T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

KABLOSUZ AĞLAR ÜZERİNDEN GERÇEKLEŞTİRİLEN SİBER TEHDİTLERİN MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE AĞ ADLİ BİLİŞİM ANALİZİNİN GERÇEKLEŞTİRİLMESİ

İmran KAÇAN

Yüksek Lisans Tezi

ADLİ BİLİŞİM MÜHENDİSLİĞİ ANABİLİM DALI

TEMMUZ 2023

T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Adli Bilişim Mühendisliği Anabilim Dalı

Yüksek Lisans Tezi

KABLOSUZ AĞLAR ÜZERİNDEN GERÇEKLEŞTİRİLEN SİBER TEHDİTLERİN MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE AĞ ADLİ BİLİŞİM ANALİZİNİN GERÇEKLEŞTİRİLMESİ

Tez Yazarı
İmran KAÇAN

Danışman
Doç. Dr. Fatih ERTAM

TEMMUZ 2023
ELAZIĞ

T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Adli Bilişim Mühendisliği Anabilim Dalı

Yüksek Lisans Tezi

Başlığı:	Kablosuz Ağlar Üzerinden Gerçekleştirilen Siber Tehditlerin Makine Öğrenmesi Yöntemleri ile Ağ Adli Bilişim Analizinin Gerçekleştirilmesi
Yazarı:	İmran KAÇAN
İlk Teslim Tarihi:	12.06.2023
Savunma Tarihi:	11.07.2023

TEZ ONAYI

Fırat Üniversitesi Fen Bilimleri Enstitüsü tez yazım kurallarına göre hazırlanan bu tez aşağıda imzaları bulunan jüri üyeleri tarafından değerlendirilmiş ve akademik dinleyicilere açık yapılan savunma sonucunda OYBİRLİĞİ ile kabul edilmiştir.

İmza

Danışman: Doç. Dr. Fatih ERTAM — Onayladım
Fırat Üniversitesi, Teknoloji Fakültesi

Başkan: Dr. Öğr. Üyesi Fahrettin Burak DEMİR — Onayladım
Bandırma Onyedi Eylül Üniversitesi, Mühendislik ve Doğa Bilimleri Fakültesi

Üye: Dr. Öğr. Üyesi Mustafa KAYA — Onayladım
Fırat Üniversitesi, Teknoloji Fakültesi

Bu tez, Enstitü Yönetim Kurulunun ...../...../20..... tarihli toplantısında tescillenmiştir.

İmza

Prof. Dr. Burhan ERGEN
Enstitü Müdürü

BEYAN

Fırat Üniversitesi Fen Bilimleri Enstitüsü tez yazım kurallarına uygun olarak hazırladığım "Kablosuz Ağlar Üzerinden Gerçekleştirilen Siber Tehditlerin Makine Öğrenmesi Yöntemleri ile Ağ Adli Bilişim Analizinin Gerçekleştirilmesi" Başlıklı Yüksek Lisans Tezimin içindeki bütün bilgilerin doğru olduğunu, bilgilerin üretilmesi ve sunulmasında bilimsel etik kurallarına uygun davrandığımı, kullandığım bütün kaynakları atıf yaparak belirttiğimi, maddi ve manevi desteği olan tüm kurum/kuruluş ve kişileri belirttiğimi, burada sunduğum veri ve bilgileri unvan almak amacıyla daha önce hiçbir şekilde kullanmadığımı beyan ederim.

11.07.2023

İmran KAÇAN

ÖNSÖZ

Günümüzde teknolojinin gelişmesi ile beraber, ağ teknolojileri de ciddi oranda gelişmiştir. Gelişen ağ teknolojileri beraberinde bu sistemlere karşı yapılan çeşitli siber saldırıları da getirmiştir. Bu tez çalışmasında kablosuz ağlara yapılan saldırıların sistem üzerindeki etkisinin analiz edilmesi amaçlanmıştır.

Bu tez çalışmasının gerçekleştirilmesinde her türlü yol gösterici olan, bilgi birikimiyle desteklerini esirgemeyen danışmanım Sayın Doç. Dr. Fatih Ertam'a, lisans eğitimim süresince bana kattıkları tüm değerler için Fırat Üniversitesi Adli Bilişim Mühendisliğinde görevli olan hocalarıma ve her zaman destekleri ile yanımda olan, hiçbir zaman bana olan güvenlerinin eksikliğini hissetmediğim aileme teşekkür ederim.

Bu tez çalışması, Fırat Üniversitesi Bilimsel Araştırma Projeleri Koordinasyon Birimi (FÜBAP) tarafından **TEKF.23.12** protokol numaralı proje ile desteklenmiştir.

İmran KAÇAN
ELAZIĞ, 2023

İÇİNDEKİLER

Sayfa

ÖZET

Kablosuz Ağlar Üzerinden Gerçekleştirilen Siber Tehditlerin Makine Öğrenmesi Yöntemleri ile Ağ Adli Bilişim Analizinin Gerçekleştirilmesi

İmran KAÇAN

Yüksek Lisans Tezi

FIRAT ÜNİVERSİTESİ
Fen Bilimleri Enstitüsü

Adli Bilişim Mühendisliği Anabilim Dalı

Temmuz 2023, Sayfa: xi + 40

Teknolojinin gelişmesi ile beraber teknolojik cihazlara yapılan saldırılar da artış göstermiştir. Artan siber saldırılar ile beraber kullandığımız cihazlar her an tehdit altında olabilmektedir. Birbiri ile iletişim halinde bulunan bu cihazlardan herhangi birine tehdit unsurunun erişmesi, ağda ki tüm cihazları tehlikeye sokabilmektedir. Ağ ortamlarına bağlı olan cihazların sayısının fazlalığı göz önünde bulundurularak bu çalışma ele alınmıştır. Bu çalışmada olası bir siber saldırıda sistemlerde oluşabilecek ağ trafiği oluşturularak, elde edilen trafikleri, farklı sınıflandırma algoritmalarını kullanarak eğitmek amaçlanmıştır.

Oluşturulan test ortamında farklı siber saldırılar gerçekleştirilmiş, siber saldırılar sırasında elde edilen ağ paketlerinin özellikleri çıkartılmış ve yaygın kullanılan sınıflandırma algoritmaları ile eğitim işlemleri gerçekleştirilmiştir. Yapılan işlemler sonucunda kullanılan sınıflandırma algoritmaları arasında en iyi başarı oranı sağlayan sınıflandırma algoritması belirlenmiştir.

Anahtar Kelimeler: Ağ Adli Bilişimi, Siber Güvenlik, Sınıflandırma, Makine Öğrenmesi

ABSTRACT

Cyberspace Over Wireless Networks Performing Network Forensics Analysis of Threats with Machine Learning Methods

İmran KAÇAN

Master's Thesis

FIRAT UNIVERSITY
Graduate School of Natural and Applied Sciences

Department of Digital Forensic Engineering

July 2023, Pages: xi + 40

With the development of technology, attacks on technological devices have increased. With increasing cyber attacks, the devices we use can be under threat at any time. An attacker's access to any of these devices that are in communication with each other can endanger all devices in the network. Considering the high number of devices connected to network environments, this study has been handled. In this study, it is aimed to train the traffic obtained by using different classification algorithms by creating the network traffic that may occur in the systems in a possible cyber attack.

Different cyber attacks were carried out in the created test environment, the characteristics of the network packets obtained during the cyber attacks were extracted and training operations were carried out with commonly used classification algorithms. As a result of the processes, the classification algorithm that provides the best success rate among the classification algorithms used was determined.

Keywords: Network Forensic, Cyber Security, Classification, Machine Learning

ŞEKİLLER LİSTESİ

Sayfa

TABLOLAR LİSTESİ

Sayfa

KISALTMALAR

ACK	: Acknowledge
DOS	: Denial of Service Attack
E-POSTA	: Elektronik Posta
FP	: False Positive
FN	: False Negative
GPS	: Global Positioning System
HTTP	: Hyper Text Transfer Protocol
HTTPS	: Secure Hyper Text Transfer Protocol
ICMP	: Internet Control Message Protocol
IP	: Internet Protocol Address
MAC	: Media Access Control
MiTM	: Man in The Middle
NFS	: Network File System
OSI	: Open Systems Interconnection
SMB	: Server Message Block
SYN	: Sunchronize
SVC	: Support Vector Classifier
SVM	: Support Vector Machines
TP	: True Positive
TN	: True Negative
TCP	: Transmission Control Protocol
UDP	: User Datagram Protocol
USB	: Universal Serial Bus
WEB	: World Wide Web

1. GİRİŞ

Siber dünyada birbirine bağlı olan sistemler, saldırılara karşı açık halde olabilmektedirler. Bu saldırıları örneklendirmek gerekirse; ortadaki adam saldırısı, hizmet reddi saldırısı, dağıtılmış hizmet reddi saldırısı, kötü amaçlı yazılım enjeksiyonu gibi saldırıları örnek olarak verebilmek mümkündür [1].

Gelişen teknoloji ile beraber internet kullanımı da ciddi oranda artmıştır. Çevremizde bulunan birçok cihaz ağa bağlıdır. Ayrıca gelişen teknoloji ile beraber dijital ortamlarda işlenen suçlarda da artış olmuştur.

Gelişen teknoloji altyapısı ile beraber, bu yapının ciddi oranda yaygınlaşmış olması bu sistemlerin risklerle de karşıya karşıya olması anlamına gelmektedir. Teknolojik yapıların artışı insan hayatını kolaylaştırmanın yanında, tehdit unsurlarının da birçok kaynağa erişim sağlamasına ve sistem hatalarından oluşabilecek sistem açıklarına erişim sağlayıp, istismar etmesine olanak tanımaktadır [2]. Bu sebeple, teknolojik cihazların ve ağ teknolojilerinin kullanılmaya başlamasıyla bu sistemlere zarar vermek niyetiyle kullanan kişilerinde ortaya çıkmasıyla bilişim suçları olarak adlandırılan suç tipi ortaya çıkmıştır [3].

Bilişim suçlarında, elektronik/manyetik alanlara kaydedilebilen veriler ise dijital delil olarak adlandırılmaktadır. Dijital delillerin farklı tipleri bulunmaktadır. Bunlardan bazılarını sıralamak gerekirse;

- Veri dosyaları,
- Fotoğraflar,
- Videolar,
- Sunucu kayıt dosyaları,
- İnternet geçmişi,
- Web sayfaları,
- Kayıt logları şeklinde sıralanabilir. Bu verilerin büyük bir çoğunluğu ağ üzerinden de iletilmektedir.

Önemli analizden biri de ağ analizidir. Ağ analizi yapmak için kullanılan programların/araçların genel amacını ağ trafiğini dinlemek, dinleme esnasında gelen ve giden paketleri yakalayarak ağ analizi yapmak şeklinde tanımlayabilmek mümkündür.

Ağ; yazıcı gibi kaynakları paylaşmak, dosya alışverişi yapmak veya elektronik iletişime izin vermek için iki veya daha fazla bilgisayar, sunucu, ağ cihazları gibi aygıtlardan oluşmaktadır [4].

Potansiyel saldırılardan korunmak için cihazlara ve saldırı türlerine hâkim olunmalıdır. Ayrıca en etkili ağ güvenlik yönteminin ağa erişimi yönetmek olduğu söylenebilir [5]. Bu güvenlik önlemlerinin alınmasında ağa bağlı olan kullanıcı profilleri iyi bilinmelidir. Kullanıcı profillerine

hâkim olmak, yetkilendirme işlemlerini yaparken ağ yöneticilerinin işini kolaylaştırmaktadır. Kullanıcı profiline uygun yetkilendirme yapıldığı takdirde ağı yönetmek daha kolay olacaktır. Her kullanıcının, her ağa erişimi olmamalıdır. Yalnızca yetkili kullanıcıların ağ kaynaklarına erişimi sağlanarak, oluşabilecek kötü niyetli aktivitelerin ağa erişimi engellenmelidir. Kullanılmayan portlar kapatılmalı, kullanıcıların ihtiyacına göre yetkilendirilmiş yapılar açık bırakılmalıdır. Kimlik doğrulama yöntemleri aktif edilmelidir. Bunlarla beraber saldırı türlerine hâkim olunarak, güvenlik politikaları belirlenmelidir.

Tüm bunlar göz önünde bulundurulduğunda adli ağ analizinin önemi göz önüne çıkmaktadır. Bu alanda güvenlik önlemleri arttırılmalı ve eğitimli insan sayısı arttırılmalıdır.

Bu tez çalışmasında ise ağ üzerinde yapılabilecek saldırıların uygulaması gerçekleştirilerek; sistem üzerinde oluşturduğu etkiyi, sınıflandırma algoritmalarını da kullanarak yorumlamaya çalışılmıştır. Bu sınıflandırma çalışmaları sonucunda çıkan sonuçlar ile saldırıların davranışları gözlemlenebilmektedir.

Bu tez çalışması 6 bölümden oluşmaktadır. Çalışmanın ikinci bölümünde adli bilişim yöntemlerine ve ağ adli bilişiminin detaylı incelemesine yer verilmiştir. Tez çalışmasının üçüncü bölümünde kablosuz ağlara yönelik siber saldırılar detaylandırılmıştır. Bir sonraki bölüm olan dördüncü bölümde ise; siber güvenlikte makine öğrenmesi uygulamalarına, makine öğrenmesi tekniklerine, sınıflandırma algoritmalarına, performans metriklerine ve sisteme ilişkin metot ile beraber gerçekleştirilen uygulamalara yer verilmiştir. Çalışmanın beşinci bölümünde elde edilen bulgulara yer verilirken, çalışmanın son bölümünde sonuçlara ve önerilere yer verilmiştir.

2. ADLİ BİLİŞİM

Adli bilişim; bilişim yoluyla gerçekleşen eylemlerde, olay yerinin incelenmesi, dijital cihazların toplanması, korunması, analiz edilmesi ve elde edilen sonuçların adli makamlara raporlanarak sunulması gibi geniş bir süreci kapsamaktadır [6].

Adli bilişimin temel amaçları maddeler halinde aşağıdaki gibi sıralanabilir;

- Materyallerin, delil olarak sunulmasına yardımcı olmak amacıyla kurtarılması, analiz edilmesi ve korunmasıdır.
- Suç teşkil edecek nedenleri ve suçlu/suçluların kimliğinin ortaya konmasına yardımcı olmak.
- Elde edilen delillerin kanıt niteliğinin bozulmayacak şekilde muhafaza edilmesi için gerekli prosedürlerin uygulanması.
- Kanıtların toplanması ve silinen verilerin kurtarılması.
- Delillerin korunması.
- İnceleme sonrası süreç ile ilgili raporun hazırlanması ve sunulması [7].

Günümüzde dijital cihazların yaygınlaşmasıyla beraber adli bilişime olan ihtiyaçta artmıştır. Dijital delil niteliğinde olan bu cihazların korunması önem arz etmektedir, çünkü elde edilen dijital deliller kolayca bozulabileceği gibi kolayca değiştirilebilirler. Adli bilişim süreci dört adımda sıralanmaktadır;

Tanımlama/Koruma: Olay yeri korunarak, delil toplama işlemleri gerçekleştirilirken alanın görüntülerinin alınması, toplanan dijital delillerinin etiketlenerek belgelenmesi süreci olarak tanımlanabilir.

İnceleme: Verileri analiz aşamasına geçmeden önce delil bütünlüğünü bozmamak adına adli kopya (imaj) alma aşaması olarak tanımlanabilir.

Analiz: Alınan adli kopyanın detaylı olarak analiz edildiği aşamadır.

Raporlama: Analiz sonrası elde edilen, delil niteliği taşıyabilecek verilerin raporlandığı adımdır. Rapor anlaşılır bir dille yazılmalıdır.

Adli bilişim süreçlerinde karşılaşılan zorluklar şu şekilde sunulabilir;

- Bilgisayarların ve internet erişiminin yaygınlaşması,
- Saldırı için kullanılan araçların kolay kullanılabilirliği,
- Verilerin fiziksel ortamlarda bulunmasının yanı sıra, sanal ortamda da dijital verilerin bulunması,
- Veri miktarlarının ciddi oranda artmış olması [7].

2.1. Adli Bilişim Yöntemleri

Teknolojinin gelişmesiyle beraber delil elde edilebilecek alanlarda artmıştır. Bu gelişim beraberinde inceleme alanlarını da çeşitlendirmiştir. Bu nedenle adli bilişim çeşitli alt alanlarda incelenmektedir. Bu alanları şu şekilde sıralamak mümkündür;

- Bilgisayar Adli Bilişimi
- Ağ Adli Bilişimi
- Mobil Adli Bilişimi
- GPS Adli Bilişimi
- Medya Araçları Adli Bilişimi
- Sosyal Ağ Adli Bilişimi
- Bulut Adli Bilişimi

2.1.1. Bilgisayar Adli Bilişimi

Belirli cihazlardan delilleri bir mahkemede sunulmaya uygun bir şekilde toplamak ve korumak için analiz tekniklerinin uygulanmasıdır. Bilgisayar adli bilişiminin amacı, cihazda tam olarak ne olduğunu ve bu olaylardan kimin sorumlu olduğunu bulmak için planlanmış bir araştırma yapmak ve belgelenmiş bir kanıt zincirini sürdürme işlemidir.

Bilgisayar adli bilişimi esas olarak bilgileri yasal işlemlerde kabul edilebilir hale getirerek veri elde etme işlemidir [8].

2.1.2. Ağ Adli Bilişimi

Ağın incelenmesi ve kötü niyetli faaliyetlerde bulunduğundan şüphelenilen bir ağ üzerinden geçen trafiğin analiz edilmesi işlemidir. İnternet ağının büyümesiyle beraber ağ adli bilişiminin de önemi artmıştır.

Ağ adli bilişimi ile dosya aktarımları, mesajlar ve web tarama geçmişi gibi verilerin tamamı elde edilebilmektedir [9].

2.1.3. Mobil Adli Bilişimi

Günümüzde mobil cihazların belleklerinin ciddi oranda artmış olması, cihazlarda yapılan işlemlerin mesajlaşma, video kaydı alma, fotoğraf çekme, ses kaydı alabilme gibi gelişmiş özelliklere sahip olması yönüyle delil elde etmede önemli yapılar haline gelmiştir [10].

2.1.4. GPS Adli Bilişimi

Bugün hemen hemen her akıllı telefonda GPS alıcısı bulunmaktadır. Akıllı telefonların dışında da GPS cihazlarını görmek mümkündür. Bunları; GPS cihazları, otomobiller, havacılık ve denizcilik cihazları olarak sıralamak mümkündür.

GPS cihazlarından elde edilebilecek veriler şu şekilde sıralanabilir;

- Favori yerler,
- Rotalar,
- Kullanıcının bulunduğu konumlar,
- Geçmişte bulunduğu yerler [11].

2.1.5. Medya Araçları Adli Bilişimi

USB bellekler, harici diskler ve dijital müzik oynatıcıları günlük hayatta yaygın olarak kullanılmaktadır. Bu cihazlarda kayıtlara ait zaman damgaları da tutulmaktadır. Bu cihazlarda bulunan veriler delil niteliği taşıyabileceği için adli bilişim açısından önem arz etmektedir [10].

2.1.6. Sosyal Ağ Adli Bilişimi

Sosyal ağların hızlı büyümesi, suç faaliyetlerinde artışa neden olmuştur. Sosyal ağlardan elde edilebilecek çok sayıda veri bulunmaktadır. Burada mühim olan hangi verilerin nereden bulunacağına hâkim olmaktır [12].

Sosyal ağlara örnek vermek gerekirse; bloglar, forumlar, Facebook, Instagram, Linkedin ve internet sözlükleri bu kapsama girmektedir [10].

2.1.7. Bulut Adli Bilişimi

Bulut bilişim kuruluşlara fayda sunmakla beraber tehlikeleri de beraberinde getirmiştir. Bulut bilişimde sanal sunucularla beraber fiziksel sunucularda incelenmelidir. İncelemeler esnasında sunucularda birçok kullanıcının verisi bulunduğu için inceleme yapılacak alanların belirlenmesi zaman açısından avantaj sağlayacaktır [13].

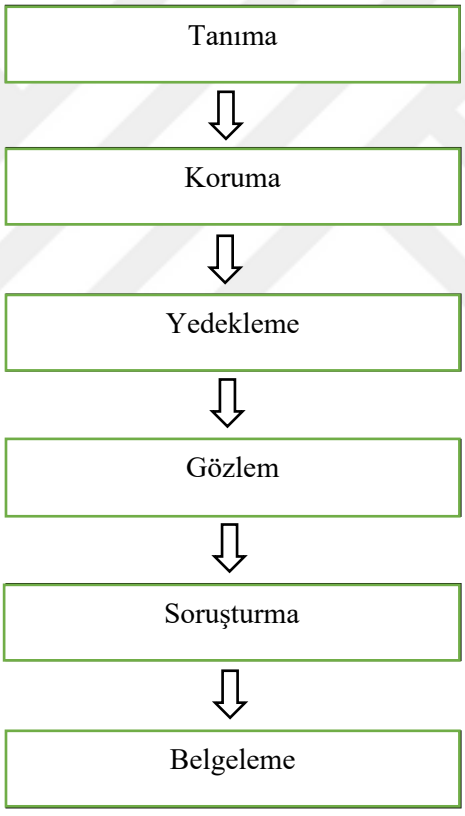
2.2. Ağ Adli Bilişiminin Detaylı İncelemesi

Ağ adli bilişimi, siber suçları daha iyi anlamak ve önlemek amacıyla ağ trafiğinin toplanması ve analizi ile ilgilenen adli bilişimin alt kümesidir. Ağ tabanlı hizmetlerinde artmasıyla beraber ağ adli bilişiminin de önemi artmıştır.

Ağ adli bilişimiyle, mesaj, dosya aktarımı, e-posta verileri ve web tarama geçmiş gibi verilerin alınabilmesi mümkündür.

İncelemeler sırasında verilerin anlamlandırılabilmesi açısından protokollerin bilinmesi önemlidir.

Ağ adli incelemesi sırasında takip edilmesi gereken altı adım bulunmaktadır. Bu adımlar olayın durumuna göre genişletilebilir. Bu adımlar Şekil 2.1 de gösterilmiştir;



Şekil 2.1 Ağ Adli Bilişimi İnceleme Adımları [9]

İnceleme esnasında birçok zorluk ile karşı karşıya kalınabilir. Bu zorluklardan birkaçı;

- Süreç sırasında üretilen verinin yönetilmesi,
- Adres sahteciliği,
- Veri bütünlüğü,
- Veri gizliliği,
- Veri depolama olarak sıralanabilir.

Zorluklar ile beraber birçok avantajı da vardır. Bunları sıralamak gerekirse;

- Güvenlik tehditlerini belirlemeye yardımcı olur,
- Güvenlik açıklarını belirlemeye yardımcı olur,
- Ağ kesinti süresinin azaltılmasına yardımcı olur,
- Yapılan iyileştirmeler sayesinde ağ kaynakları daha iyi kullanılabilir [9].

Ağ üzerinden elde edilen deliller uçucu ve dinamik deliller içermektedir, bu sebeple incelemeler proaktif bakış açısı ile yapılmaktadır.

Ağ analizinde; güvenlik amacıyla ağ üzerindeki anormal trafik tespiti ve müdahalesi gerçekleştirilmesinin önemli nedenlerinden biri; kötü niyetli kişilerce, saldı sonrası log kayıtlarının silinmesi ihtimalidir. Bazı analizlerde, log kayıtlarının silinmesi durumunda elde ki tek delil ağ trafiği olabilmektedir [14].

Ağ analizi sırasında, ağ paketlerinin anlamlandırılabilmesi için uygulama ve ağ protokollerine hâkim olunması araştırmacının işini kolaylaştıracaktır. Bunlardan bahsetmek gerekirse;

- Web protokolleri (http, https)
- Dosya aktarım protokolleri (SMB, NFS)
- E-posta protokolleri
- Ağ protokolleri

Adli ağ analizi yalnızca saldırı tespit işlemleri için yapılmamaktadır. Ağlardaki performans, güvenlik ve politika sorunlarını çözmek için de kullanılabilmektedir. Bunları daha detaylı olarak listelemek gerekirse;

- Güvenlik saldırılarının tespiti,
- Belli aralıklarla performans sorunlarını giderme,
- Bilişim teknolojileri ve insan kaynakları politikalarına uyum için kullanıcı etkinliğini izleme,
- Veri sızıntılarının kaynağını belirleme,

şeklinde sıralanabilir [15].

Ağ tabanlı soruşturmalar gerçekleştirilirken süreci daha başarılı sürdürebilmek için saldırılardan önce yapılabilecek adımlar;

İzlenecek süreç belirlenmelidir. İncelemelerin gerçekleştirilebilmesi için ağ paketlerinin yakalanmış olması gerekmektedir. Ağ paketlerinin yakalanması, depolanması ilkeleri için prosedürler belirlenmeli ve uygulanmalıdır.

Olaylar karşısından uygulanacak planlar önceden belirlenmelidir. Olay yönetiminin planmış olması, saldırı anında verilecek tepki süresini kısaltacak ve kısalan tepki süresi ile beraber saldırının oluşturacağı etkileri en aza indirmeye yardımcı olacaktır.

Kişisel yetenekler geliştirilmelidir. Ağ paketlerini yorumlama, dosyaları yakalama ve kötü niyetli olayları tespit etme, ağ ve uygulama protokolleri hakkında bilgi sahibi olmayı gerektirmektedir [16].Gelişen yetenekle olayın aydınlatılması sırasında hız kazandıracaktır ve olayı doğru yorumlamak için önem arz etmektedir.

2.2.1. Ağ Analizinde Kullanılabilecek Araçlar

Ağ üzerinde gerçekleşen olayların aydınlatılması için çeşitli araçlar kullanılmaktadır. Birçok araç, ağ trafiğinin gerçek zamanlı olarak incelememize izin vermektedir. Bu gerçek zamanlı izleme, önemli ölçüde insan gücü ve donanım kaynağı gerektirmektedir. Çoğunlukla gerçek zamanlı inceleme yerine tüm trafiği arşivleyip, gerektiği zaman taramalar yaparak incelemek daha pratik olmaktadır [17]. Ağ adli bilişimi için kullanılabilecek ücretsiz araçların bazıları grafik arabirimine sahipken, çoğu komut satırında çalışmaktadır.

Ağ analizleri yapılmadan önce; özellikle büyük boyutlardaki paketler ile çalışılacaksa, veriler filtreleme yoluyla azaltılmalıdır [16]. Filtreleme sonrası yapılacak analiz daha etkili olacaktır.

Ağ analizinde kullanılabilecek bazı araçlara bölümün devamında yer verilmiştir.

2.2.2. Wireshark

Wireshark ağ trafiğinin, bir grafik arayüz aracılığıyla izlenmesini sağlayan bir programdır. Uygulama, kurulu olduğu bilgisayar üzerinden anlık ağ trafiğinin izlenmesine olanak tanımasının yanı sıra daha önce kaydedilmiş dosyaların incelenmesini yapmak da mümkündür.

Program ayrıca özet bilgi elde etmemize de olanak tanımaktadır. Bu özet bilgiler dosya ismi, yakalanan paket sayısı, paket yakalama işleminin süresi gibi bilgiler sunmaktadır.

Bu tez çalışmasında ağ paketlerini toplamak amacıyla wireshark aracı kullanılmıştır.

2.2.3. Tcpdump

Tcpdump paket analiz aracıdır. Komut satırında çalışmaktadır. Tarama yapan cihazın bağlı olduğu ağ üzerinden iletilen/alınan paketleri kaydetme ve inceleme imkânı sağlamaktadır [18].

2.2.4. NetworkMiner

Windows işletim sistemlerinde kullanılabilecek adli ağ analiz aracıdır. Bu araç ile ağ üzerinde herhangi bir trafik oluşturmadan işletim sistemlerini, oturumları, açık port bilgilerini algılamak için kullanılabilecek bir paket yakalama aracı olarak tanımlayabilmek mümkündür [19].

2.2.5. Argus

Bir veri ağı trafiği akışında görülen tüm ağ sistemlerinin durumunu ve performansını izlemek, raporlamak için kullanılabilecek bir araçtır. Bu araç, bağlantı, kapasite, kayıp ve gecikme gibi verileri raporlamaya olanak tanımaktadır [20].

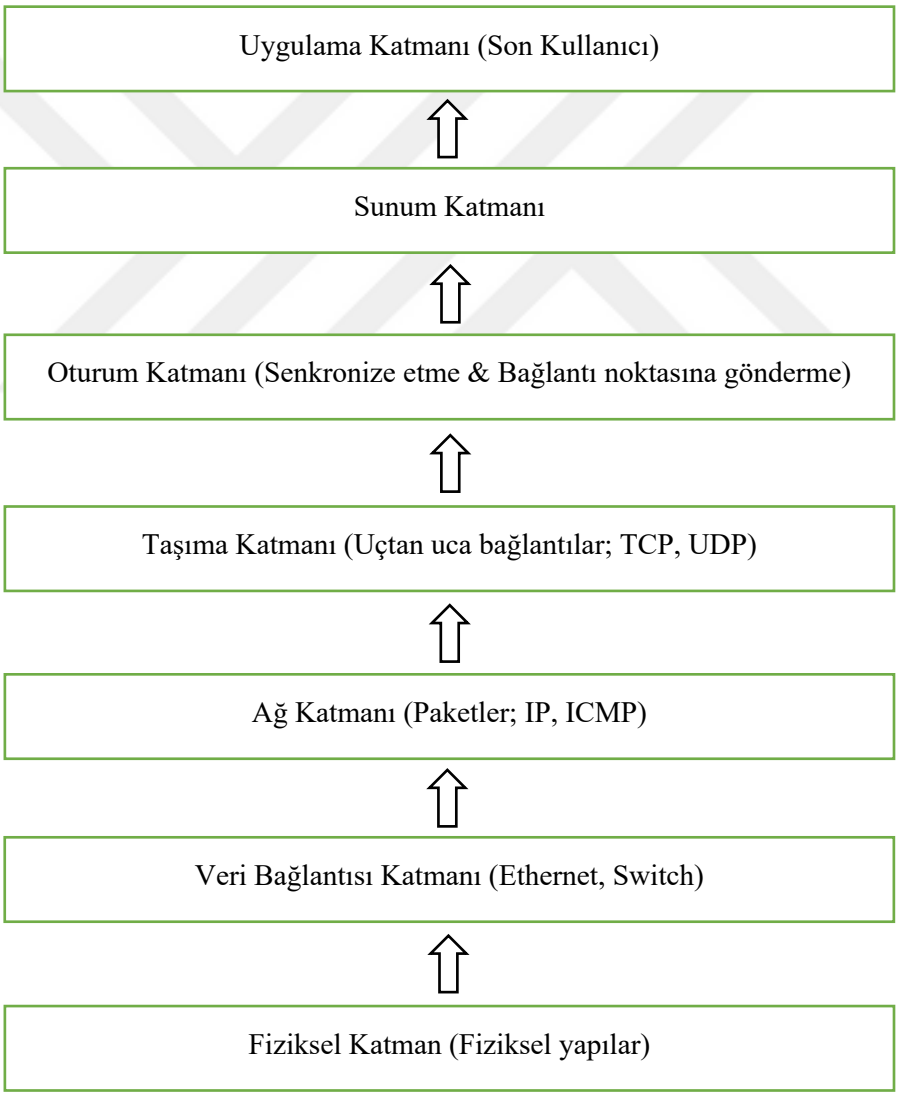
2.2.6. DoHlyzer

Https trafiğini algılayan bir ağ akış aracıdır. Python kullanılarak çalıştırılabilmektedir, bu nedenle birçok platformda kullanılabilmektedir. Bu araç ile ağ paketleri yakalanabileceği gibi, önceden alınmış pcap dosyalarının da incelenmesine olanak tanımaktadır [21].

3. KABLOSUZ AĞLARA YÖNELİK SİBER SALDIRILAR

Taşınılabilirliğin öneminin artmasıyla kablosuz ağların kullanımı da artış göstermiştir. Taşınılabilirlik ile beraber ağları genişletmek için de kablosuz erişim noktaları kullanılarak ağlar genişletilmektedir [22]. Genişletilen bu ağlara birçok kullanıcı bağlanmaktadır. Çok sayıda kullanıcının kullandığı, avantajlı kullanım sağlayan kablosuz ağlar, siber saldırılara açık olan yapılardır [23]. Bu yapıların güvenli kabul edilmesi için; özgünlük, gizlilik, bütünlük ve kullanılabilirlik şartlarını sağlaması gerekmektedir [24].

Siber saldırılar farklı katmanlarda gerçekleşmektedir. Katmanlar ve protokollere ilişkin yapı Şekil 3.1 de verilmiştir.



Şekil 3.1 OSI Modeli [23]

Şekil 3.1 de verilen katmanlardan fiziksel katman; veri paketlerinin cihazlar arasında aktarılmasını sağlamaktadır. Bu aktarım elektrik veya ışık sinyalleri ile gerçekleştirilmektedir. Diğer katmanlar, fiziksel katmana bağımlı şekilde çalışmaktadır.

Bir sonraki katman olan veri bağlantısı katmanı, fiziksel bağlantı katmanından gelen bitleri bir cihazdan diğerine aktarırken, bitleri paketlere ayırmaktadır. Bu katman, cihazların fiziksel adreslerinin kontrollerini sağlayarak veri paketlerini doğru noktalara iletir. Veri iletim işlemi bit bit gerçekleşmektedir.

İletişim kurmak için IP protokolünü kullanan ağ katmanı, verilerin iletilmesi sırasında adresleme ve yönlendirme işlemlerini gerçekleştirmektedir.

Taşıma katmanında ise veriler segmentlere ayrılır ve hedefte tekrardan birleştirilir. Bu katmanda, taşıma katmanı başlığı eklenerek verilerin doğru yere iletilmesi sağlanır.

Oturum katmanında; uygulamalar arasında gerekli oturumların açılması, yönetilmesi ve sonlandırılması yapılan katmandır. Veri alışverişinin tam olarak sağlanabilmesi için, veri aktarımı sırasında oturumun yeterince süre açık kalması bu katmanda sağlanır. Bu katmanda bağlantı sağlanırken, sistemler çift yönlü iletişim başlatabilmektedir. Veri aktarım işlemi tamamlandıktan sonra sistem kaynaklarının gereğinden fazla harcanmaması için oturum sonlandırılır.

Sunum katmanında, veri paketlerinin alışverişi yapılırken hangi protokollerin kullanılacağı belirlenmektedir. Bu katmanda veri dönüştürme işlemi de gerçekleştirilmektedir. Uygulama katmanına aktarım için veriler uygun formatlara dönüştürülür.

Uygulama katmanı, son kullanıcılara görüntülenen katmandır. Bu katmanda kullanıcılar veri girişi sağlayabilmektedir. Bu katman kullanıcıların en net görüntülemeyi sağlayabildiği katmandır. Bu katman e-posta, anlık mesajlaşma, dosya aktarımı benzeri uygulamalar ile iletişime müsaade eden çeşitli protokoller içerir [25].

Tablo 3.1 de fiziksel katmanda yapılabilecek başlıca tehdit türleri verilmiştir.

Tablo 3.1 Fiziksel Katmanda Yapılabilecek Tehdit Türleri

Fiziksel Saldırılar	**Tanım**
Eavesdropping	Özel bilgilerin gözetimi
Jamming	İletilen bilgilerin bozulması
Side-channel Attacks	Gizli bilgilerin keşfi
Random Interference	Tehdit unsurlarının hedef sisteme müdahalesi
Timing Atak	Anahtar verileri elde etmek için şifreleme/şifre çözme tekniklerinin çalıştırılması için gereken sürenin hesaplanması

Tablo 3.2 de bağlantı katmanında yapılabilecek başlıca tehdit türleri verilmiştir.

Tablo 3.2 Bağlantı Katmanında Yapılabilecek Tehdit Türleri

Veri Katmanı Saldırıları	**Tanım**
MAC Spoofing	MAC adresi sahteciliği
Identity Theft	Kullanıcın MAC adresinin çalınması
Man in the Middle	Kullanıcıların ağının arasına girme
Network Injection	Sahte ağ paketlerinin enjeksiyonu
Mac Flooding	Ağ anahtarı güvenliğinin engellenmesi

Tablo 3.3 de ağ katmanında yapılabilecek başlıca tehdit türleri verilmiştir.

Tablo 3.3 Ağ Katmanında Yapılabilecek Tehdit Türleri

Ağ Katmanı Saldırıları	**Tanım**
IP spoofing	IP adresi sahteciliği
IP hijacking	Bir kullanıcının IP adresine bürünme
Smurf atak	Bir ağı durdurmak için çok sayıda ICMP isteği gönderme
Sinkhole atak	Tehdit unsurları, baz istasyonuna verilerin alınmasını engellemek için büyük miktarda trafik oluşturmaya çalışır

Tablo 3.4 de taşıma katmanında yapılabilecek başlıca tehdit türleri verilmiştir.

Tablo 3.4 Taşıma Katmanında Yapılabilecek Tehdit Türleri

Taşıma Katmanı Saldırıları	**Tanım**
TCP flooding	Birçok sunucuya ping atılır.
UDP flooding	Çok fazla UDP paketi gönderilmesi
TCP dizisi ve tahmin saldırısı	TCP paketlerinin sırasını tahmin ederek hedef makineye TCP paketlerinin gönderilmesi

Tablo 3.5 de uygulama katmanında yapılabilecek başlıca tehdit türleri verilmiştir.

Tablo 3.5 Uygulama Katmanında Yapılabilecek Tehdit Türleri [24]

Uygulama Katmanı Saldırıları	**Tanım**
Malware atak	Tehdit unsurları, kodlama, komut dosyası oluşturma ve etkin içerik biçiminde kötü amaçlı yazılım oluşturur.
SQL Injection	Web sitelerinde sql ifadeleri çalıştırarak yetkisiz erişim elde etmeye çalışmak
Cross-site scripting	İstemci tarafı komut dosyalarını web sitelerine enjekte ederek, çeşitli erişim kontrol önlemleri atlanır
FTP bounce	Geçerli bir kullanıcının kimliğine bürünerek yasa dışı erişim elde edilir

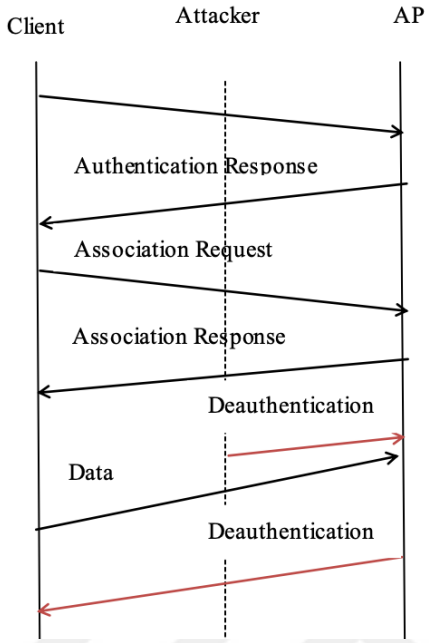
Tez çalışmasında kullanılmak üzere gerekli verileri elde etmek için hazırlanan test ortamında 6 adet kablosuz ağ saldırısı gerçekleştirilmiştir. Bu saldırılar aşağıda sunulmuştur:

1. Deauthentication Saldırısı
2. DoS
3. UDP Flood
4. ICMP Flood
5. SYN Flood
6. Ortadaki Adam (Man in The Middle).

3.1. Deauthentication

İkinci katman saldırısı olan [26] deauthentication (kimlik doğrulama) saldırılarında, bir kablosuz ağa bağlı olan cihazlara birçok paket göndererek cihazların ağdan düşürülmesi hedeflenmektedir. Bu saldırı türünü Servis Hizmet Reddi (DoS) saldırıları sınıfında düşünülebilir [27].

Cihazların bağlı oldukları ağa erişimlerini kaybettirmeye yönelik yapılan deauthentication saldırı yapısı Şekil 3.2 de verilmiştir.



Şekil 3.2 Deauthentication Saldırısı [28]

3.2. Dos Saldırısı

Bu saldırı türünde, tehdit unsuru hedef cihazlarda/sistemlerde bilgiye/servislere erişimi engellemeyi hedeflemektedir. Hedef herhangi bir cihaz olabileceği gibi, ağ bağlantısı veya site erişimi de olabilmektedir. Yapılan saldırı sonucunda hedef cihaz/sunucu servis dışı olmaktadır. Sürecin nasıl işlediğinden bahsetmek gerekirse; tehdit oluşturan makine, hedef e çok sayıda istek gönderir ve hedef makine/sunucu bu isteklere cevap verir. Bu isteklerin sayısı çok fazla olması sebebiyle hedef cihazın/sunucunun kaynakları bir süre sonra tükenir ve sistem bu süre içerisinde kullanılmaz hale gelir. Bu saldırıların belirtileri aşağıdaki gibidir;

- Sistem performansının düşmesi,
- Web sitelerine yapılan dos saldırılarında, web sitesinin bir süre kullanılamaması,
- Spam maillerde artış [29].

3.3. UDP Flood

Bu saldırı türünde temel olarak rastgele bir kaynak IP oluşturulur ve udp paketleri ana bilgisayarlar arasındaki rastgele hedeflere gönderilir [30]. Saldırıya uğraya makine;

- İlgili portu dinleyen bir uygulama olup olmadığını kontrol eder,
- Hiçbir uygulamanın portu dinlemediği görüldüğü zaman ICMP “Hedefe Ulaşılamıyor” şeklinde bir paket ile cevap verir.

Çok sayı da UDP paketi gönderildiği zaman, hedef sistem yanıt olarak çok fazla sayıda ICMP paketi göndermek durumunda kalır ve bu durum sistemin kaynaklarının tükenmesine sebep olabilmektedir. Bu durumda başka istemcilerin sisteme erişememe durumu söz konusu olabilmektedir [31].

3.4. ICMP Flood

ICMP Flood Saldırısı, hedef kullanıcının canlı olup olmadığını kontrol etmek için hedef cihaza bir echo paketi gönderilmesini sağlayan İnternet Kontrol Mesajı Protokolünden (ICMP) yararlanmaktadır. Bu saldırı türünde hedef cihaza büyük hacimlerde ping paketleri gönderilir. Bu paketler hedeften yanıt ister ve bunun sonucunda hedef ağın bant genişliği dolar. ICMP Flood saldırısı sırasında kaynak IP sahte olarak kullanılabilmektedir. Tehdit unsurunun gerçek kimliğini gizlemek için IP sahtekarlığı yaptığı durumlarda, saldırının izinin sürülmesi zorlaşmaktadır [32].

3.5. SYN Flood

Sistemler üzerinde veri alışverilerinde sunucu ve hedef arasında üçlü el sıkışması olayı gerçekleşmektedir. Üçlü el sıkışması denilen durum da hedef anahtar SYN paketini alır ve karşılık gelen kaynak tablosunda IP adresi, kaynak bağlantısı gibi bilgiler kontrol edilir. Bu işlemler tamamlandıktan sonra, SYN-ACK paketleri önceden oluşturulmuş tanımlama bilgileri ile istemciye tekrar gönderilir. En son adımda ise, anahtar ACK paketini aldığında tablo tekrar sorgulanır ve istemciden doğru tanımlama bilgisi gelip gelmediğini doğrulamak için onay numarasını kontrol eder. Tüm adımlar başarılı bir şekilde tamamlanırsa doğrulama başarılı olur [33].

SYN Flood atağında ise bu üçlü el sıkışması sırasında araya girilerek atak başlatılır. Bu saldırıda ki amaç servis hizmet reddi saldırılarında olduğu gibi sunucuya taşıyabileceğinden fazla veri göndererek sistemi yormak ve bağlantı sağlanmasını engellemektir.

3.6. Ortadaki Adam (Man in The Middle (MiTM))

Man in the Middle saldırı türünde iki bağlantı arasındaki verileri dinlemek hedeflenmektedir. Bu saldırı türünde veriler dinlenebildiği gibi veriler üzerinde değişiklikler de yapmak mümkündür. Saldırının gerçekleştirilme mantığı şu şekilde ifade edilebilir; kablosuz ağ yayını bulunan ortamlarda, bu ağa bağlı olan kişilerin verilerini dinlemek amacıyla network trafiğinin tehdit unsuru makine üzerinden geçecek şekilde yönlendirilerek, verilerin ele geçirilmesi. Bu araya girme işlemi hedef ile ağ unsurları arasında gerçekleşmektedir. Bu ağ unsurları; modem, router, sunucu veya switch olabilir [34].

MiTM saldırıları, istemcilerin bağlı bulunduğu ağdan bağlantısını kesmeden konumunu elde ettiği için tespit edilmesi zor bir saldırı türüdür [35].

4. MATERYAL VE METOT

Çalışma ortamının oluşturulması sırasında kullanılan yazılımlar seçilirken ücretli/ücretsiz birçok yazılım olduğu gözlemlenmiştir. Bunlar arasında aşağıda da ismine yer verilen ücretsiz yazılımlar bu tez çalışmasında kullanılmak üzere seçilmiştir. Kullanılan cihazlar ise uygulamaların ihtiyaçları göz önünde bulundurularak seçilmiştir.

Test ortamı için kullanılan cihazlar ve detayları şu şekildedir;

- Windows 10 Pro – x64 İşlemci – 8,00 GB RAM – Bilgisayar
- Windows 10 Home – x64 İşlemci – 4,00 GB RAM – Bilgisayar
- Tp-link – Archer C5v – AC1200 Wireless Dual Band Gigabit VoIP Router
- USB 2.0 Wireless 802.INN

Test ortamı için kullanılan yazılımlar ise şu şekildedir;

- Oracle VM VirtualBox 6.1.16
- Debian – x64 İşlemci – 2,00 GB RAM – Sanal Makine
- Wireshark 4.0.5
- CicFlowMeter

Windows 10 Pro cihaza sanal makine kurulumu gerçekleştirilerek siber saldırıların gerçekleştirileceği makine haline getirilmiştir. Sanal makine kurulumu için Oracle VM VirtualBox yazılımı seçilmiştir. Sanal makineyi kablosuz ağa bağlamak için USB 2.0 Wireless adaptör kullanılmıştır.

Uygulamaların gerçekleştirilmesi için hedef makine olarak Windows 10 Home cihazı kullanılmıştır. Ağ paketlerinin dinlenmesi için kullanılan wireshark aracı bu makine üzerine kurulup, ağ dinleme işlemleri hedef makine olan bu cihaz üzerinden gerçekleştirilmiştir.

CicFlowMeter aracı ise wireshark ile elde edilen paketlerin özellik çıkarma işlemi için kullanılmıştır.

4.1. Siber Güvenlikte Makine Öğrenmesi Uygulamaları

Makine öğrenmesini kısaca tanımlamak gerekirse; belirli algoritmalar kullanılarak verilerin ayrıştırılması, ayrıştırılan verilerin öğrenilmesi sonucunda konu hakkında belli bir yargıya veya tahmine varılması olarak açıklanması mümkündür [36].

Makine öğrenimi sayesinde verileri anlamlandırabilmemiz mümkün hale gelmektedir. Teknolojinin gelişmesi beraberinde işlenmesi gereken çok fazla veriyi de oluşturduğu anlamına gelmektedir [37]. Günümüzde siber tehdit unsurlarının da ciddi oranda arttığı göz önünde

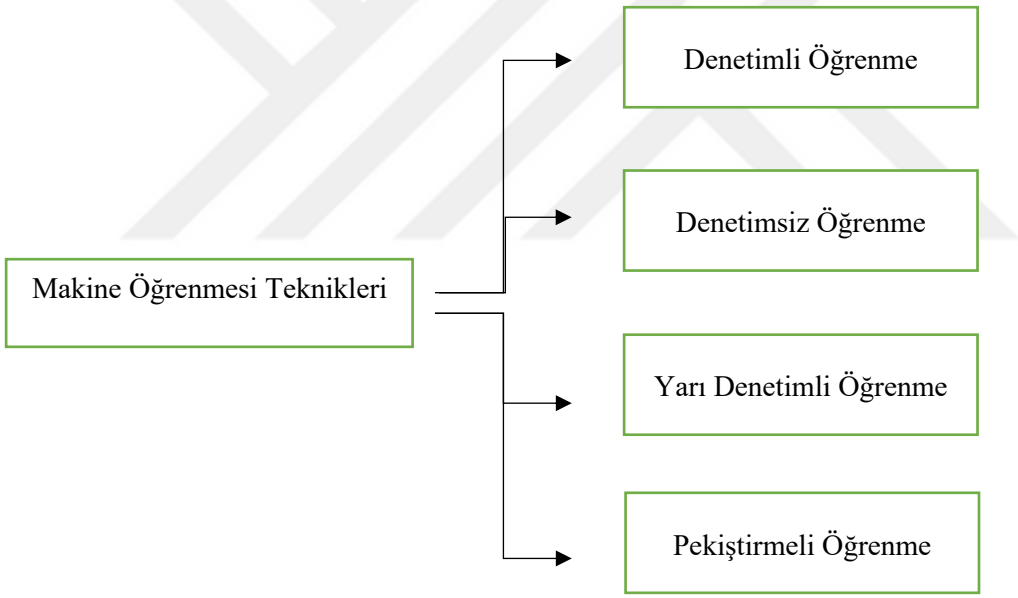
bulundurulduğunda bu alanda oluşan veri miktarının da ciddi boyutlarda olduğunu söylemek mümkündür. Bu tez çalışmasında bu alanda veriler elde edilerek makine öğrenmesi yöntemleri ile analizi gerçekleştirilmiştir.

4.1.1. Makine Öğrenmesi Teknikleri

Çevremizdeki makinelerin öğrenimlerini sağlamak amacıyla çeşitli makine öğrenmesi teknikleri kullanılmaktadır. Her makine öğrenmesi her veri için en iyi sonucu vermeyeceği için çeşitli makine öğrenmesi tekniği bulunmaktadır. Veri işleme sürecinde veriye en uygun öğrenim tekniğinin tespitinin sağlanması için ise performans metrikleri devreye girmektedir.

Makine öğrenmesine ilişkin teknikleri dört ana başlıkta toplarken, eğitim için kullanılabilecek birçok sınıflandırma algoritması bulunmaktadır.

Bu tez çalışmasında yalnızca, uygulama sırasında kullanılan sınıflandırma algoritmalarının detaylarına yer verilmiştir. Makine öğrenmesine ilişkin teknikler ise Şekil 4.1 deki gibidir.



Şekil 4.1 Makine Öğrenmesi Teknikleri

- **Denetimli Öğrenme**

Bu makine öğrenmesi türünde hedef, etiketlenmiş eğitim verisinden çıkarım yapmaktır. Denetimli öğrenmede eğitim verileri aynı zamanda eğitim örneklerini de içermektedir [38]. Yani, test aşamasında alınan çıktılar, eğitim esnasında verilen veri setlerinden edinilen bilgiler doğrultusunda oluşmaktadır.

- **Denetimsiz Öğrenme**

Bu tür modellerde veriler üzerinde test işlemleri gerçekleştikçe modelin karar verme yeteneği artmaktadır. Bu öğrenme modelinde verilen girdi verilerinden yola çıkarak veriler arasındaki ilişki öğrenilmeye çalışılmaktadır [39]. Veri sayısının ve test işlemlerinin sayısının arttırılması başarı oranını da arttıracaktır.

- **Yarı Denetimli Öğrenme**

Bu öğrenme modelinde ise etiketlenmiş veri sayısı azdır. Etiketlenmiş verilerden, etiketlenmemiş verileri tahmin işlemi gerçekleştirilmektedir. Bu öğrenme modelinde birden fazla deneme yapılarak, edinilen eğitim deneyimlerinden öğrenim gerçekleştirilerek en iyi performansı elde edilir [40].

- **Pekiştirmeli Öğrenme**

Bu tür öğrenme modellerinde en iyi sonucu elde etmek amacıyla belirli kurallar kullanılır. Bu modelde farklı birden fazla yöntem beraber kullanılmaktadır. En iyi sonuç veren işlem belirlenerek model oluşturulur [39].

4.1.2. Sınıflandırma Algoritmaları

Sınıflandırma algoritmalarını, eğitim verilerini kullanarak yeni gözlemlerin kategorisini belirlemek için kullanılan denetimli öğrenme tekniğidir. Sınıflandırma algoritmaları, veri kümesinde bulunan verilerden öğrenme işlemini gerçekleştirir ve bu öğrenimleri birkaç sınıfa/gruba sınıflandırır. Bu sınıflandırmalar, etiketler/kategoriler olarak çağrılabilir [41]. Bir sınıflandırmada temel amaç yeni bir verinin düşeceği sınıfı belirtmektir. Sınıflandırıcı olarak adlandırılan kavram, girdi olarak verilen verileri belirli bir kategoriye göre eşleyen algoritma iken; sınıflandırma modeli, eğitim için verilen girdi verilerini değerlendirerek bazı sonuçlar ortaya çıkarmaya çalışan yapılar olarak tanımlanabilmektedir. Özellik kavramını ise, gözlemlenen bir olgunun bireysel ölçülebilir bir özelliği olarak tanımlanabilir [42].

- **Naive Bayes**

Olasılıkların gözden geçirilmesini temel alan Navie Bayes teoremi ilk olarak Thomas Bayes tarafından kullanılmıştır. Bu teorem kullanılarak Navie Bayes Sınıflandırıcıları geliştirilmiştir [43]. Bu teorem kullanılarak; belli bir olay meydana geldiğinde, ikinci olayın gerçekleşme olasılığı bulunabilmektedir. Bahsedilen durumda birinci olay kanıt verilerini oluştururken, ikinci olay ise hipotezdir.

Bu sınıflandırıcının üç farklı modeli bulunmaktadır. Bu modeller;

- Gaussian,
- Multinominal
- Bernoulli şeklinde sıralanmaktadır [44].

Fazla sayıda değişken verisinin olduğu durumlarda kullanışlı bir sınıflandırıcı olduğunu söylemek mümkündür. [45] Ayrıca bu sınıflandırma algoritmasında diğer sınıflandırma algoritmalarından farklı olarak özellik verilerinin sayısı arttıkça elde edilen sonuçlar daha iyi olmaktadır [43].

Naive Bayes Sınıflandırıcısının avantajlarından bahsetmek gerekirse; Naive Bayes Sınıflandırıcısını anlamanın ve oluşturmanın kolay olduğunu söyleyebilmek mümkündür. Naive Bayes Sınıflandırıcısını büyük veri içeren örneklerde kullanırken bile hızlı bir şekilde verilerin eğitimini tamamladığını gözlemleyebilmek mümkündür.

Birçok alanda kullanılmış olan Naive Bayes Sınıflandırıcısı; insan hareketi tanıma projeleri, trafik sıkışıklığı projeleri ve tıbbi araştırma projelerinde gayet iyi sonuçlar vermiştir [45].

- **Support Vector Machines (SVM)**

Support Vector Machines (Destek Vektör Makineleri) sınıflandırma modeli regresyon analizi yapan bir modeldir [46]. Denetimli öğrenme olan destek vektör makineleri modeli, istatistiksel öğrenim teorisine dayanır.

Açıklayıcı değişkenler, doğrusal olmayan yapılar aracılığla yüksek boyutlu bir uzaya eşlenir ve ardından, her iki sınıfı da optimal olarak ayıran bir hiper düzlem oluşturulur. Bu hiper düzlem, marjları veya hiper düzlemden her sınıfın en yakın eğitim örneklerine olan mesafe toplamını maksimize ederken, sınıflandırma hatalarını en aza indirmeyi amaçlamaktadır [47].

Bu sınıflandırma yönteminde girdi olarak verilen verileri dönüştürmek için kernel fonksiyonları kullanmaktadır. Bu dönüşüm, girdi verilerinin iki sınıf arasında yüksek boyutlu bir alana dönüşüm işlemidir. Bu veri gruplarının arasında oluşan ayrım ne kadar yüksek olursa destek vektörleri makinelerinin performansı da aynı oranda artacaktır [48].

- **Gradient Boosting**

Gradient Boosting algoritması regresyon modelleri için kullanılabildiği gibi sınıflandırma modelleri için de kullanılmaktadır [49].

Gradient Boosting çalışma adımları;

Adım 1: Bir regresyon problemi çözülürken, her veri noktasının ilk tahminlerinin değerlerinin ortalaması olarak alınır. Olasılıkların günlüğü alınır ve bu değer sınıf tahmini için olasılık olarak kullanılır.

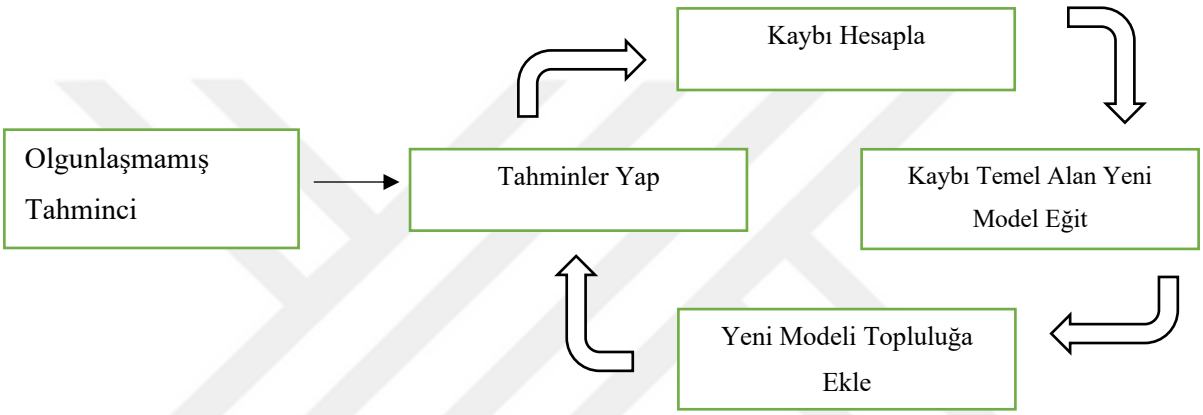
Adım 2: Tahminlerdeki kayıp değeri hesaplanır.

Adım 3: Tahmin değerleri kullanılarak yeni bir karar ağacı oluşturulur. Bu ağaç, öğrenmek için orijinal veri kümesi üzerinden eğitilir.

Adım 4: Bu yeni model topluluğa eklenir. Bu değer ile bir sonraki tahmin yapıldığında, ilk tahmin edici yeni karar ağacıyla birlikte kullanılacağı kastedilmektedir.

Adım 5: Karar ağaçları için tanımlanan sınıra ulaşılıncaya kadar veya yeni karar ağacının eklenmesine rağmen iyileştirme durana kadar ikinci adımdan itibaren tekrarlanır.

Aşağıdaki Şekil 4.2 deki diyagramda gradient boosting algoritmasının iş akışı gösterilmektedir:



Şekil 4.2 Gradient Boosting [50]

K-Nearest Neighbors

Bir sınıflandırma algoritması olan K-Nearest Neighbors kavramsal olarak anlaşılması en kolay sınıflandırma algoritmalarından biridir.

K-Nearest Neighbors sınıflandırma algoritmasında, örnek verilerin veri özniteliklerinin sayısı olduğu n boyutlu bir uzayda çizilmektedir. N-boyutlu uzaydaki her nokta, sınıf değeri ile etiketlenmektedir. Sınıflandırılmamış bir verinin sınıflandırılmasını keşfetmek için, n boyutlu uzayda çizilir ve en yakın k veri noktasının sınıf etiketleri not edilir. Genellikle k bir tek sayıdır. En yakın k veri noktası arasında maksimum sayıda oluşan sınıf, yeni veri noktasının sınıfı olarak alınır. Yani, karar k komşu noktanın oylanmasıyla verilmektedir. K-Nearest Neighbors sınıflandırma algoritmasının avantajlarından biri, paralel işleme uygun olmasıdır [51].

4.1.3. Performans Metrikleri

Sınıflandırma modelinin her bir sınıf için performansının değerlendirilmesi istenilen durumlarda, doğruluk dışında sınıf tabanlı performans metrikleri kullanabilir.

Tablo 4.1 deki ölçümler F1 puanı denklemleri için yapı taşı görevi görecektir. Tabloda verilmiş olan karışıklık matrisini kullanarak tez çalışmasında yapılan uygulamalarda da kullanılan sınıflandırıcılarda yapılan hatalar ile ilgili fikir edinilmesine olanak tanımaktadır [52].

Tablo 4.1 İki Sınıflı Karışıklık (Confusion) Matrisi

GERÇEK DEĞERLER		TAHMİN EDİLEN DEĞERLER	
	Classes	Positive (1)	Negative (0)
	Positive (1)	TP (True Positive)	FN (False Negative)
	Negative (0)	FP (False Positive)	TN (True Negative)

- **True Positive (TP)**

Gerçekte de, sınıflandırıcı tarafından edilen tahminde de pozitif olan örnekler olarak tanımlanır.

- **False Positive (FP)**

Gerçekte negatif olan, sınıflandırıcı tahmini olarak pozitif değer döndüren örnekler olarak tanımlanır.

- **False Negative (FN)**

Gerçekte pozitif olan, sınıflandırıcı tahmini olarak negatif değer döndüren örnekler olarak tanımlanır.

- **True Negative (TN)**

Gerçekte de, sınıflandırıcı tarafından edilen tahminde de negatif olan örnekler olarak tanımlanır [52].

Tablo 4.1 den yola çıkarak precision değeri denklem-1 de ki gibi hesaplanırken, recall değerinin hesaplanışına denklem-2 de yer verilmiştir. Daha detaylandırmak gerekirse; precision değeri ögelerin kesrinin pozitif olarak tahmin edilen birimlerinin toplam sayısına bölünmesidir, başka bir deyişle tahmin edilen pozitiflerin sütun toplamıdır.

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{4.1}$$

$$Recall = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4.2}$$

Precision, modelin pozitif olduğunu söyler ve pozitif olan birimlerin oranını ifade etmektedir.

Recall ise gerçek pozitif ögelerin, pozitif olarak sınıflandırılan birimlerin toplam sayısına bölünmesiyle elde edilmektedir. Özellikle yanlış negatif, model tarafından negatif olarak etiketlenmiş öğelerdir fakat aslında olumludurlar.

Recall, modelin pozitif sınıf için öngörücü doğruluğunu ölçer; sezgisel olarak modelin veri kümesindeki tüm pozitif birimleri bulma yeteneğini ölçmektedir.

Accuracy (Doğruluk), çok sınıflı sınıflandırmadaki en popüler metriklerden biridir. Denklem-3 de doğruluk değerinin hesaplanışı verilmiştir.

$$Accuracy = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}} \tag{4.3}$$

Accuracy, modelin tüm veri kümesi üzerinde ne kadar doğru tahmin yapıldığının genel bir ölçüsünü vermektedir.

F1 Skoru, sınıflandırma modelinin performansını karışıklık matrisinden başlayarak değerlendirir ve aşağıdaki 4 numaralı denklemde görüldüğü üzere Precision ve Recall ölçümlerinin harmonik kavramı altında toplar,

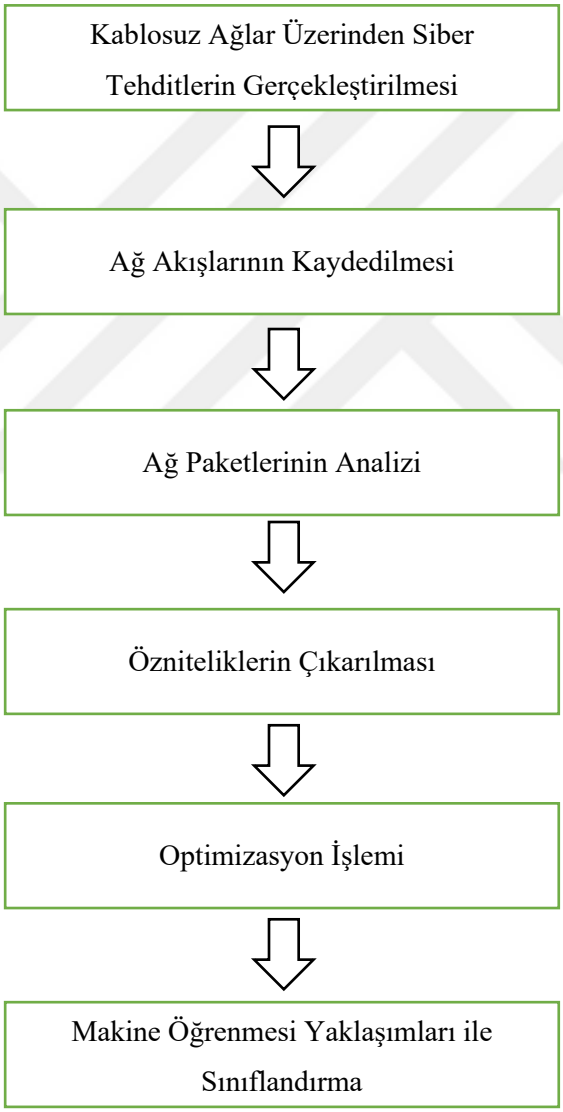
$$\text{F1 Score} = \left(\frac{2}{precision^{-1}+recall^{-1}} \right) = 2^* \left(\frac{\text{precision}*\text{recall}}{precision+recall} \right) \tag{4.4}$$

F1 skorunun formülü, precision ve recall arasında ağırlıklı ortalama olarak yorumlanabilir. F1 skorunun en yüksek değeri 1 iken, en düşük değeri 0 dır [53].

4.2. Sisteme İlişkin Metot

Bu tez çalışması için planlanan sistemin ilk adımı olarak kablosuz ağlara siber tehditlerin gerçekleştirilmesi planlanmıştır. Bir sonraki adım da ise siber tehditler gerçekleştirilirken oluşan ağ akışının kaydedilmesi adımı bulunmaktadır. Kaydedilen bu ağ paketlerinin analizi ise oluşturulan metodun üçüncü adımını oluşturmaktadır. Sonraki adımlar da ise, özniteliklerin çıkartılarak, optimizasyon işleminden sonra makine öğrenmesi yaklaşımları ile sınıflandırılması bulunmaktadır.

Oluşturulan çalışmanın sistemine ilişkin metot önerisi Şekil 4.3 deki gibidir.



Şekil 4.3 Önerilen Metot

Çalışmanın detayları ise şu şekildedir;

Deauthentication Saldırısı gerçekleştirilirken öncelikle hedef makinede wireshark aracı çalıştırılmıştır. Bu araç ile saldırı esnasında cihazda oluşan ağ trafiği kaydedilmiştir.

Saldırı gerçekleştirilirken öncelikle "airmon-ng start wlan0" komutu ile ağ monitör moda alınmıştır. Sonraki adım olarak etraftaki ağları taramak amacıyla "airodump-ng wlan0" komutu kullanılmıştır. "airodump-ng –channel *kanal numarası* –bssid *modem mac adresi* wlan0" ve "airoplay-ng –death *istenilen miktarda paket uzunluğu* -a *saldırılacak modem BSSID* -c *hedef mac adresi* wlan0" komutu çalıştırıldan sonra saldırı başlamıştır. Bir süre sonra hedef cihazın bant genişliği dolduğu için hedef cihaz ağdan kopmuştur.

Dos saldırısı gerçekleştirilirken hedef cihazımızda oluşacak ağ paketlerini kaydetmek amacıyla öncelikle wireshark aracı çalıştırılmıştır. Yapılan saldırı sonucunda hedef cihazın bant genişliği dolduğu için hedef cihaz ağ üzerinde işlemler yapamadığı gibi bir süre sonra ağdan düşmüştür.

UDP Flood, ICMP Flood ve SYN Flood saldırıları esnasında ağ trafiğinin kaydedilmesi amacıyla her saldırı öncesinde wireshark aracı başlatılmıştır. Bu saldırılarda hedef cihaza ayrı ayrı olmak üzere; UDP, ICMP ve SYN paketleri gönderilmiştir. Çok sayıda gönderilen paketler sebebiyle hedef sistemin bant genişliği dolmuştur. Bant genişliğinin dolması sebebiyle hedef sistem ağ üzerinde yaptığı işlemlerini gerçekleştirememiştir ve hizmet dışı kalmıştır.

Man in the Middle saldırısına başlanmadan önce wireshark çalıştırılarak ağ üzerinde oluşan paketler kayıt altına alınmıştır. Saldırı gerçekleştirilirken hedef makinenin verileri sunucuya gitmeden önce tehdit oluşturan makineden geçirilmesi hedeflenmiştir. Bu sayede hedef makinenin ağ üzerinden yaptığı işlemler sunucuya gitmeden önce tehdit oluşturan makineden geçmektedir. Bu saldırıyı gerçekleştirmek için kullanılan komutlar aşağıdaki gibidir;

- "apt-get install bettercap"
- "bettercap"
- "net.probe on"
- "net.sniff on"
- "arp.spoof on"
- "set net.sniff.local true"
- "set arp.spoof.fullduplex true"
- "set arp.spoof.targets *Hedef Makine Ip Adresi* "
- "dns.spoof on"
- "set dns.spoof.all true"

Wireshark aracı ile alınan pcap dosyalarından özellik çıkarmak amacıyla CICFlowMeter aracı kullanılmıştır. Bu araç sayesinde saldırılar sonucu alınan pcap dosyalarından sınıflandırma işlemi için gerekli olan özellik çıkarma işlemi tamamlanmıştır. Çıkarılan verilere ilişkin bilgiler Tablo 4.2 de verilmiştir.

Tablo 4.2 Çıkarılan Özellikler

Özellik	**Tanım**	**Özellik**	**Tanım**
src.port	Kaynak bağlantı noktası	pkt.len.min	Akışın sahip olduğu minimum uzunluk
dst.port	Hedef bağlantı noktası	pkt.len.max	Akışın sahip olduğu maksimum uzunluk
Protocol	Protokol	pkt.len.avg	Akışın ortalama uzunluğu
f1.dur	Akış (f1) süresi	pkt.len.std	Akışın sahip olduğu standart sapma uzunluğu
tot.fw.pk	İleri yön (FWD) paketleri toplamı	pkt.len.va	Paketin minimum varış süresi
tot.bw.pk	Geriye doğru yön (BWD) paketleri toplamı	fin.cnt	FIN içeren paket sayısı
tot.1.fw.pkt	FWD paketlerinin boyutu	syn.cnt	SYN içeren paket sayısı
fw.pkt.1.max	BWD deki toplam paket boyutu	rst.cnt	RST içeren paket sayısı
fw.pkt.1.max	Maksimum FWD paket boyutu	pst.cnt	PUSH içeren paket sayısı
fw.pkt.1.min	Minimum FWD paket boyutu	ack.cnt	ACK içeren paket sayısı
fw.pkt.1.avg	FWD paketlerinin ortalama boyutu	urg.cnt	URG içeren paket sayısı
fw.pkt.1.std	FWD paketinin standart sapma boyutu	cwe.cnt	CWE içeren paket sayısı
bw.pkt.1.max	Maksimum BWD paket boyutu	ece.cnt	ECE içeren paket sayısı
bw.pkt.1.min	Minimum BWD paket boyutu	down.up.ratio	Yükleme ve indirme oranı
bw.pkt.1.avg	BWD paketlerinin ortalama boyutu	pkt.size.avg	Ortalama paket boyutu

Tablo 4.2 (Devamı)

bw.pkt.1.std	BWD paketinin standart sapma boyutu	fw.seg.avg	FWD'de izlenen ortalama boyut
fl.byt.s	Bayt akış hızı	bw.seg.avg	BWD'de izlenen ortalama boyut
fl.pkt.s	Paket akış hızı	fw.byt.blk.avg	FWD'de ortalama baytların toplu iş oranı
fl.iat.avg	Akış arasındaki ortalama süre	fw.pkt.blk.avg	FWD'de ortalama paketlerin parti hızı
fl.iat.std	İki akış arasındaki standart sapma süresi	fw.blk.rate.avg	Ortalama FWD toplu oranı sayısı
fl.iat.max	Akış arasındaki maksimum süre	bw.byt.blk.avg	BWD'de ortalama baytların toplu iş oranı
fl.iat.min	Akış arasındaki minimum süre	bw.pkt.blk.avg	BWD'de ortalama paketlerin parti hızı
fw.iat.tot	İki FWD paketi arasındaki süre	bw.blk.rate.avg	Ortalama BWD toplu oranı sayısı
fw.iat.avg	İki FWD paketi arasındaki ortalama süre	subfl.fw.pk	FWD'deki bir alt akışta bulunan ortalama paket sayısı
fw.iat.std	İki FWD paketi arasındaki standart sapma süresi	subfl.fw.byt	FWD'deki bir alt akışta bulunan ortalama bayt sayıları
fw.iat.max	İki FWD paketi arasındaki maksimum süre	subfl.bw.pk	BWD'deki bir alt akışta bulunan ortalama paket sayısı
fw.iat.min	İki FWD paketi arasındaki minimum süre	subfl.bw.byt	BWD'deki bir alt akışta bulunan ortalama
bw.iat.tot	İki BWD paketi arasındaki süre	fw.win.byt	İlk pencerede iletilen bayt sayıları
bw.iat.avg	İki BWD paketi arasındaki ortalama süre	bw.win.byt	İlk pencerede geriye gönderilen bayt sayıları

Tablo 4.2 (Devamı) [54]

bw.iat.std	İki BWD paketi arasındaki standart sapma süresi	fw.act.pkt	Minimum 1 bayt TCP yüküne sahip FWD paketlerinin sayısı
bw.iat.max	İki BWD paketi arasındaki maksimum süre	fw.seg.min	FWD'de izlenen minimum segment boyutu
bw.iat.min	İki BWD paketi arasındaki minimum süre	atv.avg	Bir akışın boşta kalmadan önce aktif olduğu ortalama süre
fw.psh.flag	FWD paketleri için PSH bayrağının ayarlanma sayısı (UDP için 0)	atv.std	Bir akışın boşta kalmadan önce aktif olduğu standart sapma süresi
bw.psh.flag	BWD paketleri için PSH bayrağının ayarlanma sayısı (UDP için 0)	atv.max	Bir akışın boşta kalmadan önce aktif olduğu maksimum süre
fw.urg.flag	FWD paketleri için URG bayrağının ayarlanma sayısı (UDP için 0)	atv.min	Bir akışın boşta kalmadan önce aktif olduğu minimum süre
bw.urg.flag	BWD paketleri için URG bayrağının ayarlanma sayısı (UDP için 0)	idl.avg	Bir akışın etkin hale gelmeden önce boşta kaldığı ortalama süre
fw.hdr.len	Başlıklar için kullanılan FWD toplam bayt sayısı	idl.std	Bir akışın aktif hale gelmeden önce boşta kaldığı standart sapma süresi
bw.hdr.len	Başlıklar için kullanılan BWD toplam bayt sayısı	idl.max	Akışın aktif hale gelmeden önce maksimum boşta kaldığı süre
fw.pkt.s	İleriye aktarılan saniye başına paket sayısı	idl.min	Akışın aktif hale gelmeden önce minimum boşta kaldığı süre
bw.pkt.s	Geriye doğru iletilen saniye başına paket sayısı		

Tablo 4.2 de çıkarılan özellikleri verilen CICFlowMeter, pcap dosyalarından Bitflowlar oluşturan ve bu akışlardan özellik çıkaran açık kaynak bir araçtır [55], [56].

Saldırılar sonrası elde edilen paketlerin özellik çıkarma işlemleri tamamlandıktan çeşitli sınıflandırma algoritmalarıyla işleme sokulmuştur. Bu algoritmalar; Naive Bayes, Gradient Boosting, Support Vector Machines ve K-Nearest Neighbors dir.

Sınıflandırma işlemi gerçekleştirilirken öncelikle algoritmalar için gerekli olan paketler içeri aktarılmıştır.

Sonraki adımda veri seti dosyasını okuma işlemi gerçekleştirildikten sonra, datamızda sayısal olmayan sütunlar bulunduğu için; bu sütunları filtreleme işlemi gerçekleştirilmiştir. numeric_columns listesine, verideki sayısal olmayan sütunların adı eklenir. Daha sonra, data_numeric değişkenine sadece sayısal sütunları içeren bir veri çerçevesi atanır.

Bir sonraki adım olarak bağımsız değişleri (x) ve hedef değişkeni (y) ayırma işlemi gerçekleştirilmiştir. x değişkenine, data_numeric veri çerçevesinin son sütunu hariç tüm sütunlar atanır ve son olarak y değişkenine ise data_numeric veri çerçevesinin son sütunu atanmaktadır.

Veri kümesi (x ve y) eğitim ve test setlerine ayrıldıktan sonra, train_test_split fonksiyonu kullanılarak ise x ve y veri kümesi, %80 eğitim ve %20 test oranında rastgele seçilen veriler ile bölünmektedir. random_state parametresi ile ise veri setinin rastgele bölünmesi veya rastgele bir süreçte tekrarlanabilirliği için kullanılmıştır.

Model oluşturma işlemi her bir sınıflandırma algoritması için gerçekleştirilmiştir ve eğitim işlemi gerçekleştirilmiştir. Eğitim veri seti (x_train ve y_train) üzerinden eğitilmiştir.

Eğitilen model kullanılarak test veri kümesi (x_test) üzerinden tahmin işlemi yapılmıştır.

Tahmin sonuçları (y_pred) ve gerçek hedef değerleri (y_test) karşılaştırılarak kullanılan modelin doğruluğu hesaplanarak ekrana yazdırılmış ve karşılaştırma işlemi için gerekli olan veriler elde edilmiştir.

Support Vector Machines sınıflandırma algoritmasının kodları için gerekli olan paketler eklenmiş, csv dosyasını okuma ve sayısal olmayan sütunları filtreleme işlemi gerçekleştirilmiştir.

Sonraki adımda ise bağımsız değişleri (x) ve hedef değişkeni (y) ayırma işlemi gerçekleştirilmektedir. x değişkenine, data_numeric veri çerçevesinin son sütunu hariç tüm sütunlar atanır ve son olarak y değişkenine ise data_numeric veri çerçevesinin son sütunu atanmaktadır. Ayrıca kategorik değişkenler sayısal değerlere dönüştürülmüştür.

Veri kümesi (x ve y) eğitim ve test setlerine ayrıldıktan sonra, train_test_split fonksiyonu kullanılarak x ve y veri kümesi, %80 eğitim ve %20 test oranında rastgele seçilen veriler ile bölünmektedir. Kullanılan random_state parametresi ise veri setinin rastgele bölünmesi veya rastgele bir süreçte tekrarlanabilirliği için kullanılmıştır.

Model oluşturma ve doğruluk skoru hesaplama işlemleri yapıldıktan sonra Support Vector Machines sınıflandırma işlemi tamamlanmıştır.

Gradient Boosting sınıflandırma algoritmasının kodları için gerekli olan paketler içeri aktarılmıştır.

Sonraki adımda csv dosyasını okuma işlemi gerçekleştirildikten sonra, datamızda sayısal olmayan sütunlar bulunduğu için; bu sütunları filtreleme işlemi gerçekleştirilmiştir. numeric_columns listesine, verideki sayısal olmayan sütunları adı eklenir. Daha sonra, data_numeric değişkenine sadece sayısal sütunları içeren bir veri çerçevesi atanır.

Bağımsız değişleri (x) ve hedef değişkeni (y) ayırma işlemi gerçekleştirilmiştir. x değişkenine, data_numeric veri çerçevesinin son sütunu hariç tüm sütunlar atanır ve son olarak y değişkenine ise data_numeric veri çerçevesinin son sütunu atanmaktadır.

Kategorik değişkenleri sayısal değerlere dönüştürme işlemi yapıldıktan sonra veri kümesi (x ve y) eğitim ve test setlerine ayrılmaktadır. train_test_split fonksiyonu kullanılarak ise x ve y veri kümesi, %80 eğitim ve %20 test oranında rastgele seçilen veriler ile bölünmektedir. Kullanılan random_state parametresi ise veri setinin rastgele bölünmesi veya rastgele bir süreçte tekrarlanabilirliği için kullanılmıştır.

Model oluşturma ve eğitme işlemleri tamamlandıktan sonra performans metrikleri hesaplanmıştır.

K-Nearest Neighbors sınıflandırma algoritmasının kodları için gerekli olan paketler içeri aktarıldıktan sonra csv dosyasını okuma işlemi gerçekleştirilmiştir. Datamızda sayısal olmayan sütunlar bulunduğu için; bu sütunları filtreleme işlemi gerçekleştirilmiştir. numeric_columns listesine, verideki sayısal olmayan sütunları adı eklenir. Daha sonra, data_numeric değişkenine sadece sayısal sütunları içeren bir veri çerçevesi atanmıştır. Sonraki adımda bağımsız değişleri (x) ve hedef değişkeni (y) ayırma işlemi gerçekleştirilmiştir. x değişkenine, data_numeric veri çerçevesinin son sütunu hariç tüm sütunlar atanmıştır ve son olarak y değişkenine ise data_numeric veri çerçevesinin son sütunu atanmaktadır. Kategorik değişkenleri sayısal değerlere dönüştürme işlemi tamamlandıktan sonra, veri kümesi (x ve y) eğitim ve test setlerine ayrılmaktadır. train_test_split fonksiyonu kullanılarak ise x ve y veri kümesi, %80 eğitim ve %20 test oranında rastgele seçilen veriler ile bölünmektedir. Kullanılan random_state parametresi ise veri setinin rastgele bölünmesi veya rastgele bir süreçte tekrarlanabilirliği için kullanılmıştır.

Model oluşturma ve eğitme işlemleri tamamlandı ve sınıflandırma işlemlerinin de tamamlanmasıyla, sınıflandırma işlemleri sonucu elde edilen performans değerleri ile en iyi performansı veren sınıflandırma algoritması gözlemlenmiştir.

5. BULGULAR VE TARTIŞMA

Yapılan çalışmalar sonucunda her bir sınıflandırma için performans değerleri (metrikleri) hesaplanmıştır.

Sınıflandırma işlemi için kullanılan verilerin, veri miktarına ilişkin değerler Tablo 5.1 de verilmiştir. Veri miktarlarının dağılımı, elde edilen performans değerlerinin yorumlanmasında etkili olmuştur.

Tablo 5.1 Veri Miktarları

Saldırı Adı	Veri Miktarı
Deauthentication	122
Dos Atak	5430
ICMP Flood	143
SYN Flood	140
Man in The Middle	2161

Navie Bayes sınıflandırma algoritmasının doğruluk değeri 0.74 olup, diğer performans metriklerine ilişkin değerler Tablo 5.2 de verilmiştir.

Tablo 5.2 Navie Bayes Algoritması Performans Değerleri

	Precision	Recall	F1-Score
Deauthentication	0.00	0.00	0.00
Dos Atak	0.91	0.99	0.95
ICMP Flood	0.00	0.00	0.00
SYN Flood	0.11	0.50	0.18
Man in The Middle	1.00	0.32	0.48

Gradient Boosting sınıflandırma algoritmasının doğruluk değeri 0.91 olup, diğer performans metriklerine ilişkin değerler Tablo 5.3 de verilmiştir.

Tablo 5.3 Gradient Boosting Algoritması Performans Değerleri

	Precision	Recall	F1-Score
Deauthentication	0.33	0.25	0.29
Dos Atak	0.98	0.99	0.99
ICMP Flood	0.00	0.00	0.00
SYN Flood	0.33	0.50	0.40
Man in The Middle	0.91	0.86	0.89

Support Vector Machine sınıflandırma algoritmasının doğruluk değeri 0.84 olup, diğer performans metriklerine ilişkin değerler Tablo 5.4 de verilmiştir.

Tablo 5.4 Support Vector Machine Algoritması Performans Değerleri

	Precision	Recall	F1-Score
Deauthentication	0.00	0.00	0.00
Dos Atak	0.96	0.86	0.90
ICMP Flood	0.00	0.00	0.00
SYN Flood	0.00	0.00	0.00
Man in The Middle	0.69	0.96	0.80

K-Nearest Neighbors sınıflandırma algoritmasının doğruluk değeri 0.85 olup, diğer performans metriklerine ilişkin değerler Tablo 5.5 de verilmiştir.

Tablo 5.5 K-Nearest Neighbors Algoritması Performans Değerleri

	Precision	Recall	F1-Score
Deauthentication	0.00	0.00	0.00
Dos Atak	0.93	0.93	0.93
ICMP Flood	0.00	0.00	0.00
SYN Flood	0.00	0.00	0.00
Man in The Middle	0.79	0.82	0.80

Tüm sınıflandırma algoritmalarının performans değerleri Tablo 5.6 de verilmiştir.

Tablo 5.6 Sınıflandırma Algoritmalarının Performans Değerlerinin Karşılaştırması

	Accuracy	Precision	Recall	F1-Score
Navie Bayes	0.74	0.40	0.36	0.32
Support Vector Machines	0.84	0.51	0.52	0.51
Gradient Boosting	0.91	0.33	0.36	0.34
K-Nearest Neighbors	0.85	0.44	0.35	0.34

Yukarıdaki tabloda da görüldüğü üzere en yüksek doğruluk değeri Gradient Boosting sınıflandırma algoritmasında alınırken, en düşük değer Navie Bayes sınıflandırma algoritmasından alınmıştır. Support Vector Machines sınıflandırma algoritması doğruluk değeri, K-Nearest Neighbors sınıflandırma algoritmasının doğruluk değeri de bu sınıflandırma algoritmalarından daha yüksek çıkmıştır.

6. SONUÇLAR

Oluşturulan test ortamında kablosuz ağlar üzerinde, sistemleri tehdit edebilecek olası saldırılar gerçekleştirilmiştir. Yapılan bu siber tehditler ile beraber ağ akışı kaydedilmiştir. Buradaki amaç, olası siber tehditlerin oluşması sırasında sistemlerde oluşabilecek trafiği canlandırarak bu ağ paketlerini kaydederek, kaydedilen ağ paketlerinin analizini sağlayabilmektir. Farklı saldırılar, ağda farklı sonuçlara yol açmaktadır. Bu ortaya çıkan sonuçlara ağın gösterdiği tepkinin farklı olması sebebiyle bir sonraki adım olan öznitelik çıkarma adımında, her bir saldırı için, farklı veriler elde etmemize olanak tanımaktadır.

Bir sonraki adımda, oluşturulan test ortamından elde edilen ağ paketlerinin özniteliklerinin çıkartılarak bir sonraki adım olan optimizasyon işlemine ve makine öğrenmesi yaklaşımları ile sınıflandırma adımlarına hazır hale getirilmiştir. Öznitelikleri çıkarma adımında kullanılan CICFlowMeter ile her saldırı sonucunda elde edilen verilerde farklı verimlilikte öznitelik çıkartılmıştır. Bunun sebebi CICFlowMeter in bazı saldırılarda saldırı paketlerini özelliklerine ayıramamasından kaynaklanmaktadır.

Sınıflandırma işlemi için farklı sınıflandırma algoritmaları kullanılarak, kullanılan sınıflandırma algoritmaları sonucunda veri setinin elde ettiği skorlar toplanmıştır. Doğruluk skoru Navie Bayes, Support Vector Machine, Gradient Boosting ve K-Nearest Neighbors algoritmaları için sırasıyla şu şekildedir; 0.74, 0.84, 0.91 ve 0.85. K-Nearest Neighbors ve Support Vector Machines algoritmalarının skorları çok yakın çıkarken, en yüksek ve en düşük skorların elde edildiği sınıflandırma algoritmaları arasındaki skor farkı fazladır. Accuracy (Doğruluk) değeri ne kadar yüksekse, algoritmanın genel anlamda doğru sınıflandırma yaptığı olarak değerlendirilse de, siber saldırılar için sınıflandırmalarda dengesizlikler oluşabileceği için diğer performans metriklerinin de göz önünde bulundurulması gerekmektedir.

Precision (Hassasiyet) değeri ise Navie Bayes, Support Vector Machine, Gradient Boosting ve K-Nearest Neighbors algoritmaları için sırasıyla şu şekildedir; 0.40, 0.51, 0.33 ve 0.44 tür. En yüksek sonuç elde edilen Support Vector Machine; diğerlerine göre daha yüksek değer alarak, dört algoritma arasında en yüksek hassasiyet değerine sahip olarak, yanlış pozitif sonuçlarının en aza indirildiği algoritma olarak değerlendirilebilir. Saldırıların tespitinde, yüksek hassasiyet değerleri önem taşımaktadır.

Recall (Duyarlılık) değeri ise Navie Bayes, Support Vector Machine, Gradient Boosting ve K-Nearest Neighbors algoritmaları için sırasıyla şu şekildedir; 0.36, 0.52, 0.36 ve 0.35 tir. Yüksek duyarlılık değeri saldırıların doğru olarak tespit edildiği anlamına gelebileceği gibi, değerin yüksek olması yanlış pozitiflerin artabileceği sonucunu da ortaya çıkarabilmektedir.

F1 Skoru Navie Bayes, Support Vector Machine, Gradient Boosting ve K-Nearest Neighbors algoritmaları için sırasıyla şu şekildedir; 0.32, 0.51, 0.34 ve 0.34 şeklindedir. Bu skor değeri precision (hassasiyet) ile recall (duyarlılık) arasındaki dengeyi ölçmektedir. Support Vector Machines algoritmasından elde edilen değer diğer üç algoritmaya göre yüksek çıkarken, Navie Bayes, Gradient Boosting ve K-Nearest Neighbors algoritmaları için oldukça yakın sonuçlar elde edilmiştir.

Yapılan işlemler sonucunda sınıflandırma işlemleri sonrasında en iyi skor Gradient Boosting sınıflandırma algoritmasından elde edilirken, en düşük skor değerini ise Navie Bayes sınıflandırma algoritmasında elde etmiştir.

ÖNERİLER

Bazı saldırıların özellik çıkarma işlemi sırasında özellik çıkarıcı olarak kullanılan CICFlowMeter, saldırı paketlerini özelliklerine ayıramadığı için yetersiz kalmıştır. Bu durum ile ilgi daha iyi bir özellik çıkarıcı yazılım yazılması veya halihazırda daha iyi performans veren bir yazılımın kullanılarak daha kapsamlı bir çalışma yapılması önerilmektedir.

Ayrıca yapılan siber saldırıların sayısı ve çeşitliliği arttırılarak çalışmanın genişletilmesi, farklı sınıflandırma algoritmaları da kullanılarak daha iyi skor elde eden sınıflandırma algoritmalarının tespit edilmesi önerilmektedir.

Son olarak, gelişen ve gelişmekte olan teknoloji ile beraber siber tehdit unsurlarının da arttığı göz önünde bulundurularak bilinçli insan kitlesinin arttırılması, bu hususta farkındalık çalışmalarıyla beraber bilimsel çalışmaların da arttırılması önerilmektedir.

KAYNAKLAR

[1] Wazid, M., Das, A. K.; Chamola, V. and Park, Y. (2022). "Uniting cyber security and machine learning: Advantages, challenges and future research," *ICT Express*, vol. 8, no. 3, pp. 313–321, Sep. 2022, doi: 10.1016/J.ICTE.2022.04.007.

[2] Gönen, S., Ulus, H. İ. and Yılmaz, E. N. (2016). "Bilişim Alanında İşlenen Suçlar Ve Kişisel Verilerin Korunması," *Bilişim Teknolojileri Dergisi*, vol. 9, no. 3, Sep. 2016, doi: 10.17671/btd.90710.

[3] Ermeydan, D. (2018). "Türk Ceza Kanunu'nda Bilişim Suçları," 2018

[4] Dr. Winkelman, R. "Chapter 1: What is a Network?" https://fcit.usf.edu/network/chap1/chap1.htm (accessed Jun. 01, 2023).

[5] Atalay, N. S., Doğan, Ş., Akbal, E. and Tuncer, T. (2019). "Adli Bilişim Alanında Ağ Analizi," *BEU Journal of Science*, 2019, Accessed: Jun. 01, 2023.

[6] Gökalp, Ö. M. (2021). "Bilgisayar Ağları ve Adli Bilişim," 2021, doi: 10.13140/RG.2.2.30695.98727.

[7] Williams, L. (2023). "What is Digital Forensics? History, Process, Types, Challenges," https://www.guru99.com/digital-forensics.html (accessed Jun. 02, 2023).

[8] Lutkevich, B. (2021). "What is Computer Forensics (Cyber Forensics)?," https://www.techtarget.com/searchsecurity/definition/computer-forensics (accessed Jun. 02, 2023).

[9] "What is Network Forensics? - GeeksforGeeks," Mar. 15, (2022). https://www.geeksforgeeks.org/what-is-network-forensics/ (accessed Jun. 02, 2023).

[10] Yavuz, O. (2019) "Dijital Delillerde Adli Bilişim - Fordefence - Adli Bilişim Laboratuvarı," Jul. 18, 2019. https://fordefence.com/dijital-delillerde-adli-bilisim/ (accessed Jun. 02, 2023).

[11] "GPS Forensics and. Location Tracking | Envista Forensics." https://www.envistaforensics.com/services/digital-forensics-services/location-forensics/gps-forensics/ (accessed Jun. 02, 2023).

[12] Dodt, C. (2018) "Computer Forensics: Introduction to Social Network Forensics | Infosec Resources," Feb. 28, 2018. https://resources.infosecinstitute.com/topic/computer-forensics-introduction-social-network-forensics/ (accessed Jun. 02, 2023).

[13] Shivam, K. (2020) "Cloud forensics. Cloud computing is the future. This… | by Kumar Shivam | Medium," Apr. 11, 2020. https://kumarshivam-66534.medium.com/cloud-forensics-be18e14230de (accessed Jun. 02, 2023).

[14] Palmer, G.; "Ağ adli bilişimi - Vikipedi." https://tr.wikipedia.org/wiki/A%C4%9F_adli_bili%C5%9Fimi (accessed Jun. 02, 2023).

[15] Staff, I. (2014). “Top Five Things You Should Know About Network Forensics | IT Business Edge,” Jan. 30, 2014. https://www.itbusinessedge.com/it-management/top-five-things-you-should-know-about-network-forensics/ (accessed Jun. 02, 2023).

[16] Mounteer, J. (2016). “Network Forensics 101 | Info Security Advisor,” Jul. 05, 2016. https://infosecurityadvisor.wordpress.com/2016/07/05/network-forensics-101/#more-583 (accessed Jun. 02, 2023).

[17] Corey, V., Peterman, C.; Shearin, S.; Greenberg, M. S. and Bokkelen, J. V.; “Network Forensics Analysis.”

[18] “Tcpdump Nedir?” https://www.beyaz.net/tr/guvenlik/makaleler/tcpdump_nedir.html (accessed Jun. 02, 2023).

[19] “NETRESEC - Network Forensics and Network Security Monitoring.” https://www.netresec.com/ (accessed Jun. 02, 2023).

[20] “Argus – SecTools Top Network Security Tools.” https://sectools.org/tool/argus/ (accessed Jun. 02, 2023).

[21] “ahlashkari/DoHLyzer: DoHlyzer is a DNS over HTTPS (DoH) traffic flow generator and analyzer for anomaly detection and characterization.” https://github.com/ahlashkari/DoHlyzer (accessed Jun. 02, 2023).

[22] Gezgin, D. M. and Buluş, E.; “Kablosuz Erişim Noktalarına Yapılan DoS Saldırıları,” pp. 83–89.

[23] Berksoy, M. (2019). “OSI Modeli Nedir? OSI Katmanları 1 (Detaylı Anlatım),” 2019. https://teknotower.com/osi-modeli-nedir-detayli/ (accessed Jun. 21, 2023).

[24] Kadhim, A. N. and Sadkhan, S. B. (2021). “Security Threats in Wireless Network Communication-Status, Challenges, and Future Trends,” in *2021 International Conference on Advanced Computer Applications, ACA 2021*, Institute of Electrical and Electronics Engineers Inc., pp. 176–181. doi: 10.1109/ACA52198.2021.9626810.

[25] “OSI Katmanları, Osi Modeli ve Katmanlı İletişim » Uzman Posta,” May 07, 2023. https://uzmanposta.com/blog/osi-katmanlari/ (accessed Jun. 21, 2023).

[26] Cossa, D.; “The Dangers of Deauthentication Attacks in an Increasingly Wireless World.”

[27] Çelik, A. (2020). “Wi-Fi Deauthentication Attacks & Prevention,” Aug. 09, 2020. https://anilcelik.medium.com/tr-wi-fi-deauthentication-attacks-prevention-f0d550feff16 (accessed Jun. 03, 2023).

[28] Cheema, R., Bansal, D. and Sofat, S. (2011). “Deauthentication/Disassociation Attack: Implementation and Security in Wireless Mesh Networks,” 2011.

[29] Gezgin, D. M. and Buluş, E. (2012). "Kablosuz Ağların Güvenlik Açıklarının Eğitim Amaçlı İncelenmesi için Uygulama Tasarımı," 2012.

[30] Bansal, J. C. (2022). "Algorithms for Intelligent Systems Series Editors: Proceedings of International Conference on Communication and Computational Technologies," Sep. 2022.

[31] "UDP Saldırısı - Vikipedi." https://tr.wikipedia.org/wiki/UDP_Sald%C4%B1r%C4%B1s%C4%B1 (accessed Jun. 04, 2023).

[32] Harshita, (2017). "Detection and Prevention of ICMP Flood DDOS Attack," Mar. 2017.

[33] Shen, Z. Y., Su, M. W., Cai, Y. Z. and Tasi, M. H. (2021). "Mitigating SYN Flooding and UDP Flooding in P4-based SDN," in *2021 22nd Asia-Pacific Network Operations and Management Symposium, APNOMS 2021*, Institute of Electrical and Electronics Engineers Inc., Sep. 2021, pp. 374–377. doi: 10.23919/APNOMS52696.2021.9562660.

[34] "Ortadaki Adam (MITM) Saldırısı Nedir?" https://www.beyaz.net/tr/guvenlik/makaleler/ortadaki_adam_mitm_saldirisi_nedir.html (accessed Jun. 03, 2023).

[35] Thankappan, M., Rifà-Pous, H., and Garrigues, C. (2022). "Multi-Channel Man-in-the-Middle attacks against protected Wi-Fi networks: A state of the art review," *Expert Systems with Applications*, vol. 210. Elsevier Ltd, Dec. 30, 2022. doi: 10.1016/j.eswa.2022.118401.

[36] Aylak, B. L., Oral, O. and Yazici, K. (2021). "Using artificial intelligence and machine learning applications in logistics," *El-Cezeri Journal of Science and Engineering*, vol. 8, no. 1. TUBITAK, pp. 74–93, 2021. doi: 10.31202/ecjse.776314.

[37] "Siber Güvenlikte Makine Öğrenmesi - BGA Cyber Security - Siber Güvenlik Çözümleri." https://www.bgasecurity.com/2018/01/siber-guvenlikte-makine-ogrenmesi/ (accessed Jun. 07, 2023).

[38] Baloğlu, İ. (2021) "Android Tabanlı Mobil Cihazlardaki Sohbet İçeriklerinin Yapay Zeka Yöntemleri ile Sınıflandırılması," Elazıg, Jul. 2021.

[39] Koca, B. (2022). "Bilişim Teknolojileri Öğretmenlerinin Scratch Yazılımına İlişkin Özyeterlilik İnançlarının Makine Öğrenmesi ve Derin Öğrenme Yöntemleri ile Sınıflandırılması," Aksaray.

[40] Şengül, Z. (2022). "Makine Öğrenmesi Algoritmalarını Kullanarak Bitcoin Fiyat Tahmini," Edirne.

[41] "Classification Algorithm in Machine Learning - Javatpoint." https://www.javatpoint.com/classification-algorithm-in-machine-learning (accessed Jun. 02, 2023).

[42] Garg, R. (2018). "7 Types of Classification Algorithms," https://analyticsindiamag.com/7-types-classification-algorithms/ (accessed Jun. 02, 2023).

[43] Çinarer, G. (2021). "Görüntü İşleme Teknikleriyle Beyin Tümörlerinin Tespiti ve Sınıflandırma Algoritmalarıyla Analizi" , Kırıkkale

[44] Güneş, A. G. (2023). "Tele-Bankacılık için Potansiyel Müşteri Tahmininde Sınıfladırma Algoritmalarının Analizi,"

[45] Keskin, A. K. (2018). "Makine Öğrenmesi Sınıflandırma Algoritmalarının İncelenmesi"

[46] Bati, F. (2020). "Makine Öğrenmesi Sınıflandırma Algoritmaları Kullanılarak Meme Kanseri Tahmini"

[47] Robles-Velasco, A., Cortés, P., Muñuzuri, J. and Onieva, L. (2020). "Prediction of pipe failures in water supply networks using logistic regression and support vector classification," *Reliab Eng Syst Saf*, vol. 196, Apr. 2020, doi: 10.1016/j.ress.2019.106754.

[48] Ağan, Y. (2023). "Patlama Kaynaklı Yer Sarsıntısı Tahmininde Uyarlamalı Bulanık Çıkarım Sistemi (ANFIS), Destek Vektör Makineleri (SVM) ve Gauss Süreç Regresyonu (GPR) Tekniklerinin Kullanımı," İstanbul, Mar. 2023.

[49] Akca, E. (2022). "Satış Tahminlemesinde Hibrit Bir Yaklaşım: PESTEL, RFM, Gradient Boosting,"

[50] Nabeel, M.; "What is gradient boosting?" https://www.educative.io/answers/what-is-gradient-boosting?utm (accessed Jun. 04, 2023).

[51] Pandya, V. J. (2017). "Comparing Handwritten Character Recognition by AdaBoostClassifier and KNeighborsClassifier," in *Proceedings - 2016 8th International Conference on Computational Intelligence and Communication Networks, CICN 2016*, Institute of Electrical and Electronics Engineers Inc., Oct. 2017, pp. 271–274. doi: 10.1109/CICN.2016.59.

[52] Torun, H. (2022). "Karışıklık Matrisi (Confusion Matrix)," Oct. 01, 2022. https://hakan.io/karisiklik-matrisi-confusion-matrix/ (accessed Jun. 08, 2023).

[53] Grandini, M., Bagli, E. and Visani, G.; "Metrics for Multi-Class Classification: an Overview," Aug. 2020,

[54] Kilincer, I. F., Ertam, F. and Sengur, A. (2022). "A comprehensive intrusion detection framework using boosting algorithms," *Computers and Electrical Engineering*, vol. 100, May 2022, doi: 10.1016/j.compeleceng.2022.107869.

[55] Lashkari, A. H., Gil, G. D., Mamun, M. S. I. and Ghorbani, A. A. (2017). "Characterization of tor traffic using time based features," in *ICISSP 2017 - Proceedings of the 3rd International Conference on Information Systems Security and Privacy*, SciTePress, 2017, pp. 253–262. doi: 10.5220/0006105602530262.

[56] Lashkari, A. H.; “CICFlowMeter/ReadMe.txt at master · ahlashkari/CICFlowMeter · GitHub.” https://github.com/ahlashkari/CICFlowMeter/blob/master/ReadMe.txt (accessed Jun. 08, 2023).

ÖZGEÇMİŞ

İmran KAÇAN

[redacted]

AKADEMİK FAALİYETLER

Bildiriler:

1) Kacan İ., Ertam F.; (2022), Adli Bilişimde Bilgi Toplamanın Önemi ve Kullanılan Araçlar, Sözlü Sunum, *Uluslararası Bilişim Kongresi (IIC2022)*, Batman

Projeler:

1) Kacan İ.; (2023), Fırat Üniversitesi Bilimsel Araştırma Projeleri (FÜBAP) TEKF.23.12 protokol numaralı Yüksek Lisans Tezi Projesi, 2023.