**KOCAELİ ÜNİVERSİTESİ**

**MÜHENDİSLİK FAKÜLTESİ**

**BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ**

**ARAŞTIRMA PROJESİ**

**MAKİNE ÖĞRENİMİ ALGORİTMALARI**

**KULLANARAK AĞA İZİNSİZ GİRİŞ TESPİTİ**

**MUSTAFA KARA**

**5,2 cm**

**MEHMET FATİH ALICI**

İÇİNDEKİLER

[ÖZET iii](#_Toc123598067)

[Abstract iii](#_Toc123598068)

[1. GİRİŞ 1](#_Toc123598069)

[2. KULLANILAN YAYGIN YÖNTEMLER 2](#_Toc123598070)

[2.1. Güvenlik Duvarı (Firewall) 2](#_Toc123598071)

[2.2. Güvenlik Yazılımları (Antivirus Yazılımları) 2](#_Toc123598072)

[2.3. Parola Güçlendirme 2](#_Toc123598073)

[2.4. İzin Verilen Cihazlar Listesi 2](#_Toc123598074)

[2.5. Güncel Sistemler 3](#_Toc123598075)

[2.6. İzleme ve Anlık Tespit Sistemleri 3](#_Toc123598076)

[2.7. Makine Öğrenimi (Machine Learning) 3](#_Toc123598077)

[3. VERİ KÜMESİ 4](#_Toc123598078)

[3.1. İçerdiği Bilgiler 4](#_Toc123598079)

[3.2. Literatürde Yaygın Diğer Veri Setleri 5](#_Toc123598080)

[4. ARAŞTIRMA PLANLAMASI 7](#_Toc123598081)

[4.1. Makine Öğrenmesinin Yeri 7](#_Toc123598082)

[4.2. Sınıflandırma Algoritmaları 7](#_Toc123598083)

[4.3. Algoritmanın Optimizasyonu ve Doğruluk Değeri 8](#_Toc123598084)

[5. YÖNTEM 10](#_Toc123598085)

[KAYNAKLAR 13](#_Toc123598086)

# ÖZET

Bilgisayar ağlarına yapılan izinsiz girişler ve tehditler, günümüzde oldukça yaygın olarak karşılaşılan güvenlik sorunlarıdır. Bu tür izinsiz girişler, bir kişinin veya bir grup kişinin, bir ağa giriş yapması veya ağdaki bilgiye ulaşması için izin almadan ağa giriş yapmasına denir. Bu tür izinsiz girişler, ağda bulunan bilgilere erişim, ağı kullanıma uygun olmayan yöntemlerle kullanım, ağda bulunan bilgileri değiştirme veya silme gibi faaliyetleri de içerebilir.

Tehditler, bilgisayar ağlarını ve bilgisayarları hedef alan, güvenlik açıklarını kullanarak ağlara ve bilgisayarlara zarar veren saldırılardır. Tehditler, birçok farklı şekilde ortaya çıkabilir ve farklı amaçlar için kullanılabilir. Örneğin, bir hacker, bir ağa izinsiz giriş yaparak, ağdaki bilgilere erişmek isteyebilir. Diğer bir tehdit türü olarak, bir malware (kötü amaçlı yazılım) ağdaki bilgisayarlara bulaşarak, bu bilgisayarları kontrol etmeye çalışabilir. Bu saldırıların arkasında maddi zarar veya elde edilmeye çalışılacak bir kazanç sebebiyle de yapılabilir.

Günümüzde, bilgisayar ağlarına yapılan izinsiz girişler ve tehditlerle mücadele etmek için birçok farklı yöntem ve araç kullanılmaktadır. Örneğin, ağ güvenliğini artırmak için güvenlik duvarı (firewall) ve güvenlik yazılımları (antivirus yazılımları) kullanılabilir. Ayrıca, kullanıcıların ağa giriş yaparken kullandıkları parolaların güçlü olması, ağa izinsiz girişlerin önüne geçebilir. Ayrıca, ağda bulunan bilgisayarların ve diğer cihazların güncel olması, tehditlerden korunmada önemli bir rol oynar.

Bu projede amaç daha spesifik olarak bilgisayar ağlarına izinsiz giriş veya denemeleri yapan cihazları tespit yöntemlerinden biri olan makine öğrenmesi yolu ile veri kümeleri paylaşan sitelerden veya kuruluşlardan bu bilgileri kullanarak analizini sağlayıp sonuca ulaşmak hedeflenmektedir.

# Abstract

Illegal access to computer networks and threats are common security problems encountered today. Such illegal access refers to a person or group of people accessing

or attempting to access a network without permission. This can include accessing information on the network, using the network in an unauthorized manner, altering or deleting information on the network.

Threats are attacks that target computer networks and computers, using vulnerabilities in security to cause harm to networks and computers. Threats can take many forms and can be used for various purposes. For example, a hacker may attempt to illegally access a network to access information on it. Another type of threat is malware, which can infect computers on a network and attempt to control them. These attacks may be carried out for the purpose of causing financial damage or seeking a gain.

There are many methods and tools used to combat illegal access to computer networks and threats. For example, firewall and security software (antivirus software) can be used to increase network security. Strong passwords for users accessing the network can also help prevent illegal access. Keeping computers and other devices on the network up to date is also important in protecting against threats.

This project specifically aims to use machine learning to identify devices attempting illegal access or attempts on wireless networks, using data from data set sharing sites or organizations, and analyzing this information to reach a conclusion.

# GİRİŞ

Kablosuz ağlara yapılan izinsiz girişler, günümüzde yaygın olarak karşılaşılan bir güvenlik sorunudur. Özellikle, kablosuz ağların yaygın kullanılmasıyla birlikte, bu tür izinsiz girişlerin sayısı da artmıştır. Kablosuz ağlara yapılan izinsiz girişler, ağa izinsiz giriş yaparak bilgiye erişme, ağı kullanıma uygun olmayan yöntemlerle kullanma ve ağdaki bilgileri değiştirme veya silme gibi faaliyetleri içerebilir.

Kablosuz ağlara yapılan izinsiz girişler, ağların güvenliğini tehdit eder ve ağdaki bilgilere ve cihazlara zarar verebilir. Bu nedenle, kablosuz ağlara yapılan izinsiz girişlerin önlenmesi ve mücadele edilmesi küresel anlamda şirketler, devletler, özel kurum ve kuruluşlar için gözden geçirilmesi gereken bir sorundur.

# KULLANILAN YAYGIN YÖNTEMLER

Kablosuz ağlara yapılan izinsiz girişlerin tespitini yakalamak için, aşağıdaki yöntemler ve teknolojiler kullanılabilir.

## Güvenlik Duvarı (Firewall)

Güvenlik duvarı, bir ağın giriş ve çıkışını denetleyen bir araçtır ve izinsiz girişleri engellemek için kullanılır. Güvenlik duvarı, ağ üzerinden gelen ve giden tüm verilere bir filtre uygular ve belirli kurallara uygun olmayan verilere izin vermez. Bu sayede, izinsiz girişler engellenir.

## Güvenlik Yazılımları (Anti virüs Yazılımları)

Güvenlik yazılımları, bilgisayar ve ağları kötü amaçlı yazılımlardan (malware) korur. Bu yazılımlar, bilgisayar ve ağlarda bulunan dosyaları tarar ve kötü amaçlı yazılım tespit ettiklerinde, bunları temizler veya engeller. Böylece, ağ ve bilgisayarların güvenliği sağlanır ve izinsiz girişler engellenir.

## Parola Güçlendirme

Kullanıcıların ağa giriş yaptıklarında kullandıkları parolaların güçlü olması, izinsiz girişleri önleyebilir. Örneğin, parolaların uzunluğunun artırılması ve parolalarda rakam, sembol ve büyük/küçük harflerin kullanılması, parolaların güçlenmesine yardımcı olabilir.

## İzin Verilen Cihazlar Listesi

Bir ağın izinsiz girişlerini önlemek için, ağa sadece belirli cihazların girişine izin verilebilir. Bu şekilde, ağa sadece izin verilen cihazlar tarafından erişim sağlanabilir ve diğer cihazların ağa girişine izin verilmez.

## Güncel Sistemler

Ağda bulunan bilgisayarların ve diğer cihazların güncel olması, tehditlerden korunmada önemli bir rol oynar. Örneğin, bilgisayarlarda ve diğer cihazlarda bulunan işletim sistemlerinin güncel olması ve güvenlik güncellemelerinin yüklü olması, ağ ve bilgisayarların güvenliğini artırır ve izinsiz girişleri engeller. Ayrıca, ağda bulunan cihazların sistem yazılımlarının (örneğin, tarayıcı yazılımları, sürücü yazılımları) güncel olması da aynı şekilde önemlidir.

## İzleme ve Anlık Tespit Sistemleri

Ağ üzerinde anlık olarak izleme ve tespit sistemleri kullanılabilir. Bu sistemler, ağ üzerinde belirli bir davranışın tespit edildiğinde, anlık olarak uyarı mesajları göndererek izinsiz girişleri önler.

## Makine Öğrenimi (Machine Learning)

Makine öğrenimi, veri setlerini kullanarak, sistemlerin kendisini öğrenmesine ve bu öğrenmeleri kullanarak, izinsiz girişleri tespit etmesine yardımcı olabilir. Bu teknolojinin kullanımı, izinsiz girişlerin tespit edilmesinde önemli bir rol oynayabilir.

# VERİ KÜMESİ

UNSW-NB15, ağ güvenliği konusunda yapılan bir araştırmada kullanılan bir data setidir. Bu data set, New South Wales Üniversitesi'nden (UNSW) geliştirilmiştir ve ağ girişlerinin tespiti gibi güvenlik sorunlarının çözümünde kullanılmak üzere tasarlandı.

UNSW-NB15 data seti, bir ağa yapılan 9 cinsiyet ayrımı olmayan, 49 başarılı ve başarısız ağ girişini içeren 2,5 milyon kayıt içerir. Bu kayıtlar, ağ girişlerini tespit etmek için kullanılabilecek özellikleri (örneğin, paketlerin zaman damgası, protokol türü, paket boyutu gibi) içerir. UNSW-NB15 data seti, ağ girişlerinin tespiti için kullanılan makine öğrenimi modellerinin eğitimi ve değerlendirilmesinde yaygın olarak kullanılmaktadır [1].

UNSW-NB15 data seti hakkında daha fazla bilgi için aşağıdaki referansı inceleyebilirsiniz:

## İçerdiği Bilgiler

UNSW-NB15 veri kümesi, Avustralya Güvenlik Merkezi (ACCS) tarafından 2015 yılında yaptırılmış bir güvenlik veri kümesidir. Bu veri kümesi, bir üniversite ağında toplanmış olan veri trafiği verilerin toplanmasıyla elde edilmiştir. Veri kümesi, 2015 yılının dokuz haftası süresince toplanmış, toplamda yaklaşık 2,5 milyon kayıt içermektedir. Her bir kayıt, bir bağlantıyı temsil etmektedir. Bu veri setinde dokuz farklı tipte saldırı türü mevcuttur. Bu saldırıların adı sırayla; Fuzzers, Analysis, Backdoors, DoS, Exploits, Generic, Reconnaissance, Shellcode ve Worms olarak geçmektedir [2].

UNSW-NB15 veri kümesinde aşağıdaki özellikler bulunur:

1. ID: Birincil anahtar, her veri kaydı için benzersiz bir sayıdır.
2. Durum: Normal (N) veya saldırı (A) olarak etiketlenmiştir.
3. Durum Zamanı: Birincil anahtar olarak kullanılır ve veri kaydının zamanını belirtir.
4. Saldırı Türü: Saldırı türünü tanımlar (eğer durum 'A' ise).
5. Protokol Türü: İletişimin kullandığı protokol türünü tanımlar.
6. Sorgu Sırası: İletişimin kullandığı sırasını tanımlar.
7. Kaynak IP Adresi: İletişimin kaynak IP adresini tanımlar.
8. Kaynak Port Numarası: İletişimin kaynak port numarasını tanımlar.
9. Hedef IP Adresi: İletişimin hedef IP adresini tanımlar.
10. Hedef Port Numarası: İletişimin hedef port numarasını tanımlar.
11. Paket Boyutu: İletişimin paket boyutunu tanımlar.
12. İşlem Adı: İletişimin yapılan işlemin adını tanımlar.
13. Bayt Sayısı: İletişimin toplam bayt sayısını tanımlar.
14. Süre: İletişimin süresini tanımlar.
15. Durum Değiştirme Oranı: İletişimin durum değiştirme oranını tanımlar.
16. Saldırı İşareti: Saldırı işaretini tanımlar (eğer durum 'A' ise).

Bu özellikler, veri kümesi içindeki her veri kaydı için geçerlidir [3].

## Literatürde Yaygın Diğer Veri Setleri

CIC-IDS-2017 CICIDS, 2017 yılında New Brunswick Üniversitesi'ndeki Kanadalı Siber Güvenlik Enstitüsü tarafından oluşturulan izinsiz giriş tespit değerlendirme veri kümesidir. Bu veri kümesi, gerçek ağ trafiğini andıran hedefiyle, en yeni ve son zamanlarda oluşturulan izinsiz giriş tespit veri kümelerinden biridir. Veri yakalama süreci hem normal hem de saldırı trafiğini içeren beş gün (3-7 Temmuz 2017) sürmüştür [4].

CICIDS 2017 veri kümesinin oluşturulmasından önce, CIC ekibi kamuya açık ve en popüler izinsiz giriş tespit veri kümelerini değerlendirmiş olup, veri kümesi oluştururken izlenmesi gereken on kriteri belirlemişlerdir. Önceki veri kümelerinin hiçbiri on kriterin hepsini desteklememektedir. Bu kriterler CICIDS veri kümesinin avantajları olarak sayılır ve aşağıdakileri içerir [4].

CICIDS 2018 veri kümesinde normal davranışların tespiti, potansiyelinden dolayı çok sayıda araştırmacının odak noktası olmuştur. Ancak, gerçek dünya uygulamalarına katılımı, sistemlerin önceden dağıtımı için ciddi miktarda test, değerlendirme ve ayarlama gerektirmesi nedeniyle zorlaşmıştır. Bu sistemleri gerçek etiketli ağ izleri üzerinde, kapsamlı ve kapsamlı bir sınırlamalar ve anormal davranış kümesi ile çalıştırmak, test etme ve değerlendirme için en idealist yöntemdir [5].

Son veri kümesi, Brute-force, Heartbleed, Botnet, DoS, DDoS, Web saldırıları ve ağın içinden infiltrasyon gibi yedi farklı saldırı senaryosunu içerir. Saldırgan altyapısı, 50 makine içerir ve kurban organizasyonu, 5 departman ve 420 makine ve 30 sunucu içerir. Veri kümesi, CICFlowMeter-V3 kullanılarak yakalanan trafikten çıkarılan 80 özellik dahil olmak üzere, her makinenin ağ trafiği ve sistem kayıtlarını içerir [5].

ISCXIDS2012 (Internet Service Provider Cybersecurity and Infrastructure Security Agency 2012), bir güvenlik veri setidir ve ağ saldırılarını tespit etmek için kullanılır. Bu veri seti, İnternet Servis Sağlayıcısı Güvenlik ve Altyapı Güvenliği Ajansı tarafından oluşturulmuştur. Seçilmiş bir üniversite ve bir internet servis sağlayıcısı arasında geçen trafikten yakalanan bağlantılardan oluşur ve bu trafikteki saldırıları tespit etmek için kullanılır. Birçok güvenlik uzmanı tarafından kullanıldığı gibi ağ saldırılarının tespiti için ise literatürde kullanmış yaygın bir veri kümesidir [6].

# ARAŞTIRMA PLANLAMASI

## Makine Öğrenmesinin Yeri

Makine öğrenimi, bilgisayar ağlarına izinsiz giriş tespiti için etkili bir yöntemdir. Özellikle, sınıflandırma algoritmaları kullanılarak izinsiz girişlerin tespiti için kullanılabilir. Örneğin, bir veri seti oluşturulur ve bu veri setinde izinsiz girişler ve izinli girişler için örnekler bulunur. Daha sonra, bu veri seti kullanılarak bir sınıflandırma modeli eğitilir. Eğitilen bu model, daha sonra yeni girişlerin izinsiz mi yoksa izinli mi olduğunu tahmin edebilir [7][8].

## Sınıflandırma Algoritmaları

Sınıflandırma algoritmaları, makine öğreniminde kullanılan bir türdür ve iki veya daha fazla sınıfı birbirinden ayırmak için kullanılır. Örneğin, bilgisayar ağlarına izinsiz giriş tespiti için, izinsiz girişler ve izinli girişler gibi iki sınıf vardır ve sınıflandırma algoritmaları bu iki sınıfı birbirinden ayırmak için kullanılır [9].

Sınıflandırma algoritmaları arasında yaygın olarak kullanılanlar şunlardır

1. Logistic Regression: Bu algoritma, iki sınıflı (ikili) sınıflandırma problemlerinde kullanılır ve bir girişin hangi sınıfa ait olduğunu tahmin eder [10].
2. Decision Tree: Bu algoritma, bir karar ağacı oluşturarak girişleri sınıflara ayırır [11].
3. K Nearest Neighbors (KNN): Bu algoritma, bir girişin sınıfını, o girişe en yakın K sayıda örnekten belirler [12].
4. Random Forest (RF): Rastgele orman algoritması, adından da anlaşılacağı gibi, ormanı birçok karar ağacıyla oluşturur. Denetimli bir sınıflandırma algoritmasıdır. Yüksek yürütme hızı nedeniyle çekici bir sınıflandırıcıdır. Pek çok karar ağacı, rastgele bir orman oluşturmak için bir araya toplanır ve her bir bileşen ağacının tahminlerinin ortalamasını alarak tahminde bulunur.
5. Genellikle tek bir karar ağacından çok daha iyi tahmin doğruluğuna sahiptir. Genel olarak, ormanda ne kadar çok ağaç olursa, orman o kadar sağlam görünür [13].
6. Support Vector Machines (SVM): Bu algoritma, sınıflandırma için bir destek vektörü oluşturarak girişleri sınıflara ayırır [14][15].
7. J48 Decision Tree: Bu algoritma, bir veri kümesindeki özellikleri kullanarak bir karar ağacı oluşturur ve bu ağdaki düğümlerde bir özellik değeri belirlenir [16].
8. Artificial neural networks (ANNs), beyin yapısı ve işlevinden ilham alan hesaplamalı modellerdir. Bu modeller "nöron" adı verilen birbirine bağlı birimlerden oluşur ve bilgi işler ve iletir.

ANNs, verideki desenleri ve ilişkileri tanımak için eğitilir ve bu sayede tahminler yapabilir veya kararlar verebilir. ANNs genellikle katmanlarda düzenlenir ve girdi katmanı veriyi alır ve çıktı katmanı ise son çıktıyı veya tahmini üretir. Girdi ve çıktı katmanları arasında bir veya daha fazla "gizli" katman olabilir, bu katmanlar veriyi işler ve bir sonraki katmana iletir [17].

## Algoritmanın Optimizasyonu ve Doğruluk Değeri

Makine öğrenmesi ile yapılan sınıflandırma işleminin test edilebilmesi için literatürde birçok yöntem uygulanmaktadır. Bu yöntemlerden en çok kullanılanı ise çapraz doğrulama (cross-validation) (Kohavi, 1995) yöntemidir. 10 katlı (10-fold) çapraz doğrulama yönteminde veri seti 10 eşit parçaya ayrılmaktadır. Ayrılan veri setinin 9 parçası eğitim verisi, kalan 1 parçası ise test verisi olarak kullanılmaktadır. Daha sonra kullanılan test verisi değiştirilerek aynı işlem 10 kez tekrarlanmaktadır. Çapraz doğrulama yöntemi veri setinin doğruluğunu test etmekte ve 10 iterasyon yaparak hesaplanan değerlerin ortalamasını almaktadır. Bulunan ortalama değer ise o algoritmanın doğruluk (accuracy) değeri olarak elde edilmektedir [18].

XGBoost, hız ve performans için tasarımı yapılmış gradient boosted decision trees (eğitilmiş karar ağaçları) uygulamasıdır. Bu, makine öğrenimi yarışmalarında popüler bir algoritmadır ve birçok bu tür yarışmayı kazanmak için kullanılmıştır.

Gradient boosting, regresyon ve sınıflandırma problemleri için bir makine öğrenimi tekniğidir ve genellikle karar ağaçları gibi zayıf tahmin modellerinin ensemble'ı (bir arada) şeklinde bir tahmin modeli üretir. Algoritma, iteratif olarak zayıf modelleri eğiterek ve onları ensemble'a ekleyerek kayıp fonksiyonunu azaltmayı hedefleyen şekilde çalışır. Bireysel modeller sırasıyla eğitilir ve her model önceki modelin hatalarını düzeltmeye çalışır.

XGBoost (eXtreme Gradient Boosting), gradient boosting'in hızlı ve daha verimli olmak üzere özel olarak tasarlandığı optimize edilmiş bir versiyondur. Öteki uygulamalardan daha hızlı ve verimli olması için paralel işleme desteği, eksik değerlerin işlenmesi ve büyük ölçekli verinin işlenmesi gibi makine öğrenimi yarışmaları için uygun olan birçok özelliği vardır [19].

AdaBoost, Adaptive Boosting olarak da adlandırılan bir Machine Learning yöntemidir. Bu yöntem, bir sınıflandırıcının (örneğin, bir karar ağacı) performansını iyileştirmeyi hedefleyen bir algoritmadır. Adaboost, verilen bir eğitim kümesine dayanarak, bir dizi sınıflandırıcı oluşturur. Bu sınıflandırıcılar, ağırlıklı olarak daha zor sınıflandırılmış örnekler üzerinde önceliklidir ve her bir sınıflandırıcının doğruluğu bir öncekinden daha fazla önem kazanır. Bu sayede, Adaboost performansını iyileştirir ve daha iyi bir tahmin yapar [20].

# YÖNTEM

UNSW-NB15 veri setini seçtik çünkü literatürde daha kapsamlı olarak kullanılmaktadır. Ayrıca CIC-IDS2017 veri setine göre daha fazla saldırı türü verisi içermektedir. Orantısal olarak beşte bir oranında test, beşte dört oranında eğitim verisi şeklinde ayrılmıştır. Kullanılacak veri setine göre hem algoritma türü hem de algoritmanın performansı değişmektedir.

Yöntemin devamında bu araştırma projesi için yapılması gereken işlem veri temizleme (data cleaning) gelmektedir. UNSW-NB15 yeteri kadar sade bir veri kümesi olmasına rağmen ileride verinin lazım olan kısımlarının ayıklanması için bu işlem öncelikli olmalıdır. Devamında veri setinde bulunan etiketlerin daha pratik bir şekilde anlaşılması için özellik mühendisliği (feature engineering) kullanılabilir. Bu sayede gereksiz sınıfların kullanılması ihtimalinden kurtarılmış olunur. Bunun sebebi ise eğer gereksiz veriler mevcudiyetini korursa, algoritmanın doğruluk oranını etkileyebilir.

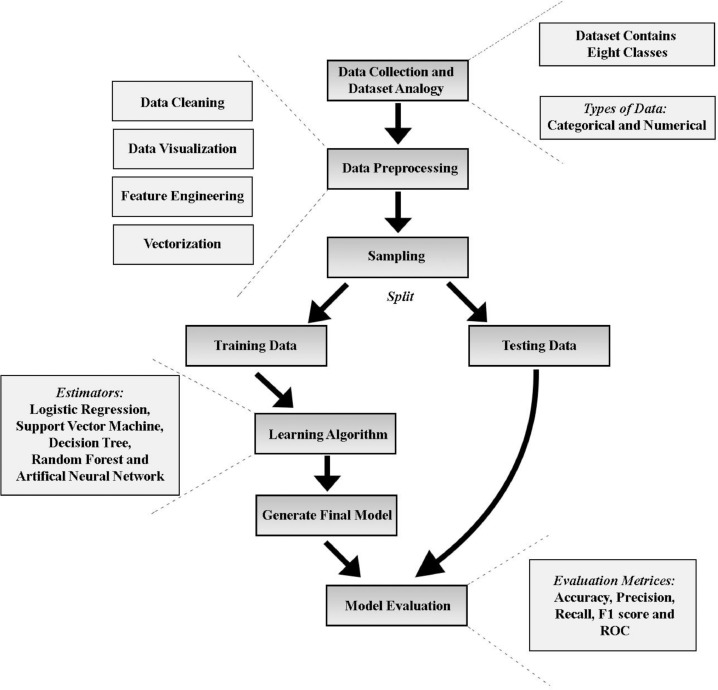
Veri ön işleme aşamasında vektörleştirme, özellikle metin verisi işleme konusunda önemli bir adımdır. Örneğin, bir makine öğrenimi modelinin tahmin etmesi gereken bir metin veri kümesi varsa, bu metinlerin kelime veya n-gramları (sözcüklerin ardışık grupları) gibi özellikleri vektörleştirilerek makine öğrenimi modeline verilebilir. Bu sayede, metin verisi sayısal bir biçime dönüştürülmüş olur ve makine öğrenimi modeli tarafından anlaşılır hale gelir.

Vektörleştirme, veri ön işleme aşamasında sıklıkla kullanılan bir tekniktir ve makine öğrenimi sistemlerinin performansını olumlu yönde etkileyebilir [21][22].

Veri örneği alınması (Sampling), veri kümesinden örnekler alınarak algoritmanın doğruluk oranı ölçülür. Örnekleme yöntemleri arasından stratejik örneklemeyi kullanacağız. Çünkü veri kümesinden seçilen örneklerin seçiminde bazı örneklere ihtiyaç olacaktır. Seçilen örnekler saldırı türleri içermelidir ki bağlantı sağlamaya çalışan cihazın anormal aktivite yapıp yapmadığı bu sayede saptanabilir.

Araştırmalarımıza göre, RF (Random Forest Algoritması) daha iyi ve kesin sonuçlar vermektedir. Fakat bu algoritmaların doğruluğu ve kesinliği veri setlerine göre farklılık göstermektedir. Seçtiğimiz veri setine göre en doğru ve kesin sonucu veren algoritmayı bulmak için ise çapraz doğrulama (cross-validation) yöntemini kullanabiliriz. Böylelikle veri setimiz için en doğru ve kesin algoritmayı bulabiliriz. Bulduğumuz algoritmanın daha iyi performans vermesi için ise Gradient Boosting Trees, XGBoost ve Adabooster yöntemlerinden en uygun olanını kullanabiliriz. Bu sayede ağa yapılan saldırılar daha hızlı saptanabilir ve bu sayede bir sonraki saldırılar daha efektif yakalanabilir ve sistemin verimliliği artar. Hatta saldırı başarıya ulaşmadan sistemin erken önlem alması sağlanabilir.

Tüm algoritmalar denendikten sonra işlemlerin sonucu grafiğe dökülüp gösterilebilir. Bu sonuçlar doğrultusunda hangi saldırıların ne kadar doğru saptanılabildiği görülebilir. Amaç ise en yüksek doğruluk oranına ulaşmaktır.



*Şekil 5.1 Ağ üzerinde saldırı ve anormallik tespiti için genel şema [24].*

# KAYNAKLAR

1. Moustafa, N., Ayad, A., ve Guizani, M. (2015). UNSW-NB15: A Comprehensive Data Set for Network Intrusion Detection Systems (UNSW-NB15 Network Data). arXiv:1608.02576 [cs.CR].
2. https://research.unsw.edu.au/projects/unsw-nb15-dataset, Erişim Tarihi: 25.12.2022
3. UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set) | IEEE Conference Publication | IEEE Xplore
4. https://www.unb.ca/cic/datasets/ids-2017.html, Erişim Tarihi: 20.12.2022.
5. https://www.unb.ca/cic/datasets/ids-2018.html, Erişim Tarihi: 20.12.2022.
6. C. D. Ziegler, K. L. Gurney, and P. L. Zador, Intrusion Detection Evaluation Dataset, Idaho National Laboratory, Idaho Falls, ID, USA, Tech. Rep. INL/EXT-12-15254, 2012.
7. A. Z. Al-Shargie, A. A. Ghalwash, and S. K. Halgamuge, "Intrusion detection in computer networks: A review," Information Sciences, vol. 429, pp. 1-18, 2018
8. S. N. Sriram, K. M. S. Santhanam, and S. S. Iyengar, "A review of feature selection methods for intrusion detection," Journal of Network and Computer Applications, vol. 36, no. 1, pp. 1-13, 2013.
9. https://www.sciencedirect.com/science/article/pii/S2542660519300241 Erişim Tarihi: 25.12.2022
10. https://machinelearningmastery.com/logistic-regression-for-machine-learning/ Erişim Tarihi: 25.12.2022
11. https://machinelearningmastery.com/decision-trees-for-machine-learning/ Erişim Tarihi: 25.12.2022
12. https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d01761 Erişim Tarihi: 25.12.2022
13. Tin Kam Ho, "Random decision forests," Proceedings of 3rd International Conference on Document Analysis and Recognition, 1995, pp. 278-282 vol.1, doi: 10.1109/ICDAR.1995.598994.
14. Cortes, C., Vapnik, V. Support-vector networks. Mach Learn 20, 273–297 (1995).
15. L. Wang, Ed., “Support Vector Machines: Theory and Applications,” Studies in Fuzziness and Soft Computing, 2005, doi: 10.1007/b95439.
16. O. K. Sahingoz, C. B. ÇEBİ, F. S. BULUT, H. FIRAT, and G. KARATAŞ, “Saldırı Tespit Sistemlerinde Makine Öğrenmesi Modellerinin Karşılaştırılması,” Erzincan Üniversitesi Fen Bilimleri Enstitüsü Dergisi. Erzincan Universitesi Fen Bilimleri Ensitusu Dergisi, Dec. 31, 2019. doi: 10.18185/erzifbed.573648.
17. E. Gelenbe and Y. Yin, “Deep Learning with Dense Random Neural Networks,” Man-Machine Interactions 5, pp. 3–18, Sep. 2017, doi: 10.1007/978-3-319-67792-7\_1.
18. https://towardsdatascience.com/cross-validation-in-machine-learning-729bf65c141 Erişim Tarihi: 25.12.2022
19. https://machinelearningmastery.com/xgboost-for-imbalanced-classification/ Erişim Tarihi: 25.12.2022
20. https://machinelearningmastery.com/boosting-and-adaboost-for-machine-learning/ Erişim Tarihi: 25.12.2022
21. https://machinelearningmastery.com/prepare-text-data-machine-learning-scikit-learn/ Erişim Ta rihi: 25.12.2022
22. https://www.geeksforgeeks.org/vectorization-in-machine-learning/ Erişim Tarihi: 25.12.2022
23. https://towardsdatascience.com/sampling-for-machine-learning-319b658c60ab Erişim Tarihi: 25.12.2022
24. Mahmudul Hasan, Md. Milon Islam, Md Ishrak Islam Zarif, M.M.A. Hashem,Attack and anomaly detection in IoT sensors in IoT sites using machine learning approaches, Internet of Things, Volume 7, 2019, 100059, ISSN 2542-6605