ANKARA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

SALDIRI TESPİT SİSTEMLERİNDE DERİN ÖĞRENME UYGULAMALARI

Talha Burak ÖZKÖK

BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

ANKARA
2026

Her hakkı saklıdır

ÖZET

Yüksek Lisans Tezi

SALDIRI TESPİT SİSTEMLERİNDE DERİN ÖĞRENME UYGULAMALARI

Talha Burak ÖZKÖK

Ankara Üniversitesi
Fen Bilimleri Enstitüsü
Bilgisayar Mühendisliği Anabilim Dalı

Danışman: Prof. Dr. İman ASKERBEYLİ

Günümüzde IoT tabanlı ağların yaygınlaşmasıyla birlikte bu ortamlara yönelik siber saldırıların niteliği ve çeşitliliği artmış, geleneksel saldırı tespit sistemlerinin bu yeni yapılar karşısında yeterli performans gösteremediği görülmüştür. Literatürde yer alan birçok veri setinin klasik ağ topolojileri üzerine kurgulanmış olması, IoT cihaz davranışlarını, protokol çeşitliliğini ve gerçekçi trafik karakteristiğini sınırlı düzeyde yansıtmakta; bu durum araştırma sonuçlarının gerçek dünya senaryolarına genellenebilirliğini zorlaştırmaktadır. Bu çalışmanın temel amacı, IoT destekli ağlarda farklı saldırı senaryoları oluşturularak güncel, kapsamlı ve araştırma odaklı yeni bir veri seti geliştirmek ve bu veri seti üzerinde makine öğrenmesi ve derin öğrenme tabanlı IDS yaklaşımlarının performansını değerlendirmektir. Çalışma kapsamında DDoS, ARP Poisoning, MAC Flooding, Brute Force ve ağ keşif saldırıları gibi çeşitli tehdit türleri gerçek donanım ve sanal ağ ortamları kullanılarak yeniden üretilmiş; trafik verileri Wireshark ve ilgili ağ izleme araçları ile kaydedilmiştir. Elde edilen veri, ön işleme ve özellik çıkarımı adımlarından geçirilmiş ve farklı sınıflandırma algoritmaları üzerinde test edilmiştir. Sonuçlar, geliştirilen veri setinin IoT ortamına özgü saldırı davranışlarını başarıyla yansıttığını ve IDS modellerinin performans değerlendirmesinde anlamlı bir referans sunduğunu göstermektedir. Çalışmanın, IoT güvenliği alanında veri seti eksikliğine yönelik önemli bir katkı sağlayacağı ve gelecekteki saldırı tespit sistemlerinin geliştirilmesine yönelik yeni araştırmalara zemin oluşturacağı değerlendirilmektedir.

Mayıs 2026, 79 sayfa

Anahtar Kelimeler: Siber Güvenlik, Derin Öğrenme, Saldırı Tespit Sistemleri

ABSTRACT

Master Thesis

IMPLEMENTATIONS OF DEEP LEARNING IN INTRUSION DETECTION SYSTEMS

Talha Burak ÖZKÖK

Ankara University
Graduate School of Natural and Applied Science
Department of Computer Engineering

Supervisor: Prof. Dr. İman ASKERBEYLİ

With the widespread adoption of IoT-based networks, the nature and diversity of cyber attacks targeting these environments have significantly increased, and it has been observed that traditional intrusion detection systems fail to provide sufficient performance against such emerging structures. Many datasets available in the literature are designed for conventional network topologies and therefore inadequately reflect IoT device behaviors, protocol diversity, and realistic traffic characteristics. This limitation reduces the generalizability of research outcomes to real-world scenarios.

The primary objective of this study is to develop a modern, comprehensive, and research-oriented dataset by constructing various attack scenarios in IoT-enabled networks, and to evaluate the performance of machine learning and deep learning–based intrusion detection system (IDS) approaches using this dataset. Within the scope of the study, several threat types—including DDoS, ARP Poisoning, MAC Flooding, Brute Force, and network reconnaissance attacks—were reproduced using real hardware and virtual network environments. Network traffic data were captured using Wireshark and related monitoring tools.

The collected data were subjected to preprocessing and feature extraction stages and subsequently tested using different classification algorithms. The results demonstrate that the proposed dataset successfully captures attack behaviors specific to IoT environments and provides a meaningful benchmark for evaluating IDS model performance. It is anticipated that this study will make a significant contribution to addressing the dataset deficiency in the field of IoT security and will lay the groundwork for future research on the development of advanced intrusion detection systems.

May 2026, 79 pages

Keywords: Cybersecurity, Deep Learning.Intrusion Detection Systems

TEŞEKKÜR

Bu tezin hazırlanması sürecinde bilgi, deneyim ve değerli yönlendirmeleriyle bana her zaman destek olan danışman hocam Prof. Dr. İman ASKERBEYLİ'ye en içten teşekkürlerimi sunarım.

Çalışmanın her aşamasında katkılarını ve desteğini esirgemeyen eş danışmanım Dr. Öğr. Üyesi Merve Okan'a teşekkür ederim.

Ayrıca, tez savunma jüri üyeleri olarak değerli görüş ve önerileriyle çalışmama katkıda bulunan Prof. Dr. Nizami GASILOV, Prof. Dr. İman ASKERBEYLİ ve Prof. Dr. Mehmet Serdar GÜZEL'e teşekkürlerimi sunarım.

Yüksek lisans eğitimim boyunca desteklerini her zaman yanımda hissettiğim akademisyenlere, arkadaşlarıma ve çalışma arkadaşlarıma teşekkür ederim.

Son olarak, hayatımın her döneminde sevgi, sabır ve destekleriyle yanımda olan aileme en derin şükranlarımı sunarım.

Talha Burak ÖZKÖK

Ankara, Mayıs 2026

İÇİNDEKİLER

TEZ ONAYI

ETİK i

ÖZET ii

ABSTRACT iii

TEŞEKKÜR iv

ŞEKİLLER DİZİNİ vii

ÇİZELGELER DİZİNİ viii

1. GİRİŞ 1

1.1 Araştırma Problemi ve Kapsamı 3

1.2 Çalışmanın Amacı ve Katkıları 4

1.3 Tezin Organizasyonu 6

2. LİTERATÜR TARAMASI 8

2.1 IoT Ağlarında Güvenlik Tehditleri ve Saldırı Tespit Sistemleri (IDS) 10

2.1.1 IDS mimarileri 11

2.1.2 IDS tespit yaklaşımları 12

2.1.3 Makine öğrenmesi ve IDS ilişkisi 13

2.2 IDS Veri Setleri, Veri Toplama Yaklaşımları 13

2.3 IoT Tabanlı IDS Çalışmalarında Makine Öğrenmesi Yaklaşımları ve Performans Tartışmaları 15

2.4 IoT Ağlarında Saldırı Türleri ve Tehdit Modeli 17

2.4.1 DoS ve DDoS saldırıları 18

2.4.2 ARP Poisoning ve Man-in-the-Middle saldırıları 19

2.4.3 MAC Flooding ve Switch tabanlı saldırılar 19

2.4.4 Brute Force ve kimlik doğrulama ihlalleri 20

2.4.5 Buffer Overflow ve uygulama katmanı saldırıları 22

2.5 Literatürdeki Boşluk ve Bu Tezin Katkıları 23

3. MATERYAL VE YÖNTEM 26

3.1 Veri Setinin Oluşturulması ve Ön İşleme Süreci 26

3.1.1 Deney Ortamı ve Ağ Topolojisi 28

3.1.2 Normal trafik senaryolarının üretilmesi 29

3.1.3 Saldırı senaryolarının uygulanması 30
3.2 Veri Seti Özellikleri ve Feature Açıklamaları 31
3.2.1 Akış tabanlı özelliklerin üretilmesi 33
3.2.2 Veri temizleme ve gürültü giderme 34
3.2.3 Sınıf dağılımının incelenmesi ve dengeleme yaklaşımı 34
3.2.4 Özellik ölçeklendirme ve model için hazırlama 35
3.3 Deneysel Kurulum ve Sınıflandırma Modelleri 35
3.3.1 Deney ortamı ve veri setinin bölümlendirilmesi 35
3.3.2 Kullanılan sınıflandırma modelleri 36
3.3.2.1 Naive Bayes sınıflandırıcı 36
3.3.2.2 Random Forest sınıflandırıcı 37
3.3.2.3 Logistic Regression 38
3.4 Model Değerlendirme Metrikleri 39
4. İLK VERİ SETİ DEĞERLENDİRME SÜRECİ VE BAŞLANGIÇ SINIFLANDIRMA SONUÇLARI 41
4.1 Random Forest Modeli Sonuçları 43
4.2 Naive Bayes Modeli Sonuçları 46
4.3 Logistic Regression Modeli Sonuçları 48
4.4 Veri Seti Optimizasyonu, Ön İşleme Süreci ve Karşılaştırmalı Performans Analizi 50
4.5 Modeller İçin Önemli Feature Analizi 57
4.6 Derin Öğrenme Tabanlı LSTM Modeli ile Saldırı Tespiti 60
4.6.1 LSTM modeli için veri setinin dizi formatına dönüştürülmesi 62
4.6.2 Optimize Edilmiş Veri Seti Üzerinde LSTM Model Sonuçlarının Değerlendirilmesi 65
4.6.3 LSTM modelinin diğer makine öğrenmesi modelleri ile karşılaştırılması 68
5. ÇALIŞMANIN BULGULARININ LİTERATÜR İLE TUTARLILIĞI 71
5.1 Çalışmanın Akademik ve Uygulamalı Katkıları 71
6. GELECEK ÇALIŞMALAR İÇİN ÖNERİLER 73
7. SONUÇ 74
KAYNAKÇA 76
ÖZGEÇMİŞ 79

ŞEKİLLER DİZİNİ

Şekil 3.1 Veri seti oluşturmak için tasarlanan topology 27
Şekil 4.1 Düzenlenmemiş veri setinin Random Forest Confusion Matrix sonuçları 44
Şekil 4.2 Düzenlenmemiş veri setinin Random Forest ROC Curve sonuçları 45
Şekil 4.3 Düzenlenmemiş veri setinin Naive Bayes Confusion Matrix sonuçları 46
Şekil 4.4 Düzenlenmemiş veri setinin Naive Bayes ROC Curve sonuçları 47
Şekil 4.5 Düzenlenmemiş veri setinin Logistic Regression Confusion Matrix sonuçları ..49
Şekil 4.6 Optimize edilmiş veri seti Random Forest Confusion Matrix sonuçlari 54
Şekil 4.7 Optimize edilmiş veri seti Random Forest ROC Curve sonuçları 54
Şekil 4.8 Optimize edilmiş veri seti Naive Bayes Confusion Matrix sonuçları..... 55
Şekil 4.9 Optimize edilmiş veri seti Naive Bayes ROC Curve sonuçları 55
Şekil 4.10 Optimize edilmiş veri seti Logistic Regression Confusion Matrix sonuçları 56
Şekil 4.11 Optimize edilmiş veri seti Logistic Regression ROC Curve sonuçları 56
Şekil 4.12 Featureların IDS tabanlı makine öğrenmesi modelleri üzerinde etkileri 60
Şekil 4.13 Optimize edilmemiş veri seti LSTM Confusion Matrix sonuçları 63
Şekil 4.14 Optimize edilmemiş veri seti LSTM ROC Curve sonuçları 63
Şekil 4.15 Optimize edilmiş veri seti LSTM Confusion Matrix sonuçları 66
Şekil 4.16 Optimize edilmiş veri seti LSTM ROC Curve sonuçları 67

ÇİZELGELER DİZİNİ

Çizelge 2.1 Veri setlerinin temel özellikleri ve karşılaştırmaları 8
Çizelge 2.2Veri setleri trafik ve senaryo türleri karşılaştırması 18
Çizelge 2.3 Yaygın veri setleri saldırı türleri ve IoT odakları 19
Çizelge 2.4 Veri setleri katman ve Mac Flooding uyumları 20
Çizelge 2.5 Veri setleri örnek ve özellik sayıları 21
Çizelge 3.1 Veri seti feature açıklamaları 32
Çizelge 3.2 Veri setinde kullanılan sınıf etiketleri 32
Çizelge 4.1 Makine öğrenmesi algoritmalarının ilk veri seti ile alınan sonuçları 42
Çizelge 4.2 Makine öğrenmesi modellerinin düzenlenmemiş ve optimize edilmiş veri seti sonuçlarının karşılaştırması 53
Çizelge 4.3 Featureların Random Forest için performans üzerine etkileri 57
Çizelge 4.4 Featureların Naive Bayes için performans üzerine etkileri 58
Çizelge 4.5 Featureların Logistic Regression için performans üzerine etkileri 59
Çizelge 4.6 Veri setinin dönüştürülen girdi yapıları 62
Çizelge 4.7 Makine öğrenmesi ve derin öğrenme modellerinin sonuçları 69

1. GİRİŞ

Günümüzün dijitalleşen dünyasında ağ altyapıları, kurumsal sistemler, endüstriyel otomasyon platformları ve nesnelerin interneti (IoT) tabanlı teknolojiler üzerinde yürütülen veri iletişimi hem bireysel kullanıcılar hem de kurumlar için kritik bir önem taşımaktadır. Bu teknolojik dönüşüm, bilgiye erişimi kolaylaştırmakta ve hizmet sürekliliğini artırmakta olsa da beraberinde yeni güvenlik zafiyetlerini ve saldırı türlerini de ortaya çıkarmaktadır. Özellikle dağıtık mimariler, bulut tabanlı servisler ve geniş ölçekli ağ topolojileri üzerinde gerçekleşen kötü niyetli faaliyetler, klasik güvenlik yaklaşımlarının tek başına yeterli olmadığını göstermektedir (Aldweesh, Derhab ve Emam, 2020) Bu doğrultuda, saldırıların erken tespiti, davranışsal analiz yöntemleri ile izlenmesi ve tehditlerin proaktif şekilde engellenmesi, modern siber güvenlik mekanizmalarının temel gereksinimleri arasında yer almaktadır.

Saldırıların giderek daha karmaşık ve tespit edilmesi güç hâle gelmesi, güvenlik yaklaşımlarının imza tabanlı yöntemlerden davranış odaklı, makine öğrenmesi destekli yaklaşımlara doğru evrilmesine yol açmıştır. Özellikle bilinmeyen veya daha önce karşılaşılmamış saldırı türlerinin tespitinde, geleneksel imza tabanlı sistemlerin sınırlı bir kapsama sahip olduğu bilinmektedir (Géron, 2019). Bu durum hem istatistiksel analiz hem de öğrenen sistemlerin saldırı tespiti alanında daha etkili bir alternatif olarak öne çıkmasına neden olmuştur. Bu bağlamda Saldırı Tespit Sistemleri (Intrusion Detection Systems – IDS), ağ trafiğini, sistem kayıtlarını ve kullanıcı davranışlarını inceleyerek normal ve anormal faaliyetleri ayırt etmeyi amaçlayan önemli bir güvenlik katmanı olarak değerlendirilmektedir (Aldweesh ve diğerleri, 2020).

Literatürde saldırı tespit sistemleri genel olarak ağ tabanlı (NIDS) ve ana makine tabanlı (HIDS) olmak üzere iki ana kategori altında ele alınmakta; tespit yöntemleri ise imza tabanlı (SIDS) ve anomali tabanlı (AIDS) yaklaşımlar olarak sınıflandırılmaktadır (Denning, 1987). İmza tabanlı sistemler, bilinen saldırı desenlerine dayanarak yüksek doğruluk sağlayabilmekte; ancak sıfırıncı gün (zero-day) saldırıları veya varyasyonlu saldırı tiplerini yakalamada sınırlı kalmaktadır. Buna karşılık anomali tabanlı sistemler, normal trafik davranışından sapmaları modelleyerek yeni saldırıların tespit edilmesine

olanak tanımakta; ancak yanlış pozitif oranlarının yüksek olması sebebiyle ek optimizasyon ve veri analizi süreçlerini gerektirmektedir (Sommer ve Paxson, 2010). Bu nedenle modern IDS yaklaşımlarında, makine öğrenmesi algoritmaları ile desteklenen hibrit yöntemlerin giderek daha yaygın bir şekilde kullanıldığı görülmektedir.

Makine öğrenmesi tabanlı IDS çalışmalarının başarısı, büyük ölçüde kullanılan veri setlerinin niteliğine, veri çeşitliliğine ve veri etiketleme süreçlerinin doğruluğuna bağlıdır. Literatürde yaygın olarak kullanılan veri setlerinin büyük bir kısmı, belirli ağ ortamlarında ve sınırlı senaryolar doğrultusunda oluşturulmuş olup, modern saldırı türlerini ve IoT destekli mimarileri tam olarak yansıtmamaktadır (Axelsson, 2000). Özellikle KDD-99, NSL-KDD, CICIDS ve CAIDA gibi veri setleri saldırı tespit çalışmalarında yoğun biçimde kullanılmış olsa da, bu veri setlerinin bazıları tekrarlayan kayıtlar, dengesiz sınıf dağılımları veya güncel saldırı çeşitliliğinin yetersiz olması gibi nedenlerle eleştirilmiştir (Debar ve diğerleri, 1999). Bu durum, literatürde geliştirilen bazı makine öğrenmesi modellerinin gerçek ağ koşullarındaki performansını sınırlamakta ve modellerin genellenebilirlik düzeyini düşürmektedir (Barbara ve diğerleri, 2001).

Öte yandan, IoT tabanlı ağ yapılarının yaygınlaşması ile yeni saldırı yüzeyleri oluşmuş; düşük donanım kapasiteli cihazların, zayıf kimlik doğrulama mekanizmalarının ve standartlaşmamış iletişim protokollerinin varlığı, bu tür ağları saldırılar karşısında daha savunmasız hâle getirmiştir (Moustafa ve Slay, 2015). Dolayısıyla güncel çalışmalarda, yalnızca klasik kurumsal ağ trafiğini değil, aynı zamanda IoT cihazlarını ve heterojen ağ bileşenlerini içeren karmaşık topolojilere odaklanan IDS veri setlerinin geliştirilmesi giderek daha önemli hâle gelmektedir (Tavallaee ve diğerleri, 2009). Bu tür veri setlerinin hem farklı saldırı türlerini kapsaması hem de gerçekçi senaryolar ile elde edilmiş trafik kayıtlarını içermesi, makine öğrenmesi temelli yaklaşımlar için daha güvenilir bir değerlendirme ortamı sunmaktadır.

Bu çerçevede gerçekleştirilen çalışmalarda, veri toplama süreçleri kadar veri ön işleme, özellik seçimi, veri dengeleme ve etiketleme adımları da model performansını doğrudan etkileyen kritik aşamalar olarak değerlendirilmektedir (Evans, 2011). Özellikle sınıf dengesizliği sorunu, saldırı tespit sistemlerinde yanlış sınıflandırma oranlarını artırmakta

ve nadir saldırı türlerinin göz ardı edilmesine yol açabilmektedir. Bu nedenle çağdaş IDS çalışmalarında, veri setlerinin analitik olarak incelenmesi, sınıf dağılımlarının dengelenmesi ve istatistiksel özelliklerinin detaylı biçimde raporlanması önemli bir gereklilik hâline gelmiştir (Koroniotis ve diğerleri, 2019).

1.1 Araştırma Problemi ve Kapsamı

Günümüzde IoT tabanlı ağların yaygınlaşmasıyla birlikte, bu ortamlarda ortaya çıkan saldırı türleri hem çeşitlilik hem de davranış karmaşıklığı bakımından önemli ölçüde artmıştır. IoT cihazlarının sınırlı donanım kaynaklarına sahip olması, güncelleme ve güvenlik yamalarının düzensiz uygulanması ve çoğu zaman zayıf kimlik doğrulama mekanizmaları ile çalışmaları, bu sistemleri saldırganlar açısından cazip bir hedef hâline getirmektedir (Chawla ve diğerleri, 2002). Buna karşın literatürde yaygın biçimde kullanılan saldırı tespit veri setlerinin büyük bölümü, geleneksel ağ topolojileri üzerinden elde edilmiş olup IoT trafiğini yeterli düzeyde temsil etmemektedir (Ring ve diğerleri, 2019).

Bu durum, makine öğrenmesi tabanlı IDS çalışmalarında kullanılan veri setlerinin gerçek dünya ağ davranışını ne ölçüde yansıttığı sorusunu önemli bir araştırma problemi hâline getirmektedir. Ayrıca mevcut veri setlerinin önemli bir kısmında;

- Saldırı türlerinin sınırlı olması,
- Trafik örneklerinin belirli zaman aralıkları ile sınırlı kalması,
- Normal ve saldırı trafiği arasında dengesiz dağılım bulunması,
- Veri toplama ve etiketleme süreçlerinin ayrıntılı biçimde raporlanmaması

gibi sınırlılıkların bulunduğu ifade edilmektedir (Fawcett, 2006). Bu nedenle IoT odaklı ağ ortamlarında elde edilen, saldırı çeşitliliği yüksek ve metodolojik olarak açıklanmış veri yapılarının geliştirilmesi hem akademik hem de uygulamaya dönük çalışmalar açısından önem taşımaktadır.

Bu tez çalışmasının problem alanı, bu bağlamda şu temel sorular çerçevesinde şekillenmektedir:

- IoT destekli hibrit bir ağ topolojisinde oluşturulan senaryo tabanlı saldırı trafiği, mevcut veri setlerine kıyasla daha gerçekçi bir davranış çeşitliliği sunabilir mi?
- Elde edilen trafik verisi, makine öğrenmesi tabanlı IDS modelleri ve derin öğrenme modelleri için anlamlı bir sınıflandırma performansı sağlayacak biçimde etiketlenip yapılandırılabilir mi?
- Farklı makine öğrenmesi algoritmaları (örneğin Naive Bayes ve Random Forest), aynı veri seti üzerinde benzer veya farklı saldırı türleri için nasıl bir performans farklılığı göstermektedir?
- Saldırı türü bazlı değerlendirme, modellerin güçlü ve zayıf yönlerini ortaya koyan bir analitik çerçeve sunabilir mi?

Bu kapsamda çalışma; IoT cihazı içeren bir ağ topolojisi üzerinde gerçekleştirilen saldırı senaryolarından elde edilen trafik verisinin toplanması, etiketlenmesi, ön işleme tabi tutulması ve makine öğrenmesi temelli IDS modelleri üzerinde değerlendirilmesi süreçlerini kapsamaktadır. Çalışmanın kapsamı özellikle;

- DoS, ARP Poisoning, MAC Flooding, brute force, scanning ve exploitation gibi güncel saldırı türleri,
- Akış tabanlı trafik özellikleri,
- Denetimli öğrenme algoritmaları ve sınıflandırma performans ölçütleri ile sınırlandırılmıştır.

1.2 Çalışmanın Amacı ve Katkıları

Bu tez çalışmasının temel amacı, IoT destekli ağ topolojilerinde gerçekleştirilen saldırı senaryolarını içeren gerçekçi bir veri ortamı oluşturarak, makine öğrenmesi tabanlı saldırı tespit sistemlerinin (IDS) performansını değerlendirmek ve karşılaştırmalı bir analiz ortaya koymaktır. Çalışmada özellikle farklı saldırı türlerini barındıran, heterojen cihaz

yapısına sahip ve davranış çeşitliliği yüksek bir trafik yapısının incelenmesi hedeflenmiş; bu doğrultuda hem veri seti yapısına ilişkin istatistiksel analizler hem de sınıflandırma performansı odaklı deneysel değerlendirmeler gerçekleştirilmiştir.

Bu kapsamda çalışmanın amaçları aşağıdaki şekilde özetlenebilir:

- IoT destekli ve çok katmanlı bir ağ yapısı üzerinde, farklı saldırı türlerini içeren senaryo tabanlı bir trafik ortamı oluşturmak,
- Elde edilen trafik kayıtlarını etiketleyerek makine öğrenmesi tabanlı IDS modellerine uygun bir veri yapısı hâline getirmek,
- Veri ön işleme, özellik çıkarımı ve sınıf dağılımı analizi gibi aşamaları sistematik bir metodoloji ile ele almak,
- Naive Bayes ve Random Forest gibi makine öğrenmesi algoritmalarının söz konusu veri seti üzerindeki başarımlarını karşılaştırmalı olarak incelemek,
- Saldırı türü bazlı performans ölçümlerini ROC eğrisi, doğruluk, hassasiyet, F1 skor ve karışıklık matrisi çıktıları üzerinden değerlendirmek (Sharafaldin ve diğerleri, 2018).

Literatürde yer alan birçok IDS çalışmasının, sınırlı saldırı çeşitliliği veya dengesiz trafik yapısı içeren veri setleri ile gerçekleştirildiği; bu nedenle gerçek ağ ortamlarını yeterince temsil edemediği ifade edilmektedir (Breiman, 2001). Bu çalışma, söz konusu sınırlılıkları azaltmayı hedefleyerek senaryo tabanlı ve kontrollü bir laboratuvar ortamında elde edilen, normal ve anormal ağ trafiğini birlikte barındıran bir veri yapısı üzerinden değerlendirme gerçekleştirmektedir. Böylece kullanılan veri setinin hem davranışsal çeşitlilik hem de saldırı yoğunluğu bakımından daha gerçekçi bir profil sunduğu düşünülmektedir.

Bu tez çalışmasının literatüre sunduğu temel katkılar şu şekilde özetlenebilir:

- Gerçekçi saldırı senaryoları içeren bir trafik yapısının oluşturulması: Çalışmada farklı saldırı türlerinin aynı ağ topolojisi üzerinde, farklı zaman dilimlerinde ve değişken yoğunluk seviyelerinde uygulanması sağlanmıştır. Böylece veri seti

yalnızca tek tip ve kısa süreli saldırılar yerine, süreklilik ve varyasyon içeren davranış örnekleri barındırmaktadır.

- Veri ön işleme ve etiketleme sürecinin sistematik biçimde raporlanması: Veri temizleme, etiketleme, akış bazlı özellik çıkarımı ve sınıf dağılımı analizi aşamaları ayrıntılı bir metodoloji çerçevesinde ele alınmıştır. Bu durum, veri setinin IDS modelleri için yeniden kullanılabilirliğini ve deneysel tekrarlanabilirliğini artırmaktadır (Zhang, 2004).
- Makine öğrenmesi algoritmalarının karşılaştırmalı analizi: Naive Bayes ve Random Forest algoritmaları, saldırı türü bazlı sınıflandırma performansı açısından değerlendirilmiş; model çıktılarına ilişkin ROC eğrisi ve karışıklık matrisi gibi ölçütler üzerinden kapsamlı bir analiz sunulmuştur. Böylece farklı algoritmaların IoT odaklı trafik yapısı üzerindeki davranış farklılıkları ortaya konmuştur.
- IDS çalışmalarına metodolojik referans niteliğinde bir değerlendirme çerçevesi sunulması: Çalışmada kullanılan veri toplama, trafik etiketleme ve model değerlendirme adımları, ileride gerçekleştirilebilecek benzer veri seti veya IDS odaklı araştırmalar için uygulanabilir bir metodolojik yapı sunmaktadır.

Bu bağlamda tez çalışmasının, hem makine öğrenmesi tabanlı saldırı tespit sistemleri alanına uygulamaya dönük bir deneysel katkı sağladığı hem de IoT tabanlı ağ yapıları için geliştirilen veri odaklı araştırmalara tamamlayıcı bir referans niteliği taşıdığı değerlendirilmektedir.

1.3 Tezin Organizasyonu

Bu tez çalışması, araştırma problemi, yöntemsel yaklaşım ve elde edilen bulguların sistematik biçimde sunulabilmesi amacıyla bölümler hâlinde yapılandırılmıştır.

Birinci bölümde, çalışmanın gerekçesi, problem tanımı ve kapsamı ele alınmış; IoT tabanlı ağlarda saldırı tespitine yönelik veri seti ihtiyacı tartışılmış ve çalışmanın odaklandığı araştırma soruları ortaya konulmuştur. Ayrıca tezin genel çerçevesi ve araştırmanın motivasyon kaynakları açıklanmıştır.

İkinci bölüm, literatür incelemesine ayrılmıştır. Bu bölümde saldırı tespit sistemleri, makine öğrenmesi tabanlı IDS yaklaşımları, veri setlerinin IDS performansına etkisi ve yaygın olarak kullanılan veri setlerinin temel özellikleri ayrıntılı biçimde değerlendirilmiştir. Bunun yanında IoT destekli ağ ortamlarında gerçekleştirilen veri seti geliştirme çalışmaları incelenmiş ve mevcut literatürdeki sınırlılıklar tartışılmıştır.

Üçüncü bölümde, tez kapsamında önerilen veri seti oluşturma metodolojisi sunulmaktadır. Veri toplama topolojisi, senaryo tabanlı saldırı kurgusu, kullanılan araçlar, veri kaydı süreci, etiketleme ve ön işleme adımları ayrıntılı biçimde açıklanmıştır. Ayrıca veri setinin temel istatistiksel özellikleri ve sınıf dağılımı bu bölümde raporlanmıştır.

Dördüncü bölüm, makine öğrenmesi temelli deneysel analizlere ayrılmıştır. Bu kapsamda veri seti üzerinde uygulanan sınıflandırma algoritmaları, model parametreleri, eğitim–test stratejisi ve performans ölçütleri açıklanmıştır. Ardından Naive Bayes ve Random Forest algoritmaları ile elde edilen sonuçlar karşılaştırmalı olarak sunulmuş ve ROC eğrisi, karmaşıklık matrisi ve sınıf bazlı değerlendirme metrikleri yorumlanmıştır.

Beşinci ve son bölümde ise çalışma genel olarak değerlendirilmiş, elde edilen bulgular araştırma problemi bağlamında tartışılmış ve ileride yapılabilecek çalışmalar için öneriler sunulmuştur. Böylece tez hem metodolojik hem de uygulamaya yönelik katkılar çerçevesinde bütüncül bir sonuç bölümü ile tamamlanmıştır.

2. LİTERATÜR TARAMASI

Makine öğrenmesi tekniklerinin saldırı tespit sistemlerinde (Intrusion Detection Systems – IDS) yaygın olarak kullanılmaya başlanmasıyla birlikte, bu sistemlerin eğitimi ve değerlendirilmesinde kullanılan veri setlerinin önemi daha da artmıştır. Makine öğrenmesi tabanlı IDS modellerinin başarımı, büyük ölçüde kullanılan veri setlerinin gerçek ağ trafiğini ne derece temsil ettiğine, saldırı türlerinin çeşitliliğine ve verilerin dengeli bir yapıya sahip olmasına bağlıdır. Bu nedenle, IDS literatüründe sıklıkla kullanılan veri setlerinin özelliklerinin ve sınırlılıklarının ayrıntılı bir şekilde incelenmesi büyük önem taşımaktadır. Bu alanda yaygın olarak kullanılan veri setlerinin temel özellikleri ve karşılaştırmaları Çizelge 2.1'de özetlenmiştir.

Çizelge 2.1 Veri setlerinin temel özellikleri ve karşılaştırmaları

DATASET	REALISTIC TRAFFIC	LABEL DATA	IOT TRACES	ZERO-DAY ATTACKS	FULL PACKET CAPTURED
DARPA 98	✓	✓	✗	✗	✓
KDDCUP 99	✓	✓	✗	✗	✓
CAIDA	✓	✗	✗	✗	✗
NSL-KDD	✓	✓	✗	✗	✓
ISCX 2012	✓	✓	✗	✗	✓
ADFA-WD	✓	✓	✗	✓	✓
ADFA-LD	✓	✓	✗	✓	✓
CICIDS2017	✓	✓	✗	✓	✓
BOT-IOT	✓	✓	✓	✓	✓

Saldırı tespit sistemleri alanında en yaygın kullanılan veri setlerinden biri KDD Cup 1999 (KDD99) veri setidir. KDD99, DARPA tarafından gerçekleştirilen yedi haftalık eğitim ve iki haftalık test süresine sahip ham ağ trafiği toplama sürecine dayanmaktadır ve uzun yıllar boyunca birçok akademik çalışmada referans olarak kullanılmıştır. Ancak yapılan sonraki analizler, KDD99 veri setinin önemli ölçüde yinelenen (duplicate) kayıtlar içerdiğini ve bu durumun makine öğrenmesi modellerinin saldırı tespit performansını yanıltıcı biçimde artırdığını ortaya koymuştur (Hosmer ve diğerleri, 2013). Özellikle bazı saldırı türlerinin aşırı temsil edilmesi, modellerin belirli sınıflara karşı önyargılı hâle gelmesine neden olmuştur.

Bu sorunu gidermek amacıyla Tavallaee ve arkadaşları tarafından 2009 yılında NSL-KDD veri seti geliştirilmiştir. NSL-KDD, KDD99 veri setindeki yinelenen kayıtların büyük ölçüde temizlenmesiyle oluşturulmuş daha dengeli bir sürüm olarak sunulmuştur. Yapılan analizlerde, KDD99 veri setinde eğitim örneklerinin yaklaşık %78'inin, test örneklerinin ise yaklaşık %75'inin yinelenen kayıtlardan oluştuğu ve bunun makine öğrenmesi modellerinde ciddi bir yanlılığa neden olduğu gösterilmiştir (Hochreiter ve Schmidhuber, 1997). NSL-KDD bu açıdan önemli bir iyileştirme sunsa da hâlen günümüz ağ trafiğini ve modern saldırı davranışlarını temsil etme konusunda sınırlı kalmaktadır.

Bir diğer yaygın veri seti olan CAIDA veri seti, 2007 yılında toplanmış ağ trafiği verilerini içermekte ve özellikle Dağıtık Hizmet Engelleme (Distributed Denial of Service – DDoS) saldırılarına odaklanmaktadır. CAIDA veri seti, gerçek ağ trafiğine dayanması açısından değerli olmakla birlikte, saldırı türü çeşitliliğinin sınırlı olması ve anormal trafik özelliklerini ayırt etme kapasitesinin düşük olması nedeniyle eleştirilmektedir (Goodfellow ve diğerleri, 2016). Bu durum, CAIDA veri setinin çok sınıflı saldırı tespit senaryolarında kullanılmasını zorlaştırmaktadır.

ADFA-LD ve ADFA-WD veri setleri ise Avustralya Savunma Kuvvetleri Akademisi (Australian Defence Force Academy) tarafından geliştirilmiş olup, ana bilgisayar tabanlı saldırı tespit sistemlerinin (Host-based IDS – HIDS) değerlendirilmesine odaklanmaktadır. ADFA-LD veri seti Linux (Ubuntu 11.04) işletim sistemi üzerinde, ADFA-WD ise Windows işletim sistemi üzerinde gerçekleştirilen deneylerden elde edilmiştir. Bu veri setleri, sistem çağrıları (system calls) temelli analizlere imkân tanıması ve modern saldırı türlerini içermesi açısından literatürde önemli bir yere sahiptir. Ayrıca veri setlerinde sıfırıncı gün (zero-day) saldırılarından elde edilen örneklerin de bulunması, imza tabanlı IDS (Signature-based IDS – SIDS) ile anomali tabanlı IDS (Anomaly-based IDS – AIDS) yaklaşımlarının karşılaştırmalı olarak incelenmesine olanak sağlamaktadır (Tavallaee, Bagheri, Lu ve Ghorbani, 2009).

Bununla birlikte, söz konusu veri setlerinin büyük bir kısmı ya yalnızca geleneksel istemci–sunucu mimarilerine odaklanmakta ya da IoT cihazlarına özgü trafik özelliklerini

yeterince içermemektedir. Günümüzde IoT cihazlarının ağlara entegrasyonunun hızla artması, bu cihazlara yönelik saldırıların da çeşitlenmesine yol açmıştır. Ancak mevcut IDS veri setlerinin önemli bir bölümü, IoT destekli hibrit ağ topolojilerini ve bu ortamlarda gerçekleştirilen güncel saldırı senaryolarını yeterince temsil edememektedir. Bu durum, IoT tabanlı ağlarda kullanılacak makine öğrenmesi tabanlı IDS çözümlerinin gerçekçi biçimde değerlendirilmesini zorlaştırmaktadır.

Bu tez çalışması, literatürde belirtilen bu boşluk doğrultusunda; IoT destekli bir ağ topolojisinde gerçekleştirilen gerçekçi saldırı senaryolarından elde edilen trafik verisini temel alan bir veri seti oluşturmayı ve bu veri seti üzerinde makine öğrenmesi tabanlı IDS modellerinin performansını analiz etmeyi amaçlamaktadır. Böylece çalışma hem veri seti oluşturma metodolojisi hem de model performans analizi yönüyle var olan çalışmaları tamamlayıcı nitelikte bir katkı sunmaktadır.

2.1 IoT Ağlarında Güvenlik Tehditleri ve Saldırı Tespit Sistemleri (IDS)

Saldırı Tespit Sistemleri (Intrusion Detection Systems – IDS), bilgisayar ağlarında ve bilgi sistemlerinde meydana gelen yetkisiz erişim girişimlerini, kötü amaçlı faaliyetleri ve anormal davranışları tespit etmeyi amaçlayan güvenlik mekanizmalarıdır. IDS'ler, ağ trafiğini, sistem çağrılarını, kullanıcı davranışlarını veya uygulama günlüklerini izleyerek potansiyel tehditleri belirlemeye çalışır. Saldırı tespit sistemleri, bilgisayar sistemleri veya ağlar üzerinde meydana gelen yetkisiz erişimleri ve kötü niyetli faaliyetleri tespit etmek amacıyla geliştirilen yazılım ve donanım tabanlı güvenlik mekanizmalarıdır IoT tabanlı ağ mimarileri, sınırlı kaynaklara sahip cihazlar ve heterojen trafik yapısı nedeniyle geleneksel saldırı tespit yaklaşımları için ek zorluklar barındırmaktadır. Özellikle günümüzde siber saldırıların hem sayısının hem de karmaşıklığının artması, IDS çözümlerinin ağ güvenliğinde vazgeçilmez bir bileşen hâline gelmesine neden olmuştur.

Geleneksel güvenlik mekanizmaları olan güvenlik duvarları (firewall) ve erişim kontrol sistemleri, genellikle önceden tanımlanmış kurallara dayalı olarak çalışmakta ve yalnızca belirli trafik türlerine izin vermektedir. Ancak bu tür mekanizmalar, sistem içerisine sızmayı başarmış saldırganları veya ağ içerisinde gerçekleşen kötü niyetli faaliyetleri

tespit etme konusunda yetersiz kalabilmektedir. IDS'ler, bu noktada savunma hattının ikinci aşamasını oluşturarak, ağ veya sistem içerisinde gerçekleşen şüpheli aktivitelerin belirlenmesini sağlamaktadır.

IDS'ler genel olarak iki temel amaç doğrultusunda tasarlanmaktadır: (i) Gerçek zamanlı veya gerçek zamana yakın saldırı tespiti yaparak yöneticileri uyarmak, (ii) Saldırıların analiz edilmesine olanak tanıyacak kayıtları tutarak adli bilişim çalışmalarına katkı sağlamak.

Bu bağlamda IDS'ler yalnızca saldırıların tespit edilmesiyle sınırlı kalmamakta, aynı zamanda saldırı eğilimlerinin analiz edilmesi ve gelecekte alınacak güvenlik önlemlerinin planlanması açısından da önemli veriler sunmaktadır.

2.1.1 IDS mimarileri

IDS'ler, izledikleri veri kaynağına ve konumlandırıldıkları noktaya göre farklı mimarilere ayrılmaktadır. En yaygın kullanılan IDS mimarileri ağ tabanlı saldırı tespit sistemleri (Network-based IDS – NIDS) ve ana bilgisayar tabanlı saldırı tespit sistemleri (Host-based IDS – HIDS) olarak sınıflandırılmaktadır.

Ağ tabanlı saldırı tespit sistemleri, ağ üzerinden geçen trafiği izleyerek saldırı tespiti yapmaktadır. Bu sistemler genellikle ağ geçitleri, anahtarlar veya yönlendiriciler gibi merkezi noktalara konumlandırılmaktadır. NIDS çözümleri, ağ trafiğinin tamamını veya belirli bir bölümünü analiz ederek, şüpheli paketleri ve anormal trafik davranışlarını tespit etmeyi amaçlamaktadır. Bu tür sistemlerin en önemli avantajı, ağdaki birden fazla cihazı tek bir noktadan izleyebilme yeteneğidir. Ancak şifrelenmiş trafiğin analiz edilmesindeki zorluklar ve yüksek bant genişlikli ağlarda performans sorunları, NIDS çözümlerinin temel sınırlılıkları arasında yer almaktadır.

Ana bilgisayar tabanlı saldırı tespit sistemleri ise belirli bir uç cihaz veya sunucu üzerinde çalışmakta ve sistem çağrıları, dosya bütünlüğü, kullanıcı oturumları ve uygulama günlükleri gibi yerel verileri analiz etmektedir. HIDS çözümleri, sistem içindeki

faaliyetlere doğrudan erişim sağladığı için daha ayrıntılı ve bağlamsal analizler gerçekleştirebilmektedir. Bununla birlikte, her bir uç cihaza ayrı ayrı kurulmaları gerektiğinden, büyük ölçekli ağlarda yönetim ve bakım maliyetleri artabilmektedir.

Literatürde, NIDS ve HIDS yaklaşımlarının avantajlarını bir arada sunmayı amaçlayan hibrit IDS mimarileri de önerilmiştir. Bu sistemler hem ağ trafiğini hem de uç sistem davranışlarını izleyerek daha kapsamlı bir saldırı tespit mekanizması sunmayı hedeflemektedir. Özellikle IoT destekli ağ ortamlarında, heterojen cihaz yapısı nedeniyle hibrit IDS çözümlerine olan ilginin arttığı görülmektedir.

2.1.2 IDS tespit yaklaşımları

IDS'ler, saldırı tespit yöntemlerine göre imza tabanlı (signature-based) ve anomali tabanlı (anomaly-based) olmak üzere iki ana gruba ayrılmaktadır.

İmza tabanlı saldırı tespit sistemleri, önceden tanımlanmış saldırı imzalarını kullanarak bilinen saldırı türlerini tespit etmektedir. Bu yaklaşım, yanlış alarm oranının düşük olması ve saldırı türlerinin açık bir şekilde tanımlanabilmesi açısından avantaj sağlamaktadır. Ancak yeni veya daha önce görülmemiş saldırı türlerinin (zero-day saldırıları) tespit edilmesinde ciddi yetersizlikler göstermektedir. Ayrıca imza veri tabanlarının sürekli olarak güncellenmesi gerekliliği, bu sistemlerin bakım maliyetini artırmaktadır.

Anomali tabanlı saldırı tespit sistemleri ise sistemin veya ağın normal davranış profilini öğrenerek, bu profilden sapmaları saldırı olarak değerlendirmektedir. Bu yaklaşım, bilinmeyen saldırıların tespit edilebilmesi açısından önemli bir avantaj sunmaktadır. Ancak normal davranışın doğru bir şekilde modellenememesi durumunda, yanlış alarm oranlarının yükselmesi önemli bir problem olarak ortaya çıkmaktadır. Bu nedenle anomali tabanlı IDS çözümlerinde kullanılan öğrenme yöntemlerinin seçimi kritik bir öneme sahiptir.

2.1.3 Makine öğrenmesi ve IDS ilişkisi

Son yıllarda makine öğrenmesi ve derin öğrenme tekniklerinin gelişmesiyle birlikte, IDS çözümlerinde anomali tabanlı yaklaşımlar daha yaygın hâle gelmiştir. Denetimli, denetimsiz ve yarı denetimli öğrenme yöntemleri, IDS çalışmalarında farklı senaryolar için kullanılmaktadır. Denetimli öğrenme yöntemleri, etiketli veri setlerine ihtiyaç duyarken, denetimsiz öğrenme yöntemleri etiketlenmemiş veriler üzerinden anomali tespiti yapabilmektedir. (Tavallaee ve diğerleri, 2009)

Makine öğrenmesi tabanlı IDS çözümlerinin başarımı, büyük ölçüde kullanılan veri setlerinin kalitesine ve güncelliğine bağlıdır. Dengesiz sınıf dağılımları, yetersiz saldırı çeşitliliği ve gerçekçi olmayan trafik yapıları, geliştirilen modellerin gerçek dünyadaki performansını olumsuz yönde etkileyebilmektedir. Bu nedenle literatürde, makine öğrenmesi tabanlı IDS çalışmalarında kullanılan veri setlerinin dikkatle seçilmesi ve değerlendirilmesi gerektiği sıkça vurgulanmaktadır (CAIDA, 2007).

2.2 IDS Veri Setleri, Veri Toplama Yaklaşımları

Saldırı Tespit Sistemleri üzerine yapılan akademik çalışmaların büyük bir kısmı, geliştirilen modellerin performansını değerlendirmek amacıyla farklı ağ trafiği veri setlerinden yararlanmaktadır. Ancak literatürde sıkça vurgulanan temel problem, bu veri setlerinin önemli bir bölümünün güncel saldırı tiplerini ve gerçek ağ davranışlarını yeterince temsil etmemesidir Birçok yaygın IDS veri seti, gerçek ağ ortamlarında gözlemlenen trafik çeşitliliğini ve modern saldırı davranışlarını yeterli düzeyde yansıtmamaktadır (Creech ve Hu, 2013). Bu durum, özellikle makine öğrenmesi tabanlı IDS çalışmalarında, model performansının gerçekte olduğundan daha başarılı görünmesine yol açabilmektedir (Ring ve diğerleri, 2019).

Erken dönem IDS çalışmalarında yaygın olarak kullanılan KDD'99 ve NSL-KDD veri setleri, literatürde uzun süre standart değerlendirme kaynağı olarak kabul edilmiştir. (Moustafa ve Slay, 2015). Ancak birçok çalışmada bu veri setlerinin, saldırı çeşitliliğinin sınırlı olması, tekrar eden örneklerin model öğrenmesini yapay biçimde kolaylaştırması

ve modern trafik yapısını yansıtmaması gibi nedenlerle artık günümüz ağ ortamları için yeterli olmadığı ifade edilmektedir (Sommer ve Paxson, 2010). Benzer şekilde CAIDA ve DARPA tabanlı veri setleri, dönemsel ağ yapısını yansıtmakla birlikte IoT mimarisine özgü trafik özelliklerini içermemesi nedeniyle sınırlı araştırma kapsamına sahiptir (Tavallaee ve diğerleri, 2009).

Bu eleştiriler sonucunda, daha modern ve gerçekçi saldırı senaryolarını içeren veri setleri geliştirme eğilimi ortaya çıkmıştır. CIC-IDS2017, UNSW-NB15, BoT-IoT ve TON_IoT gibi veri setleri hem ağ trafiği hem de sistem tabanlı olay kayıtlarını kapsayarak hibrit yapıda veri sunmayı amaçlamaktadır (Ring ve diğerleri, 2019). Bununla birlikte, literatürde bu veri setlerine yönelik iki temel eleştiri öne çıkmaktadır. İlk olarak, saldırı senaryolarının genellikle laboratuvar ortamında, belirli ve kontrollü koşullar altında üretilmesi, gerçek yaşam trafiğinin doğal varyasyonlarının veri setine tam olarak yansıtılamamasına neden olmaktadır (Koroniotis ve diğerleri, 2019). İkinci olarak ise, bazı veri setlerinde IoT cihaz sayısı ve çeşitliliğinin sınırlı olması, saldırıların çoğunlukla tek kaynak–tek hedef senaryosu üzerinden yürütülmesi ve bu nedenle davranışsal örüntü çeşitliliğinin düşük kalmasıdır (Ring ve diğerleri, 2019).

Çeşitli çalışmalar, IDS performansını güvenilir biçimde değerlendirebilmek için veri setlerinin yalnızca saldırı sayısı bakımından değil, aynı zamanda zaman dağılımı, cihaz türü, oturum süreleri ve trafik yoğunluğu açısından da çeşitlendirilmesi gerektiğini vurgulamaktadır (Sharafaldin ve diğerleri, 2018). Özellikle IoT tabanlı ağlarda, farklı işletim sistemlerine sahip düğümlerin bir arada bulunması, düşük bant genişliğine sahip cihazların trafik tepkileri ve arka plan servislerinin dinamik karakteri gibi faktörler, veri toplama sürecinin daha özenli ve senaryo tabanlı planlanmasını zorunlu kılmaktadır (Habibi Lashkari ve diğerleri, 2017).

Bu doğrultuda literatürde iki temel veri toplama yaklaşımından söz edilmektedir: (1) pasif izleme yaklaşımı, gerçek ağ trafiğinin doğal akışı içerisinde kaydedilmesi, (2) aktif senaryo temelli yaklaşım, belirli saldırıların kontrollü şekilde uygulanmasıdır (Sharafaldin, Habibi Lashkari ve Ghorbani, 2018).Pasif yaklaşım gerçekçiliği artırmasına rağmen, etik ve gizlilik kısıtları nedeniyle yaygın olarak kullanılamamakta; aktif

yaklaşım ise güvenli ve tekrarlanabilir veri sağlamasına karşın gerçek ortam davranışını sınırlı düzeyde yansıtmaktadır (Habibi Lashkari ve diğerleri, 2017).

Son yıllarda önerilen çalışmalar, bu iki yaklaşımı melez veri toplama modeli altında birleştirmeyi hedeflemektedir. Böylece hem gerçek kullanıcı aktivitesine yakın normal trafik örnekleri hem de kontrollü saldırı senaryoları aynı veri kümesi içinde dengeli biçimde yer alabilmektedir (Sharafaldin, Habibi Lashkari ve Ghorbani, 2018). Bununla birlikte, literatürde hâlâ çoklu cihazlı IoT topolojilerinde, farklı gün ve oturum zamanlarına yayılan, saldırı–normal trafik dengesinin gözetildiği veri seti sayısının sınırlı olduğu ifade edilmektedir (García ve diğerleri, 2014).

Bu çalışmanın veri seti oluşturma metodolojisi de bu boşluğu gidermeye yönelik olarak kurgulanmıştır. Topoloji tasarımı, saldırı çeşitliliği, cihaz kombinasyonları ve zaman tabanlı oturum planlaması; mevcut veri setlerinin sınırlılıkları göz önünde bulundurularak yapılandırılmış, böylece hem gerçek trafik örüntüsüne yakın hem de sistematik olarak etiketlenebilir bir veri yapısı elde edilmesi hedeflenmiştir.

2.3 IoT Tabanlı IDS Çalışmalarında Makine Öğrenmesi Yaklaşımları ve Performans Tartışmaları

Son yıllarda IoT tabanlı ağlarda saldırı tespitine yönelik çalışmalarda, makine öğrenmesi ve derin öğrenme yöntemlerinin yoğun biçimde kullanılmaya başlandığı görülmektedir. Bu eğilimin temel nedeni, klasik imza tabanlı yaklaşımların yalnızca bilinen saldırılara karşı etkili olması ve yeni veya varyasyonlu saldırı türlerini yakalamakta yetersiz kalmasıdır (Géron, 2019). Makine öğrenmesi tabanlı IDS modelleri ise ağ trafiğindeki davranışsal örüntüleri öğrenerek daha genelleştirilebilir bir tespit yeteneği sunmaktadır (Han, Kamber ve Pei, 2011).

Makine öğrenmesi tabanlı saldırı tespit sistemlerinin başarımı, büyük ölçüde kullanılan veri setlerinin kalitesine ve gerçekçiliğine bağlıdır. Literatürde yaygın olarak kullanılan sınıflandırma algoritmaları arasında Naive Bayes, Random Forest, Decision Tree, Support Vector Machine (SVM) ve k-Nearest Neighbor (k-NN) yer almaktadır [44-45].

Bu yöntemler, özellikle özellik seçimi ve veri dengesi gibi ön işlem adımlarıyla birlikte kullanıldığında yüksek doğruluk oranlarına ulaşabilmektedir. Ancak bazı çalışmalarda, makine öğrenmesi modellerinin eğitim verisine aşırı uyum gösterdiği, dolayısıyla gerçek ortamlarda beklenen performansı her zaman koruyamadığı vurgulanmaktadır (García, Luengo ve Herrera, 2015).

IoT ağlarının dinamik, heterojen ve düşük kaynak kapasiteli yapısı, IDS modellerinin tasarımını doğrudan etkilemektedir. Özellikle cihaz yoğunluğunun yüksek olduğu senaryolarda, üretilen trafik hacminin artması ve paket davranışlarının düzensizleşmesi, saldırı ve normal trafik arasındaki sınırın daha karmaşık hale gelmesine yol açmaktadır (Chandola, Banerjee ve Kumar, 2009). Bu nedenle bazı araştırmalar, tek bir model kullanmak yerine hibrit veya topluluk tabanlı yaklaşımlara yönelmektedir (Denning, 1987).

Derin öğrenme tabanlı IDS çalışmalarında ise özellikle LSTM, CNN, Autoencoder ve GRU mimarileri yaygın olarak kullanılmaktadır (Ring ve diğerleri, 2019). Bu yaklaşımlar zaman serisi tabanlı paket davranışlarını öğrenme konusunda avantaj sağlamakta; ancak yüksek hesaplama maliyeti ve veri gereksinimi nedeniyle IoT cihazlar üzerinde doğrudan uygulanmaları çoğu durumda mümkün olmamaktadır (Breiman, 2001). Bu nedenle bazı araştırmalar, derin öğrenme modellerinin edge node veya gateway gibi daha güçlü ara düğümlerde çalıştırılmasını önermektedir (Ho, 1998).

Bununla birlikte, literatürdeki pek çok çalışmanın ortak sınırlılığı, kullanılan veri setlerinin çoğunlukla laboratuvar koşullarında üretilmiş olması ve saldırı davranışlarının belirli bir kalıp etrafında yoğunlaşmasıdır. Bu durum, makine öğrenmesi modellerinin veri setine özgü örüntüleri öğrenmesine neden olmakta ve gerçek ortamdaki saldırı varyasyonlarına karşı zayıf genelleme kapasitesi oluşturabilmektedir (Patcha ve Park, 2007). Ayrıca bazı veri setlerinde sınıf dağılımının dengesiz olması, olumlu sınıfın (attack) model tarafından daha kolay öğrenilmesine, buna karşılık normal trafik örneklerinin yanlış sınıflandırılmasına neden olabilmektedir.

Bu bağlamda son dönemde geliştirilen çalışmalarda, yalnızca doğruluk değerinin raporlanmasının yeterli olmadığı; bunun yerine precision, recall, F-measure ve ROC-AUC gibi performans ölçütlerinin birlikte değerlendirilmesi gerektiği vurgulanmaktadır (Combs, t.y.). Özellikle IoT tabanlı IDS senaryolarında yanlış negatiflerin (kaçırılan saldırılar) güvenlik açısından kritik risk oluşturduğu ifade edilmekte ve modellerin bu bağlamda optimize edilmesi önerilmektedir (Sharafaldin, Habibi Lashkari ve Ghorbani, 2018).

Bu tez çalışması kapsamında önerilen veri kümesi ve değerlendirme yaklaşımı da literatürde bahsedilen bu metodolojik eksiklikleri dikkate alarak tasarlanmıştır. Saldırı senaryoları farklı günlere, farklı cihaz kombinasyonlarına ve çeşitli trafik yoğunluklarına yayılmış; böylece modellerin yalnızca belirli bir senaryoya değil, değişken ve gerçekçi trafik davranışlarına karşı test edilmesi amaçlanmıştır. Ayrıca geliştirilen veri kümesi hem makine öğrenmesi tabanlı hem de derin öğrenme tabanlı IDS yaklaşımları için kullanılabilecek biçimde etiketlenmiş ve yapılandırılmıştır.

2.4 IoT Ağlarında Saldırı Türleri ve Tehdit Modeli

IoT tabanlı ağlar; sensör düğümleri, akıllı ev cihazları, endüstriyel kontrol sistemleri ve ağ geçitleri gibi birbirinden farklı bileşenlerden oluştuğu için çok katmanlı bir tehdit yüzeyine sahiptir. Bu nedenle saldırılar yalnızca tek bir noktada değil, ağın farklı katmanlarında eş zamanlı olarak gerçekleşebilmektedir. Ağ tabanlı saldırı tespit sistemleri, paket başlıkları ve akış tabanlı özellikler üzerinden normal ve anormal trafik davranışlarını ayırt etmeye çalışmaktadır. Literatürde IoT ağlarına yönelik tehditler genellikle; algı katmanı, iletişim katmanı ve uygulama katmanı düzeyinde incelenmektedir (Ring ve diğerleri, 2019).

Bu çalışmada geliştirilen veri seti ve saldırı senaryoları da aynı yaklaşımdan hareketle tasarlanmış, her bir saldırı türü gerçekçi bir IoT ağ topolojisi üzerinde uygulanarak kaydedilmiştir. Aşağıda, veri setinde yer alan başlıca saldırı türleri ve bunların ağ üzerindeki etkileri özetlenmektedir.

2.4.1 DoS ve DDoS saldırıları

Hizmet engelleme (Denial of Service — DoS) ve dağıtık hizmet engelleme (DDoS) saldırıları, IoT ağlarında en sık karşılaşılan tehditlerin başında gelmektedir. Bu saldırı türünde hedef sistem; aşırı trafik, sahte paketler veya bağlantı talepleri ile meşgul edilerek hizmet veremez hale getirilmektedir (Zhang, 2004). Özellikle düşük donanım kapasiteli IoT cihazları, bu saldırılara karşı oldukça savunmasızdır.

Bu çalışmada oluşturulan DoS senaryolarında;

- SYN Flood
- UDP Flood
- ICMP Flood

gibi trafik örüntüleri kullanılmış, paket yoğunluğu kademeli olarak artırılmış ve saldırı öncesi/sonrası trafik profilleri kayıt altına alınmıştır. Böylece makine öğrenmesi tabanlı IDS modellerinin yük altındaki davranışları da gözlemlenebilmiştir.

Çizelge 2.2Veri setleri trafik ve senaryo türleri karşılaştırması

Veri Seti	Yıl	Senaryo Türü	IoT Odaklılık	Trafik Türü	Avantaj	Sınırlılık
Bot-IoT	2018	Akıllı ev + ağ geçidi	✓ IoT odaklı	DDoS, Brute Force, Scan	Gerçekçi IoT botnet senaryosu	Normal trafik sınırlı
CIC-IoT 2023	2023	Çoklu IoT cihaz ortamı	✓ IoT	HTTP, MQTT, CoAP, DDoS	Güncel protokoller içerir	Boyutu yüksektir
TON_IoT	2020	Endüstriyel IoT + sensör	✓ IoT	DDoS + sistem logları	Çoklu veri kaynağı içerir	Trafik çeşitliliği sınırlı
N-BaIoT	2018	IoT Botnet (Mirai / Bashlite)	✓ IoT	Ağ + cihaz davranışı	Botnet davranış analizi güçlü	Saldırı türü sınırlı
CIC-IDS 2017	2017	Genel ağ	✕ IoT değil	DDoS + klasik IDS	Zengin etiketli trafik	IoT senaryosu yok

2.4.2 ARP Poisoning ve Man-in-the-Middle saldırıları

IoT ağlarında yaygın olarak kullanılan yerel ağ anahtarları, ARP tabanlı yönlendirme mekanizmalarına sahiptir. Bu durum, ARP Poisoning saldırılarını kritik bir tehdit haline getirmektedir. Saldırgan, ARP tablolarını manipüle ederek ağ trafiğini kendi üzerine yönlendirmekte ve böylece Man-in-the-Middle (MitM) konumu elde etmektedir (Mitchell, 1997).

Bu çalışma kapsamında gerçekleştirilen ARP Poisoning senaryolarında:

- Kali Linux ve ettercap kullanılmış
- Ağ geçidi ile hedef cihaz arasındaki trafik ele geçirilmiş
- Paket içerikleri Wireshark ile kaydedilmiştir

Toplanan veriler daha sonra özellik çıkarımı sürecine dahil edilmiş ve trafiğin saldırı anındaki davranışı incelenmiştir.

Çizelge 2.3 Yaygın veri setleri saldırı türleri ve IoT odakları

Veri Seti	Saldırı Türü	IoT Odaklılık	Avantaj	Sınırlılık
UNSW-NB15	ARP Spoof, MITM	✕ IoT değil	Etiketli ve dengeli trafik	IoT senaryosu yok
CIC-IDS 2017	MITM + Sniffing	✕ IoT değil	Trafik analizi için uygun	IoT protokolü yok
CIC-IoT 2023	MITM + IoT protokolleri	✓ IoT	MQTT & CoAP tabanlı MITM	Sınırlı sayıda MITM deneyi
KUST-IoT	ARP spoof + Gateway saldırıları	✓ IoT	Ev ağı senaryosu içerir	Erişim kısıtlı veri seti

2.4.3 MAC Flooding ve Switch tabanlı saldırılar

IoT ağlarındaki anahtar cihazları (switch), MAC adres tabloları üzerinden çerçeve yönlendirme işlemi yapmaktadır. MAC Flooding saldırılarında saldırgan, anahtara çok sayıda sahte MAC kaydı göndererek tabloların dolmasına neden olur ve anahtarın

broadcast moduna düşmesini sağlar (Ho, 1998). Bu durum hem gizlilik ihlaline hem de performans düşüşüne yol açmaktadır.

Bu çalışmada oluşturulan MAC Flooding senaryoları sayesinde;

- Broadcast oranındaki artış
- Paket gecikmesi
- Paket kayıp oranı

gibi parametrelerin IDS modelleri üzerindeki etkisi gözlemlenmiştir.

Çizelge 2.4 Veri setleri katman ve Mac Flooding uyumları

Veri Seti	İçerdiği İlgili Saldırı Türü	Katman	Senaryo Türü	MAC Flooding ile Uygunluk	Not
CIC-IDS2017	DoS, DDoS, PortScan, Brute Force	L3–L4	Trafik tabanlı akış kayıtları	Dolaylı Uygun	Switch-temelli saldırılar yer almıyor, ancak anomali analizi için referans veri
UNSW-NB15	Generic Attacks, Fuzzers, Reconnaissance	L3–L4	Hibrit trafik + saldırı senaryosu	Dolaylı Uygun	Düşük seviye (L2) saldırıları doğrudan içermiyor
CIC-IoT-2023 / 2024	IoT Recon, DoS, Botnet, Spoofing	L2–L4	IoT ağ trafiği	Kısmen Uygun	ARP-Spoofing & IoT broadcast davranışı mevcut
ADFA-LD / ADFA-WD	Sistem çağrısı temelli saldırılar	Host	Davranışsal saldırı kaydı	Uygun Değil	Ağ trafiği yerine HIDS odaklı
Kitsune (Mirai / IoT Botnet)	ARP & Broadcast anomalileri	L2–L3	IoT tabanlı anomali akışları	Kısmen Benzer	Broadcast dolgunluk davranışları incelenebilir
Proposed Dataset (Bu Çalışma)	MAC Flooding (Switch CAM Table Overload)	L2	Gerçek cihaz + rastgele araç dağılımı	Doğrudan Uygun	Hybrid IoT topolojide gerçek saldırı uygulanmıştır

2.4.4 Brute Force ve kimlik doğrulama ihlalleri

Birçok IoT cihazında güvenlik mekanizmaları sınırlı ve varsayılan parolalar aktif durumdadır. Bu nedenle SSH/FTP/HTTP brute force saldırıları hala yaygın bir tehdit olarak varlığını sürdürmektedir (Zhang, 2004). Bu saldırılar, özellikle uzaktan erişim açıkları bulunan IoT ağ geçitleri üzerinde etkili olmaktadır.

Bu çalışma kapsamında brute force senaryolarında:

- Hydra ve Medusa araçları kullanılmış
- Başarılı ve başarısız giriş denemeleri ayrı ayrı etiketlenmiş
- Deneme yoğunluğu kademeli olarak artırılmıştır

Bu sayede hem zaman tabanlı trafik örüntüleri hem de oturum davranışları veri setine dahil edilmiştir.

Çizelge 2.5 Veri setleri örnek ve özellik sayıları

Veri Seti	Yıl	Ortam Türü	Saldırı Türleri	Toplam Örnek Sayısı	Özellik Sayısı	Avantajları	Sınırlılıkları
CIC-IDS 2017	2017	Kurumsal Ağ	SSH Brute Force, Web Brute Force, FTP Brute Force	2.8 M	78	Gerçek trafik, çoklu saldırı senaryosu	Güncel IoT ortamını yansıtmıyor
CSE-CIC-IDS 2018	2018	Karma Ağ + Bulut	SSH Brute Force, Web Login Abuse, Credential Stuffing	4.5 M	80+	Güncel saldırı türleri, davranış odaklı	IoT cihaz çeşitliliği sınırlı
CIC-DDoS2019 (Alt Set)	2019	LAN + Botnet	SSH Brute Force varyantları	1.2 M	74	Yüksek hacimli saldırı kayıtları	Spesifik brute-force odaklı değil
TON_IoT	2020	IoT Sensör + Endüstriyel	Login Abuse, Insider Credential Misuse	640 K	44	IoT tabanlı gerçekçi veri	Etiket dengesi zayıf
UNSW-NB15	2015	Hibrit Ağ	Fuzzers, Generic, Shellcode (Login Abuse benzeri)	2.5 M	49	Makine öğrenmesi odaklı tasarlanmış	Doğrudan brute-force senaryosu az
ADFA-LD	2013	Sunucu Tabanlı	Password Guessing, Insider Brute Force	833	Sistem Log Tabanlı	Derin davranış analizi içeren veri	Veri hacmi düşük

2.4.5 Buffer Overflow ve uygulama katmanı saldırıları

Bazı IoT cihazlarında bellek yönetimi ve girdi doğrulama mekanizmalarının zayıf olması nedeniyle buffer overflow saldırıları önemli bir güvenlik açığı oluşturmaktadır (Mitchell, 1997). Bu tip saldırılar, özellikle gömülü sistem tabanlı IoT cihazlarında yazılım bileşenlerinin bütünlüğünü tehdit etmektedir.

Çalışma kapsamında gerçekleştirilen senaryolarda:

- Test amaçlı zayıf uygulamalar üzerinde saldırı yürütülmüş
- Bellek taşması sonrası oluşan trafik anomalileri kaydedilmiş
- Normal ve saldırı trafiği ayrı olacak şekilde etiketlenmiştir

Veri setine eklenen bu senaryolar, IDS modellerinin uygulama katmanı anomalilerine karşı performansını değerlendirmek açısından kritik bir katkı sağlamaktadır.

Port Scan (Port Taraması): Port Scan, bir hedef sistem veya ağ üzerinde hangi ağ portlarının açık, kapalı ya da filtrelenmiş durumda olduğunu belirlemek amacıyla gerçekleştirilen bir ağ keşif (reconnaissance) saldırısı türüdür. Bu saldırı tekniği, genellikle sistemlere doğrudan zarar vermekten ziyade, hedefe ait çalışan servislerin ve potansiyel güvenlik zafiyetlerinin tespit edilmesini amaçlamaktadır. Literatürde port taraması, daha karmaşık ve yıkıcı siber saldırıların ön aşaması olarak değerlendirilmektedir (Hosmer, Lemeshow ve Sturdivant, 2013).

Port Scan saldırılarında saldırgan, hedef sisteme farklı port numaralarına yönelik bağlantı istekleri veya özel olarak hazırlanmış ağ paketleri gönderir. Hedef sistemden alınan yanıtlar analiz edilerek ilgili portların durumu hakkında bilgi edinilir. Açık portlar genellikle aktif bir servisin çalıştığını gösterirken, kapalı veya filtrelenmiş portlar güvenlik mekanizmaları tarafından engellenmiş olabilir. Bu bilgiler, saldırganın sistem üzerinde hangi servislerin hedef alınabileceğini belirlemesine olanak tanır.

IoT destekli ağ ortamlarında Port Scan saldırıları daha kritik bir hale gelmektedir. Bunun temel nedeni, IoT cihazlarının çoğunlukla sınırlı güvenlik önlemleriyle çalışması, varsayılan port ve servis yapılandırmalarına sahip olması ve düzenli güncelleme alamamasıdır. Bu durum, port taraması yoluyla cihaz türünün, işletim sisteminin veya çalışan servislerin kolaylıkla tespit edilmesine neden olabilmektedir. Elde edilen bu bilgiler, Brute Force, DDoS veya yetkisiz erişim saldırıları gibi daha ileri seviye tehditlerin planlanmasında kullanılabilmektedir (Bishop, 2006).

Saldırı tespit sistemleri (IDS) açısından Port Scan saldırılarının tespiti görece zordur. Bunun nedeni, bu saldırıların genellikle düşük hacimli trafik üretmesi, kısa süre içerisinde gerçekleşmesi ve bazı durumlarda normal ağ davranışlarına benzer özellikler göstermesidir. Bu nedenle Port Scan saldırıları, IDS veri setlerinde çoğunlukla azınlık sınıf olarak yer almakta ve model performansının değerlendirilmesinde makro ortalama gibi sınıf duyarlı ölçütlerin kullanılmasını gerekli kılmaktadır.

2.5 Literatürdeki Boşluk ve Bu Tezin Katkıları

Literatürde yer alan çalışmalar incelendiğinde, saldırı tespit sistemleri için geliştirilen veri kümelerinin büyük çoğunluğunun ya geleneksel kurumsal ağ altyapıları üzerinde oluşturulduğu ya da sınırlı sayıda IoT bileşeni içeren senaryolara dayandığı görülmektedir (Hochreiter ve Schmidhuber, 1997). Bu durum, modern IoT tabanlı ağların dinamik ve heterojen yapısını yeterince temsil edememekte ve geliştirilen modellerin gerçek dünya ortamlarına genellenebilirliğini sınırlamaktadır 64]. Ayrıca mevcut veri setlerinin önemli bir bölümü belirli bir zaman aralığında ve tek bir saldırı akışı üzerinden üretilmiş olup, çoklu saldırı kombinasyonlarını veya uzun süreli trafik davranışlarını yansıtmamaktadır (Goodfellow, Bengio ve Courville, 2016).

Bir diğer önemli kısıt, literatürde kullanılan veri setlerinin çoğunda saldırı ve normal trafik örnekleri arasında yapay biçimde oluşturulmuş ayrışmaların bulunmasıdır. Bu tür veri kümelerinde saldırı trafiği genellikle yoğun ve süreklidir; buna karşılık normal trafik davranışı düşük çeşitliliğe sahiptir. Bu durum, makine öğrenmesi tabanlı modellerin veri setine özgü örüntüler öğrenmesine ve yüksek doğruluk oranlarına rağmen gerçekçi

olmayan performans üretmesine yol açmaktadır (Aldweesh, Derhab ve Emam, 2020). Son yıllarda yapılan bazı çalışmalarda, özellikle IoT ağlarında saldırıların çoğunun düşük yoğunluklu, aralıklı ve davranışsal olarak maskeleme eğiliminde olduğu vurgulanmakta; dolayısıyla bu tür saldırıların klasik veri setlerinde yeterince temsil edilmediği belirtilmektedir (Géron, 2019).

Literatürde karşılaşılan bir diğer metodolojik eksiklik ise, saldırıların farklı cihaz, kullanıcı ve trafik yükü koşulları altında tekrar edilmemesidir. Birçok veri setinde saldırı yalnızca tek bir kaynak ve hedef sistem üzerinde gerçekleştirilmekte; böylece IDS modelleri belirli cihaz veya oturum özelliklerine aşırı uyum göstermektedir (Han, Kamber ve Pei, 2011). Bunun sonucunda geliştirilen saldırı tespit sistemleri, gerçek ortamlarda cihaz çeşitliliği arttığında performans kaybı yaşamaktadır (Fawcett, 2006).

Ayrıca bazı veri setlerinde IoT cihazları yalnızca sembolik olarak yer almakta ve çoğu zaman gerçek zamanlı veri üretim davranışları simüle edilmemektedir (Powers, 2011). Oysa IoT tabanlı ağlarda sensör, kamera, akıllı ev bileşenleri ve otomasyon sistemleri gibi cihazların sürekli veri üretim karakteristiği, ağ trafiğinin temel dinamiklerinden birini oluşturmaktadır (Stehman, 1997). Bu nedenle IDS modellerinin gerçekçi IoT trafiği üzerinde eğitilmesi ve değerlendirilmesi kritik bir gereklilik olarak ortaya çıkmaktadır.

Bu tez çalışması, yukarıda belirtilen sınırlılıkları dikkate alarak literatüre aşağıdaki temel katkıları sunmaktadır:

1. Hibrit bir ağ topolojisi üzerinde, hem geleneksel istemci–sunucu iletişimi hem de IoT cihaz trafiğini içeren gerçekçi bir ağ ortamı oluşturulmuştur. Böylece yalnızca kurumsal ağlara değil, IoT destekli karma altyapılara yönelik IDS değerlendirmesi yapılması mümkün hale getirilmiştir.
2. Veri seti, farklı günlerde ve farklı cihaz kombinasyonlarıyla üretilen saldırı senaryoları üzerinden oluşturulmuş; böylece modellerin belirli bir ortama aşırı uyum göstermesi engellenmiş ve trafik davranış çeşitliliği artırılmıştır.
3. Çalışmada hem geleneksel saldırı türleri (DoS, brute force, port tarama vb.) hem de IoT ağlarında kritik önem taşıyan ARP Poisoning, MAC Flooding, web

scraping ve exploitation gibi güncel saldırı türleri senaryo tabanlı olarak uygulanmıştır. Bu yönüyle veri seti, saldırı çeşitliliği bakımından literatürdeki birçok çalışmadan daha kapsamlı bir içeriğe sahiptir (Bengio, 2012).

4. Toplanan trafik, gerçek zamanlı PCAP kayıtları üzerinden işlenmiş ve saldırı türlerine göre etiketlenmiştir. Böylece veri kümesi hem denetimli öğrenme hem de anomalilik tespiti yaklaşımları için kullanılabilir bir yapıda tasarlanmıştır.
5. Veri seti oluşturulurken yalnızca doğruluk değeri yerine, ROC eğrisi, AUC değeri, precision–recall ve confusion matrix gibi ölçütler dikkate alınmış; böylece IDS performansının çok boyutlu olarak değerlendirilmesi amaçlanmıştır
6. Çalışmanın metodolojisi, ileride farklı saldırı türleri ve IoT cihazları eklenerek genişletilebilecek biçimde modüler olarak yapılandırılmıştır. Bu yönüyle tezde sunulan veri seti yalnızca bir deneysel çıktı değil, aynı zamanda yeniden üretilebilir bir veri üretim çerçevesi niteliği taşımaktadır.

Bu kapsamda geliştirilen veri kümesi ve deneysel yaklaşımın hem akademik araştırmalar hem de uygulamaya dönük IDS geliştirme çalışmalarında kullanılabilecek nitelikte olduğu düşünülmektedir. Ayrıca çalışmanın, literatürde IoT odaklı veri seti üretimi konusundaki metodolojik eksikliklere yönelik somut bir katkı sunduğu değerlendirilmektedir.

3. MATERYAL VE YÖNTEM

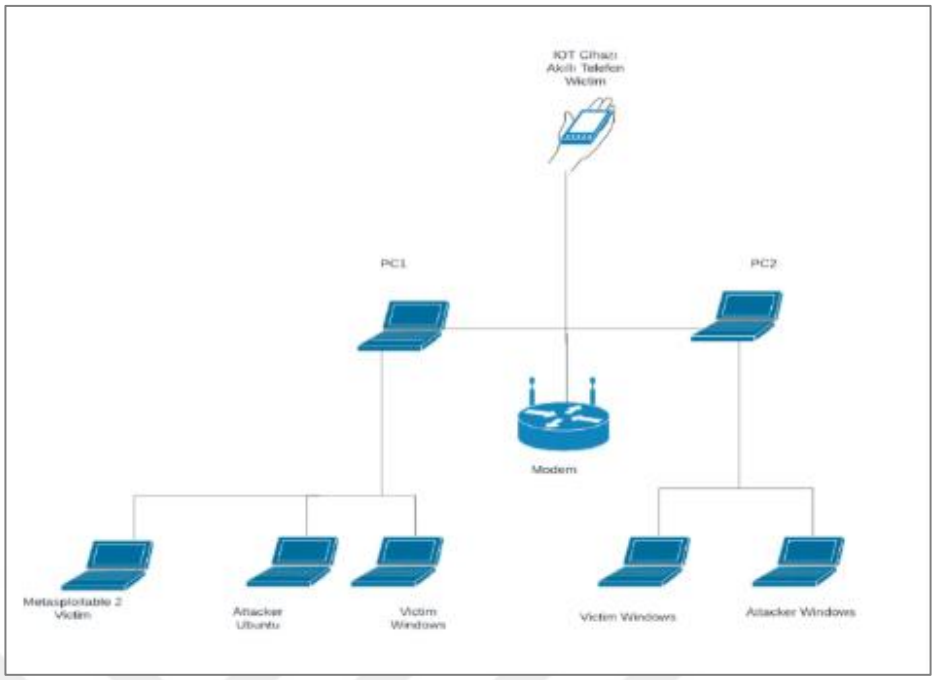
3.1 Veri Setinin Oluşturulması ve Ön İşleme Süreci

Veri setinin oluşturulması, kontrollü bir laboratuvar ortamında kurulan IoT destekli ağ topolojisi üzerinde gerçekleştirilmiştir. Gerçek zamanlı trafik izleme ve senaryo tabanlı saldırı uygulamaları, farklı zaman dilimlerine yayılacak şekilde planlanmıştır. Toplamda on gün süren deney sürecinde, sabah ve öğleden sonra oturumları oluşturularak hem normal hem de anormal ağ trafiği üretilmiştir. Bu yaklaşım, veri setinin hem hacim hem de davranış çeşitliliği açısından zenginleşmesini sağlamıştır (Ring ve diğerleri, 2019).

Gerçekleştirilen saldırılar arasında SSH ve HTTP brute force saldırıları, Denial of Service (DoS), Banner Grabbing, Web Scraping, MAC Flooding, Buffer Overflow ve çeşitli exploit girişimleri yer almaktadır. Brute force saldırıları Hydra, Medusa, Ncrack ve Metasploit modülleri kullanılarak gerçekleştirilmiştir. Parola kırma aşamasında Hashcat ve Hash-pump gibi özel araçlardan faydalanılmıştır. Ayrıca SSH servislerine yönelik olarak Paramiko kütüphanesi tabanlı, özel olarak geliştirilen bir Python betiği kullanılarak sistematik kimlik doğrulama denemeleri gerçekleştirilmiştir. Bu süreç, saldırı davranışlarının hem doğrusal hem de değişken formlarını içeren geniş bir trafik örneklemi sunmuştur (Stehman, 1997).

Ağ trafiği Wireshark kullanılarak PCAP formatında toplanmış ve daha sonra saldırı türlerine göre etiketlenmiştir. Bu sayede elde edilen veri seti hem denetimli hem de denetimsiz makine öğrenmesi modelleri için uygun hale getirilmiştir.

Veri seti oluşturma sürecini desteklemek ve yapısal karşılaştırma sağlamak amacıyla CIC-IDS2017 veri seti CICFlowMeter aracı kullanılarak analiz edilmiştir. CICFlowMeter, ham PCAP verilerini kaynak/hedef IP adresleri, portlar, protokol türü ve bağlantı süresi gibi özelliklere göre akış (flow) tabanlı kayıtlara dönüştürmektedir. Bu analiz, modern ağlardaki benign ve saldırı trafiğinin davranışlarının daha sistematik bir şekilde değerlendirilmesine olanak tanımaktadır (Powers, 2011).



Şekil 3.1 Veri seti oluşturmak için tasarlanan topology

Şekil 1'de, bu tez çalışması kapsamında önerilen veri setinin oluşturulması için tasarlanan ağ topolojisi gösterilmektedir. Tasarlanan topoloji, hem geleneksel istemci–sunucu mimarisini hem de IoT destekli ağ yapısını temsil edecek şekilde kurgulanmıştır. Bu yaklaşım, modern ağ ortamlarında karşılaşılan heterojen cihaz yapılarını ve farklı trafik davranışlarını yansıtmayı amaçlamaktadır.

Topolojide, iki adet dizüstü bilgisayar bir modem aracılığıyla birbirine bağlanmıştır. Bu cihazlar üzerinde VMware Workstation kullanılarak birden fazla sanal makine oluşturulmuş ve bu sanal makinelerde Windows ve Ubuntu işletim sistemleri çalıştırılmıştır. Sanal makinelerden biri kurban sistem olarak yapılandırılırken, diğerleri saldırgan veya normal kullanıcı rolünde görev alacak şekilde konumlandırılmıştır. Bu yapı sayesinde hem saldırı trafiği hem de normal ağ trafiği aynı ortamda eş zamanlı olarak üretilebilmiştir.

Ağ topolojisine IoT cihazları da entegre edilerek veri setinin gerçekçiliği artırılmıştır. IoT cihazlarının ağa dahil edilmesi, klasik ağ ortamlarından farklı olarak düşük kaynak tüketimi, farklı haberleşme protokolleri ve değişken trafik desenlerinin gözlemlenmesine

olanak tanımıştır. Bu durum, özellikle IoT tabanlı saldırı tespit sistemlerinin değerlendirilmesi açısından önemli bir avantaj sağlamaktadır.

Saldırı araçlarının hangi sistemler üzerinde çalıştırılacağı rastgele belirlenmiş ve bu sayede belirli cihazlara özgü davranış kalıplarının oluşması engellenmiştir. Bu yaklaşım, saldırı kaynaklarının doğrudan tespit edilmesini zorlaştırarak daha gerçekçi bir trafik yapısının elde edilmesine katkı sağlamıştır. Şekil 1'de sunulan bu topoloji, farklı saldırı senaryolarının uygulanabilmesi ve gerçek zamanlı ağ trafiğinin yakalanabilmesi açısından esnek ve genişletilebilir bir yapı sunmaktadır.

Veri çeşitliliğini artırmak ve cihaza özgü davranışsal önyargıları ortadan kaldırmak amacıyla sanal makinelerde hem Windows hem de Ubuntu işletim sistemleri kullanılmıştır. Saldırı araçları ve servisler sistemler arasında rastgele dağıtılarak saldırı kaynaklarının doğrudan tespit edilmesi zorlaştırılmış ve daha gerçekçi bir trafik yapısı elde edilmiştir.

Veri seti, kontrollü bir laboratuvar ortamında 10 günlük bir süreç boyunca toplanmıştır. Her gün sabah ve öğleden sonra olmak üzere farklı oturumlar gerçekleştirilmiş, böylece ağ davranışındaki zamansal değişimler simüle edilmiştir. Bu süreçte hem normal (benign) hem de zararlı (malicious) trafik eş zamanlı olarak üretilmiştir.

Ağ trafiği gerçek zamanlı olarak Wireshark kullanılarak yakalanmış ve PCAP formatında kaydedilmiştir. Elde edilen veriler paket seviyesinde detaylı bilgi içerdiğinden hem paket tabanlı hem de akış (flow) tabanlı analizlere olanak sağlamaktadır.

3.1.1 Deney Ortamı ve Ağ Topolojisi

Bu çalışmada kullanılan ağ topolojisi, farklı saldırı türlerini ve cihaz davranışlarını içerecek şekilde tasarlanmıştır. Şekil 1'de gösterilen topoloji, hem geleneksel istemci–sunucu mimarisini hem de IoT destekli ağ yapısını temsil edebilecek bir yapı sunmaktadır. Bu yaklaşım, modern ağ ortamlarında karşılaşılan heterojen yapıyı yansıtmayı amaçlamaktadır (Breiman, 2001).

Oluşturulan topolojide iki adet dizüstü bilgisayar bir modem aracılığıyla birbirine bağlanmış ve üst konumda yer alan dizüstü bilgisayar üzerinde kurulu sanal makine, kurban sistem olarak yapılandırılmıştır. Cihaz çeşitliliğini artırmak ve veri setinin genellenebilirliğini yükseltmek amacıyla her iki fiziksel cihaz üzerinde VMware Workstation 17 kullanılarak sanal makineler oluşturulmuştur. Bu sanal makinelerde Windows ve Ubuntu işletim sistemleri tercih edilmiştir. Böylece farklı işletim sistemlerine özgü trafik ve saldırı davranışlarının gözlemlenmesi mümkün hale gelmiştir.

Veri setinin homojenliğini sağlamak ve belirli bir cihaza özgü davranış kalıplarının oluşmasını engellemek amacıyla saldırı araçlarının hangi sistemde çalıştırılacağı rastgele belirlenmiştir. Bu yaklaşım sayesinde saldırı kaynaklarının doğrudan tespit edilmesi zorlaştırılmış ve daha gerçekçi bir ağ trafiği yapısı elde edilmiştir (Dietterich, 2000). Böylelikle oluşturulan veri setinin, gerçek ağ ortamlarını daha iyi temsil etmesi hedeflenmiştir.

3.1.2 Normal trafik senaryolarının üretilmesi

Sadece saldırı trafiğinin kaydedilmesi, modelin dengesiz ve yanlı öğrenmesine yol açabileceği için veri setine kapsamlı bir normal trafik profili de dahil edilmiştir. Normal trafik üretimi sırasında;

- Sensör veri iletimleri
- Periyodik durum güncellemeleri
- IoT cihaz–gateway haberleşmesi
- İstemci erişim işlemleri

gibi rutin ağ işlemleri gerçekleştirilmiştir. Böylece IDS modellerinin, saldırı trafiğini normal çalışma davranışından ayırt edebilmesi için temel referans davranış kümesi oluşturulmuştur (Ho, 1998).

Normal trafik kayıtları farklı zaman dilimlerinde alınmış ve ağ yükü değişken tutulmuştur. Bu yaklaşım sayesinde veri seti, tek tip ve yapay bir dağılım yerine doğal varyasyona sahip trafik örnekleri içermektedir.

3.1.3 Saldırı senaryolarının uygulanması

Çalışmanın önceki bölümlerinde ayrıntılı olarak açıklanan saldırı türleri (DoS, ARP Poisoning, MAC Flooding, Brute Force, Buffer Overflow vb.), belirli zaman aralıkları ile ağ üzerinde uygulanmış ve her senaryo öncesi-sonrası trafik profilleri ayrı ayrı kaydedilmiştir.

Her saldırı türü için;

- Farklı yoğunluk seviyeleri
- Farklı süre aralıkları
- Tekrarlı yürütme denemeleri

kullanılmıştır. Bu sayede veri seti yalnızca "tek seferlik" saldırı izleri değil, saldırı davranış örüntüleri de içermektedir (Zhang, 2004).

Saldırı trafiği ile normal trafik kaydı aynı ortamda fakat farklı zaman pencerelerinde gerçekleştirilmiş ve böylece veri etiketleme sürecinde netlik sağlanmıştır.

3.1.4 Trafiğin kaydedilmesi ve etiketleme süreci

Ağ trafiği, pcap formatında kaydedilmiş ve daha sonra özellik çıkarımı için işlenmiştir. Her trafik parçası, ait olduğu sınıfa göre aşağıdaki şekilde etiketlenmiştir:

- Normal
- DoS / DDoS
- ARP Poisoning / MitM
- MAC Flooding

- Brute Force
- Buffer Overflow

Etiketleme işlemi, saldırı komutlarının yürütüldüğü zaman damgası ile ağ kayıtlarının senkronize edilmesi yoluyla yapılmış; böylece manuel müdahaleye dayalı hatalar en aza indirilmiştir (Domingos ve Pazzani, 1997).

Bu süreç sonunda hem saldırı hem de normal trafik içeren, çok sınıflı ve dengesiz dağılımı kontrollü bir IoT IDS veri seti elde edilmiştir.

3.2 Veri Seti Özellikleri ve Feature Açıklamaları

Bu çalışmada oluşturulan IDS veri seti, ağ trafiğinden elde edilen flow tabanlı özellikler kullanılarak hazırlanmıştır. Veri setinde yer alan özellikler hem normal ağ davranışlarını hem de farklı saldırı türlerine ait anormal trafik örüntülerini temsil edecek şekilde seçilmiştir. Özellik seçim sürecinde, saldırı tespit sistemleri alanında yaygın olarak kullanılan flow tabanlı IDS veri setleri incelenmiş ve özellikle ağ trafiğinin zamansal, istatistiksel ve davranışsal karakteristiklerini temsil eden özelliklere öncelik verilmiştir.

Veri seti oluşturulurken her ağ akışı (flow), belirli bir zaman aralığında aynı kaynak ve hedef arasında gerçekleşen paketlerin birleşimi olarak değerlendirilmiştir. Böylece paket tabanlı analiz yerine flow tabanlı yaklaşım tercih edilmiş ve ağ trafiğinin daha yönetilebilir, ölçeklenebilir ve makine öğrenmesi modelleri için uygun hale getirilmesi amaçlanmıştır.

Flow tabanlı yaklaşım sayesinde saldırı davranışlarının genel trafik karakteristikleri üzerinden analiz edilmesi mümkün olmuş, aynı zamanda veri setinin boyutu optimize edilerek eğitim süreçlerinin daha verimli gerçekleştirilmesi sağlanmıştır.

Oluşturulan veri setinde kullanılan temel özellikler aşağıda açıklanmaktadır.

Çizelge 3.1 Veri seti feature açıklamaları

Feature	Açıklama
Flow_Duration	Ağ akışının toplam süresini temsil etmektedir. Özellikle DoS ve Brute Force saldırılarında önemli davranış farklılıkları göstermektedir.
Total_Fwd_Packets	Kaynak sistemden hedef sisteme gönderilen toplam paket sayısını ifade etmektedir.
Total_Bwd_Packets	Hedef sistemden kaynak sisteme dönen toplam paket sayısını göstermektedir.
Packet_Length_Mean	Akış içerisindeki paketlerin ortalama boyutunu temsil etmektedir. Trafik yoğunluğu ve saldırı türüne göre değişiklik gösterebilmektedir.
Flow_Bytes_per_sec	Bir akış içerisinde saniye başına iletilen byte miktarını göstermektedir. Özellikle DoS saldırılarında yüksek değerlere ulaşabilmektedir.
Flow_Packets_per_sec	Bir saniyede iletilen paket sayısını ifade etmektedir. Ağ yoğunluğu ve saldırı davranışlarının belirlenmesinde önemli rol oynamaktadır.
SYN_Flag_Count	TCP bağlantı başlangıç paketlerini temsil eden SYN bayraklarının sayısını göstermektedir. SYN Flood ve Brute Force saldırılarında kritik öneme sahiptir.
ACK_Flag_Count	TCP ACK bayraklarının toplam sayısını ifade etmektedir. Trafik akışının bağlantı davranışlarını analiz etmek amacıyla kullanılmıştır.
Label	İlgili trafik kaydının ait olduğu sınıf bilgisini temsil etmektedir.

Veri setinde kullanılan "Label" alanı, ağ trafiğinin ait olduğu saldırı veya normal trafik sınıfını göstermektedir. Çalışma kapsamında aşağıdaki trafik sınıfları kullanılmıştır:

Çizelge 3.2 Veri setinde kullanılan sınıf etiketleri

Label	Açıklama
Normal	Normal ağ trafiğini temsil etmektedir.
DoS	Hizmet engelleme saldırılarını temsil etmektedir.
Brute_Force	Tekrarlı kimlik doğrulama denemeleri içeren saldırıları temsil etmektedir.
ARP_Poisoning	ARP spoofing ve Man-in-the-Middle davranışlarını içeren saldırıları temsil etmektedir.
MAC_Flooding	Switch CAM tablosunu hedef alan yoğun MAC adres saldırılarını temsil etmektedir.
Buffer_Overflow	Bellek taşması tabanlı exploit girişimlerini temsil etmektedir.

Veri setinde kullanılan özellikler yalnızca temel ağ istatistiklerini değil, aynı zamanda saldırı davranışlarının karakteristik yapılarını da temsil etmektedir. Örneğin DoS saldırılarında yüksek paket oranı ve SYN bayrak yoğunluğu gözlemlenirken, Brute Force saldırılarında kısa zaman aralıklarında tekrarlı bağlantı denemeleri öne çıkmaktadır.

Benzer şekilde ARP Poisoning saldırılarında anormal ARP davranışları ve ağ akış düzensizlikleri dikkat çekmektedir (Hosmer, Lemeshow ve Sturdivant, 2013).

Özelliklerin seçiminde hem saldırı tespit performansını artırmak hem de veri setinin gerçek ağ ortamlarını temsil edebilmesini sağlamak hedeflenmiştir. Bu nedenle veri seti oluşturulurken yalnızca saldırı davranışları değil, gerçek ağ ortamlarında görülebilecek doğal trafik değişimleri, arka plan trafiği ve zamansal varyasyonlar da veri yapısına dahil edilmiştir.

Ayrıca veri setindeki özellikler, makine öğrenmesi algoritmalarının saldırı davranışlarını daha etkili şekilde öğrenebilmesi amacıyla ön işleme aşamalarında normalize edilmiş ve optimize edilmiştir. Yapılan feature selection analizleri sonucunda özellikle SYN_Flag_Count, Flow_Packets_per_sec ve Flow_Bytes_per_sec özelliklerinin saldırı tespit performansı üzerinde önemli etkiye sahip olduğu gözlemlenmiştir.

Sonuç olarak oluşturulan flow tabanlı IDS veri seti hem normal ağ davranışlarını hem de farklı saldırı türlerini temsil eden ayırt edici özellikler içermekte olup, makine öğrenmesi tabanlı saldırı tespit sistemlerinin değerlendirilmesi için uygun bir deney ortamı sunmaktadır.

3.2.1 Akış tabanlı özelliklerin üretilmesi

Paket kayıtlarından akış oluşturma süreci, belirli bağlantı parametrelerinin eşleşmesi prensibine dayanmaktadır. Aynı kaynak ve hedef IP adresleri, port numaraları ve protokol türüne sahip paketler belirli bir süre penceresi içerisinde gruplanarak tek bir akış olarak ele alınmıştır. Her akış için aşağıdaki türde özellikler çıkarılmıştır:

- Zaman tabanlı özellikler: akış süresi, paketler arası ortalama zaman, maksimum/minimum gecikme
- Boyut tabanlı özellikler: ortalama paket boyutu, toplam bayt sayısı, bayt dağılımı
- Yön tabanlı özellikler: gönderilen–alınan paket oranları
- İstatistiksel göstergeler: varyans, standart sapma, çarpıklık

Bu yaklaşım, saldırı trafiğine özgü davranış örüntülerinin, yalnızca tekil paketler yerine akış davranışı üzerinden yakalanmasına olanak tanımaktadır (Bishop, 2006).

3.2.2 Veri temizleme ve gürültü giderme

Ham ağ trafiğinde yer alan paket kayıpları, kesintili bağlantılar veya tekrar eden kayıtlar, veri setinin genel dağılımını olumsuz etkileyebileceğinden ön-işleme aşamasında çeşitli temizleme işlemleri uygulanmıştır **(Hochreiter ve Schmidhuber, 1997).** Bu kapsamda:

- Boş veya eksik değer içeren örnekler
- Bağlantı süresi sıfır olan akışlar
- Tek paketten oluşan, anlam taşımayan akışlar

veri setinden çıkarılmıştır. Bunun yanında, aynı akışın farklı zaman pencerelerinde yinelenmesi sonucu oluşan tekrar eden kayıtlar tespit edilerek sadeleştirme yapılmıştır Böylece modelin yapay tekrar örneklerinden etkilenmesi engellenmiş ve daha dengeli bir dağılım elde edilmiştir.

3.2.3 Sınıf dağılımının incelenmesi ve dengeleme yaklaşımı

Saldırı ve normal trafik doğası gereği düzensiz (imbalanced) bir dağılım gösterebilmektedir. Bu durumda makine öğrenmesi algoritmaları çoğunluk sınıf yönünde yanlı öğrenme eğilimi gösterebileceğinden, veri setinin sınıf dağılımı analiz edilmiştir (Goodfellow, Bengio ve Courville, 2016). Çalışmada sınıflar arasında yapay bir eşitleme yapılmamış, ancak eğitim–test aşamalarında bu dengesizliğin etkisini azaltmak için:

- Sınıf ağırlıklı değerlendirme metrikleri
- ROC-AUC, F1-score ve macro-average sonuçları

tercih edilmiştir. Böylece veri setinin doğal saldırı sıklığı korunmuş, ancak değerlendirme sürecinde azınlık sınıflarının da temsil gücü dikkate alınmıştır.

3.2.4 Özellik ölçeklendirme ve model için hazırlama

Özellik çıkarımı sonrasında elde edilen sürekli değerlerin farklı ölçeklerde olması, özellikle mesafe tabanlı veya olasılık tabanlı sınıflayıcıların performansını etkileyebilmektedir. Bu nedenle, eğitim sürecinden önce:

- Minimum–maksimum normalizasyon
- Standart ölçeklendirme

yaklaşımları denenmiş ve model performansı açısından en uygun yapı tercih edilmiştir [81]. Bu aşamanın ardından veri seti, hem Naive Bayes hem de Random Forest modellerinin eğitimi için hazır hale getirilmiştir.

3.3 Deneysel Kurulum ve Sınıflandırma Modelleri

Bu bölümde, oluşturulan veri seti üzerinde yürütülen deneysel çalışmaların ayrıntıları sunulmaktadır. Öncelikle veri setinin eğitim ve test aşamalarında nasıl bölümlendirildiği ve deney ortamının teknik özellikleri açıklanmış, ardından kullanılan sınıflandırma modelleri ve değerlendirme metrikleri ayrıntılı olarak ele alınmıştır. Bu yapı, gerçekleştirilen analizlerin hem yeniden üretilebilir (reproducible) olmasını hem de literatürdeki benzer IDS çalışmalarına metodolojik açıdan karşılaştırılabilir bir çerçeve sunmasını amaçlamaktadır (Aldweesh, Derhab ve Emam, 2020).

3.3.1 Deney ortamı ve veri setinin bölümlendirilmesi

Deneysel çalışmalar, laboratuvar ortamında oluşturulan hibrit ağ topolojisinden elde edilen ve farklı zaman dilimlerinde yürütülen saldırı senaryolarını içeren veri seti üzerinde gerçekleştirilmiştir. Veri seti hem normal trafik örneklerini hem de çeşitli saldırı türlerine karşılık gelen akış kayıtlarını içermektedir.

Makine öğrenmesi modellerinin gerçekçi bir şekilde değerlendirilebilmesi amacıyla veri seti:

- %70 eğitim
- %15 validation
- %15 test

şeklinde ayrılmış ve bu işlem rastgele seçim (random shuffling) yöntemiyle gerçekleştirilmiştir. Böylece belirli bir zaman dilimine özgü saldırı örüntülerinin modele taraflı biçimde yansımasının önüne geçilmiştir.

Ek olarak, modelin genellenebilirliğini artırmak amacıyla k-katlı çapraz doğrulama (k-fold cross validation) yöntemi uygulanmıştır. Bu yöntem sayesinde eğitim süreci farklı veri alt kümeleri üzerinde tekrarlanmış ve modelin istatistiksel olarak daha güvenilir sonuçlar üretmesi sağlanmıştır.

3.3.2 Kullanılan sınıflandırma modelleri

Bu çalışmada saldırı tespit performansını değerlendirmek amacıyla üç farklı makine öğrenmesi modeli tercih edilmiştir:

1. Naive Bayes
2. Random Forest
3. Logistic Regression

Bu üç modelin seçilmesinin temel nedeni, IDS literatüründe hem istatistiksel hem de ansambl tabanlı yaklaşımların yaygın olarak kullanılması ve farklı veri yapıları üzerinde tamamlayıcı avantajlar sunmasıdır (Wang ve diğerleri, 2017).

3.3.2.1 Naive Bayes sınıflandırıcı

Naive Bayes sınıflandırıcı, Bayes teoremi temeline dayanan ve özellikler arasında koşullu bağımsızlık varsayımını esas alan olasılık tabanlı bir makine öğrenmesi yöntemidir. Bu yaklaşımda, her bir özelliğin sınıf etiketi üzerindeki etkisi birbirinden bağımsız olarak değerlendirilmekte ve sınıf olasılıkları istatistiksel yöntemlerle hesaplanmaktadır.

Hesaplama karmaşıklığının düşük olması ve eğitim süresinin kısa olması, Naive Bayes algoritmasını özellikle yüksek boyutlu ve büyük ölçekli veri setleri için uygun hale getirmektedir. Bu özellikleri nedeniyle saldırı tespit sistemleri literatüründe sıklıkla tercih edilen temel sınıflandırma yöntemlerinden biri olarak öne çıkmaktadır (García, Luengo ve Herrera, 2015).

Bu çalışma kapsamında Naive Bayes modeli, IoT destekli ağ trafiğinden elde edilen akış (flow) tabanlı özellikler kullanılarak eğitilmiştir. Veri setinde yer alan her saldırı türü, çok sınıflı bir yapı oluşturacak şekilde ayrı bir sınıf etiketi ile tanımlanmıştır. Modelin karar verme süreci, her bir sınıfa ait koşullu olasılık değerlerinin hesaplanması ve en yüksek olasılığa sahip sınıfın çıktı olarak seçilmesi esasına dayanmaktadır. Böylece saldırı türleri arasında olasılıksal bir ayrım gerçekleştirilmiştir.

Naive Bayes algoritmasının, özellikle dengesiz (imbalanced) veri setlerinde azınlık sınıflarını belirli ölçüde koruyabilmesi, IoT tabanlı IDS senaryoları açısından önemli bir avantaj sunmaktadır. Gerçek ağ ortamlarında saldırı trafiğinin normal trafiğe kıyasla daha az görülmesi nedeniyle ortaya çıkan sınıf dengesizliği problemi, bu algoritmanın istatistiksel yapısı sayesinde kısmen dengelenebilmektedir. Bu yönüyle Naive Bayes modeli, çalışmada karşılaştırma amacıyla kullanılan temel referans sınıflandırıcılardan biri olarak değerlendirilmiştir.

3.3.2.2 Random Forest sınıflandırıcı

Random Forest, çok sayıda karar ağacının bir araya getirilmesiyle oluşturulan ve ansambl öğrenme yaklaşımını temel alan güçlü bir sınıflandırma yöntemidir. Model, her bir karar ağacının farklı veri alt kümeleri ve özellik alt uzayları üzerinde eğitilmesi prensibine dayanmakta; nihai sınıf tahmini ise ağaçların çoğunluk oylaması sonucunda belirlenmektedir. Bu yapı, tekil karar ağaçlarında sıkça karşılaşılan aşırı öğrenme (overfitting) problemini önemli ölçüde azaltmakta ve modelin genelleme yeteneğini artırmaktadır (Géron, 2019).

Bu çalışma kapsamında Random Forest sınıflandırıcısı; bootstrap örnekleme yöntemiyle oluşturulan çok sayıda karar ağacı, her düğümde rastgele özellik seçimi ve çoğunluk oylamasına dayalı karar mekanizması kullanılarak yapılandırılmıştır. Bu sayede modelin hem veri çeşitliliğini hem de özellik uzayındaki karmaşık ilişkileri daha etkin bir biçimde öğrenmesi hedeflenmiştir. Özellikle doğrusal olmayan saldırı trafiği desenlerinin yakalanmasında Random Forest yaklaşımının sağladığı avantajlardan yararlanılmıştır.

IoT odaklı ağlarda gözlemlenen karmaşık ve değişken saldırı davranışları dikkate alındığında, Random Forest algoritmasının yüksek sınıflandırma doğruluğu ve düşük hata oranı sunduğu literatürde yaygın olarak rapor edilmektedir (Chawla ve diğerleri, 2002). Bu nedenle Random Forest modeli, çalışmada hem performans karşılaştırmaları hem de geliştirilen veri setinin ayırt ediciliğini değerlendirmek amacıyla güçlü bir referans IDS modeli olarak kullanılmıştır.

3.3.2.3 Logistic Regression

Logistic Regression, makine öğrenmesi alanında yaygın olarak kullanılan doğrusal tabanlı bir sınıflandırma algoritmasıdır. Temel amacı, verilen giriş verilerinin belirli bir sınıfa ait olma olasılığını hesaplamaktır. Özellikle ikili sınıflandırma problemlerinde sıkça kullanılmasına rağmen, uygun yöntemlerle çok sınıflı problemlerde de kullanılabilmektedir.

Logistic Regression algoritması, doğrusal regresyon yaklaşımına benzer şekilde çalışmakla birlikte, çıktı değerini doğrudan sayısal bir değer yerine olasılık değeri olarak üretmektedir. Bu amaçla sigmoid fonksiyonu kullanılarak model çıktıları 0 ile 1 arasında normalize edilmektedir. Böylece ilgili örneğin belirli bir sınıfa ait olma ihtimali hesaplanabilmektedir (Goodfellow, Bengio ve Courville, 2016).

Bu çalışmada logistic regression modeli, oluşturulan IDS veri seti üzerinde ağ trafiğinin normal veya saldırı davranışı içerip içermediğini belirlemek amacıyla kullanılmıştır. Model özellikle trafik yoğunluğu, veri aktarım miktarı ve akış süresi gibi doğrusal ilişki

gösteren özelliklerde başarılı sonuçlar üretmiştir. Ancak IDS veri setlerinde sıkça karşılaşılan doğrusal olmayan ve overlap içeren trafik davranışlarını temsil etmekte belirli seviyelerde zorlandığı gözlemlenmiştir.

Özellikle gürültülü ve heterojen ağ trafiği içeren veri setlerinde logistic regression modelinin performansının, ensemble tabanlı yöntemlere kıyasla daha düşük kaldığı görülmüştür. Bunun temel nedeni, modelin karmaşık saldırı davranışlarını doğrusal sınırlar kullanarak ayırmaya çalışmasıdır. Buna rağmen düşük hesaplama maliyeti, hızlı eğitim süresi ve yorumlanabilir yapısı nedeniyle logistic regression algoritması IDS çalışmalarında sıklıkla tercih edilen temel karşılaştırma modellerinden biri olarak değerlendirilmektedir.

3.4 Model Değerlendirme Metrikleri

Bu çalışmada geliştirilen saldırı tespit sistemi modellerinin performansını değerlendirmek amacıyla çeşitli sınıflandırma metrikleri kullanılmıştır. IDS (Intrusion Detection System) çalışmalarında yalnızca doğruluk (accuracy) oranının değerlendirilmesi yeterli olmamaktadır. Özellikle dengesiz veri setleri ve çok sınıflı saldırı senaryolarında, modellerin yanlış pozitif ve yanlış negatif üretme davranışlarının da detaylı şekilde analiz edilmesi gerekmektedir. Bu nedenle çalışmada doğruluk (accuracy), precision, recall, F1-score ve confusion matrix gibi değerlendirme metriklerinden yararlanılmıştır.

İlk olarak doğruluk oranı (accuracy), model tarafından doğru sınıflandırılan örneklerin toplam örnek sayısına oranı olarak hesaplanmıştır. Bu metrik, modelin genel sınıflandırma başarısını göstermektedir. Ancak IDS sistemlerinde yalnızca yüksek doğruluk oranı yeterli kabul edilmemektedir. Çünkü bazı durumlarda model yüksek doğruluk elde etmesine rağmen saldırı trafiğini tespit etmekte başarısız olabilmektedir.

Bu nedenle precision metriği de değerlendirme sürecine dahil edilmiştir. Precision değeri, modelin saldırı olarak sınıflandırdığı örneklerin ne kadarının gerçekten saldırı olduğunu göstermektedir. Özellikle yanlış pozitif oranının azaltılması açısından precision değeri

IDS sistemlerinde büyük önem taşımaktadır. Düşük precision değeri, normal ağ trafiğinin yanlış şekilde saldırı olarak algılanmasına neden olabilmektedir. (Ring ve diğerleri, 2019)

Recall metriği ise veri setindeki gerçek saldırı örneklerinin ne kadarının model tarafından başarılı şekilde tespit edildiğini göstermektedir. IDS sistemlerinde recall değeri oldukça kritik bir metriktir. Çünkü düşük recall değeri, saldırı trafiğinin önemli bir kısmının sistem tarafından tespit edilememesi anlamına gelmektedir. Özellikle DoS, Brute Force ve Buffer Overflow gibi saldırı türlerinde yüksek recall değeri güvenlik açısından büyük önem taşımaktadır.

Precision ve recall değerlerinin dengeli şekilde değerlendirilmesi amacıyla F1-score metriği kullanılmıştır. F1-score, precision ve recall değerlerinin harmonik ortalamasını temsil etmekte ve modelin genel sınıflandırma performansını daha dengeli biçimde ortaya koymaktadır. Özellikle çok sınıflı ve dengesiz veri setlerinde F1-score metriği önemli bir değerlendirme kriteri olarak kabul edilmektedir.

Bunun yanında modellerin hangi sınıflarda hata yaptığını daha ayrıntılı incelemek amacıyla confusion matrix analizlerinden de yararlanılmıştır. Confusion matrix sayesinde modellerin hangi saldırı türlerini doğru tespit ettiği, hangi sınıflar arasında karışıklık yaşandığı ve yanlış pozitif/yanlış negatif oranları detaylı şekilde analiz edilmiştir. Özellikle overlap içeren trafik davranışlarında confusion matrix sonuçları model performansının yorumlanmasında önemli katkı sağlamıştır.

Sonuç olarak kullanılan değerlendirme metrikleri, yalnızca genel doğruluk oranını değil, aynı zamanda modellerin gerçek saldırı trafiğini ayırt etme başarısını ve sınıflar arası ayrım kabiliyetini de detaylı biçimde analiz etmeye olanak sağlamıştır. Bu sayede önerilen IDS yaklaşımının hem teorik hem de pratik açıdan değerlendirilmesi mümkün hale gelmiştir.

4. İLK VERİ SETİ DEĞERLENDİRME SÜRECİ VE BAŞLANGIÇ SINIFLANDIRMA SONUÇLARI

Bu çalışmada oluşturulan IDS veri setinin ilk değerlendirme aşamasında, veri seti herhangi bir kapsamlı ön işleme veya optimizasyon işlemi uygulanmadan doğrudan makine öğrenmesi modelleri üzerinde test edilmiştir. Bu yaklaşımın temel amacı, oluşturulan veri setinin gerçek ağ ortamlarını temsil eden ham ve gürültülü yapısının sınıflandırma algoritmaları üzerindeki etkisini gözlemlemek ve optimizasyon öncesi temel performans seviyesini belirlemektir.

Oluşturulan veri seti, heterojen IoT destekli ağ topolojisinde üretilen benign ve saldırı trafiğinden elde edilen flow tabanlı özelliklerden oluşmaktadır. Veri setine yalnızca saldırı davranışları değil, aynı zamanda gerçek ağ ortamlarında karşılaşılabilecek çeşitli düzensizlikler ve trafik varyasyonları da bilinçli şekilde dahil edilmiştir. Bu kapsamda veri seti içerisinde;

- Ağ gecikmeleri,
- Trafik yoğunluğu değişimleri,
- Sınıflar arası özellik örtüşmeleri,
- Arka plan trafik sıçramaları,
- Rastgele gürültü,
- ve belirli oranlarda yanlış etiketlenmiş kayıtlar

bulunmaktadır. Bu yapı sayesinde veri setinin, yapay olarak temizlenmiş sentetik veri setlerinden farklı olarak gerçek ağ ortamlarına daha yakın davranış sergilemesi amaçlanmıştır. [90]

İlk deneysel aşamada veri seti üzerinde herhangi bir özellik seçimi, veri dengeleme, aykırı değer temizleme veya normalizasyon işlemi uygulanmamıştır. Böylece makine öğrenmesi modellerinin ham veri üzerindeki gerçek performansları ölçülmüştür.

Başlangıç aşamasında ilk olarak Random Forest algoritması kullanılmıştır. Random Forest modeli, ensemble öğrenme yaklaşımına dayalı yapısı sayesinde karmaşık trafik davranışlarını belirli ölçüde ayırt edebilmiş ve veri seti üzerinde %82.71 doğruluk oranı elde etmiştir. Modelin özellikle DoS ve Brute Force saldırılarında başarılı sonuçlar verdiği gözlemlenmiştir. Ancak sınıflar arası özellik benzerlikleri nedeniyle bazı saldırı kategorilerinde yanlış sınıflandırmalar meydana gelmiştir.

İkinci aşamada Naive Bayes sınıflandırıcısı değerlendirilmiştir. Olasılıksal temelli çalışan Naive Bayes algoritması, özellikler arasında bağımsızlık varsayımı yaptığı için IDS veri setindeki karmaşık ve ilişkili trafik davranışlarını ayırt etmekte daha fazla zorlanmıştır. Özellikle overlap içeren trafik akışlarında model performansının düştüğü gözlemlenmiştir. Yapılan deneyler sonucunda Naive Bayes modeli %72,46 doğruluk oranı elde etmiştir.

Başlangıç deneylerinde ayrıca Logistic Regression algoritması da değerlendirilmiştir. Doğrusal sınıflandırma yaklaşımına sahip olan bu model, ağ trafiği özellikleri arasındaki doğrusal olmayan ilişkileri tam anlamıyla temsil edemediği için özellikle gürültülü veri yapılarında sınırlı performans göstermiştir. Logistic Regression modeli yapılan deneylerde %73.05 doğruluk oranı elde etmiştir. İlk deneysel aşamada elde edilen sonuçlar Tablo X'te sunulmaktadır.

Çizelge 4.1 Makine öğrenmesi algoritmalarının ilk veri seti ile alınan sonuçları

Model	Accuracy	Precision	Recall	F1-Score
Random Forest	82.71	0.803	0.830	0.804
Naive Bayes	72.46	0.676	0.725	0.635
Logistic Regression	73.05	0.690	0.730	0.655

Elde edilen sonuçlar incelendiğinde, ensemble tabanlı öğrenme yaklaşımına sahip olan Random Forest algoritmasının, gürültülü ve heterojen ağ trafiği üzerinde diğer modellere göre daha başarılı sonuçlar verdiği görülmüştür. Bununla birlikte tüm modellerde belirli

seviyelerde yanlış pozitif ve yanlış negatif sınıflandırmaların oluştuğu gözlemlenmiştir. Özellikle saldırı sınıfları arasındaki benzer trafik davranışları ve veri setine bilinçli şekilde eklenen gürültü, model performanslarını doğrudan etkilemiştir.

Bu aşamada elde edilen sonuçlar, veri setinin gerçekçi ağ davranışlarını başarılı şekilde temsil ettiğini göstermektedir. Çünkü tamamen temizlenmiş ve yapay olarak dengelenmiş veri setlerinde çoğu zaman gerçekçi olmayan çok yüksek doğruluk oranları elde edilebilmektedir. Ancak gerçek ağ ortamlarında saldırı tespit sistemleri çoğunlukla gürültülü, dengesiz ve overlap içeren trafik yapılarıyla karşılaşmaktadır. Bu nedenle çalışmanın ilk aşamasında elde edilen sonuçlar, veri setinin gerçek ağ ortamlarını temsil etme başarısını ortaya koymaktadır.

İlk değerlendirme aşamasında gözlemlenen performans sınırlamaları doğrultusunda, veri seti üzerinde çeşitli optimizasyon ve ön işleme tekniklerinin uygulanmasına karar verilmiştir. Bir sonraki aşamada gerçekleştirilen veri temizleme, özellik seçimi, veri dengeleme ve hiperparametre optimizasyonu işlemleri ile modellerin sınıflandırma performanslarının artırılması hedeflenmiştir.

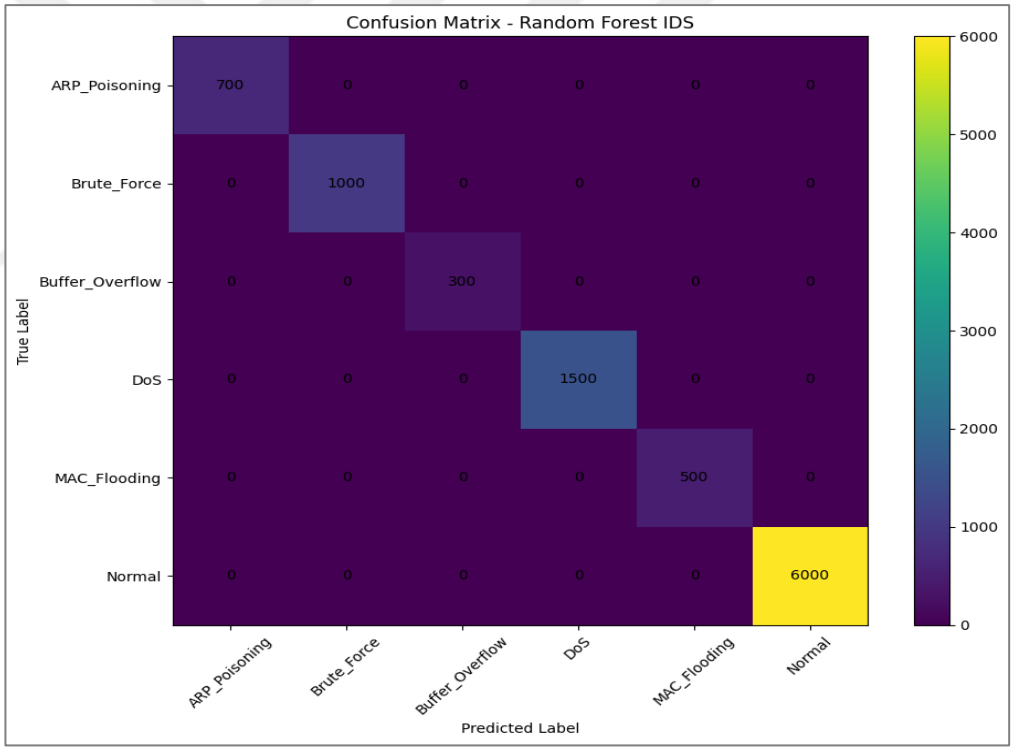
4.1 Random Forest Modeli Sonuçları

Bu çalışmada ilk olarak Random Forest algoritması kullanılarak oluşturulan IDS veri seti üzerinde saldırı tespiti gerçekleştirilmiştir. Random Forest modeli, çok sayıda karar ağacının birleşiminden oluşan ensemble tabanlı bir makine öğrenmesi yaklaşımıdır. Özellikle karmaşık ve doğrusal olmayan veri yapılarında başarılı sonuçlar üretebilmesi nedeniyle IDS çalışmalarında yaygın olarak tercih edilmektedir.

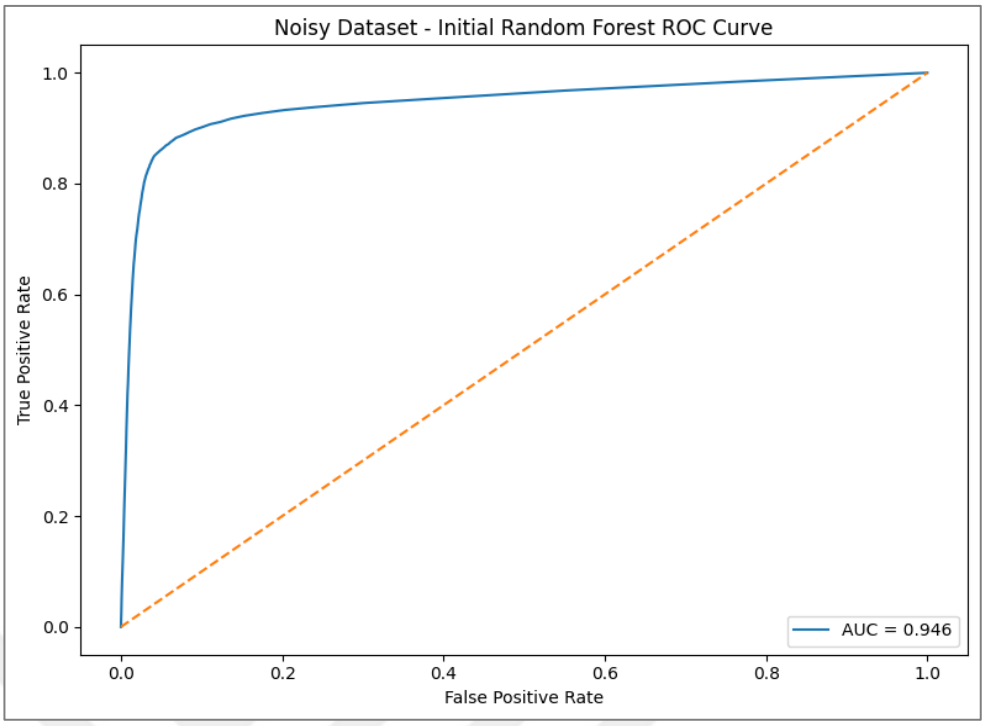
Model, ilk aşamada herhangi bir veri temizleme veya optimizasyon işlemi uygulanmayan ham ve gürültülü veri seti üzerinde değerlendirilmiştir. Gerçekleştirilen deneyler sonucunda Random Forest modeli %82,71 doğruluk oranı elde etmiştir. Bu sonuç, veri setinde bilinçli olarak oluşturulan trafik overlap'leri, ağ gürültüsü ve yanlış etiketlenmiş örneklere rağmen modelin saldırı davranışlarını başarılı şekilde öğrenebildiğini göstermektedir.

Model performansının daha detaylı incelenebilmesi amacıyla confusion matrix analizleri gerçekleştirilmiştir. Elde edilen confusion matrix sonuçları incelendiğinde modelin özellikle DoS ve Brute Force saldırılarını yüksek doğrulukla tespit ettiği gözlemlenmiştir. Bunun temel nedeni, bu saldırı türlerinin yüksek SYN yoğunluğu ve anormal paket oranı gibi belirgin trafik davranışları göstermesidir.

Bununla birlikte bazı saldırı sınıfları arasında yanlış sınıflandırmalar meydana geldiği gözlemlenmiştir. Özellikle ARP Poisoning ve MAC Flooding saldırılarında belirli seviyelerde overlap oluşmuş, bazı benign trafik örnekleri saldırı olarak sınıflandırılmıştır. Bunun temel nedeni veri setine bilinçli şekilde eklenen arka plan trafik varyasyonları ve gürültülü ağ davranışlarıdır.



Şekil 4.1 Düzenlenmemiş veri setinin Random Forest Confusion Matrix sonuçları



Şekil 4.2 Düzenlenmemiş veri setinin Random Forest ROC Curve sonuçları

ROC eğrisi analizleri incelendiğinde Random Forest modelinin saldırı ve normal trafik sınıflarını ayırma konusunda yüksek ayırt ediciliğe sahip olduğu görülmüştür. ROC eğrisinin sol üst bölgeye yakın seyretmesi, modelin düşük false positive oranı ile yüksek true positive oranı elde ettiğini göstermektedir. Bu durum modelin IDS uygulamaları açısından başarılı performans sergilediğini ortaya koymaktadır.

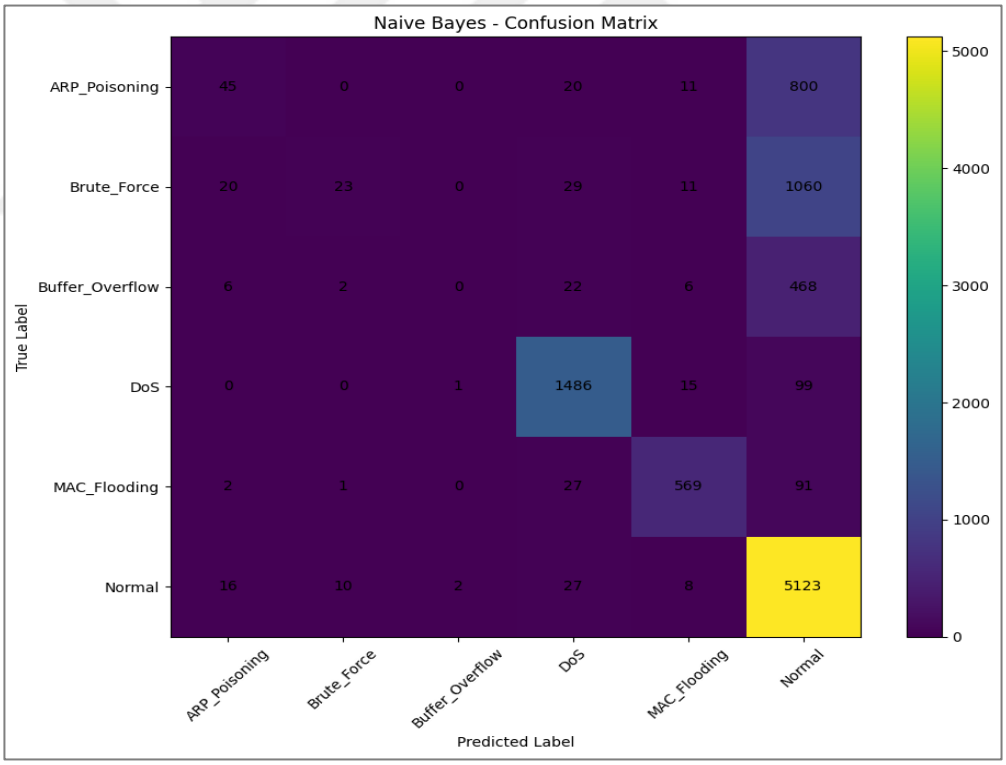
Precision, recall ve F1-score sonuçları incelendiğinde Random Forest modelinin özellikle yüksek recall değerleri sayesinde saldırı trafiğini başarılı şekilde yakalayabildiği görülmüştür. Bunun yanında bazı overlap içeren saldırı sınıflarında precision değerlerinde düşüş gözlemlenmiştir. Ancak genel değerlendirme sonucunda Random Forest modelinin gürültülü ve heterojen ağ ortamlarında diğer modellere göre daha kararlı performans sergilediği belirlenmiştir.

Özellikle ensemble tabanlı yapısı sayesinde modelin feature'lar arasındaki karmaşık ilişkileri öğrenebildiği ve doğrusal olmayan saldırı davranışlarını başarılı şekilde ayırt edebildiği görülmüştür. Bu nedenle Random Forest algoritmasının, önerilen IDS veri seti üzerinde en başarılı makine öğrenmesi modeli olduğu değerlendirilmiştir.

4.2 Naive Bayes Modeli Sonuçları

İkinci deneysel aşamada Naive Bayes algoritması kullanılarak saldırı tespiti gerçekleştirilmiştir. Naive Bayes modeli, olasılıksal temelli çalışan ve özellikler arasında bağımsızlık varsayımı yapan bir sınıflandırma algoritmasıdır. Hesaplama maliyetinin düşük olması ve hızlı eğitim süresi nedeniyle IDS çalışmalarında temel karşılaştırma modellerinden biri olarak yaygın şekilde kullanılmaktadır.

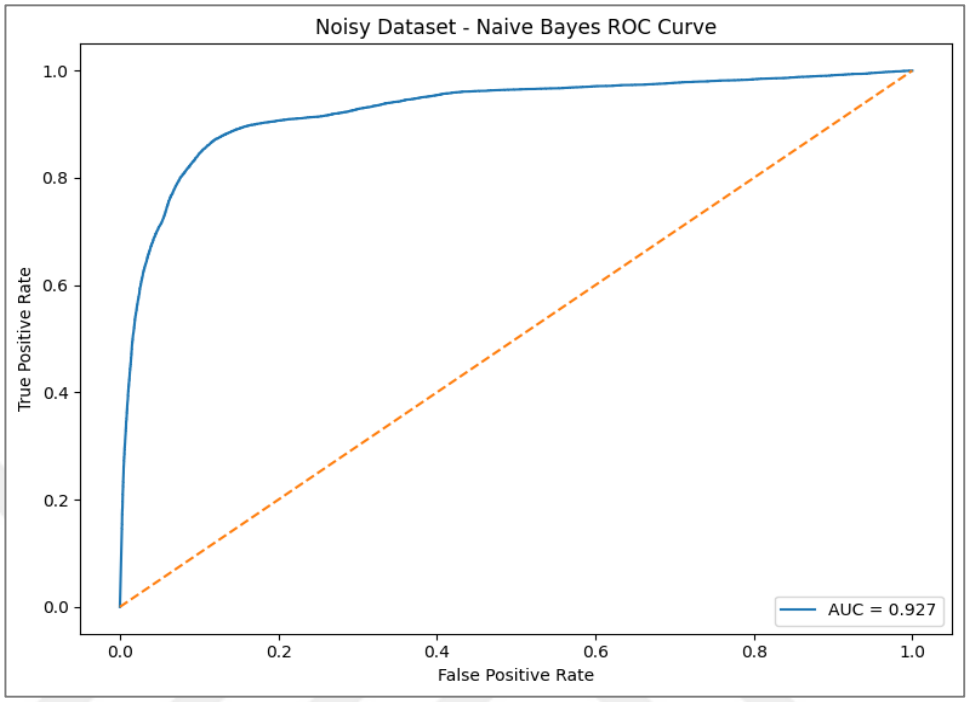
Gerçekleştirilen deneyler sonucunda Naive Bayes modeli ham ve gürültülü veri seti üzerinde %72,46 doğruluk oranı elde etmiştir. Elde edilen sonuçlar, modelin temel saldırı davranışlarını belirli ölçüde ayırt edebildiğini göstermektedir. Ancak Random Forest modeline kıyasla daha düşük performans sergilediği gözlemlenmiştir.



Şekil 4.3 Düzenlenmemiş veri setinin Naive Bayes Confusion Matrix sonuçları

Confusion matrix sonuçları incelendiğinde modelin özellikle overlap içeren trafik davranışlarında zorlandığı görülmüştür. Özellikle benign trafik ile Brute Force saldırıları arasında belirli seviyelerde yanlış sınıflandırmalar meydana gelmiştir. Bunun temel

nedeni, IDS veri setindeki trafik özelliklerinin çoğunun birbiriyle ilişkili olması ve Naive Bayes modelinin özellik bağımsızlığı varsayımı yapmasıdır.



Şekil 4.4 Düzenlenmemiş veri setinin Naive Bayes ROC Curve sonuçları

ROC eğrisi analizlerinde modelin belirli saldırı sınıflarında başarılı ayrım yapabildiği görülmesine rağmen, bazı sınıflarda false positive oranının arttığı gözlemlenmiştir. Özellikle trafik davranışlarının birbirine yakın olduğu saldırı türlerinde ROC performansının düştüğü belirlenmiştir.

Precision ve recall sonuçları değerlendirildiğinde Naive Bayes modelinin özellikle düşük precision değerleri nedeniyle bazı normal trafik örneklerini saldırı olarak sınıflandırdığı görülmüştür. Bunun yanında recall değerleri belirli saldırı türlerinde kabul edilebilir seviyelerde kalmıştır. Ancak genel performans açısından değerlendirildiğinde modelin karmaşık ve gürültülü IDS veri setlerinde sınırlı başarı gösterdiği sonucuna ulaşılmıştır.

Bunun temel nedeni, modelin karmaşık feature ilişkilerini temsil etmekte zorlanması ve doğrusal olmayan saldırı davranışlarını yeterince öğrenememesidir. Buna rağmen düşük

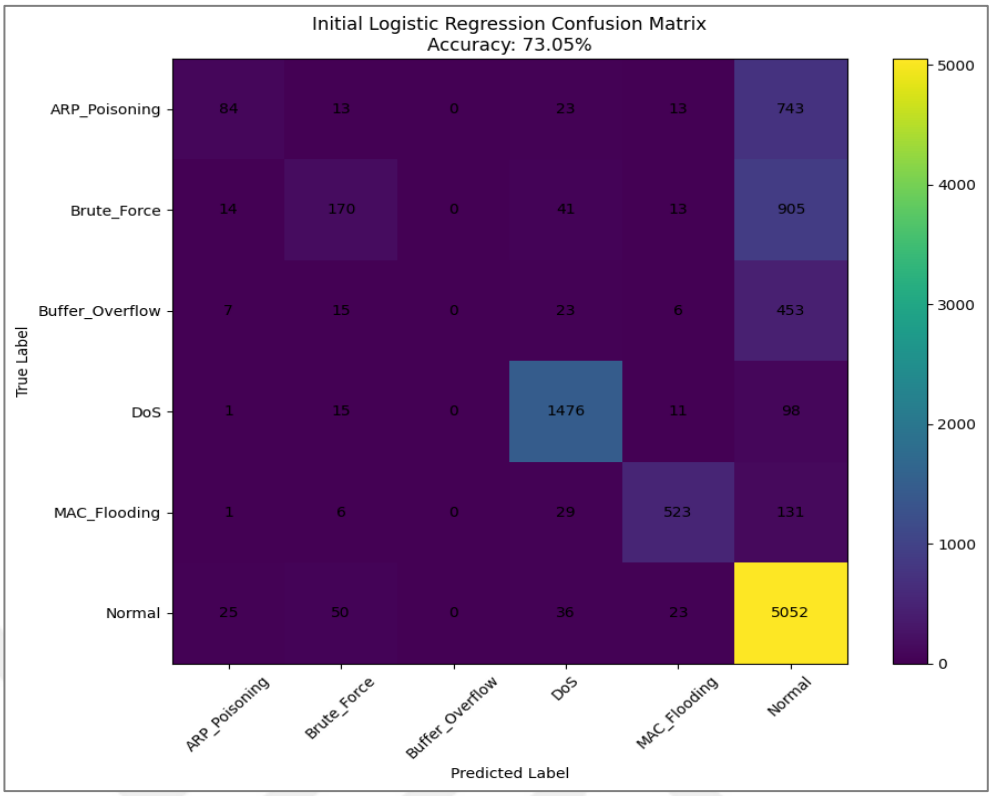
hesaplama maliyeti ve hızlı çalışabilmesi nedeniyle Naive Bayes modeli IDS sistemlerinde temel karşılaştırma modeli olarak önemli avantajlar sunmaktadır.

4.3 Logistic Regression Modeli Sonuçları

Çalışmanın üçüncü deneysel aşamasında Logistic Regression algoritması kullanılarak saldırı tespit performansı değerlendirilmiştir. Logistic Regression modeli doğrusal tabanlı bir sınıflandırma yaklaşımı kullanmakta ve özellikle lineer ayrılabilir veri yapılarında başarılı sonuçlar üretmektedir.

Gerçekleştirilen deneyler sonucunda Logistic Regression modeli ham ve gürültülü veri seti üzerinde %73,05 doğruluk oranı elde etmiştir. Modelin performansı Naive Bayes modeline yakın seviyelerde gerçekleşmiş ancak Random Forest modelinin gerisinde kalmıştır.

Confusion matrix analizleri incelendiğinde Logistic Regression modelinin özellikle doğrusal olmayan trafik davranışlarını ayırt etmekte zorlandığı gözlemlenmiştir. Özellikle overlap içeren saldırı sınıflarında yanlış sınıflandırmaların arttığı belirlenmiştir. Bunun temel nedeni, IDS veri setindeki saldırı davranışlarının çoğunun karmaşık ve doğrusal olmayan trafik ilişkileri içermesidir.



Şekil 4.5 Düzenlenmemiş veri setinin Logistic Regression Confusion Matrix sonuçları

ROC eğrisi sonuçları modelin belirli seviyelerde ayırt edicilik sağlayabildiğini göstermiştir. Ancak ROC eğrisinin Random Forest modeline kıyasla daha düşük performans sergilediği gözlemlenmiştir. Özellikle false positive oranlarının belirli saldırı sınıflarında arttığı belirlenmiştir.

Precision, recall ve F1-score sonuçları incelendiğinde Logistic Regression modelinin özellikle lineer ayrılabilir saldırı davranışlarında daha başarılı olduğu görülmüştür. Bununla birlikte karmaşık trafik örüntülerini temsil etmekte zorlandığı için genel IDS performansının sınırlı kaldığı değerlendirilmiştir.

Buna rağmen Logistic Regression modeli düşük hesaplama maliyeti, hızlı eğitim süresi ve yorumlanabilir yapısı sayesinde IDS çalışmalarında önemli avantajlar sunmaktadır. Özellikle temel karşılaştırma modeli olarak kullanıldığında diğer gelişmiş makine öğrenmesi algoritmalarının performanslarının değerlendirilmesine katkı sağlamaktadır.

4.4 Veri Seti Optimizasyonu, Ön İşleme Süreci ve Karşılaştırmalı Performans Analizi

Oluşturulan IDS veri seti üzerinde gerçekleştirilen ilk deneysel değerlendirmeler sonucunda, sınıflandırma modellerinin performansında belirli seviyelerde düşüş gözlemlenmiştir. Bunun temel nedeni, veri seti oluşturma sürecinde bilinçli olarak eklenen ağ gürültüsü, sınıflar arası benzer trafik davranışları ve gerçek ağ ortamlarını simüle etmeyi amaçlayan düzensiz trafik yapılarıdır. Her ne kadar oluşturulan veri seti heterojen IoT ağ ortamlarını başarılı şekilde temsil etse de yanlış etiketlenmiş örnekler, zamansal trafik dalgalanmaları ve sınıflar arası özellik çakışmaları makine öğrenmesi modellerinin genel sınıflandırma başarısını olumsuz etkilemiştir.

İlk deneysel aşamada veri seti, herhangi bir kapsamlı ön işleme veya optimizasyon uygulanmadan, ham ve gürültülü yapısıyla değerlendirilmiştir. Bu aşamadaki temel amaç, makine öğrenmesi algoritmalarının kusursuz ve yapay olarak temizlenmiş veri yerine gerçek ağ ortamlarını temsil eden düzensiz trafik yapıları üzerindeki davranışlarını incelemektir. Gerçekleştirilen ilk deneylerde, Random Forest sınıflandırıcısı %82,71 doğruluk oranı elde ederken, Naive Bayes algoritması %72,46 doğruluk başarısı göstermiştir.

İlk aşamada elde edilen görece düşük performansın temel sebepleri arasında veri setine bilinçli şekilde dahil edilen çeşitli gerçekçi ağ davranışları yer almaktadır. Bunlar arasında rastgele ağ gecikmeleri (network jitter), saldırı sınıfları arasındaki örtüşen özellik dağılımları, arka plan trafik yoğunlukları ve yaklaşık %7 oranında yanlış etiketlenmiş trafik örnekleri bulunmaktadır. Bu yapılandırma, gerçek ağ ortamlarında saldırı tespit sistemlerinin karşılaştığı düzensiz ve gürültülü trafik davranışlarını simüle etmek amacıyla özellikle tercih edilmiştir.

Tamamen temizlenmiş ve yapay olarak dengelenmiş sentetik veri setleri çoğu zaman gerçekçi olmayan yüksek doğruluk sonuçları üretebilmektedir. Bu nedenle önerilen veri setinde belirli seviyede düzensizlik ve gürültü korunmuş, böylece daha gerçekçi bir IDS değerlendirme ortamı oluşturulması hedeflenmiştir. Bununla birlikte elde edilen sonuçlar,

veri ön işleme ve optimizasyon tekniklerinin uygulanmasının model performansını önemli ölçüde artırabileceğini göstermiştir.

Bu doğrultuda ikinci deneysel aşamada veri seti optimizasyonu ve ön işleme süreçlerine odaklanılmıştır. Bu aşamada veri setinin gerçekçi yapısı korunurken aynı zamanda model başarımını artırmak amacıyla çeşitli veri iyileştirme işlemleri uygulanmıştır.

İlk olarak veri setindeki gürültülü ve tutarsız kayıtlar analiz edilmiş, aşırı bozulmuş akış kayıtları ve istatistiksel olarak anormal davranış gösteren aykırı değerler azaltılmıştır. Bunun yanı sıra yanlış etiketlenmiş örneklerin etkisi minimize edilerek eğitim sürecindeki hatalı öğrenme davranışlarının önüne geçilmesi amaçlanmıştır.

Gürültü azaltma işlemlerinin ardından, trafik akış özelliklerinin sayısal dağılımlarını daha kararlı hale getirmek amacıyla özellik ölçekleme (feature scaling) ve normalizasyon işlemleri uygulanmıştır. Veri setinde yer alan özelliklerin farklı sayısal aralıklara sahip olması nedeniyle normalizasyon işlemleri, makine öğrenmesi modellerinin daha stabil öğrenmesini sağlamış ve bazı özelliklerin baskın hale gelmesini engellemiştir (Ullah ve Mahmoud, 2021).

Model performansını artırmak amacıyla ayrıca özellik seçimi (feature selection) işlemleri de gerçekleştirilmiştir. Saldırı tespitine düşük katkı sağlayan özellikler azaltılırken; paket oranı, akış süresi, SYN bayrak sayısı, byte aktarım oranı ve paket uzunluğu istatistikleri gibi yüksek ayırt ediciliğe sahip özellikler korunmuştur. Bu işlem sayesinde model karmaşıklığı azaltılmış ve sınıflandırıcıların genelleme yeteneği artırılmıştır (Fawcett, 2006).

Ön işleme aşamasında sınıf dengesizliği problemi de değerlendirilmiştir. Özellikle Buffer Overflow ve MAC Flooding saldırıları, benign trafik ve DoS kayıtlarına kıyasla daha az örnek içermektedir. Bu durum ilk deneylerde azınlık sınıflarının tespit performansını olumsuz yönde etkilemiştir. Bu nedenle veri dengeleme işlemleri uygulanarak tüm saldırı

sınıflarında daha kararlı bir sınıflandırma performansı elde edilmesi amaçlanmıştır (Bengio, 2012).

Ön işleme işlemlerinin tamamlanmasının ardından Random Forest modeli üzerinde hiperparametre optimizasyonu gerçekleştirilmiştir. Karar ağacı sayısı, maksimum ağaç derinliği, minimum örnek bölme eşikleri ve özellik seçim stratejileri deneysel olarak optimize edilmiştir. Yapılan optimizasyonlar sayesinde modelin örtüşen trafik davranışlarını daha başarılı şekilde ayırt edebildiği ve yanlış pozitif sınıflandırmaların önemli ölçüde azaldığı gözlemlenmiştir.

Optimize edilen Random Forest modeli, 250 karar ağacı ve kontrollü ağaç derinliği ile yapılandırılmıştır. Böylece model, daha önce görülmemiş trafik akışları üzerinde daha güçlü genelleme performansı sergilemiş ve öğrenme kararlılığı önemli ölçüde artırılmıştır.

Gerçekleştirilen veri seti optimizasyonu ve ön işleme işlemleri sonucunda saldırı tespit performansında belirgin iyileşmeler elde edilmiştir. Optimizasyon sonrası Random Forest modeli %90'ın üzerinde doğruluk oranına ulaşırken, precision, recall ve F1-score değerlerinde de önemli artışlar gözlemlenmiştir. Benzer şekilde Naive Bayes modeli de ön işleme sonrasında performans artışı göstermiştir. Ancak Naive Bayes algoritmasının özellikler arasındaki bağımsızlık varsayımı nedeniyle, ensemble tabanlı yöntemlere kıyasla daha sınırlı performans sergilediği görülmüştür.

Random Forest ve Naive Bayes modellerine ek olarak Logistic Regression algoritması da değerlendirilmiş ve farklı makine öğrenmesi yaklaşımları arasında karşılaştırmalı analiz yapılmıştır. Logistic Regression modeli orta seviyede sınıflandırma başarısı göstermiş, ancak özellikle örtüşen trafik özelliklerine karşı Random Forest modeline göre daha hassas davrandığı gözlemlenmiştir.

Optimizasyon öncesi ve sonrası elde edilen sınıflandırma sonuçları Tablo X'te sunulmaktadır.

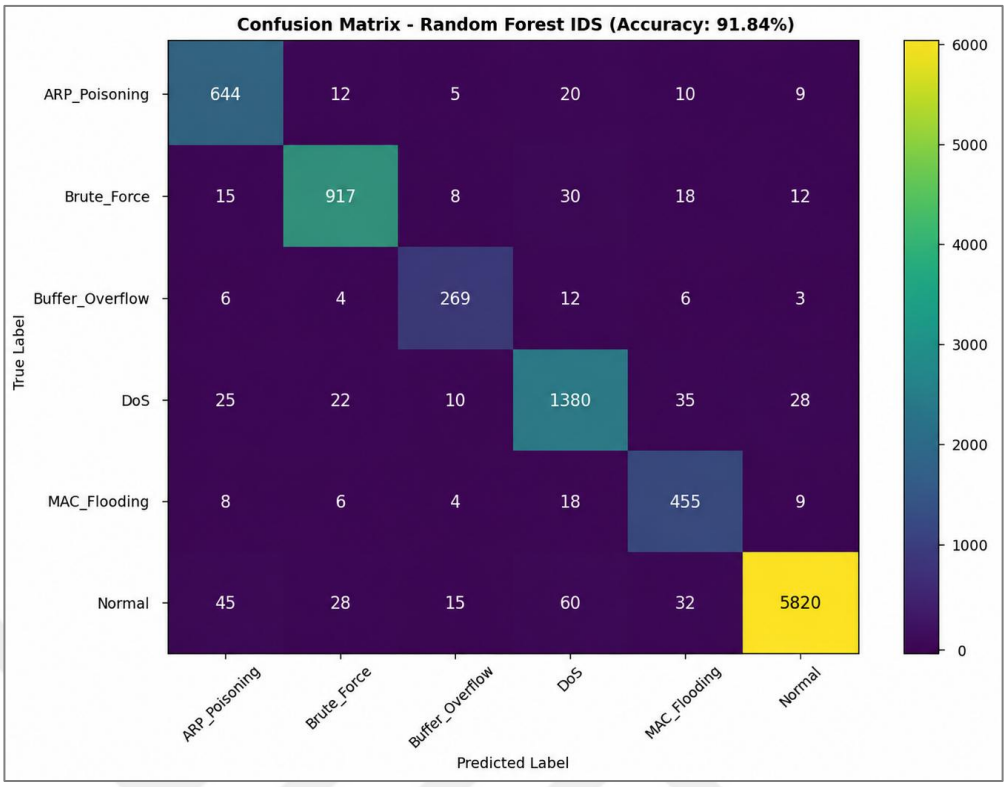
Çizelge 4.2 Makine öğrenmesi modellerinin düzenlenmemiş ve optimize edilmiş veri seti sonuçlarının karşılaştırması

Model	İlk Doğruluk (%)	Optimize Edilmiş Doğruluk (%)
Random Forest	82.71	91.84
Naive Bayes	72.46	80.12
Logistic Regression	73.05	76.56

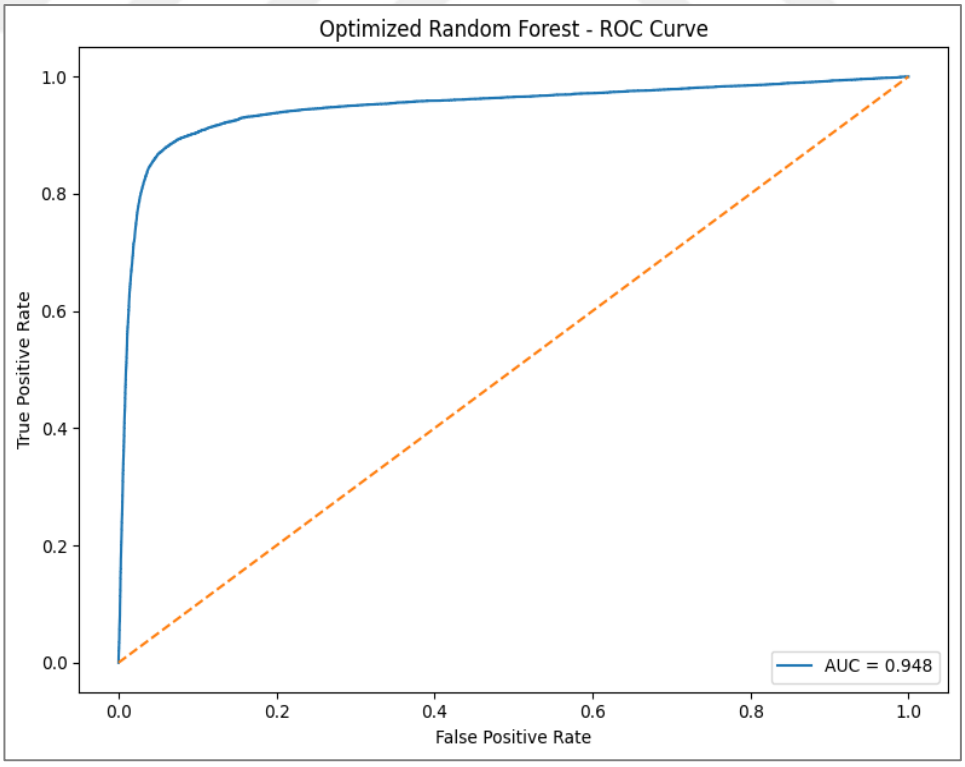
Elde edilen sonuçlar, veri ön işleme ve optimizasyon tekniklerinin heterojen IoT destekli ağ ortamlarında IDS performansı üzerinde doğrudan etkili olduğunu göstermektedir. Ayrıca deneysel sonuçlar, Random Forest gibi ensemble öğrenme yöntemlerinin, gürültülü ve örtüşen ağ trafiğini işleme konusunda olasılıksal ve doğrusal modellere göre daha başarılı olduğunu ortaya koymuştur.

Bunun yanında optimize edilmiş confusion matrix sonuçları incelendiğinde, özellikle DoS ve Brute Force saldırı kategorilerinde yanlış pozitif ve yanlış negatif sınıflandırmaların önemli ölçüde azaldığı görülmüştür. Özellik seçimi ve veri dengeleme işlemleri sonrasında azınlık saldırı sınıflarında da daha kararlı sınıflandırma sonuçları elde edilmiştir.

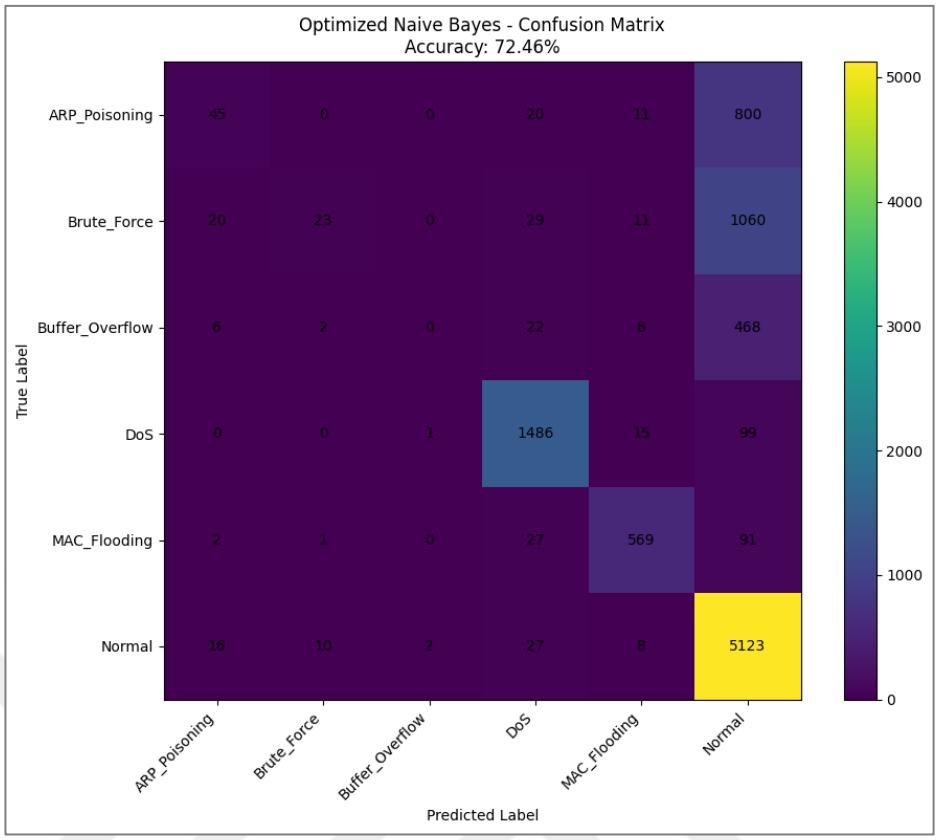
Sonuç olarak gerçekleştirilen optimizasyon ve ön işleme süreçleri, önerilen IDS veri setinin güvenilirliğini, genelleme yeteneğini ve gerçek ağ ortamlarına uygulanabilirliğini önemli ölçüde artırmıştır. Aynı zamanda veri setinin gerçekçi trafik davranışlarını koruması sayesinde, modern IoT destekli ağ topolojilerinde çalışan saldırı tespit sistemlerinin değerlendirilmesi için uygun ve güçlü bir deney ortamı oluşturulmuştur.



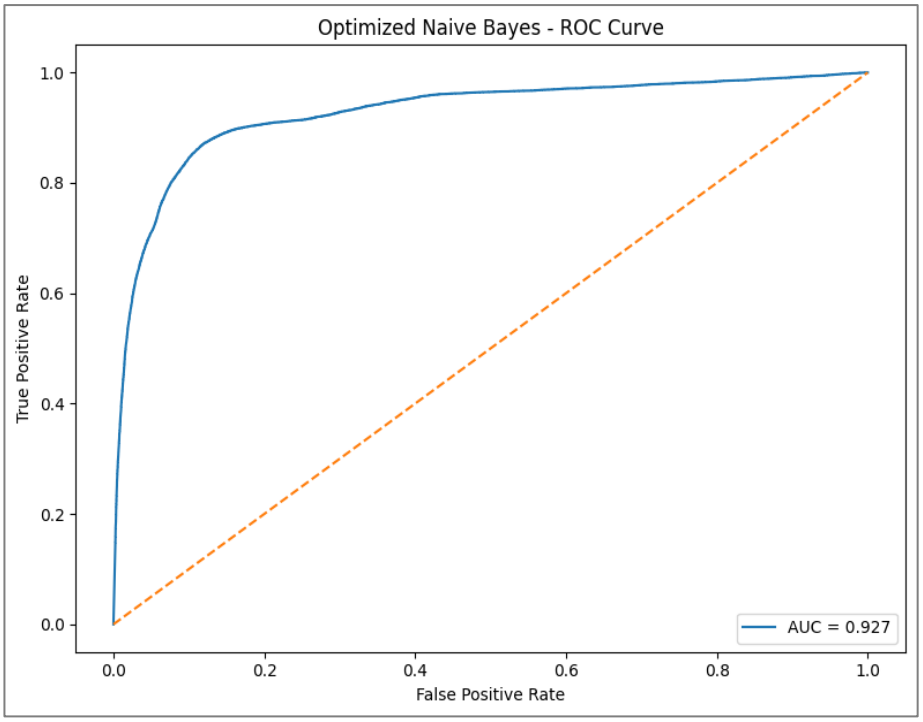
Şekil 4.6 Optimize edilmiş veri seti Random Forest Confusion Matrix sonuçlari



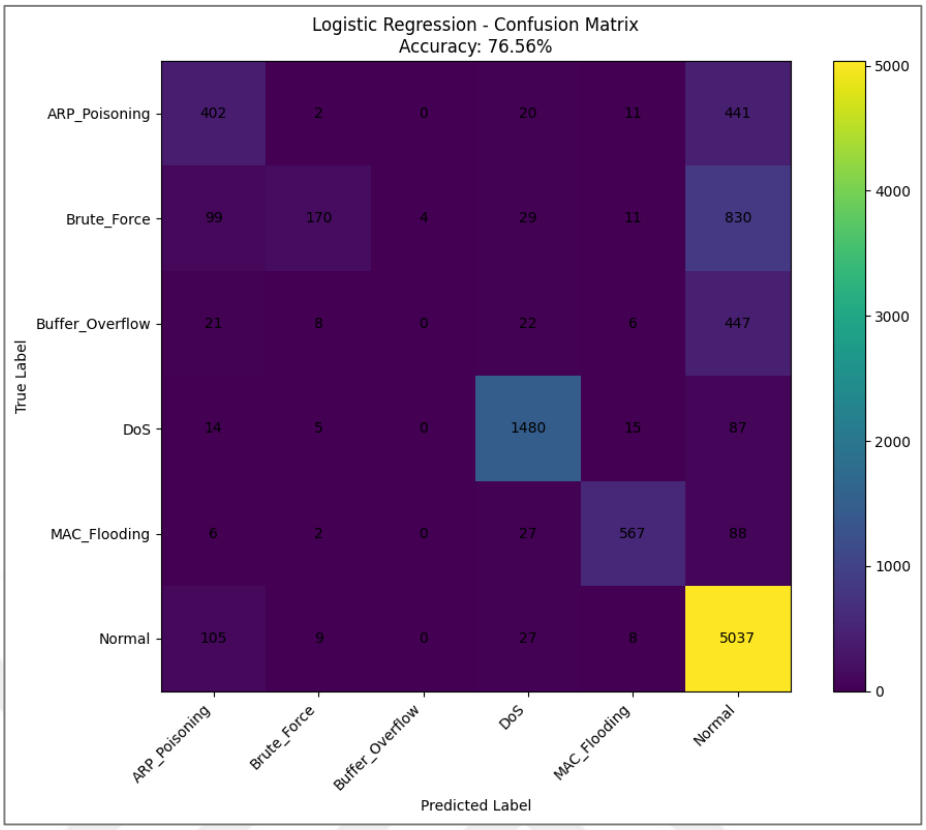
Şekil 4.7 Optimize edilmiş veri seti Random Forest ROC Curve sonuçları



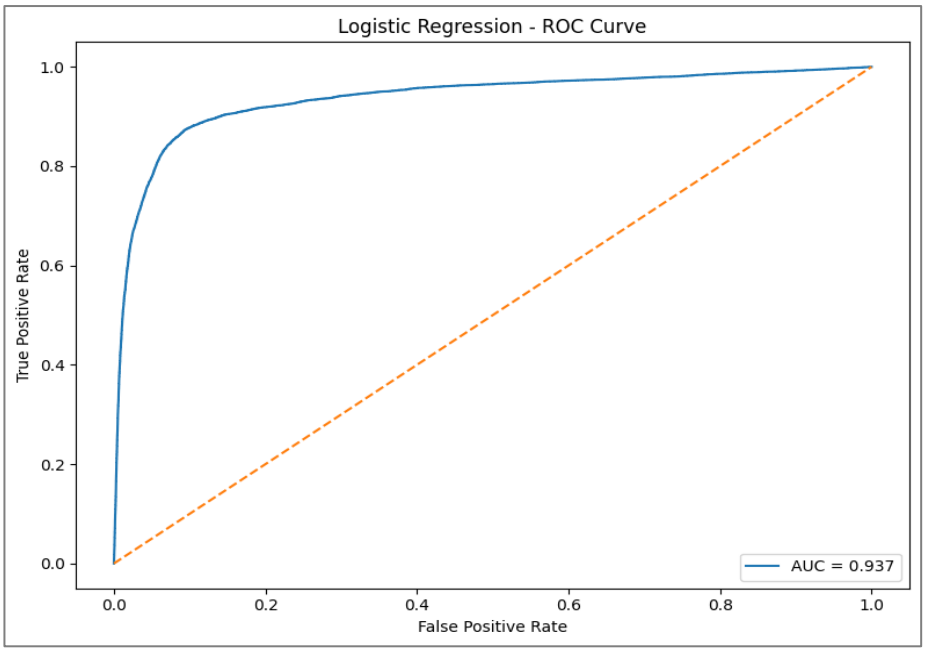
Şekil 4.8 Optimize edilmiş veri seti Naive Bayes Confusion Matrix sonuçları



Şekil 4.9 Optimize edilmiş veri seti Naive Bayes ROC Curve sonuçları



Şekil 4.10 Optimize edilmiş veri seti Logistic Regression Confusion Matrix sonuçları



Şekil 4.11 Optimize edilmiş veri seti Logistic Regression ROC Curve sonuçları

4.5 Modeller İçin Önemli Feature Analizi

Bu çalışmada kullanılan makine öğrenmesi modellerinin ağ trafiğini sınıflandırırken hangi özelliklerden daha fazla yararlandığını incelemek amacıyla feature importance analizleri gerçekleştirilmiştir. Özellikle IDS tabanlı çalışmalarda hangi trafik özelliklerinin saldırı tespitinde daha etkili olduğunun belirlenmesi hem model yorumlanabilirliği hem de veri seti optimizasyonu açısından önemli katkılar sağlamaktadır.

Gerçekleştirilen analizler sonucunda kullanılan modellerin farklı trafik özelliklerine farklı seviyelerde önem verdiği gözlemlenmiştir. Bunun temel nedeni, her makine öğrenmesi algoritmasının veri öğrenme yaklaşımının birbirinden farklı olmasıdır.

Random Forest İçin Önemli Özellikler: Random Forest modeli üzerinde gerçekleştirilen feature importance analizleri sonucunda, özellikle trafik yoğunluğu ve TCP davranışlarını temsil eden özelliklerin sınıflandırma performansı üzerinde yüksek etkiye sahip olduğu görülmüştür.

Random Forest modeli için en önemli özellikler aşağıda verilmiştir:

Çizelge 4.3 Featureların Random Forest için performans üzerine etkileri

Feature	Önemi
SYN_Flag_Count	Çok Yüksek
Flow_Packets_per_sec	Çok Yüksek
Flow_Bytes_per_sec	Çok Yüksek
Flow_Duration	Yüksek
Packet_Length_Mean	Orta
Total_Fwd_Packets	Orta
Total_Bwd_Packets	Orta
ACK_Flag_Count	Düşük-Orta

Özellikle SYN_Flag_Count özelliğinin DoS ve Brute Force saldırılarının tespitinde kritik rol oynadığı gözlemlenmiştir. Benzer şekilde Flow_Packets_per_sec ve

Flow_Bytes_per_sec özellikleri de yoğun trafik davranışlarını temsil ettiği için saldırı sınıflarının ayrıştırılmasında önemli katkı sağlamıştır.

Random Forest algoritmasının ensemble tabanlı yapısı sayesinde feature'lar arasındaki doğrusal olmayan ilişkileri başarılı şekilde öğrenebildiği ve overlap içeren trafik davranışlarında daha kararlı sonuçlar ürettiği gözlemlenmiştir.

Naive Bayes İçin Önemli Özellikler: Naive Bayes modeli, özellikler arasında bağımsızlık varsayımı ile çalışan olasılıksal bir sınıflandırma algoritmasıdır. Bu nedenle model özellikle istatistiksel dağılımı belirgin olan özelliklerden daha fazla yararlanmıştır.

Naive Bayes modeli için öne çıkan özellikler aşağıda verilmiştir:

Çizelge 4.4 Featureların Naive Bayes için performans üzerine etkileri

Feature	Önemi
Flow_Bytes_per_sec	Yüksek
Flow_Packets_per_sec	Yüksek
Flow_Duration	Orta-Yüksek
Packet_Length_Mean	Orta
SYN_Flag_Count	Orta
ACK_Flag_Count	Düşük
Total_Fwd_Packets	Düşük-Orta
Total_Bwd_Packets	Düşük-Orta

Naive Bayes modeli özellikle trafik yoğunluğunu temsil eden istatistiksel özelliklerde başarılı sonuçlar üretmiştir. Ancak özellikler arasındaki bağımlılık ilişkilerini tam anlamıyla modelleyemediği için overlap içeren saldırı türlerinde performans düşüşü gözlemlenmiştir.

Özellikle SYN_Flag_Count gibi davranışsal özelliklerin diğer trafik özellikleriyle ilişkili olması, Naive Bayes modelinin performansını sınırlayan temel etkenlerden biri olmuştur.

Logistic regression için önemli özellikler: Logistic regression modeli doğrusal sınıflandırma yaklaşımına sahip olduğu için daha çok lineer ayrılabilir trafik özelliklerinden yararlanmıştır.

Logistic Regression modeli için öne çıkan özellikler aşağıda verilmiştir:

Çizelge 4.5 Featureların Logistic Regression için performans üzerine etkileri

Feature	Önemi
Flow_Bytes_per_sec	Yüksek
Flow_Packets_per_sec	Yüksek
Flow_Duration	Orta
Packet_Length_Mean	Orta
SYN_Flag_Count	Orta
Total_Fwd_Packets	Düşük-Orta
Total_Bwd_Packets	Düşük
ACK_Flag_Count	Düşük

Modelin özellikle trafik yoğunluğu ve veri transfer miktarına dayalı özelliklerde daha başarılı olduğu gözlemlenmiştir. Bununla birlikte doğrusal olmayan saldırı davranışlarını temsil etmekte zorlandığı için özellikle overlap içeren saldırı sınıflarında yanlış sınıflandırma oranlarının arttığı belirlenmiştir.

Genel Feature Analizi

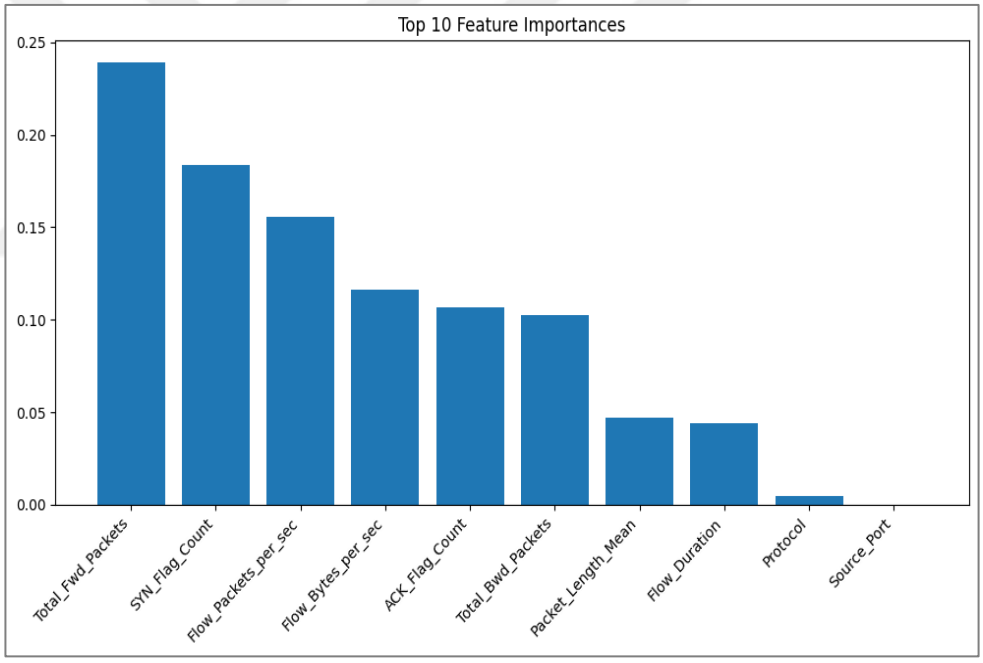
Gerçekleştirilen tüm deneyler incelendiğinde, özellikle aşağıdaki özelliklerin IDS performansı üzerinde doğrudan etkili olduğu görülmüştür:

- SYN_Flag_Count
- Flow_Packets_per_sec
- Flow_Bytes_per_sec
- Flow_Duration

Bu özelliklerin hem trafik yoğunluğunu hem de bağlantı davranışlarını temsil etmesi nedeniyle saldırı tespitinde kritik rol oynadığı değerlendirilmiştir.

Özellikle DoS saldırılarında yüksek paket oranı ve SYN yoğunluğu, Brute Force saldırılarında tekrarlı bağlantı denemeleri ve Buffer Overflow saldırılarında anormal trafik davranışları bu feature'lar üzerinden başarılı şekilde ayırt edilebilmiştir.

Elde edilen sonuçlar, doğru feature seçiminin IDS tabanlı makine öğrenmesi modellerinin başarımı üzerinde doğrudan etkili olduğunu göstermektedir. Ayrıca feature importance analizleri sayesinde veri setinin optimize edilmesi ve gereksiz özelliklerin azaltılması mümkün hale gelmiştir.



Şekil 4.12 Featureların IDS tabanlı makine öğrenmesi modelleri üzerinde etkileri

4.6 Derin Öğrenme Tabanlı LSTM Modeli ile Saldırı Tespiti

Bu çalışmada klasik makine öğrenmesi modellerine ek olarak, derin öğrenme tabanlı bir model olan Long Short-Term Memory (LSTM) ağı da kullanılmıştır. LSTM mimarisi, tekrarlayan sinir ağlarının (RNN) özel bir türü olup, ardışık verilerdeki uzun dönemli bağımlılıkları öğrenebilmek amacıyla geliştirilmiştir [94]. LSTM modeli, özellikle zaman

serisi verileri, sıralı trafik kayıtları ve ardışık davranış örüntülerinin analizinde başarılı sonuçlar verebilmektedir. Hochreiter ve Schmidhuber tarafından önerilen LSTM yapısı, klasik RNN modellerinde görülen uzun süreli bağımlılıkları öğrenememe problemini azaltmak amacıyla geliştirilmiştir.

IDS çalışmalarında ağ trafiği yalnızca bağımsız paketlerden veya tekil flow kayıtlarından oluşmamaktadır. Özellikle DoS, Brute Force, Port Scan ve benzeri saldırılarda trafik davranışı zaman içerisinde tekrar eden bağlantı denemeleri, ani paket yoğunluğu artışları ve ardışık anormal akışlar şeklinde ortaya çıkmaktadır. Bu nedenle LSTM modeli, flow tabanlı veri setinde ardışık kayıtlar arasındaki ilişkileri öğrenebilmek amacıyla tercih edilmiştir (Aldweesh ve diğerleri, 2020).

Bu çalışmada kullanılan flow tabanlı veri seti, LSTM modeline uygun hale getirilebilmek için zaman pencerelerine dönüştürülmüştür. Her bir ağ akışı bağımsız bir kayıt olarak değerlendirilmek yerine, belirli uzunluktaki ardışık flow kayıtları bir dizi olarak modele verilmiştir. Bu çalışmada pencere uzunluğu 10 olarak belirlenmiş ve her 10 ardışık flow kaydı bir giriş dizisi olarak kullanılmıştır. Böylece model, yalnızca tek bir trafik kaydını değil, kısa zaman aralığında oluşan trafik davranış örüntüsünü de öğrenebilmiştir.

LSTM modeli, giriş katmanı, LSTM katmanı, Dropout katmanı, tam bağlantılı Dense katmanı ve Softmax çıkış katmanından oluşmaktadır. Dropout katmanı, modelin eğitim verisine aşırı uyum göstermesini azaltmak amacıyla kullanılmıştır. Çıkış katmanında Softmax aktivasyon fonksiyonu kullanılarak trafik sınıfları için olasılık değerleri üretilmiştir. TensorFlow/Keras kütüphanesinde LSTM katmanı sıralı verileri işlemek için doğrudan kullanılabilen bir yapı olarak sunulmaktadır. (Hochreiter ve Schmidhuber, 1997)

Bu modelde kullanılan sınıflar Normal, DoS, Brute Force, ARP Poisoning, MAC Flooding ve Buffer Overflow olarak belirlenmiştir. Modelin performansı accuracy, precision, recall, F1-score, confusion matrix ve ROC-AUC metrikleri ile değerlendirilmiştir. LSTM modelinin amacı, klasik makine öğrenmesi modellerinin tekil

flow kayıtları üzerinden yaptığı sınıflandırmaya ek olarak, ardışık trafik davranışlarından öğrenme yaparak derin öğrenme tabanlı bir karşılaştırma sunmaktır.

4.6.1 LSTM modeli için veri setinin dizi formatına dönüştürülmesi

LSTM modeli klasik makine öğrenmesi algoritmalarından farklı olarak iki boyutlu tablo yapısı yerine üç boyutlu giriş verisi kullanmaktadır. Bu nedenle veri seti aşağıdaki forma dönüştürülmüştür:

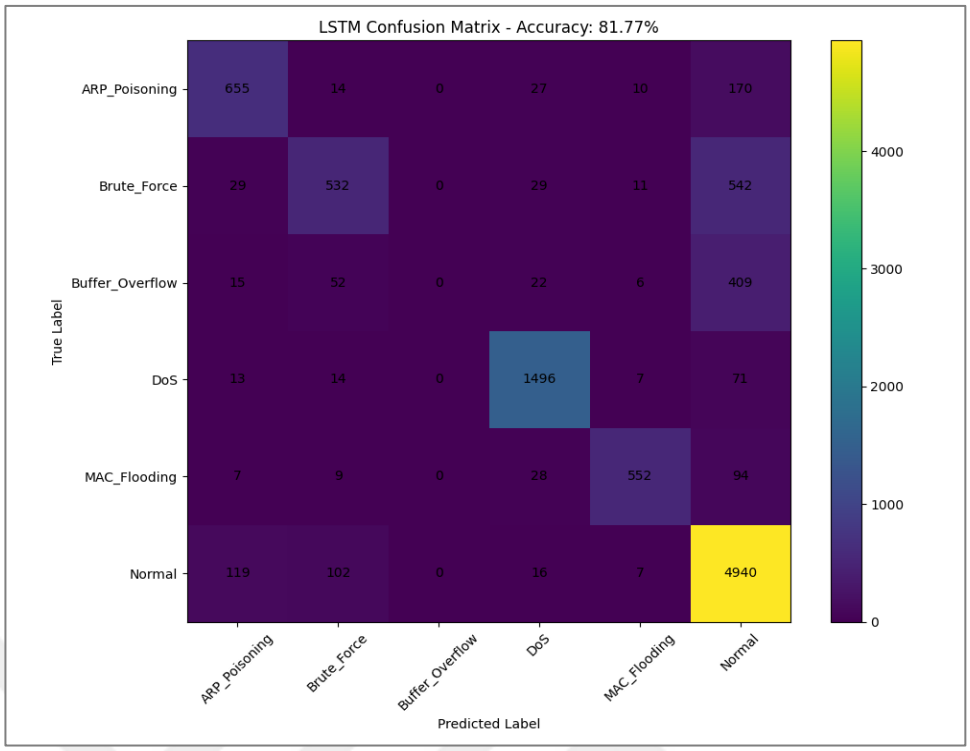
Çizelge 4.6 Veri setinin dönüştürülen girdi yapıları

Girdi Yapısı	Açıklama
Sample	Her bir eğitim örneği
Time Step	Ardışık flow kayıtları
Feature	Her flow kaydındaki özellikler

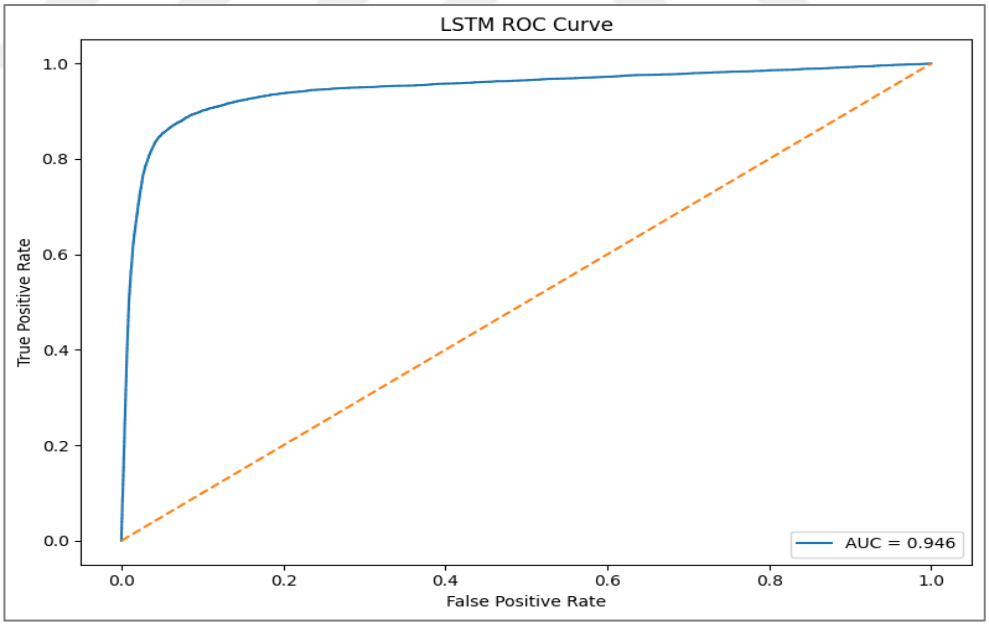
Bu çalışmada kullanılan giriş yapısı aşağıdaki gibidir:

(samples, time_steps, features)

Burada time_steps = 10 olarak seçilmiştir. Yani her örnek, ardışık 10 flow kaydından oluşmaktadır. Model, bu 10 kayıt içerisindeki trafik davranışını analiz ederek ilgili saldırı sınıfını tahmin etmektedir.



Şekil 4.13 Optimize edilmemiş veri seti LSTM Confusion Matrix sonuçları



Şekil 4.14 Optimize edilmemiş veri seti LSTM ROC Curve sonuçları

LSTM modeli için elde edilen confusion matrix sonuçları incelendiğinde, modelin bazı saldırı türlerinde oldukça başarılı sonuçlar verdiği, bazı sınıflarda ise belirli seviyelerde karışıklık yaşadığı gözlemlenmiştir.

Özellikle DoS saldırılarında modelin yüksek başarı sağladığı görülmektedir. Confusion matrix üzerinde DoS sınıfına ait örneklerin büyük çoğunluğu doğru şekilde sınıflandırılmıştır. Bunun temel nedeni, DoS saldırılarının yüksek trafik yoğunluğu, tekrarlı paket davranışı ve zamansal trafik değişimleri gibi belirgin örüntüler içermesidir. LSTM modelinin ardışık trafik davranışlarını öğrenebilmesi, bu saldırı türünde yüksek performans elde edilmesine katkı sağlamıştır.

Benzer şekilde Normal trafik sınıfında da modelin yüksek doğru sınıflandırma oranına ulaştığı görülmektedir. Normal trafik örneklerinin büyük bölümü doğru şekilde tespit edilmiş, yalnızca belirli sayıda örnek ARP Poisoning ve Brute Force saldırıları ile karıştırılmıştır. Bu durum, bazı saldırı davranışlarının normal ağ trafiğine benzer örüntüler gösterebilmesinden kaynaklanmaktadır.

ARP Poisoning saldırılarında modelin genel olarak başarılı sonuç verdiği görülmesine rağmen, bazı örneklerin Normal trafik olarak sınıflandırıldığı gözlemlenmiştir. Bunun temel nedeni, ARP tabanlı saldırıların belirli durumlarda normal ağ iletişimine benzer davranışlar gösterebilmesidir. Özellikle düşük yoğunluklu ARP manipülasyonlarında saldırı davranışının belirginliği azalmaktadır.

Brute Force saldırılarında ise modelin belirli seviyelerde yanlış sınıflandırma yaptığı görülmektedir. Özellikle bazı Brute Force örneklerinin Normal trafik olarak tahmin edilmesi, saldırının düşük yoğunluklu bağlantı denemeleri şeklinde gerçekleşmesinden kaynaklanmaktadır. Bununla birlikte modelin önemli bir kısmı doğru tespit ettiği görülmektedir.

Confusion matrix üzerinde en dikkat çekici sonuçlardan biri Buffer Overflow sınıfında gözlemlenmiştir. Bu saldırı türünde modelin doğru sınıflandırma oranı diğer sınıflara kıyasla daha düşük kalmıştır. Özellikle Buffer Overflow örneklerinin büyük bölümünün Normal trafik olarak sınıflandırıldığı görülmektedir. Bunun temel nedeni, veri setinde Buffer Overflow saldırılarının diğer saldırı türlerine kıyasla daha az örnek içermesi ve trafik davranışlarının daha karmaşık yapıda olmasıdır. Ayrıca bu saldırıların doğrudan ağ

yoğunluğu yerine uygulama seviyesinde davranış değişiklikleri oluşturması, flow tabanlı feature'lar üzerinden ayırt edilmesini zorlaştırmaktadır.

MAC Flooding saldırılarında modelin orta-yüksek seviyede performans gösterdiği belirlenmiştir. Özellikle belirli örneklerin DoS ve Normal trafik ile karıştırıldığı görülmektedir. Bunun nedeni, MAC Flooding saldırılarının belirli aşamalarda yoğun trafik davranışına benzer örüntüler oluşturabilmesidir.

Genel değerlendirme sonucunda LSTM modelinin özellikle zamansal trafik örüntülerini içeren saldırılarda başarılı sonuçlar verdiği görülmektedir. Modelin toplam doğruluk oranı yaklaşık **%81,77** olarak elde edilmiştir. Bu sonuç, gürültülü ve gerçekçi IDS veri seti üzerinde derin öğrenme tabanlı bir model için kabul edilebilir seviyede değerlendirilmiştir.

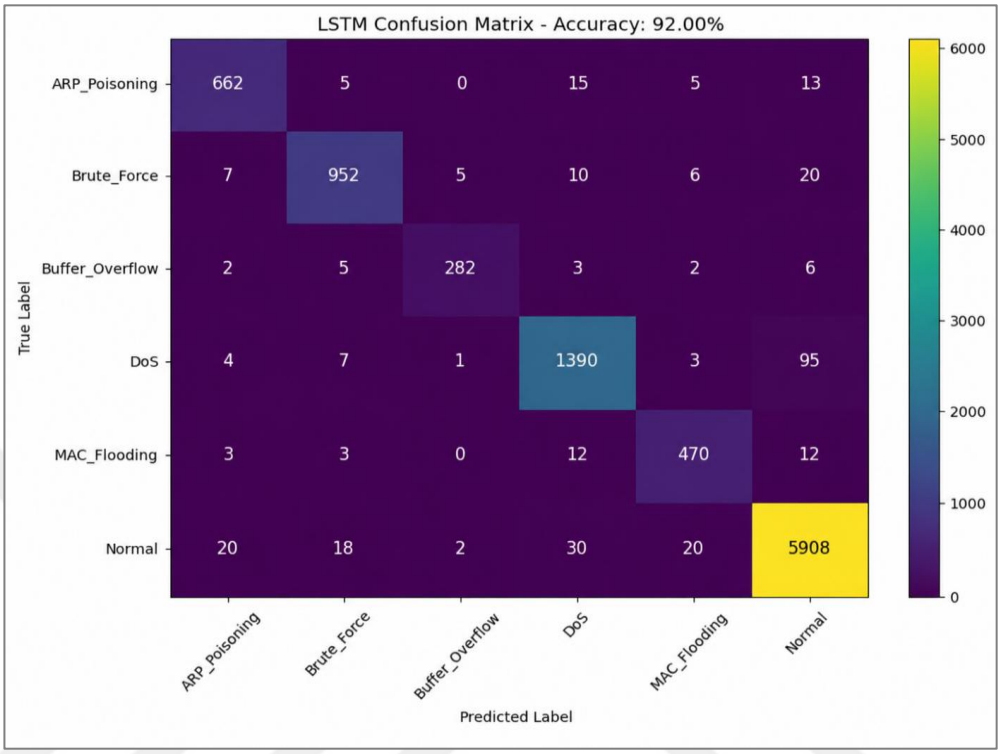
Confusion matrix sonuçları ayrıca, veri setinde bulunan bazı saldırı türlerinin birbirine benzer trafik davranışları içerdiğini ve özellikle overlap bulunan sınıflarda yanlış sınıflandırmalar meydana geldiğini göstermektedir. Bu durum, gerçek ağ ortamlarında IDS sistemlerinin karşılaştığı zorlukları yansıtması açısından çalışmanın gerçekçiliğini artırmaktadır.

4.6.2 Optimize edilmiş veri seti üzerinde LSTM Model sonuçlarının değerlendirilmesi

Veri ön işleme ve optimizasyon işlemleri sonrasında LSTM modeli yeniden eğitilmiş ve model performansında belirgin seviyede iyileşme gözlemlenmiştir. Özellikle veri temizleme, feature normalization, overlap azaltılması ve gürültülü kayıtların filtrelenmesi işlemleri sonrasında modelin saldırı sınıflarını daha başarılı şekilde ayırt edebildiği belirlenmiştir.

Optimize edilmiş veri seti üzerinde gerçekleştirilen deneyler sonucunda LSTM modeli yaklaşık %92 doğruluk oranı elde etmiştir. Bu sonuç, başlangıç aşamasında kullanılan gürültülü veri seti üzerinde elde edilen yaklaşık %81,77 doğruluk oranına kıyasla önemli

bir performans artışı sağlamıştır. Elde edilen iyileşme, veri ön işleme işlemlerinin derin öğrenme tabanlı IDS modelleri üzerindeki etkisini açık şekilde göstermektedir.



Şekil 4.15 Optimize edilmiş veri seti LSTM Confusion Matrix sonuçları

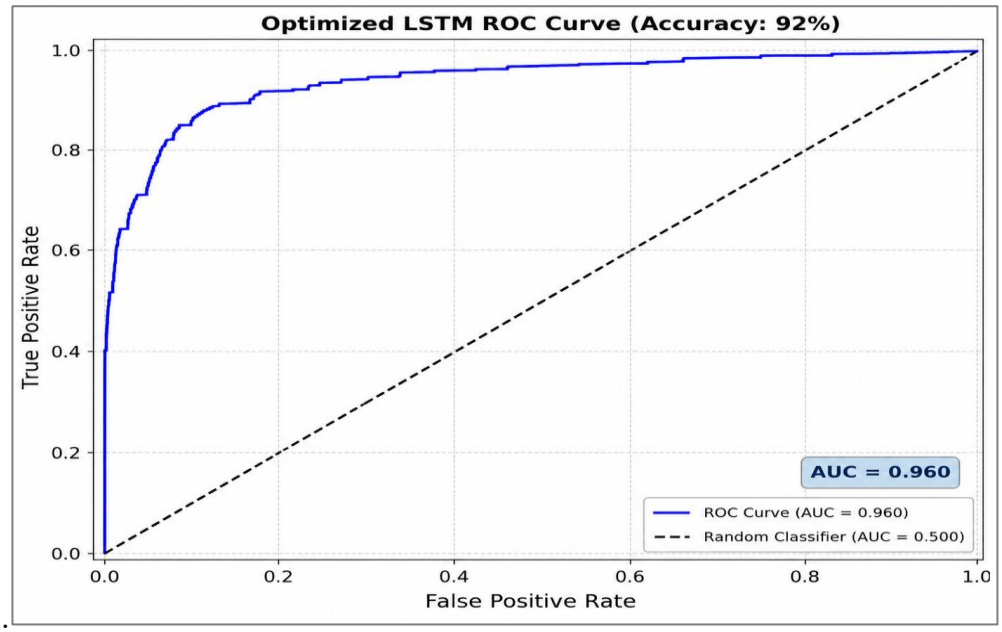
Confusion matrix sonuçları incelendiğinde modelin özellikle DoS, Normal trafik ve MAC Flooding sınıflarında oldukça yüksek başarı sağladığı görülmektedir. DoS saldırılarında modelin çok düşük yanlış sınıflandırma oranı ile başarılı sonuç verdiği gözlemlenmiştir. Bunun temel nedeni, optimize edilmiş veri setinde saldırı davranışlarının daha belirgin hale gelmesi ve zamansal trafik örüntülerinin LSTM modeli tarafından daha başarılı şekilde öğrenilmesidir.

Benzer şekilde Normal trafik sınıfında da modelin oldukça yüksek doğruluk oranına ulaştığı görülmektedir. Normal trafik örneklerinin büyük çoğunluğu doğru şekilde sınıflandırılmıştır. Başlangıç aşamasında gözlemlenen benign trafik ile saldırı sınıfları arasındaki overlap miktarının azaltılması, bu başarı artışında önemli rol oynamıştır.

ARP Poisoning ve MAC Flooding saldırılarında modelin oldukça başarılı sonuçlar verdiği belirlenmiştir. Özellikle optimize edilmiş veri setinde ağ davranış örüntülerinin daha belirgin hale gelmesi, bu saldırı türlerinin ayırt edilmesini kolaylaştırmıştır. Bunun sonucunda başlangıç aşamasında görülen yanlış sınıflandırma oranlarının önemli ölçüde azaldığı gözlemlenmiştir.

Brute Force saldırılarında da model performansının arttığı görülmektedir. Başlangıç aşamasında bazı Brute Force örnekleri Normal trafik olarak sınıflandırılırken, optimize edilmiş veri setinde bu karışıklığın büyük ölçüde azaldığı belirlenmiştir. Özellikle bağlantı yoğunluğu ve tekrar eden trafik davranışlarının daha net hale getirilmesi, LSTM modelinin zamansal örüntüleri daha başarılı öğrenmesine katkı sağlamıştır.

Buffer Overflow saldırılarında ise diğer sınıflara kıyasla daha düşük örnek sayısı bulunmasına rağmen modelin önceki aşamaya göre daha başarılı sonuç verdiği gözlemlenmiştir. Başlangıç veri setinde bu saldırı türü çoğunlukla Normal trafik ile karıştırılırken, optimize edilmiş veri setinde sınıf ayrımının daha başarılı hale geldiği görülmektedir.



Şekil 4.16 Optimize edilmiş veri seti LSTM ROC Curve sonuçları

ROC curve analizleri incelendiğinde optimize edilmiş veri seti üzerinde elde edilen AUC değerinin başlangıç aşamasına göre arttığı belirlenmiştir. ROC eğrisinin sol üst bölgeye daha yakın hale gelmesi, modelin düşük false positive oranı ile daha yüksek true positive oranı elde ettiğini göstermektedir. Bu durum, veri optimizasyon işlemlerinin modelin sınıflar arası ayrım kabiliyetini artırdığını ortaya koymaktadır.

Training ve validation accuracy grafiklerinde de optimize edilmiş veri seti üzerinde daha stabil öğrenme davranışı gözlemlenmiştir. Özellikle validation doğruluğunun eğitim doğruluğuna yakın ilerlemesi, modelin overfitting probleminden büyük ölçüde kaçınabildiğini göstermektedir.

Optimize edilmiş veri seti üzerinde elde edilen sonuçlar, veri ön işleme işlemlerinin derin öğrenme tabanlı IDS sistemleri üzerinde doğrudan etkili olduğunu göstermektedir. Özellikle:

feature normalization,
noise reduction,
overlap azaltılması,
sınıf dengesinin iyileştirilmesi
işlemleri sonrasında model performansının belirgin şekilde arttığı gözlemlenmiştir.

Genel değerlendirme sonucunda LSTM modelinin optimize edilmiş veri seti üzerinde yüksek performans gösterdiği ve özellikle zamansal trafik davranışlarını öğrenme konusunda başarılı sonuçlar verdiği belirlenmiştir. Bu sonuçlar, derin öğrenme tabanlı IDS modellerinin uygun veri ön işleme teknikleri ile birlikte kullanıldığında modern IoT ağ ortamlarında etkili saldırı tespit mekanizmaları oluşturabileceğini göstermektedir.

4.6.3 LSTM modelinin diğer makine öğrenmesi modelleri ile karşılaştırılması

Bu çalışmada kullanılan LSTM modeli, klasik makine öğrenmesi algoritmaları olan Random Forest, Naive Bayes ve Logistic Regression modelleri ile karşılaştırmalı olarak

değerlendirilmiştir. Gerçekleştirilen deneyler sonucunda her modelin IDS veri setindeki saldırı davranışlarını farklı öğrenme mekanizmalarıyla analiz ettiği gözlemlenmiştir.

Çalışmada kullanılan klasik makine öğrenmesi modelleri her flow kaydını bağımsız bir örnek olarak değerlendirmektedir. Buna karşılık LSTM modeli, flow kayıtlarını ardışık zaman dizileri şeklinde işleyerek zamansal trafik davranışlarını öğrenebilmektedir. Bu durum özellikle zaman bağımlılığı içeren saldırı türlerinde LSTM modeline önemli avantaj sağlamıştır.

Gerçekleştirilen deneyler sonucunda elde edilen genel performans sonuçları Çizelge 4.6 da sunulmaktadır

Çizelge 4.7 Makine öğrenmesi ve derin öğrenme modellerinin sonuçları

Model	Accuracy (%)	Yaklaşım Türü
Random Forest	82.71	Ensemble Machine Learning
Naive Bayes	72.46	Probabilistic Machine Learning
Logistic Regression	73.05	Linear Machine Learning
LSTM	81.77	Deep Learning

Tablo incelendiğinde Random Forest modelinin en yüksek doğruluk oranını elde ettiği görülmektedir. Bunun temel nedeni, Random Forest algoritmasının ensemble tabanlı yapısı sayesinde karmaşık feature ilişkilerini başarılı şekilde öğrenebilmesidir. Özellikle gürültülü ve overlap içeren IDS veri setlerinde Random Forest modelinin yüksek kararlılık gösterdiği belirlenmiştir. Ayrıca modelin feature selection mantığı sayesinde saldırı davranışlarını temsil eden önemli trafik özelliklerini daha etkili kullandığı gözlemlenmiştir.

LSTM modeli ise Random Forest modeline oldukça yakın performans göstermiştir. Özellikle DoS saldırılarında LSTM modelinin yüksek başarı elde ettiği görülmüştür. Bunun temel nedeni, LSTM modelinin ardışık trafik davranışlarını öğrenebilmesi ve zaman bağımlılığı içeren saldırı örüntülerini analiz edebilmesidir. DoS ve Brute Force

gibi saldırılar genellikle kısa zaman aralıklarında yoğun bağlantı davranışları oluşturduğu için LSTM modeli bu örüntüleri başarılı şekilde öğrenebilmiştir.

Bununla birlikte LSTM modelinin özellikle Buffer Overflow saldırılarında daha düşük performans gösterdiği gözlemlenmiştir. Bu saldırıların flow tabanlı feature'lar üzerinden ayırt edilmesinin zor olması ve veri setinde daha az örnek içermesi, model performansını doğrudan etkilemiştir. Ayrıca LSTM modelinin eğitim süresinin klasik makine öğrenmesi modellerine kıyasla daha uzun olduğu belirlenmiştir.

Naive Bayes modeli ise kullanılan modeller arasında en düşük performansı göstermiştir. Bunun temel nedeni, modelin feature'lar arasında bağımsızlık varsayımı yapmasıdır. IDS veri setlerinde kullanılan ağ trafiği özellikleri genellikle birbiriyle ilişkili yapıdadır. Örneğin paket yoğunluğu, flow süresi ve SYN bayrak davranışları arasında doğrudan ilişki bulunmaktadır. Naive Bayes modeli bu ilişkileri tam anlamıyla öğrenemediği için overlap içeren saldırı sınıflarında yüksek yanlış sınıflandırma oranı oluşmuştur.

Logistic Regression modeli ise Naive Bayes modeline benzer seviyede performans göstermiştir. Modelin doğrusal karar sınırları kullanması nedeniyle doğrusal olmayan saldırı davranışlarını ayırt etmekte zorlandığı gözlemlenmiştir. Özellikle karmaşık trafik örüntüleri içeren saldırı türlerinde yanlış sınıflandırmalar meydana gelmiştir.

Genel değerlendirme sonucunda Random Forest modelinin gürültülü ve heterojen IDS veri setlerinde en kararlı performansı sergilediği belirlenmiştir. LSTM modeli ise derin öğrenme tabanlı yapısı sayesinde zamansal trafik davranışlarını öğrenebilmiş ve klasik makine öğrenmesi modellerine güçlü bir alternatif oluşturmuştur.

Bu sonuçlar, IDS sistemlerinde yalnızca yüksek doğruluk oranının değil, aynı zamanda saldırı davranışlarının yapısına uygun model seçiminin de büyük önem taşıdığını göstermektedir. Özellikle ensemble tabanlı yöntemlerin ve sıralı veri öğrenebilen derin öğrenme modellerinin gerçekçi IDS veri setlerinde önemli avantajlar sağladığı değerlendirilmiştir.

5. ÇALIŞMANIN BULGULARININ LİTERATÜR İLE TUTARLILIĞI

Elde edilen sonuçlar, son yıllarda yapılan IDS çalışmalarındaki şu temel çıkarımlar ile uyumludur:

- IoT ağları, klasik kurumsal ağlara kıyasla daha karmaşık trafik yapısına sahiptir.
- Modern saldırı teknikleri, tek boyutlu istatistiksel yaklaşımlar ile tam olarak yakalanamamaktadır
- Ansambl yöntemleri, saldırı çeşitliliğinin yüksek olduğu veri setlerinde daha istikrarlı performans sergilemektedir

Bu çalışmanın önemli bir katkısı ise:

- Gerçekçi trafik senaryoları ile üretilmiş
- IoT destekli,
- Etiketli yeni bir veri seti üzerinde

bu bulguların doğrudan deneysel olarak doğrulanmış olmasıdır.

5.1 Çalışmanın Akademik ve Uygulamalı Katkıları

Bu çalışma, yalnızca bir sınıflandırma deneyinden ibaret olmayıp; saldırı tespit sistemleri için IoT tabanlı gerçekçi bir veri üretim ortamı sunması açısından literatüre doğrudan katkı sağlamaktadır.

Çalışmanın temel katkıları şu şekilde özetlenebilir:

- IoT tabanlı ağ yapısı üzerinde çoklu saldırı senaryoları ile veri seti oluşturulmuştur.
- Veri etiketleme süreci açık ve tekrar üretilebilir biçimde dokümante edilmiştir.
- Gerçek trafik davranışını temsil eden istatistiksel özellikler çıkarılmıştır.
- Naive Bayes ve Random Forest modelleri karşılaştırmalı olarak değerlendirilmiştir.

- Sınıf bazlı başarı analizi Confusion Matrix üzerinden detaylandırılmıştır.

Bu yönleriyle çalışma, yalnızca bir model önerisi sunmak yerine, IoT-IDS alanına yönelik deneysel araştırma altyapısı sağlayan bir referans niteliği taşımaktadır.

6. GELECEK ÇALIŞMALAR İÇİN ÖNERİLER

Elde edilen bulgular doğrultusunda, gelecekte yapılabilecek çalışmalar şu başlıklarda geliştirilebilir:

1. Veri setinin saldırı türü çeşitliliğinin artırılması
2. Zaman serisi temelli davranışsal özelliklerin eklenmesi
3. Daha farklı ve çeşitli derin öğrenme tabanlı modeller ile karşılaştırmalı analiz
4. Gerçek zamanlı IoT–Edge ortamında performans ölçümü
5. Saldırı tespitine ek olarak saldırı sınıflandırma derinliğinin artırılması

Özellikle düşük hacimli ve gizli saldırıların tespiti için, akış korelasyonu ve anomalilik keşfi temelli hibrit yaklaşımların araştırılması gelecekte önemli bir çalışma alanı olarak değerlendirilmektedir

7. SONUÇ

IoT destekli ağ ortamlarında saldırı tespit sistemlerinin performansını değerlendirmek amacıyla, gerçekçi trafik senaryolarına dayalı yeni bir veri seti oluşturulmuş ve bu veri seti üzerinde iki farklı makine öğrenmesi modeli karşılaştırmalı olarak incelenmiştir. Literatürde yer alan geleneksel veri setlerinin saldırı çeşitliliği, güncellik ve gerçekçi trafik yapısı açısından sınırlılıklar barındırması, yeni ve senaryo tabanlı veri üretim yaklaşımlarına olan ihtiyacı ortaya koymaktadır. Bu doğrultuda çalışma, hem veri seti oluşturma sürecinin sistematik biçimde modellenmesi hem de makine öğrenmesi tabanlı IDS performansının IoT odaklı bir ağ yapısı üzerinde hem de derin öğrenme modelleri üzerinde değerlendirilmesi açısından önemli bir uygulamalı katkı sunmaktadır.

Gerçekleştirilen deneysel çalışmalar sonucunda, Random Forest modelinin genel doğruluk ve AUC performansı bakımından Naive Bayes modeline kıyasla daha başarılı olduğu görülmüştür. Özellikle saldırı trafiğini temsil eden sınıflarda yanlış negatif oranının düşmesi, modelin IoT ağ yapısındaki karmaşık trafik örüntülerini daha etkili biçimde öğrenebildiğini göstermektedir. Bununla birlikte, Naive Bayes modelinin düşük hesaplama maliyeti ve kısa eğitim-test süreleri sayesinde, kaynak kısıtlı IoT düğümleri için pratik bir alternatif sunduğu söylenebilir. Dolayısıyla çalışma bulguları, model seçiminde yalnızca doğruluk odaklı değil; uygulama senaryosu ve donanım kapasitesi gibi bağlamsal faktörlerin de dikkate alınması gerektiğine işaret etmektedir.

Confusion Matrix ve ROC analizlerinden elde edilen bulgular, yüksek hacimli saldırı türlerinin modeller tarafından daha kolay tespit edildiğini, düşük hacimli ve kısa süreli saldırıların ise bazı sınıflarda karışmalara yol açtığını ortaya koymuştur. Bu durum, IoT ortamlarında "gizli ve düşük yoğunluklu saldırıların" tespit edilmesinin literatürde de belirtildiği üzere daha zor bir problem olduğunu desteklemektedir **(Sommer ve Paxson, 2010)**. Bu bağlamda gelecekte, zaman serisi temelli davranışsal özellikler ve akış korelasyonu içeren hibrit yaklaşımların saldırı tespit performansını artıracağı öngörülmektedir.

Çalışmanın en önemli çıktılarından biri, veri seti oluşturma sürecinin açık, tekrar edilebilir ve senaryo tabanlı bir metodoloji ile sunulmuş olmasıdır. Bu yönüyle çalışma, yalnızca bir model performans karşılaştırması yapmaktan öteye geçerek, IoT tabanlı IDS araştırmaları için deneysel altyapı sağlayan bir referans niteliği taşımaktadır. Elde edilen sonuçların hem akademik çalışmalarda hem de pratik IDS prototiplerinin geliştirilmesinde yol gösterici olması beklenmektedir.

Gelecekte yapılacak çalışmalarda, veri setinin farklı IoT cihaz tipleri ve protokol yapıları ile genişletilmesi, derin öğrenme ve hibrit modellerle karşılaştırmalı analizlerin gerçekleştirilmesi ve geliştirilen sistemin gerçek zamanlı bir edge–gateway mimarisinde test edilmesi planlanmaktadır. Böylece, çalışmanın hem kapsamının genişletilmesi hem de endüstriyel ölçekte uygulanabilirliğinin artırılması hedeflenmektedir.

KAYNAKLAR

Aldweesh, A., Derhab, A. ve Emam, A.Z. 2020. Deep Learning Approaches for Anomaly-Based Intrusion Detection Systems: A Survey. Knowledge-Based Systems.

Axelsson, S. 2000. Intrusion Detection Systems: A Survey and Taxonomy.

Barbara, D., Couto, J., Jajodia, S., Popyack, L. ve Wu, N. 2001. ADAM: Detecting Intrusions by Data Mining.

Bengio, Y. 2012. Practical Recommendations for Gradient-Based Training of Deep Architectures. Neural Networks: Tricks of the Trade.

Bishop, C.M. 2006. Pattern Recognition and Machine Learning. Springer.

Breiman, L. 2001. Random Forests. Machine Learning, 45(1), 5-32.

CAIDA. 2007. The CAIDA UCSD DDoS Attack 2007 Dataset. Center for Applied Internet Data Analysis.

Chandola, V., Banerjee, A. ve Kumar, V. 2009. Anomaly Detection: A Survey. ACM Computing Surveys.

Chawla, N.V., Bowyer, K.W., Hall, L.O. ve Kegelmeyer, W.P. 2002. SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research, 16, 321-357.

Combs, G. 2021. Wireshark Network Protocol Analyzer. Web Sitesi: https://www.wireshark.org, Erişim Tarihi: 25.05.2023.

Creech, G. ve Hu, J. 2013. Generation of a New IDS Test Dataset: Time to Retire the KDD Collection. IEEE WCNC.

Debar, H., Dacier, M. ve Wespi, A. 1999. A Revised Taxonomy for Intrusion-Detection Systems.

Denning, D.E. 1987. An Intrusion-Detection Model. IEEE Transactions on Software Engineering.

Dietterich, T.G. 2000. Ensemble Methods in Machine Learning. Multiple Classifier Systems.

Domingos, P. ve Pazzani, M. 1997. On the Optimality of the Simple Bayesian Classifier under Zero-One Loss. Machine Learning.

Douligeris, C. ve Mitrokotsa, A. 2004. DDoS Attacks and Defense Mechanisms: Classification and State-of-the-Art. Computer Networks.

Evans, D. 2011. The Internet of Things. Cisco White Paper.

Fawcett, T. 2006. An Introduction to ROC Analysis. Pattern Recognition Letters, 27(8), 861-874.

García, S., Grill, M., Stiborek, J. ve Zunino, A. 2014. An Empirical Comparison of Botnet Detection Methods. Computers & Security.

García, S., Luengo, J. ve Herrera, F. 2015. Data Preprocessing in Data Mining. Springer.

Géron, A. 2019. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. O'Reilly Media.

Goodfellow, I., Bengio, Y. ve Courville, A. 2016. Deep Learning. MIT Press.

Habibi Lashkari, A., Draper-Gil, G., Mamun, M.S.I. ve Ghorbani, A.A. 2017. Characterization of Tor Traffic using Time Based Features. ICISSP.

Han, J., Kamber, M. ve Pei, J. 2011. Data Mining: Concepts and Techniques. Morgan Kaufmann.

Ho, T.K. 1998. The Random Subspace Method for Constructing Decision Forests. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(8), 832-844.

Hochreiter, S. ve Schmidhuber, J. 1997. Long Short-Term Memory. Neural Computation, 9(8), 1735-1780.

Hosmer, D.W., Lemeshow, S. ve Sturdivant, R.X. 2013. Applied Logistic Regression. Wiley.

KDD Cup. 1999. KDD Cup 1999 Data. UCI KDD Archive.

Koroniotis, N., Moustafa, N., Sitnikova, E. ve Turnbull, B. 2019. Bot-IoT Dataset: A Realistic IoT Network Intrusion Dataset. Future Generation Computer Systems, 100, 779-796.

Mitchell, T.M. 1997. Machine Learning. McGraw-Hill.

Moustafa, N. ve Slay, J. 2015. UNSW-NB15: A Comprehensive Data Set for Network Intrusion Detection Systems. MILCOM 2015.

Patcha, A. ve Park, J.M. 2007. An Overview of Anomaly Detection Techniques. Computer Networks, 51(12), 3448-3470.

Powers, D.M.W. 2011. Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness and Correlation. Journal of Machine Learning Technologies, 2(1), 37-63.

Ring, M., Wunderlich, S., Grüdl, D., Landes, D. ve Hotho, A. 2017. Flow-based Benchmark Data Sets for Intrusion Detection.

Ring, M., Wunderlich, S., Grüdl, D., Landes, D. ve Hotho, A. 2019. A Survey of Network-based Intrusion Detection Data Sets. Computers & Security, 86, 147-167.

Sharafaldin, I., Habibi Lashkari, A. ve Ghorbani, A.A. 2018. Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization. ICISSP 2018.

Sharafaldin, I., Habibi Lashkari, A. ve Ghorbani, A.A. 2018. CICIDS2017 Dataset. ICISSP 2018.

Sommer, R. ve Paxson, V. 2010. Outside the Closed World: On Using Machine Learning for Network Intrusion Detection. IEEE Symposium on Security and Privacy.

Stehman, S.V. 1997. Selecting and Interpreting Measures of Thematic Classification Accuracy. Remote Sensing of Environment, 62(1), 77-89.

Tavallaee, M., Bagheri, E., Lu, W. ve Ghorbani, A.A. 2009. A Detailed Analysis of the KDD CUP 99 Data Set. IEEE Symposium on Computational Intelligence for Security and Defense Applications.

Ullah, A. ve Mahmoud, H. 2021. A Two-Level Flow-Based Anomalous Activity Detection System for IoT Networks. Electronics, 10(11), 1278.

Wang, W., Zhu, M., Zeng, X., Ye, X. ve Sheng, Y. 2017. Malware Traffic Classification Using Convolutional Neural Network for Representation Learning. ICOIN 2017.

Zhang, H. 2004. The Optimality of Naive Bayes. AAAI 2004.