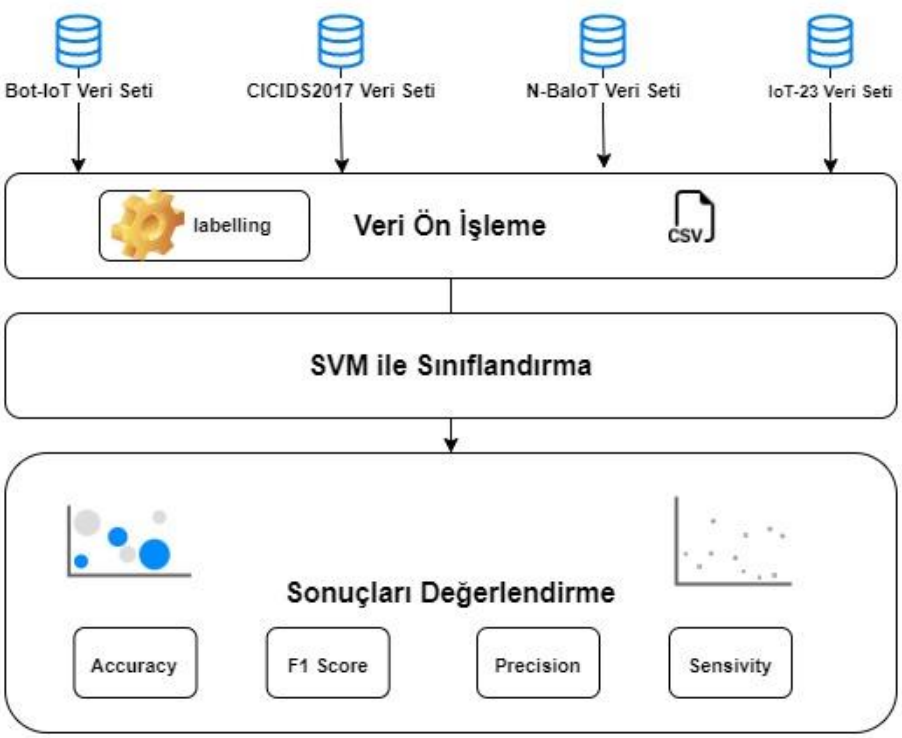
IoT-23 veri setinde (küçük sürüm) 52,5 GB veri bulunmaktadır. Bu veriler labeled file tipinde 23 ayrı sette heterojen olarak kaydedilmiştir. Çalışma kapsamımızda bu dosyalardan tüm veri tiplerini kapsayacak şekilde yeni bir veri seti oluşturulmuştur. Sadeleştirilmiş formatta veriler csv dosya formatına dönüştürülmüş ve hatalı veriler temizlenmiştir. Test edilen verilere ait analiz sonuçları Tablo 4'te bulunmaktadır.

3.2. Uygulama Detayları

Çalışmanın tüm aşamaları PyCharm (2021.3.3) IDE'si üzerinde, Python (3.9) programlama dili kullanılarak yapılmıştır. Modelin çalıştırılması için; sklearn (1.0.2), scipy (1.7.3) matplotlib (3.5.1) , numpy (1.22.3) , pandas (1.4.3) , seaborn (0.11.2) kütüphaneleri kullanılmıştır. Donanım olarak ise Intel Core i5 CPU, 8 GB belleğe sahip bir bilgisayar üzerinde çalışmamız gerçekleştirilmiştir. Uygulamada kullanılan platform Windows 11'dir.

3.3. Kullanılan Model

Deneysel çalışmamızda Şekil 1.'de belirtildiği gibi üç temel süreçten geçilmiştir. İlk aşamada literatürde IoT botnet çalışmalarında kullanılan veri setleri incelenmiş ve bunlardan Bot-IoT, CICIDS-2017, N-BaIoT ve IoT-23 veri setleri uygulamada kullanılmak üzere seçilmiştir. Seçilen veri setleri öncelikle SVM tekniği ile sınıflandırmaya uygun hale getirilmek üzere veri ön işleme sürecine tabi tutulmuştur. Veri setindeki hatalı veriler ayıklanmış, eksik değerler, sonsuzluk ve NaN değerleri veri sınıflandırıcısının girdisiyle uyumlu olacak şekilde düzenlenmiştir. Sayısal nitelikteki türler düzenlenerek sınıflandırmaya uygun bir şekilde etiketlendirilmiştir.



Şekil 1. Önerilen Mimari

Modelin ikinci aşamasında sınıflandırma işlemi yapılmıştır. Bu adımda Scikit-learn (Scikit-learn, 2011) kütüphanesinden faydalanılmıştır. Sınıflandırma işleminde her bir veri seti için hangi sütundaki verilere göre sınıflandırma yapılacağı ve hangi sütundaki verilerin tahmin edileceği belirlenmiştir. Verilerin %75'i eğitimde, %25'i tahminde kullanılmak üzere seçilmiştir. Sınıflandırma nesnesi yaratılarak model oluşturulmuştur. Kernel değeri olarak "lineer" seçilmiştir. Daha önceki adımda ayrılan test seti kullanılarak model ile tahmin yapılmıştır.

Modelin son adımında elde edilen sonuçlar gerçek değerler ile karşılaştırılmıştır. Karşılaştırmada bir sonraki başlıkta belirtilen performans değerlendirme metrikleri ile karşılaştırılmıştır.

3.4. Performans Değerlendirme Metrikleri

Yaptığımız sınıflandırma çalışmasının sonuçları değerlendirmek amacıyla yaygın olarak kullanılan karmaşıklık matrisi (confusion matrix), doğruluk (accuracy), kesinlik (precision), duyarlılık (sensivity), F1 skor metrikleri kullanılmıştır. Bu metriklerin hesaplanmasında Gerçek Pozitif (TP), Gerçek Negatif (TN), Yalancı Pozitif (FP), Yalancı Negatif (FN) değerleri kullanılmaktadır. Bu değerlerin açılımı şu şekildedir:

- Gerçek Pozitif (TP): Doğru sınıflandırılmış pozitif sınıf sayısı
- Gerçek Negatif (TN): Doğru sınıflandırılmış negatif sınıf sayısı
- Yalancı Pozitif (FP): Yanlış sınıflandırılmış pozitif sınıf sayısı
- Yalancı Negatif (FN): Yanlış sınıflandırılmış negatif sınıf sayısı

Tablo 5. Metrik Formülasyonları

Doğruluk	(TP + TN) / (TP + TN + FP + FN)
Kesinlik	TP / (TP + FP)
Duyarlılık	TP / (TP + FN)
F1 Skor	(2 x Kesinlik x Duyarlılık) / (Kesinlik + Duyarlılık)

Kullanılan metriklere ait formülasyonlar Tablo 5.'te açıklanmıştır. Karmaşıklık matrisi ise veri setindeki gerçek veri sınıfları ile sınıflandırma işlemi sonrasında tahmin edilen sınıfların bir tablo halinde karşılaştırıldığı bir metriktir.

Şekil 2. Karmaşıklık Matrisi (2x2)

Karmaşıklık matrisinin en temel hali ikili (binary) sınıflandırma için oluşturulan 2x2 bir matristir. Değerlerin gösterimi Şekil 2.'deki gibi olmaktadır.

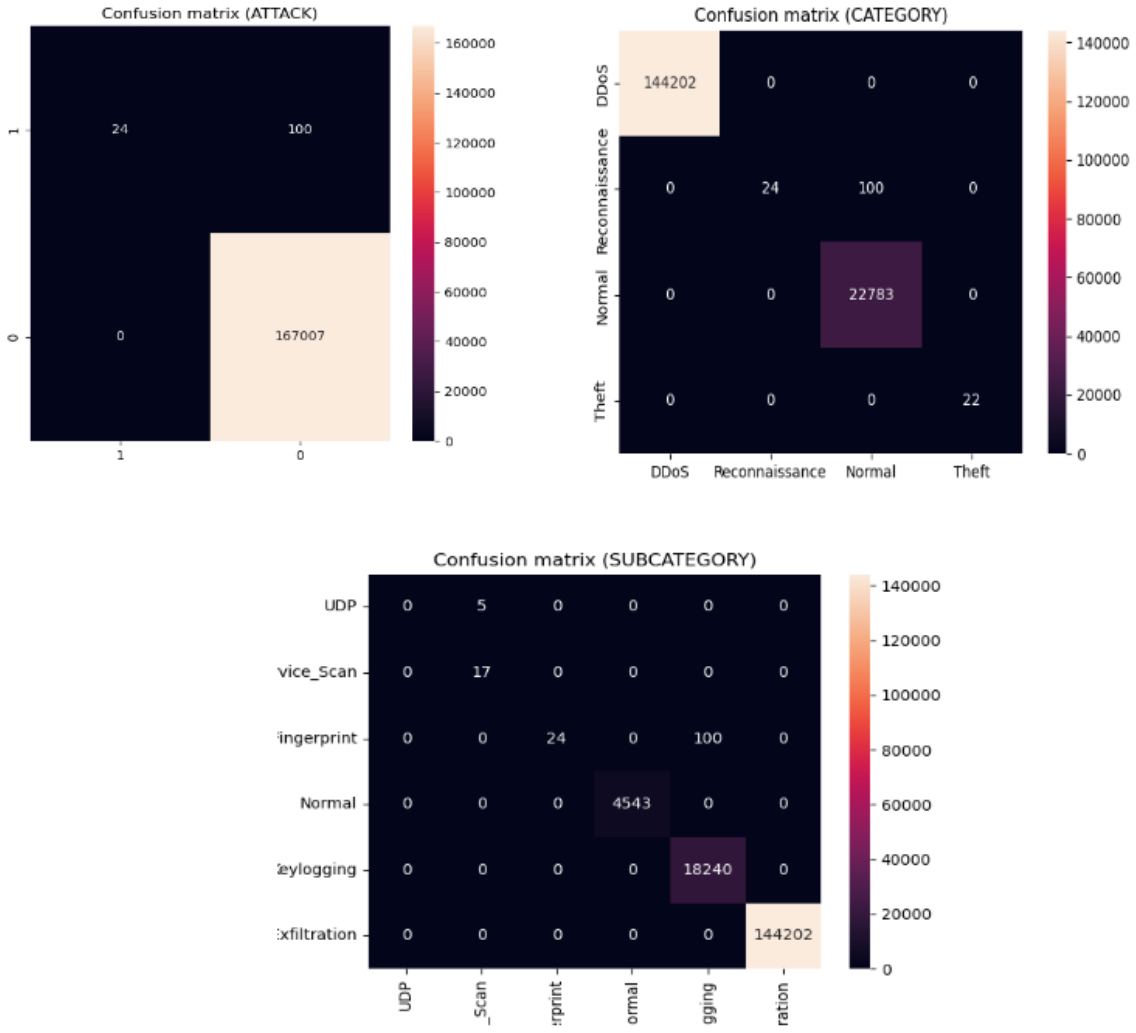
3.5. Değerlendirme Sonuçları

Çalışmamızda Bot-IoT veri setine ait 4 adet csv dosyasının sınıflandırılması yapılmıştır. İlk üç dosyada (UNSW_2018_IoT_Botnet_Full5pc_1.csv, UNSW_2018_IoT_Botnet_Full5pc_2.csv, UNSW_2018_IoT_Botnet_Full5pc_3.csv) "Attack" ve "Category" türlerinde tek çeşit sınıf olduğu için sınıflandırma yapılamamıştır. "Subcategory" türündeki sınıflandırma sonuçlarında %100 metrik değerlerine ulaşılmıştır.

Tablo 6 Bot-IoT Metrik Sonuçları

Dosya İsmi	Kayıt Sayısı	Sınıflandırma	Accuracy	F1 Score	Precision	Sensivity
UNSW_2018_IoT_ Botnet_Full5pc_4	668522	Attack	99.94%	99.92%	99.94%	99.94%
		Category	99.94%	99.92%	99.94%	99.94%
		Subcategory	99.94%	99.92%	99.94%	99.94%

Bot-IoT verisetinin sınıflandırılmasında en tutarlı sonuçlar Tablo 6.'da görüldüğü üzere dördüncü dosyada (UNSW_2018_IoT_Botnet_Full5pc_4.csv) elde edilmiştir.



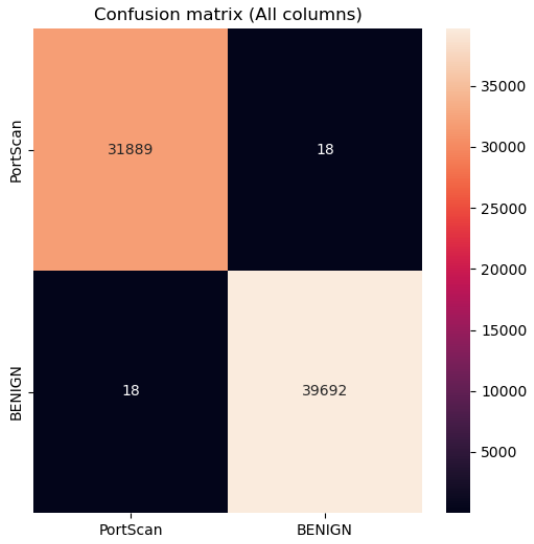
Şekil 3. Bot-IoT Karmaşıklık Matrisi

Bot-IoT veri setinin UNSW_2018_IoT_Botnet_Full5pc_4.csv dosyasındaki sınıflandırma işlemine ait sonuçların ("Attack", "Category" ve "Subcategory" türlerinde) karmaşıklık matrisleri Şekil 3'te gösterilmektedir.

Tablo 7 CICIDS2017 Metrik Sonuçları

Dosya İsmi	Bölüm	Kayıt Sayısı	Accuracy	F1 Score	Precision	Sensivity
Tuesday-WorkingHours.pcap_ISCX	Tek Parça	445909	%99.26	%99.15	%99.27	%99.26
Wednesday-workingHours.pcap_ISCX	Part 1	349999	%99.58	%99.58	%99.58	%99.58
	Part 2	342704	%99.89	%99.89	%99.89	%99.89
Thursday-WorkingHours-Morning-WebAttacks.pcap_ISCX	Tek Parça	170366	%99.16	%99.08	%99.47	%99.16
Thursday-WorkingHours-Afternoon-Infilteration.pcap_ISCX	Tek Parça	288602	%99.90	%99.90	%99.90	%99.90
Friday-WorkingHours-Morning.pcap_ISCX	Tek Parça	191033	%99.43	%99.39	%99.38	%99.43
Friday-WorkingHours-Afternoon-DDos.pcap_ISCX	Tek Parça	225745	%99.92	%99.92	%99.92	%99.92
Friday-WorkingHours-Afternoon-PortScan.pcap_ISCX	Tek Parça	286467	%99.95	%99.95	%99.95	%99.95

CICIDS2017 veri setinde beş günlük trafik akışı 8 farklı csv dosyasında bulunmaktadır. Bu veri setinin sınıflandırılması aşamasında Çarşamba gününe ait veriler iki kısımda işlenmiştir. Monday-WorkingHours.pcap_ISCX dosyasında tek türde kayıt olduğu için sınıflandırma yapılamamıştır. Sınıflandırma sonuçları Tablo 7.'de gösterilmektedir.

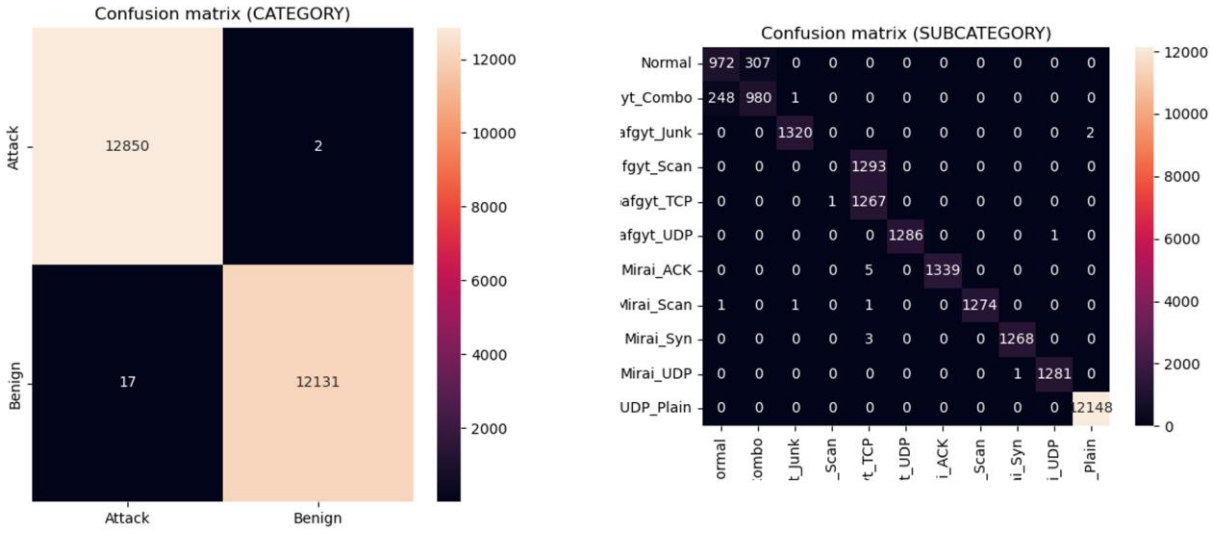


Şekil 4. CICIDS2017 Friday-WorkingHours-Afternoon-PortScan.pcap_ISCX Karmaşıklık Matrisi

CICIDS2017 veri setine ait sınıflandırma işlemlerinde en iyi sonuç Friday-WorkingHours-Afternoon-PortScan.pcap_ISCX dosyasında elde edilmiştir. Bu dosyaya ait test sonucunun karmaşıklık matrisi Şekil 4.'te verilmiştir.

Tablo 8 N-BaIoT Metrik Sonuçları

Kayıt Sayısı	Sınıflandırma	Accuracy	F1 Score	Precision	Sensivity
100000	Category	%99.92	%99.92	%99.92	%99.92
	Subcategory	%92.54	%90.84	%90.02	%92.54

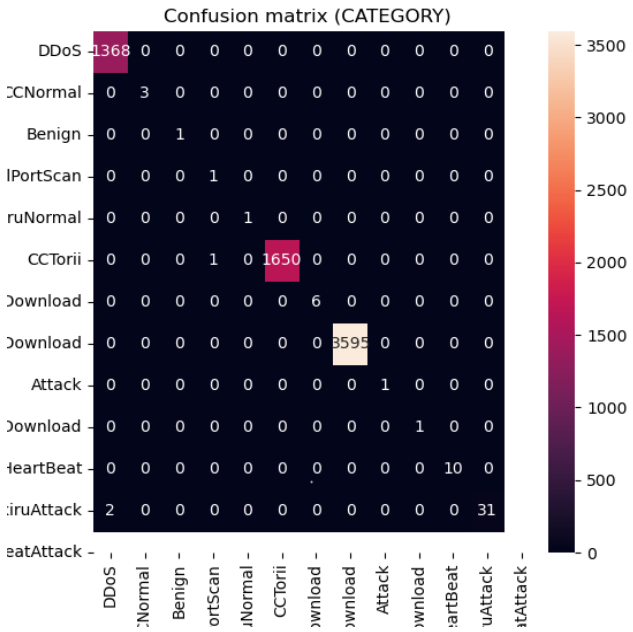


Şekil 5. N-BaIoT Karmaşıklık Matrisi

Çalışmamızda N-BaIoT veriseti için Danmini cihazına ait veriler sınıflandırılmıştır. Bu verisetindeki veri sayısı 1018298 adet olduğu için daha küçük boyuta indirgenmiş ve sınıflandırma işlemi bu alt küme üzerinde yapılmıştır. Test kapsamında 100000 veri işlenmiştir, bu verilerin 49000'i normal 51000'i ise atak türündedir. Test sonuçları Tablo 8.'de ve karmaşıklık matrisi Şekil 5'te gösterilmektedir.

Tablo 9 IoT-23 Metrik Sonuçları

Kayıt Sayısı	Sınıflandırma	Accuracy	F1 Score	Precision	Sensivity
26681	Category	%99.96	%99.96	%99.96	%99.96



Şekil 6. IoT-23 Veriseti Karmaşıklık Matrisi

IoT-23 veri setinin (küçük sürüm) verileri 23 ayrı sette heterojen olarak kaydedilmiştir. Çalışma kapsamımızda bu dosyalardan tüm veri tiplerini kapsayacak şekilde yeni bir veri seti oluşturulmuştur. Sadeleştirilmiş sürüme ait sınıflandırma sonuçları Tablo 9'da, karmaşıklık matrisi ise Şekil 6.'da gösterilmektedir.

Tablo 10 Gerçekleştirilen Çalışmanın İlgili Çalışmalarla Karşılaştırılması

	Bot-IoT	**CICIDS-2017**	**N-BaIoT**	**IoT-23**
A- Gerçekleştirilen Çalışma	%99.94	%99.95	%99.92	%99.96
B- Asadi, 2021	%99.998	X	X	X
C- Popoola vd., 2021	%100 - %99.49	X	X	X
D- Apostol vd., 2021	%99.7	X	X	X
E- Basati ve Faghih, 2021	X	%98.73 - %99.50	X	X
F- Kompougias vd., 2021	X	*	X	X
G- Alkahtani ve Aldhyani, 2021	X	X	%90.88	X
H- Song vd., 2021	X	X	%99.97	X
İ- Hussain vd., 2021	X	X	%99.95	X
J- Sahu vd., 2021	X	X	X	%96.23- %95.13
K- Abdalgawad vd., 2022	X	X	X	*

A- Gerçekleştirilen çalışmaya ait en iyi sonuç değerleridir.
B- Asadi, 2021 : SVM model ile yapılan sınıflandırmada "Full Features" sürümündeki en iyi sonuçlardır
C- Popoola vd., 2021: Binary sınıflandırma sonucunda %100; çok sınıflı sınıflandırma sonucunda %99.49 doğruluk değeri elde edilmiştir.
D- Apostol vd., 2021
E- Basati ve Faghih, 2021: Anomali tespitinde elde edilen doğruluk değeri %98.73; çok sınıflı sınıflandırma sonucunda elde edilen doğruluk değeri %99.50 verilmiştir.
F- Kompougias vd., 2021: Hesaplamalarında doğruluk metriği kullanılmamıştır.
G- Alkahtani ve Aldhyani, 2021: Danmini kapı zili üzerinde yapılan testin doğruluk metrik sonucudur.
H- Song vd., 2021: Danmini kapı zili üzerinde yapılan testlerde elde edilen en yüksek doğruluk değeridir.
İ- Hussain vd., 2021
J- Sahu vd., 2021: Kullanılan hibrit modelin zararlı kayıtları tespit etmedeki doğruluk değeri %96.23; normal kayıtları tespit etmedeki doğruluk değeri %95.13'tür
K- Abdalgawad vd., 2022: Hesaplamalarında doğruluk metriği doğrudan kullanılmamıştır. %99 F1 skor değerine ulaşılmıştır.

Tablo 10.'da çalışmamız sonucunda elde edilen en iyi doğruluk değerleri ile ilgili çalışmaların sonuçları karşılaştırılmaktadır. Karşılaştırmada kullanılan makine öğrenme yöntemleri ve test için kullanılan kayıt miktarının değişken olduğu göz önünde bulundurulmalıdır.

4. Sonuç

Çalışmamızda SVM yöntemine dayalı olarak geliştirdiğimiz model ile literatürde çokça kullanılan dört farklı IoT botnet veri seti incelenmiştir. Temel amacımız tek bir yöntem ile farklı veri setlerini analiz etmek, saldırı tiplerini sınıflandırmak ve bu veri setlerini kullanım kolaylığı açısından değerlendirmektir.

Bu veri setlerinden Bot-IoT'de %99.94 doğruluk, %99.92 F1 skor, %99.94 kesinlik ve %99.94 duyarlılık değerleri sonucuna ulaşılmıştır (UNSW_2018_IoT_Botnet_Full5pc_4 dosyasında). CICIDS-2017 veri setinde %99.95 doğruluk, %99.95 F1 skor, %99.95 kesinlik ve %99.95 duyarlılık değerleri sonucuna ulaşılmıştır (Friday-WorkingHours-Afternoon-PortScan.pcap_ISCX dosyasında). Derlenmiş N-BaIoT veri setinde kategori tipinde %99.92 doğruluk, %99.92 F1 skor, %99.92 kesinlik ve %99.92 duyarlılık; alt kategori tipinde %92.54 doğruluk, %90.84 F1 skor, %90.02 kesinlik ve %92.54 duyarlılık değerleri sonucuna ulaşılmıştır. Son olarak IoT-23 veri setinde %99.96 doğruluk, %99.96 F1 skor, %99.96 kesinlik ve %99.96 duyarlılık değerleri sonucuna ulaşılmıştır.

Metrik değerlerinin karşılaştırılması sonucunda IoT-23 verisetinin sonuçlarının en yüksek değerleri aldığı görülmektedir. Her ne kadar karşılaştırma metriklerinin sonuçları önemli bir ölçüt olsa da IoT botnet saldırılarının tespitinde kullanılacak verisetlerinin seçiminde farklı kıstasların da önemli olduğu görülmüştür. Seçim için kontrol edilecek ilk özellik verisetinin yeni saldırı türlerini içermesidir. Güncel olmayan verisetlerinin kullanımı saldırı tespit sistemlerinde yeni saldırı türlerinin yakalanmasında verimsizliğe neden olabilir. Verisetindeki kayıtların tür

açısından zengin olması önemli kıstaslardan biridir. Bu ölçüte göre değerlendirme yapıldığında IoT-23 ve N-BaIoT veri setlerinin tür içeriği açısından daha fazla kategoriye sahip olduğu görülmektedir. Ayrıca saldırı tespit sistemlerinde insan müdahalesinin hiç olmaması gerekmektedir. Veri ön işleme ve uygun dosya formatına dönüştürme işlemleri iş yüküne neden olduğu gibi gerçek ortamda karşılaşılmaması gereken durumlardır. Bu kıstasa göre karşılaştırma yapıldığında ise Bot-IoT veri setinde en az eforun sarf edildiği söylenebilir.

Kaynaklar

K. Ashton, “That ‘internet of things’ thing,” RFiD J, vol. 22, pp. 97–114, 2009, https://www.rfidjournal.com/articles/view 4986.

Cisco, Cisco Visual networking Index (VNI) global Mobile data traffic Forecast update, 2017–2022, Cisco Systems Inc., San Jose, CA, USA, 2019.

Broadcom, “Symantec Internet Security Threat Report 2019,” vol. 24, 2020, https://docs.broadcom.com/doc/istr-24-2019- en.

B. Nugraha, A. Nambiar and T. Bauschert, "Performance Evaluation of Botnet Detection using Deep Learning Techniques," 2020 11th International Conference on Network of the Future (NoF), 2020, pp. 141-149, doi: 10.1109/NoF50125.2020.9249198.

Asadi, Mehdi. (2021). Detecting IoT botnets based on the combination of cooperative game theory with deep and machine learning approaches. Journal of Ambient Intelligence and Humanized Computing. 10.1007/s12652-021-03185-x.

S. I. Popoola, B. Adebisi, M. Hammoudeh, G. Gui and H. Gacanin, "Hybrid Deep Learning for Botnet Attack Detection in the Internet-of-Things Networks," in IEEE Internet of Things Journal, vol. 8, no. 6, pp. 4944-4956, 15 March15, 2021, doi: 10.1109/JIOT.2020.3034156.

Apostol, I.; Preda, M.; Nila, C.; Bica, I. IoT Botnet Anomaly Detection Using Unsupervised Deep Learning. Electronics 2021, 10, 1876. https://doi.org/10.3390/electronics10161876

Hasan Alkahtani, Theyazn H. H. Aldhyani, "Botnet Attack Detection by Using CNN-LSTM Model for Internet of Things Applications", Security and Communication Networks, vol. 2021, Article ID 3806459, 23 pages, 2021. https://doi.org/10.1155/2021/3806459

Basati, A., Faghih, M.M. APAE: an IoT intrusion detection system using asymmetric parallel auto-encoder. Neural Comput & Applic (2021). https://doi.org/10.1007/s00521-021-06011-9

O. Kompougias et al., "IoT Botnet Detection on Flow Data using Autoencoders," 2021 IEEE International Mediterranean Conference on Communications and Networking (MeditCom), 2021, pp. 506-511, doi: 10.1109/MeditCom49071.2021.9647639.

Song, Y.; Hyun, S.; Cheong, Y.-G. Analysis of Autoencoders for Network Intrusion Detection. Sensors 2021, 21, 4294. https://doi.org/10.3390/s21134294

Hussain, Z.; Akhunzada, A.; Iqbal, J.; Bibi, I.; Gani, A. Secure IIoT-Enabled Industry 4.0. Sustainability 2021, 13, 12384. https://doi.org/10.3390/su132212384

Sahu, Amiya & Sharma, Suraj & Tanveer, M. & Raja, Rohit. (2021). Internet of Things attack detection using hybrid Deep Learning Model. Computer Communications. 176. 10.1016/j.comcom.2021.05.024.

N. Abdalgawad, A. Sajun, Y. Kaddoura, I. A. Zualkernan and F. Aloul, "Generative Deep Learning to Detect Cyberattacks for the IoT-23 Dataset," in IEEE Access, vol. 10, pp. 6430-6441, 2022, doi: 10.1109/ACCESS.2021.3140015.

Koroniotis, Nickolaos, Nour Moustafa, Elena Sitnikova, and Benjamin Turnbull. "Towards the development of realistic botnet dataset in the internet of things for network forensic analytics: Bot-iot dataset." Future Generation Computer Systems 100 (2019): 779-796.

Iman Sharafaldin, Arash Habibi Lashkari, and Ali A. Ghorbani, “Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization”, 4th International Conference on Information Systems Security and Privacy (ICISSP), Portugal, January 2018

Sebastian Garcia, Agustin Parmisano, & Maria Jose Erquiaga. (2020). IoT-23: A labeled dataset with malicious and benign IoT network traffic (Version 1.0.0) [Data set]. Zenodo. http://doi.org/10.5281/zenodo.4743746

Meidan, Y.; Bohadana, M.; Mathov, Y.; Mirsky, Y.; Shabtai, A.; Breitenbacher, D.; Elovici, Y. N-BaIoT—Network-Based Detection of IoT Botnet Attacks Using Deep Autoencoders. IEEE Pervasive Comput. 2018, 17, 12–22.

Moustafa, Nour, and Jill Slay. "UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)." Military Communications and Information Systems Conference (MilCIS), 2015. IEEE, 2015.

Ahmad, Rasheed & Alsmadi, Izzat & Alhamdani, Wasim & Tawalbeh, Loai. (2021). A comprehensive deep learning benchmark for IoT IDS. Computers & Security. 114. 102588. 10.1016/j.cose.2021.102588.

B. Nugraha, A. Nambiar and T. Bauschert, "Performance Evaluation of Botnet Detection using Deep Learning Techniques," 2020 11th International Conference on Network of the Future (NoF), 2020, pp. 141-149, doi: 10.1109/NoF50125.2020.9249198.

IoT-23 Veriseti (2022), https://www.stratosphereips.org/datasets-iot23, Erişim: 10 Temmuz 2022

CTU-13 Veriseti (2013), https://www.stratosphereips.org/datasets-ctu13, Erişim: 10 Temmuz 2022

KDD-CUP99 Veriseti (1999), http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html , Erişim: 10 Temmuz 2022

NSL-KDD Veriseti (2009), https://www.unb.ca/cic/datasets/nsl.html, Erişim: 10 Temmuz 2022

IoTID20 Veriseti (2020), https://sites.google.com/view/iot-network-intrusion-dataset/home, Erişim: 10 Temmuz 2022

Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011.