



T.C.  
ÜSKÜDAR ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ

FEN BİLİMLERİ ANABİLİM DALI  
SİBER GÜVENLİK YÜKSEK LİSANS PROGRAMI  
**YÜKSEK LİSANS TEZİ**

**MAKİNE ÖĞRENMESİ YAKLAŞIMI KULLANARAK  
SİBER E-DOLANDIRICILIK SALDIRILARININ TESPİTİ**

Murat DEMİRÇİ

Tez Danışmanı  
Dr.Öğr. Üyesi Nuri BİNGÖL

**İSTANBUL-2022**

T.C.  
ÜSKÜDAR ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ

FEN BİLİMLERİ ANABİLİM DALI  
SİBER GÜVENLİK YÜKSEK LİSANS PROGRAMI  
**YÜKSEK LİSANS TEZİ**

MAKİNE ÖĞRENMESİ YAKLAŞIMI KULLANARAK  
**SİBER E-DOLANDIRICILIK SALDIRILARININ TESPİTİ**

Murat DEMİRCİ

Tez Danışmanı  
Dr.Öğr. Üyesi Nuri BİNGÖL

İSTANBUL-2022

## ÖZET

### MAKİNE ÖĞRENMESİ YAKLAŞIMI KULLANARAK SİBER E-DOLANDIRICILIK SALDIRILARININ TESPİTİ

Günümüz Dünyasında yaşamın artık ip üzerinde olduğu gerçeği kaçınılmazdır. Pandemi sürecinde bu daha belirgin olarak anlaşılmıştır. Okuldan işe, işten alışverişe yaşamın her alanı internete taşınmıştır. Bu çerçeveden bakıldığından internetin daha güvenli bir alan haline getirilmesi ve saldırılardan belirlenmesi ve tespiti açısından çok önemli olduğu ortaya çıkmıştır.

Siber saldırıları arasında en fazla tercih edilen yöntem olan insan odaklı saldırının tespiti ve oluşturulacak veri setinin makine öğrenmesi ile sürekli güncel tutularak bu saldırının azaltılması ve farkındalık yaratması amaçlanmıştır.

**Anahtar Kelimeler:** E-Saldırı, E-dolandırıcılık, Siber Saldırı, Oltalama Yöntemleri, Makine Öğrenmesi.

## **ABSTRACT**

### **DETECTION OF CYBER FRAUD ATTACKS USING MACHINE LEARNING APPROACH**

In today's world, the fact that life is on a tightrope is inevitable. This has become more evident during the pandemic process. From school to work, from work to shopping, all areas of life have moved to the internet. From this perspective, it has emerged that making the internet a safer area and detection and detection of attacks are very important.

It is aimed to detect the human-oriented attack method, which is the most preferred method among cyber attacks, and to reduce this attack and raise awareness by keeping the data set to be created constantly up-to-date with machine learning.

**Keywords:** E-Attack, Phishing, Cyber Attack, Phishing Methods, Machine Learning.

## **TEŞEKKÜR**

Bu tez çalışmasının gerçekleştirilmesinde, değerli bilgilerini benimle paylaşan, kendisine ne zaman danışsam bana kıymetli zamanını ayırip içtenlikle cevap veren değerli bilgilerinden faydalandığım kıymetli danışman hocam Dr.Öğr. Üyesi Nuri BİNGÖL hocama teşekkürü bir borç biliyor ve şükranlarımı sunuyorum. Yine çalışmamda konu, kaynak ve yöntem açısından diğer ders hocalarımın kendi alanlarında öğretmiş oldukları konular için sonsuz teşekkürlerimi sunarım. Ayrıca tez sürecin de eşim Reyhan'a ve kızlarım Eslem ve Zeynep'e bana verdikleri destekten dolayı şükranlarımı sunarım. Son olarak okul yönetimi ve personeline teşekkürlerimi belirtmek istiyorum.



## **BEYAN FORMU**

Bu çalışmadaki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi, görsel, işitsel ve yazılı tüm bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu, kullandığım verilerde herhangi bir tahrifat yapmadığımı, yararlandığım kaynaklara bilimsel normlara uygun olarak atıfta bulunduğumu, tezimin kaynak gösterilen durumlar dışında özgün olduğunu, tarafimdan üretildiğini ve Üsküdar Üniversitesi Fen Bilimleri Enstitüsü Tez Yazım Kılavuzuna göre yazıldığını beyan ederim.



**Tarih**

**Murat DEMİRCİ**

**İmzası**

## İÇİNDEKİLER

<b>ÖZET .....</b>	<b>i</b>
<b>ABSTRACT.....</b>	<b>ii</b>
<b>TEŞEKKÜR .....</b>	<b>iii</b>
<b>BEYAN FORMU .....</b>	<b>iv</b>
<b>İÇİNDEKİLER .....</b>	<b>v</b>
<b>TABLOLAR DİZİNİ.....</b>	<b>vii</b>
<b>ŞEKİLLER DİZİNİ.....</b>	<b>viii</b>
<b>SİMGELER VE KISALTMALAR DİZİNİ .....</b>	<b>ix</b>
<b>1. GİRİŞ .....</b>	<b>1</b>
<b>2. GENEL BİLGİLER .....</b>	<b>4</b>
2.1. Literatür Taraması.....	4
2.2. En Fazla Kullanılan Siber Saldırı Yöntemleri .....	9
2.3. İnsan Odaklı Yapılan Saldırılar .....	10
2.4. Oltalama Saldırıları Hakkında .....	13
2.5. Oltalama Saldırısı Nedir?.....	13
2.6. Oltalama Saldırılarıyla Mücadele .....	16
<b>3. GEREÇ VE YÖNTEM .....</b>	<b>20</b>
3.1. Veri Madenciliği .....	20
3.2. Kullanılan Alanlar.....	21
3.3. Siber Güvenlik Alanında Kullanım Alanlar .....	21
3.4. Veri Madenciliği Adımları.....	24
3.4.1. Veri toplama .....	24
3.4.2. Veri temizleme ve dönüştürme .....	24
3.4.3. Model kurma.....	24
3.4.4. Model değerlendirme .....	24

3.4.5. Raporlama .....	25
3.4.6. Değerlendirme .....	25
3.4.7. Uygulama entegrasyonu .....	25
3.4.8. Model yönetimi.....	25
3.5. Veri Madenciliği Uygulamaları .....	25
3.5.1 Python ile veri madenciliği yöntemi.....	26
<b>4. BULGULAR .....</b>	<b>28</b>
4.1. Makina Öğrenmesi ve Oltalama Sayfalarının Tespiti.....	28
4.2. Özellik çıkartma.....	28
4.3. Modeller ve eğitim .....	29
<b>5. TARTIŞMA .....</b>	<b>36</b>
<b>6. SONUÇ VE ÖNERİLER.....</b>	<b>39</b>
<b>7. KAYNAKLAR .....</b>	<b>42</b>

## TABLOLAR DİZİNİ

	<u>Sayfa</u>
<b>Tablo 1:</b> Veri Seti.....	33
<b>Tablo 2:</b> Veri Ön İşleme .....	33
<b>Tablo 3:</b> Veri Ön İşleme .....	33
<b>Tablo 4:</b> Sonuç Listesi .....	34



## **ŞEKİLLER DİZİNİ**

	<u>Sayfa</u>
<b>Şekil 1:</b> İnternet Kullanımı .....	1
<b>Şekil 2:</b> Saldırı Oranı .....	2
<b>Şekil 3:</b> Sektörel Bazda Saldırı Türü.....	12
<b>Şekil 4:</b> Saldırı Türleri.....	13
<b>Şekil 5:</b> Oltalama Saldırısı.....	14
<b>Şekil 6:</b> Oltalama Saldırı Süreci .....	15
<b>Şekil 7:</b> Oltalama Saldırı Aktivitesi.....	16
<b>Şekil 8:</b> URL Yapısı.....	17
<b>Şekil 9:</b> URL Özellik Çıkartma.....	18
<b>Şekil 10:</b> Sahte Mail Örneği .....	19
<b>Şekil 11:</b> Veri Dağılımı.....	31
<b>Şekil 12:</b> Kolerasyon Haritası.....	32
<b>Şekil 13:</b> Karar Ağacı Sınıflandırıcı.....	33
<b>Şekil 14:</b> Rastgele Orman Sınıflandırıcı.....	34

## SİMGELER VE KISALTMALAR DİZİNİ

<b>DDOS</b>	: Distributed Denial Of Service Attack
<b>DOS</b>	: Denial Of Service Attack
<b>SQL</b>	: Structured Query Language
<b>MITM</b>	: Man In The Middle Attack
<b>URL</b>	: Uniform Resource Locator
<b>BTK</b>	: Bilgi Teknolojileri ve İletişim Kurumu
<b>OLPT</b>	: On Line Transaction Processing
<b>MAPE</b>	: Mean Absolute Percentage Error
<b>GNL</b>	: General Public License
<b>GNU</b>	: GNU's Not Unix
<b>AGPL</b>	: Affero General Public License
<b>HTML</b>	: Hypertext Markup Language
<b>CDDO</b>	: Cumhurbaşkanlığı Dijital Dönüşüm Ofisi
<b>TSE</b>	: Türk Standartları Enstitüsü
<b>SVM</b>	: Support Vector Machines
<b>NFC</b>	: Near Field Communication
<b>BEC</b>	: Business E-mail Compromise
<b>DNS</b>	: Domain Name Server
<b>USOM</b>	: Ulusal Siber Olaylara Müdahale Merkezi
<b>NIDS</b>	: Network Intrusion Detection System
<b>WEKA</b>	: Waikato Environment for Knowledge Analysis

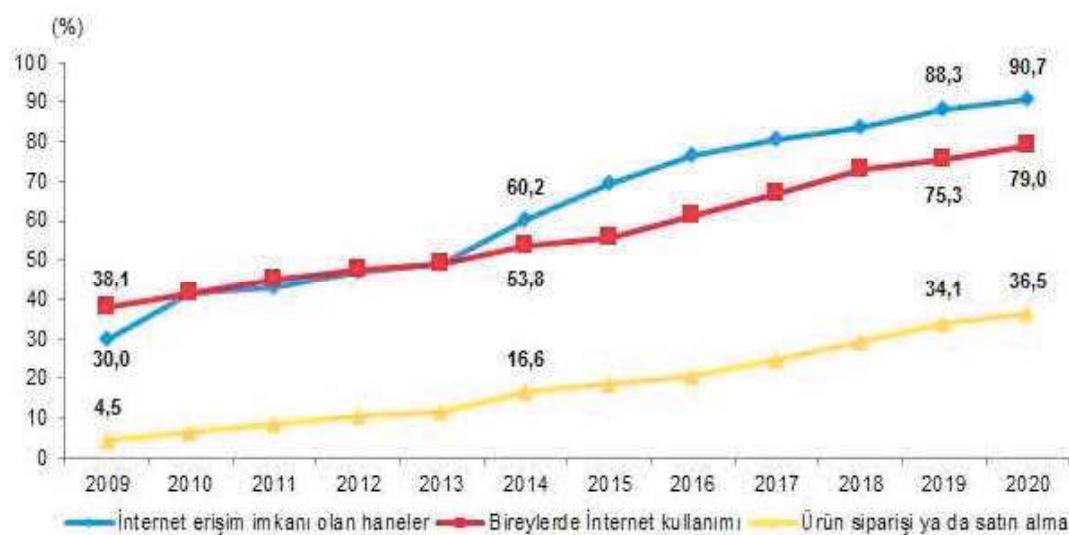
## 1. GİRİŞ

2019 yılında tüm dünyayı saran virüsle birlikte İnternetin kullanım oranı çok artmıştır. Pandemiden kaynaklı olarak insanların evlere kapanmasından dolayı alışveriş, iş hayatı, eğitim-öğretim gibi alanların hepsi internet ortamına taşınmış oldu. Bundan dolayı internet ortamında aktif olan kullanıcı sayısı da çok hızlı büyümüştür. Bunu fırsat bilen kötü niyetli saldırganlar kullanıcıların finansal bilgileri ele geçirmek, hassas kişisel bilgilere ulaşmak ve fidye yazılımları kullanarak gelir elde etmek ve sanal terörizm oluşturmayı amaçlamışlardır.

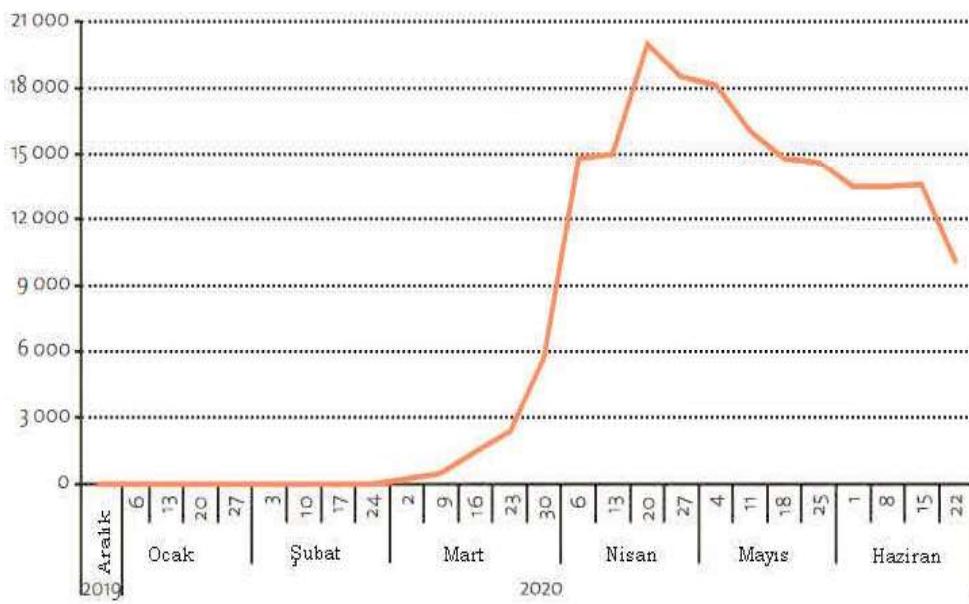
Saldırganlar kullanıcıları kandırmak ve bilgilerine erişmek için her türlü yolu kullanırlar. Bunlardan en fazla tercih edilen oltalama yöntemi ile sahte mail veya sahte internet sayfası hazırlayarak kullanıcıların zayıflıklarından faydalanan gizli bilgileri ele geçirmeyi amaçlamaktadır. Tüm bu saldırısı türlerini siber suçlar olarak nitelendirebiliriz (HEKİM ve BAŞIBÜYÜK 2013).

İnternete bağlanmak için kullandığımız tüm cihazların sanal ortama taşınması siber saldırıya maruz kalma riskini arttırmış oluyor. Şekil 1'de görüldüğü gibi son yıllarda internet kullanım oranı ve şekil 2'de saldırılarının arttığı görülmektedir.

Şekil 1: İnternet kullanımı (<https://data.tuik.gov.tr>)



Şekil 2: Saldırı oranı (<https://docs.apwg.org>)



Saldırganlar, sistemlere zarar vermek için her geçen gün daha farklı ve daha karmaşık teknikler geliştirmektedirler. Bu durumun sonucu olarak geliştirilen yeni saldırı tekniklerinin tespit edilmesi de zorlaşmaktadır. Tespit edile zayıflıkların kapatılması için yayınlanan güvenlik yamalarının da sistem yöneticileri tarafından hemen uygulanmamaktadır. Bundan dolayı, sistemlerin güncel ve etkin bir şekilde koruyabilmek ve yönetebilmek için şimdide kadar birçok yazılım ve yöntem geliştirilmiştir.

Pandemi öncesinde interneti az kullanan veya hiç kullanmayan insanların bu süreçle birlikte interneti mecbur kullanmak zorunda kaldılar. Bu durum saldırılının artmasının yanında büyük veri (Big Data) biliminin daha hızlı gelişmesine de sebep olmuştur. Daha çok ticari amaçlarla kullanılmak istenen bu verileri anlamlandırmak ve pazarlama/reklam alanlarında kullanmak için makine öğrenmesi, yapay zekâ gibi bilimlerin hızla gelişmesine neden olmuştur.

Siber suçlar arasında saldırıcıların en fazla tercih ettiği oltalama yöntemi her geçen gün daha artmakta ve farklı teknikler kullanılarak gerçekleştirilmektedir. Oltalama saldıruları hem bireyleri hem de kuruluşları tehdit etmektedir ve büyük bir problem haline gelmiştir. Günümüzde en çok karşılaşılan siber suç türü, mail üzerinden yapılan oltalama saldırısı türüdür. Saldırgan internet üzerinden kullanıcı mail bilgilerini toplayarak veya sizdirilmiş mail adreslerinin listesini satın alarak kazanç elde etmek, korkutma, bilgi isteme gibi tehdit edici unsurlara sahip, sahte bilgiler içeren mail göndererek kullanıcıların kişisel bilgilerini elde etmeyi amaçlamaktadırlar. Kullanıcı kendisine gelen

mailin sahte olduğunu fark etmez ve bu tuzağa düşerse kişisel gizli bilgilerini saldırganlara verebilir. Saldırgan bu yöntemi kullanarak, tek bir tıklamayla binlerce kişiye oltalama maili gönderebilmektedir. Bundan dolayı diğer saldırı türlerine göre daha az bir efor harcamaktadır. Çok kısa sürede çok fazla kişiye oltalama maili gönderilebilmesi durumu, çok fazla kullanıcının bu saldırısı tipiyle karşılaşması, saldırının etki yüzeyinin çok büyük olmasına neden olmaktadır. Bu nedenle bu saldırısı tipi günümüzde sıkılıkla görülmektedir.

Son dönemde sosyal medya mesajlaşma platformlarına olan ilginin çok arttığı görülmektedir. Dünya genelinde milyonlarca kullanıcıya sahip olmaları saldırganları bu alana çekmiştir. Takipçi sayısı çok olan kullanıcıların veya ünlülerin hesaplarını ele geçirmek için kullandıkları yöntemlerden birisi “*hesabınızla alakalı bir problem var*” mesajı gönderilerek linke tıklanması ve sonrasında gelen sms kodunu ve mailine gelen şartların kabul edilmesi istenmektedir.

Finansal odaklı saldırısı yöntemi olan şirket e-posta dolandırıcılığı (BEC / Business E-mail Compromise) saldırularına maruz kalınabiliyor. Bir şirket çalışanı mail yazışmalarının arasına girerek veya çalışanların mail hesaplarını ele geçirilerek ödeme hesap bilgilerin değiştğini ve yeni hesap numaralarına ödemeyi yapılması gerektiğine dair bir mail gönderip ödemelerin kendilerine yapılmasını amaçlamaktadırlar (AGAZZI 2020).

Bu saldırısı türü bazen ulaşılması gereken sistemlere erişmek için birinci basamak olarak da kullanılabilir. Bundan dolayı Devlet kurumlarının ve büyük şirketlerin siber farkındalık eğitimlerinin yanında sistemlerinin güvenlik zayıflıkları ile alakalı farklı bir göz ile kontrol edilmesi açısından siber güvenlik firmalarından yardım ve destek almaları önerilmektedir.

Cybercrime Magazin haberine göre siber saldırısı suçların, Dünya genelinde maliyetlerine bakıldığından yıldan yıla yüzde 15-20 artarak, 2015 yılında 3 trilyon ABD dolarından 2025 yılına kadar da yıllık 10,5 trilyon ABD dolarına ulaşmasını bekliyor. Siber suçların başında gelen oltalama saldıruları ile mücadele için kullanılan tekniklerin ve araçların sürekli güncellendirilmesi ve geliştirilmesi gerekmektedir.

## **2. GENEL BİLGİLER**

### **2.1. Literatür Taraması**

Bu tez çalışmasının amacı siber suçlar arasında en fazla kullanılan ve engellemesi zor olan oltalama saldırısının tespitinin yapılması adına gerçekleştirilen literatür taramasında; sahte internet sitelerinin tespiti, sahte internet sitelerinin yapısı, sahte mail adreslerinin tespiti ve URL adresinin güvenli, güvensiz ve spam listelerinin oluşturulması ve oltalama simülasyonları ile farkındalık oluşturma gibi bir çok konu ele incelenmiştir. Bu kapsamda incelenen çalışmaların çoğu mail adreslerin ve url adreslerinin makine öğrenmesi algoritmalarıyla yakalanması ve eküri değerine göre iyi sonuçlar elde etmeye çalışılmıştır. Bu çalışma da bunların yanı sıra aktif olarak web gezgini üzerinden sorgulama yapılarak web adresin sahte mi gerçek mi olduğunu sorgulama yapılarak aynı ekranda hızlıca görülebilecektir. 2010-2020 arasındaki oltalama çalışmaları kısaca özetlenmiştir.

Eryılmaz, ark. (2020) yılında yayınladıkları çalışmada iki farklı Türkçe e-posta veri kümesi üzerinde yedi farklı makine öğrenmesi algoritması kullanılarak yaramaz e-postalar tespit edilmeye çalışılmıştır. Bu algoritmaları kullanmadan önce veri kümesi üzerinde ön işlem adımları gerçekleştirilmiştir. Daha sonrasında ise öznitelik çıkarımı ve öznitelik seçimi yapılmıştır. Öznitelik seçimleri sonrasında özellik vektörü oluşturarak makinenin anlayacağı formatta değerler elde edilmiştir. Özellik vektörü makine öğrenmesi algoritmaları ile test edilerek yaramaz e-posta filtreleme işlemiyle elde edilen başarım sonuçları değerlendirilmiştir. Metin sınıflandırma çalışmalarında sıkça kullanılan filtreleme tabanlı Ki-Kare, Bilgi Kazancı ve Doküman Frekansı Eşikleme öznitelik seçme yöntemleri kullanılmaktadır. İki Türkçe e-posta veri kümesi ile bu öznitelik seçme yöntemlerinin çeşitli makine öğrenmesi sınıflandırma algoritmaları üzerinde verdiği sonuçlar incelendiğinde en başarılı sonuç Ki-Kare öznitelik seçimi ile görülmüştür.

Nazlı (2020) tez çalışmasında otomatik spam eposta filtreleme problemi üzerinde çalışılmıştır. Bazı var olan makina öğrenme algoritmaları açık bir veri seti üzerinde test edilmiş ve sonuçlar analiz edilmiştir. Geliştirilen metotlar makina öğrenme ve yazı sınıflandırma teknikleri kullanılarak geliştirildi. Değişik veri setleri ve test metotları karşılaştırıldı. Ağırıklı TF-IDF, SciKit Learn tabanlı ve Word2Vec vektörizasyonu kullanarak problem çözüm için metotlar geliştirildi. Eposta yazıları için farklı vektör gösterim metotları geliştirildi ve denetimli makina öğrenme algoritmaları ile epostalar spam veya ham olarak sınıflandırıldı. WEKA yazılım aracı kullanılarak epostaların vektör gösterimleri üzerinde makina öğrenme sınıflandırma metotları uygulandı. Sınıflandırma için Destek Vektör Mekanizması, Naive Bayes, Bayesian Ağları, J48 ve Rastgele Orman algoritmaları kullanıldı. Sınıflandırma yöntemlerinden elde ettiğimiz sonuçları karşılaştırdık ve analiz ettik. Sonuçlarımız Word2Vec vektörü ile SVM algoritmasının 300 e-posta veri kümesi için 98.33% spam algılama hassasiyeti ile en iyi performansı göstermektedir.

Awad ve Foqaha (2016) yılında yaptığı çalışmada E-postanın, günümüzün en popüler iletişim araçlarından biri olduğunu tüm dünyada bilgi paylaşımı ve alışveriş için etkin ve hızlı bir yöntem haline geldiği vurgulanmıştır. Bundan dolayı son yıllarda, kullanıcıların spam e-postaların bir sorun olduğu posta sunucularının depolanma alanını tükettiğini, kullanıcılar için zaman kaybına neden olduğunu ve ağ bant genişliğini tükettiği tespit edilmiştir. E-posta mesajlarını istenilen ve istenilmeyen posta olmayan olmak üzere iki gruba ayırmak için istenmeyen posta filtrelemeye kullanılan bir yöntem üzerinde çalışılmıştır. Makine öğrenmesi veri sınıflandırması algoritması kullanılarak istenmeyen e-posta filtreleme tekniği geliştirilmeye çalışılmıştır.

Jansson ve Rossouw (2011) yılında Nelson Mandela Üniversitesi yaptıkları çalışmada sosyal mühendislik saldırısının simülasyon yapılarak farkındalık eğitimleri ile birlikte bu saldırısının azalmasını hedeflenmiş ve Güney Afrika'da bir kurumda oltalama saldırısı tatbikatı yapılmıştır.

Tan ve ark. (2019) yılında yayınlanan çalışmada grafik teorilerindeki kavamlardan yararlanarak oltalama saldırılarını engellemek için web sayfası üzerinde köprü bağlantılarının çıkarılmasını ve ilgili lokal web sayfalarının getirilmesini içeriyyordu. Bu işlem sırasında, sayfa bağlama verileri toplandı ve web sayfasının genel köprüsünü ve ağ yapısını modelleyen bir web grafiği oluşturmak için kullanıldı. Web grafiğinden, oltalama

web sayfalarını tespit etmek için bir sınıf türetmek için grafik ölçütleri hesaplandı ve grafik özelliklerini olarak çıkarıldı. Deneysel sonuçlar, önerilen grafik özelliklerinin sınıflandırıcı olarak C4.5 kullanıldığında % 97,8 oranında iyileştirilmiş bir genel doğruluk elde ettiği ve aynı veri örneklerinden türetilen mevcut geleneksel özelliklerden daha iyi performans gösterdiği görülmüştür. Çalışmada grafik tabanlı tekniğinin, diğer oltalama yöntemleri ile kıyaslandığında daha umut verici sonuçlar elde edildiği görülmüştür. Bu nedenle, önerilen teknik, algılama performansını iyileştirmeye yönelik mevcut oltalama araştırmalarına önemli bir katkı sağladığı görülmüştür.

Akinyelu ve Adewumi (2014) yılında yayınlanan bu çalışmada, daha iyi tahmin doğruluğu ve daha az sayıda özelliğe sahip gelişmiş bir oltalama e-posta sınıflandırıcısı geliştirmek amacıyla oltalama saldırısının sınıflandırılmasında rastgele orman makine öğrenimi algoritması kullanılarak araştırılmıştır. 2000 sahte web adres ve ham e-postadan oluşan bir veri kümesinden, makine öğrenimi algoritması tarafından% 99,7'lik bir sınıflandırma doğruluğu elde edilmiştir.

Smadi ve ark. (2015) yılında yayınlanan makale, farklı e-posta bölümleriyle ilgili bir dizi özelliği çıkarılarak bir ön işleme aşamasına dayanan kimlik avı e-postalarının algılanması için akıllı bir model çalışması yapılmıştır. Çıkarılan özellikler, J48 sınıflandırma algoritması kullanılarak sınıflandırılmış. Literatürde kullanılan toplam 23 özelliği denemiştir. Eğitim, test ve doğrulama için on kat çapraz doğrulama ile birincil odak noktası olan, ön işleme aşamasına odaklanarak e-posta sınıflandırmasının genel ölçüm değerlerini geliştirilmiş ve bu alanda kullanılabilecek en iyi algoritmayı belirlemeye çalışılmıştır. Sonuç olarak, veri kümesinden özellikleri çıkarmak için ön işleme aşamasının faydaları belirlenmiştir. Model, onaylanmış bir veri kümesi için şimdije kadar kaydedilen en yüksek değer olan rastgele orman algoritması için% 98,87 doğruluk elde etmiş. On farklı sınıflandırma algoritmasının karşılaştırılması, bir dizi deney aracılığıyla bunların faydalarını ve yeteneklerini gösterilmiştir.

Baykara ve Gürel (2018) yılında yayınlanan oltalama saldırısının "Anti Phishing Simulator" adlı bir yazılım geliştirilerek, kimlik avı e-postalarının nasıl tespit edileceği hakkında bilgi verilmiş aynı zamanda bu yazılım ile mail içerikleri incelenerek phishing ve spam mailleri yakalanarak veri tabanına eklenip spam kelimelerin Bayesian algoritması ile sınıflandırılması sağlanmıştır.

Shreeram ve ark. (2010) yılında yayınlanan çalışmada Kimlik avına karşı bir kurumsal çözümün bir parçası olarak kullanılabilecek, genetik algoritma tarafından oluşturulan kural tabanlı sistemi kullanarak kimlik avı köprülerini tespit etmeye yönelik bir yaklaşım önerilmiştir. Güvenilir bir web sayfası sahibi, web ‘de şüpheli köprüler aramak için bu yaklaşımı kullanabilir. Bu yaklaşımın, kimlik avı bağlantısını meşru bağlantından ayırmak için kullanılan kuralları geliştirmek için genetik algoritma kullanılır. Bu algoritma, değerlendirme işlevi, geçiş ve mutasyon gibi parametreleri değerlendirerek, yalnızca kimlik avı bağlantılarıyla eşleşen bir kural kümesi oluşturur. Bu kural seti bir veritabanında saklanır ve kural tabanlı sistemdeki kurallardan herhangi biriyle eşleşiyorsa bir bağlantı kimlik avı bağlantısı olarak rapor edilir ve böylece sahte adreslerden korumaya çalışılmıştır.

Yearwood, ve (2010) yılında yayınlanan çalışmada kimlik avı e-postalarının analizinden kimlik avı etkinliğinin profilini çıkarmak için yeni bir yöntem önerilmiştir. Profil oluşturulurken, bir bireyin veya belirli bir kimlik avcısı grubunun etkinliğinden yararlıdır. Kimlik avı alanındaki çalışmalar genellikle kimlik avı e-postalarını tespit etmeyi amaçlar. Bu makalede, kimlik avı e-postalarının tespit edilmesinden farklı olarak profil oluşturmaya odaklanılmıştır. Profil oluşturma problemini, kimlik avı e-postalarındaki köprüleri, e-postaların özellikleri ve yapısal özellikleri ile birlikte, profil sınıfları olarak köprülerdeki (yani DNS) bilgilerinin kim olduğu gibi verilerden sınıflandırma yapılmaya çalışılmıştır. Sınıflandırıcı tahminlerine dayalı profiller oluşturarak veri kümeleri elde edilmiştir. Kimlik avı e-postalarındaki köprü bilgilerinden oluşturulan üç farklı veri kümelerinde çok etiketli sınıf tahminleri oluşturmak için bir artırma algoritması ve SVM kullanılmış tahminlerin dışında tam profillerin oluşturulması sağlanmaya çalışılmıştır.

Parekh ve ark. (2018) yılında yayınlanan bu makalenin birincil amacı, Random Forest algoritmasını kullanarak URL algılama yöntemini kullanarak kimlik avı web sitelerini tespit etmek için bir çözüm olarak bir model ortaya koymaktır. Bu modelde ayrıştırma, verilerin sezgisel sınıflandırılması, performans analizi gibi 3 ana aşama ele alınmıştır. Her aşamada daha iyi sonuçlar vermek için verilerin işlenmesi için farklı bir teknik veya algoritma kullanılmıştır.

Pais ve Rao (2018) yılında yayınlanan bu çalışmada mevcut oltalama saldıruları yöntemlerinin üstesinden gelmek için URL adresinden, kaynak kodundan ve üçüncü taraf

hizmetlerinden elde edilen sezgisel özelliklere dayanan yeni bir sınıflandırma modeli üzerinde çalışılmıştır. Model sekiz farklı makine öğrenme algoritması kullanılarak değerlendirilmiş ve bunların arasından rastgele orman sınıflandırıcısı algoritması %99,31 doğrulukla en iyi performansı göstermiştir.

Mohamed ve Visumathib (2020) yılında yayınlanan bu çalışmada saldırganların, oltalama sitesini ve mail adreslerinin daha da fazla ikna etmek için etki alanı sahtekârlığı tekniklerini kullandığını bundan dolayı e-posta hizmetlerinin politikasını ve uyarlarını ve e-posta izleme ortamını değerlendiriliyor. Oltalama web siteleri için, sahte alan adlarının tespit edilmesi ve kimlik avi sayfalarını belirlemek için dinamik ve statik analizler kullanılmıştır.

Mohammad ve ark. (2020) yılında yayınladıkları çalışmada son zamanlarda, oltalama URL'leri algılama sistemlerinin araştırıldığını ancak makine öğrenimi algoritması seçiminin olmaması nedeniyle, bu sistemlerin performansını etkilendirdiğini ortaya konulmuştur. 1. düzeydeki bireysel makine öğrenimi sınıflandırıcılarına ilişkin ayrıntılı bir inceleme ve 2. düzeyden son tahmin ve üç gerçek veri kümesi sunulmuştur. Performans, hassaslık-geri çağrıma eğrisi, AUC-ROC eğrisi, doğruluk, yanlış sınıflandırma oranı ve ortalama mutlak hata ile değerlendirilmiştir. İkili sınıflandırmada sayısal özellik kümesiyle çok sınıflı özellik kümesinde % 97,44 oranında daha yüksek doğruluk sağlanırken, performansı % 97,86 doğrulukla sağladığı görülmüştür. Yığın genelleme, çok sınıflı özellik setiyle minimum hata oranı ve % 2,142857 MAE sağlar ve bu da kimlik avi önleme araçları geliştirmenin güçlü bir temelini oluşturduğu savunulmuştur.

Shekokar ve ark. (2015) yılında yayınladığı bu yazıda, URL tabanlı ve Web sayfası benzerliğine dayalı algılamayı birleştiren bir kimlik avi algılama ve önleme yaklaşımı önerilmiştir. URL tabanlı kimlik avi tespiti, gerçek URL'nin (web sitesinin gerçekten yönlendirildiği) ve görsel URL'nin (kullanıcı tarafından görülebilen) çıkarılmasını içerir. LinkGuard Algoritması iki URL'yi analiz ederek algoritma tarafından üretilen sonuca bağlı olarak prosedür bir sonraki aşamaya geçerek URL tabanlı oltalama saldırısı algılamaya çalışılmıştır.

Abutair ve ark. (2017) yılında yayınlanan bu yazıda, durum Tabanlı Akıl Yürütmeye Kimlik Avı Algılama Sistemi araştırılmıştır. Yoğun bir şekilde eğitilmesi gereken diğer sınıflandırıcıların aksine, nispeten küçük bir veri kümesiyle yeni kimlik avi saldırısını

tespit etmek için kolayca uyarlanabildiğinden, önerilen sistem oldukça uyarlanabilir ve dinamik olduğu ortaya çıkmıştır. Sistemimizi, dengeli bir 572 kimlik avı ve yasal URL'ler üzerinde farklı senaryolar kullanarak test edilip Kimlik Avı Algılama Sisteminin doğruluğunun % 95.62'yi aştığı görülmüştür.

## 2.2. En Fazla Kullanılan Siber Saldırı Yöntemleri

Siber saldırı, bir veya birden fazla bilgisayardan karşısındaki bilgisayarlara veya ağlara yapılan veri çalmak, değiştirmek ya da yok etmek için çeşitli yöntemler kullanılarak yapılan saldırının bütününe verilen isimdir. Her geçen zaman daha büyük çaplı eylemlerle gündeme gelen siber saldırılar; kişileri, kurumları ya da devletleri hem maddi hem manevi açıdan oldukça zor durumda bırakabiliyor. Şirketlerin önemli verilerinin çalınması ve devletlerin gizli belgelerinin yayınlanması gibi dünya çapında yankı uyandıran olayların tümü siber saldırılar aracılığıyla gerçekleştiriliyor. Bu tarz saldırıları önlemek için alınan önlemler gelişikçe kötü niyetli kişiler de siber saldırı yöntemlerini çoğaltarak güçlendiriyor. Pek çok siber saldırı yöntemi bulunsa da en çok karşılaşılan saldırı türlerini iki bölümde inceleyeceğiz.

Sistemlere yapılan saldırılar, saldırganlar hedef olmadan yani ayrim gözetmeksızın olabildiğince çok sisteme veya hizmete saldırıda bulunabilirler. Sistemlerin ve hizmetlerin güvenlik zafiyetlerinden faydalananarak yaptıkları bu saldırıları yöntemlerine internet ortamında çok kolay ulaşılabilmektektir. Bu yöntemlerin en fazla tercih edilenleri aşağıda sıralanmıştır.

Malware - Kötü Amaçlı Yazılım, Dijital dünyanın en yaygın olarak kullanılan siber saldırı aracı olan bilinir. Virüs, Truva Atı ve Solucan gibi isimler bu tarz zararlı yazılımları tanımlamak için kullanılıyor. Başta bilgisayarlar olmak üzere birçok cihazlara sızdırılabilen malware çeşitleri, cihazları veya sistemlerin çalışmasını engelleyebileceği gibi kendini kopyalayarak çoğalabiliyor, gizli kalarak dinleme yapabilir ve saldırganlara erişim izni vererek uzaktan bağlanmalarına imkân verebilir.

DDoS ve DoS, açılımları Distributed Denial of Services ve Denial of Services olan DDoS ve DoS, son dönemde gelişen güvenlik önlemleri nedeniyle etki açısından azalmış olsa da hala en popüler siber saldırı yöntemleri arasında yer almaktadır. Online servislere, internet sitelerine çoklu istek göndererek çok yoğun bir trafik oluşturup çevrimiçi

servislerin veya internet sitelerinin çökmesine veya geçici olarak hizmet vermelerini engelleyebilir.

SQL Injection, veri tabanına dışarıdan komutlarda kodları veya sorguları yetkisiz bir şekilde enjekte etmek anlamında gelmektedir. Veri tabanı SQL kullanan yazılımlar ile uyumlu olacak şekilde geliştiriliyor. Kullanıcılarından veri alan web siteleri de bu verileri SQL veri tabanlarına gönderir ve kayıt eder. SQL zafiyet açıklarından yararlanan saldırgan kişiler de böylece kullanıcıların bilgilerine ulaşmış oluyorlar. Bazı durumlarda siber korsanlar, SQL kodlarını ad ve adres isteyen bir web formunda da yazabiliyor. Bu sayede kullanıcı bilgilerini çalma sürecini de çok daha hızlandırmış oluyorlar.

Man In The Middle (Ortadaki Adam), diğer bir siber saldırısı yöntemi olan MITM (Man In The Middle), kullanıcıların erişmek istedikleri hizmetler verilerin onların üzerinden geçmelerini sağlayarak pek çok bilgiyi elde edebiliyorlar. Örneğin açık alanlarda veya bu saldırısı yöntemine karşı önlem olmayan ağlarda bilmenden saldırganların ağına bağlanabilir.

Passwords Attack (Şifre Saldırısı), adından da anlaşılacağı gibi şifreleri çözerek kişisel veya kurumsal hesaplara zarar verebilen Passwords Attack, birbirinden farklı yöntemler ile gerçekleştiriliyor. Bunlardan en bilineni kaba kuvvet saldırısı (brute force) olan, zararlı yazılımlar tarafından arka arkaya farklı şifre kombinasyonları girerek mevcut/kolay şifrenin bulunmasıdır. Bu tarz bir girişimi önlemek en etkili yollarından biri de iki faktörlü kimlik doğrulama çözümü ve yöntemleri ile erişimlerin daha güvenli hale getirilmesidir.

### **2.3. İnsan Odaklı Yapılan Saldırılar**

Saldırganlar, siber saldırısı veya sanal terörizm yapacakları zaman zafiyet araştırma ve bilgi toplama aşamalarında insanların zayıflıklarından veya sistem zafiyetleri araştırırlar. Bu zafiyet ve zayıflıklardan yararlanarak bir sistemin güvenliğini çok kolay bir şekilde ele geçebilirler. Hedefli veya hedefsiz yapılan saldırılarda insan odaklı saldırılardan en fazla tercih edilen yöntemler aşağıda sıralanmıştır ( CANBEK ve SAĞIROĞLU 2007)

E-posta, çoğu saldırgan için popüler bir yöntem olmaya devam ediyor. Bir sorunu çözmenize yardımcı olmak için size ulaşan popüler bir markayı veya kurumu taklit ederek resmi görünümülü iletişim kurma, bir şifre veya diğer hesap bilgilerinizi onaylamamanızı ister. Daha karmaşık aldatıcı kimlik avi e-postaları, gönderen adresinin düzenli olarak

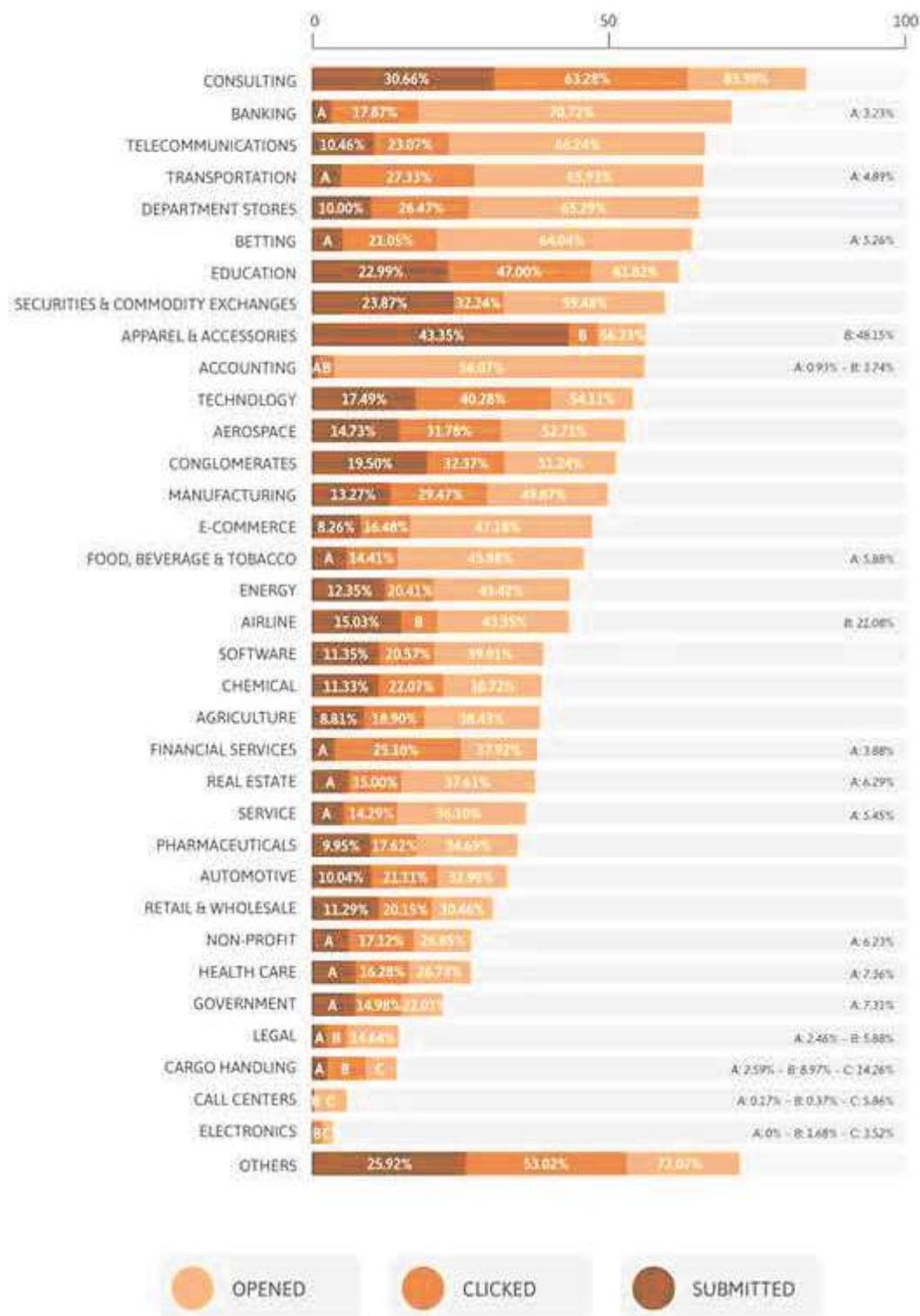
İletişim kurduğunuz kişilerin veya işletmelerin adresleriyle eşleşmesini sağlar. Cihazınıza kötü amaçlı yazılım göndermek için tasarlanmış kötü amaçlı ekler veya bağlantılar içerirler.

Telefon ve Smishing (SMS Kimlik Avı), tüm saldırılardan internet üzerinden gelmez. Birçok işletme müşterilerini bilgilendirmek için telefonla aramaktadır. Saldırganlar bu yöntemi kullanarak sanki işletme adına ariyormuş gibi bilgililerini isteyip banka hesaplarını boşaltabilmekte veya parolalarını ele geçirmektedir. Bu tarz saldırıları maalesef sıkılıkla haber kanalların da izlemektedir. Diğer taraftan saldırganlar SMS göndererek linkleri tıklamaları ve gönderdikleri dosyaları açmaları isteniyor (SOYKAN ve BAGRIYAKIN 2020).

Sosyal medya, saldırganlar Facebook veya diğer sosyal medya platformlarını kullanarak saldırıyla açık ve kolay kandırabilecekleri insanlara ulaşmak için bu platformları kullanırlar. Heyecan uyandırıcı, acil yapılması gereken bir iş varmış gibi veya bir arkadaşından geliyormuş gibi mesajlarla kurbanın hesaplarını ele geçirmeye çalışmaktadır.

İnsan odaklı saldırı türlerine oltama saldırıları olarak nitelendirilebiliriz. Son yillardaki saldırı artış oranlarına ve sektörel bazda bakıldığından bu yöntemin arttığını söyleyebiliriz.

Şekil 3: Sektörel bazda saldırı türleri (<https://www.secureworld.io>)



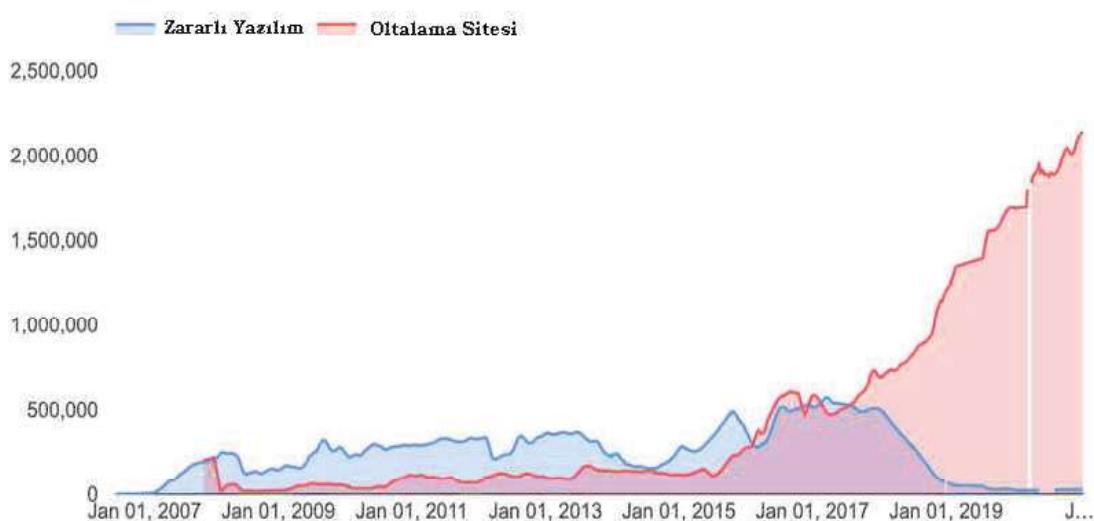
## 2.4. Oltalama Saldırıları Hakkında

2019 yılının son aylarında Covid-19 virüsünün tüm Dünyayı etkisi altına alması sonucunda gerçek Dünyada fiziksel olarak yapılan birçok sektör ve eğitim kurumları yaptıkları işleri internet ortamına yani sanal ortama taşımak zorunda kalmıştır. 2022 yılında olmamıza rağmen hala etkileri sürmektedir. Birçok alanın sanal dünyaya taşınmasından kaynaklı siber saldırılarında artmasına ve sanal terörizm olaylarına sebep olmuştur. İnsan odaklı saldırısı türü olmasından kaynaklı saldırganların en fazla tercih ettiği yöntemlerin başında gelmektedir.

Bu saldırısı yönteminin saldırganlar tarafından en fazla tercih edilmesinin sebebi şu şekilde sıralayabiliriz.

- İnsanların yoğun çalışma tempoları
- Mail ile gelen linklerin veya dosyaların zararlı olup olmadığını anlayamamaları
- Saldırı yöntemlerinin çok gelişkin ve insanların siber farkındalıklarının olmaması.
- Teknoloji okuryazarlığının az olması
- Güvenlik yazılımların bu tarz mailleri yakalamaması

**Şekil 4: Saldırı türleri (<https://www.tessian.com>)**

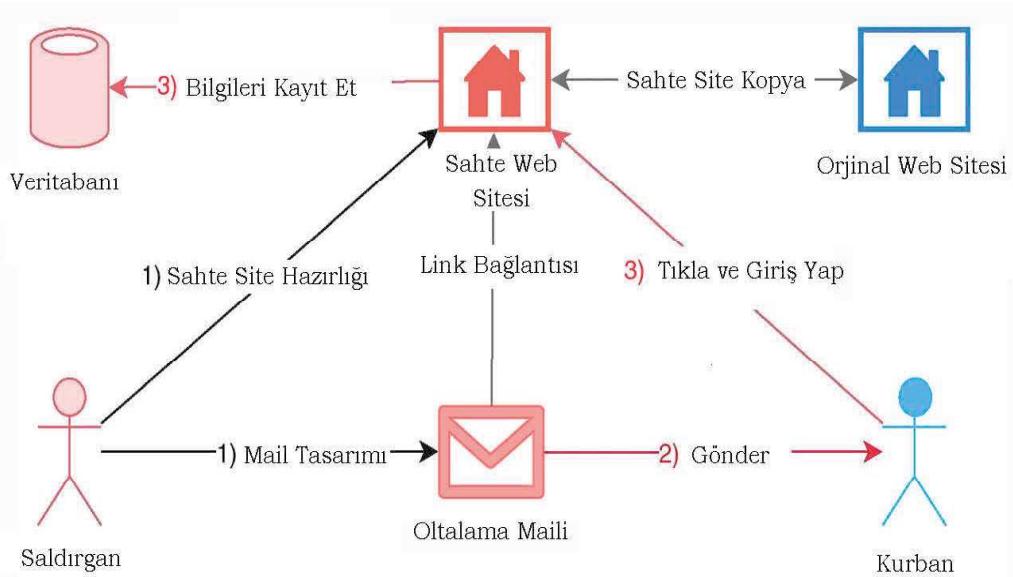


## 2.5. Oltalama Saldırısı Nedir?

Oltalama saldırısı, kurbana tanıdık bir kişi ya da kurumsal bir işletmeden geliyormuş gibi gönderilen bir e-mail ile başlıyor. E-Mail içerisinde kullanıcıyı bir link ile veya ekindeki zararlı dosya ile makinaya uzaktan bağlantı kurmaya çalışarak ele geçirmeye veya tıkladığı linkten gelen kopya sayfaya parola gibi hazzan bilgileri girmesi istenir.

Bu sayede kullanıcının hesap erişim bilgileri ya da kredi kartı gibi hassas bilgileri ele geçirmeye çalışacaktır.

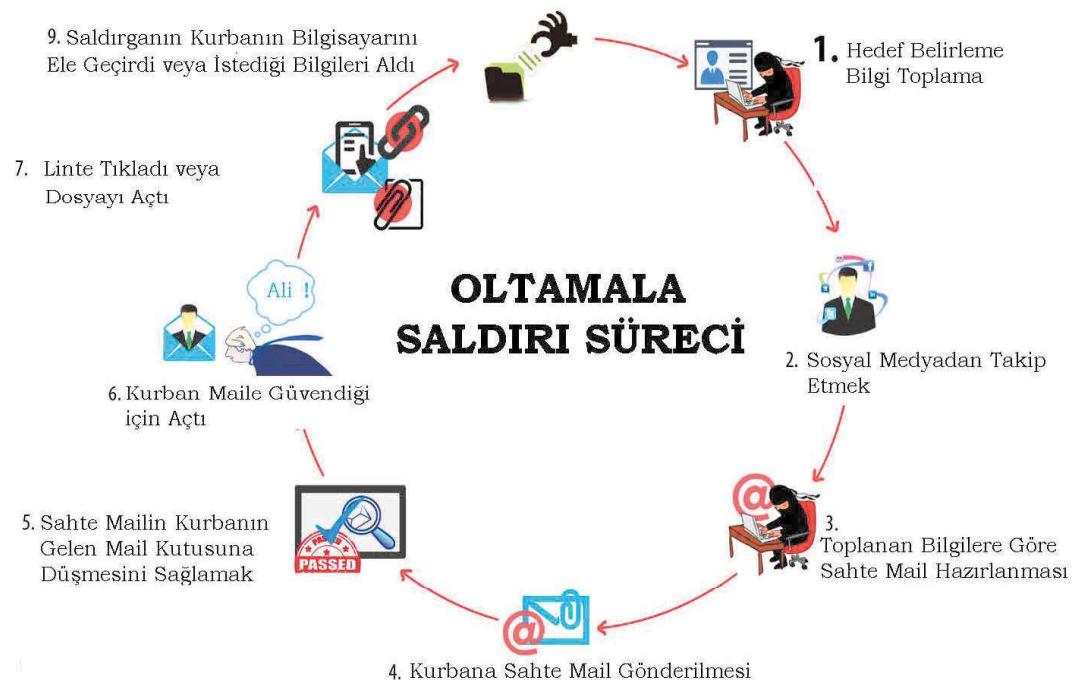
Şekil 5: Oltalama saldırısı (<https://www.researchgate.net>)



Saldırganların kullandığı bu yöntem, sosyal mühendislik temellidir. Bu teknikte e-posta yardımıyla aldatıcı linkler gönderilir ve kurbanın bu linke tıklayıp sahte web sayfasına gitmesi beklenir. Zararlı yazılım temelli saldırı tekniğinde ise kullanıcıya direkt olarak soru sorulup bilgi istenmez. Kullanıcı mail içindeki linke tıkladığında, güvenlik açığı veya zafiyetinden faydalananarak zararlı yazılım aracılığıyla kullanıcıların bilgisayarı ele geçirilmeye çalışır. Bu yöntem diğerine göre daha kapsamlıdır. Eğer zararlı yazılım çalıştırılırsa kurbanın bilgisayarı saldırganlar tarafından ele geçirilmiş hassas bilgilerin ele geçirilmesi dışında zombi bilgisayar olarak da kullanılabilir.

Şekil 6' te oltalama işleniminin genel süreci gösterilmektedir. Oltalama süreci, hedefe alınmış bireye e-mail gönderilmesi ile başlar. Bu e-mailin gönderilmesindeki maksat, mailin içerisinde bulunan linkin veya ekteki zararlı yazılımın hedefteki birey tarafından tıklanmasıdır. Bu anlamda çevrim içi oltalama faaliyetinin, geleneksel balık tutma faaliyetine benzetilmesinin sebebi de ortaya çıkmaktadır. Saldırgan e-posta içinde bulunan linki yem olarak kullanarak, kullanıcının oltanın ucunda bulunan bu yeme kanarak yakalanmasını bekler. Linke tıklanması durumunda kullanıcı sahte web sitesine yönlendirilir ve kendisinden kişisel bilgiler istenir.

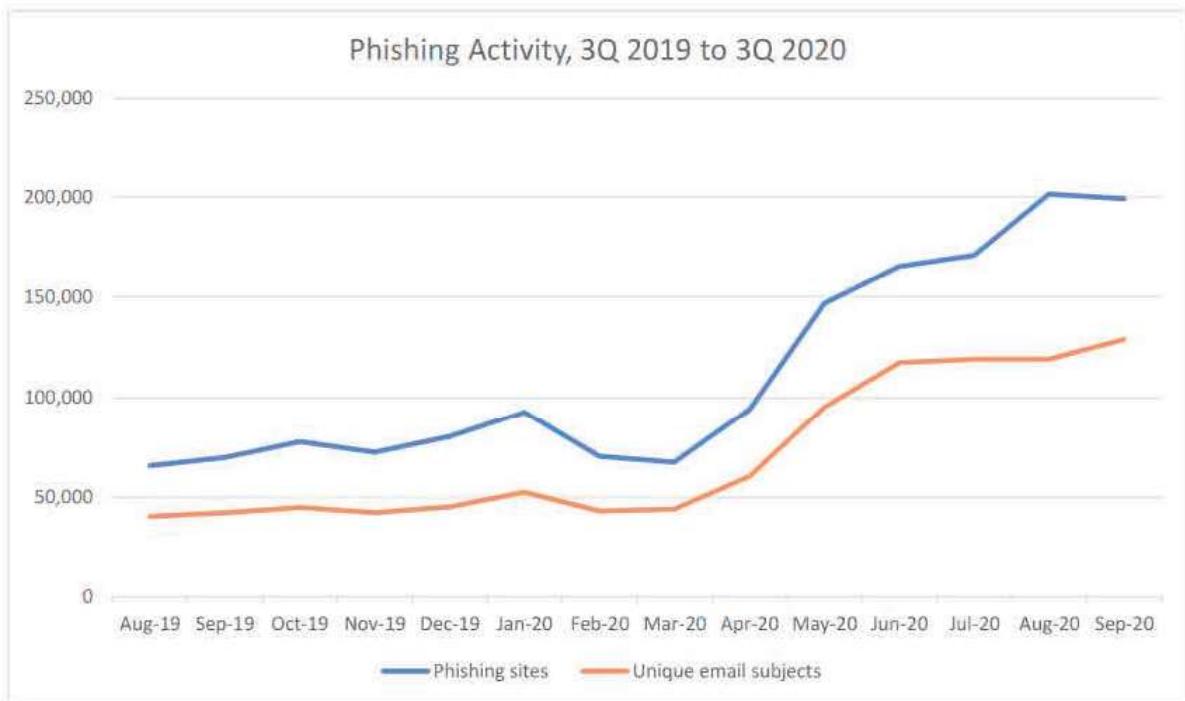
**Şekil 6: Oltalama saldırısı süreci (<https://not2phish.co.uk>)**



Oltalama saldırıları, perakendecilik, bankacılık, finans, eğitim gibi birçok alanda gerçekleştirilmekte ve kullanıcıların mahrem bilgileri elde edilip kullanıcıların büyük miktarlarda zarara uğramalarına sebebiyet vermektedir.

2019 3. çeyrek ile 2020 3. çeyrek arasında oltalama saldırısı aktivitelerine bakıldığından kendine özgü konulu maillerin gönderildiği bunun yanında oltalama sitelerinin de yükselişte olduğu görülmektedir (APWG 2020).

**Şekil 7: Oltalama saldırısı aktivitesi (<https://apwg.org>)**



## 2.6. Oltalama Saldırılarıyla Mücadele

Saldırırganların oltalama yöntemini ile daha kolay sızma yapabilmeleri ve kullanıcı üzerindeki etkilerini artması ve maddi manevi zararlar vermesi sebebiyle oldukça dikkat çekici bir çalışma alanı hâline gelmiştir. Bundan dolayı otalama saldırıları ile mücadele edip saldırıları engelleyebilmek için bazı teknikler geliştirilmiştir. Bunları bazlarını şu şekilde sıralayabiliriz.

### URL İsimlerinin Analizi

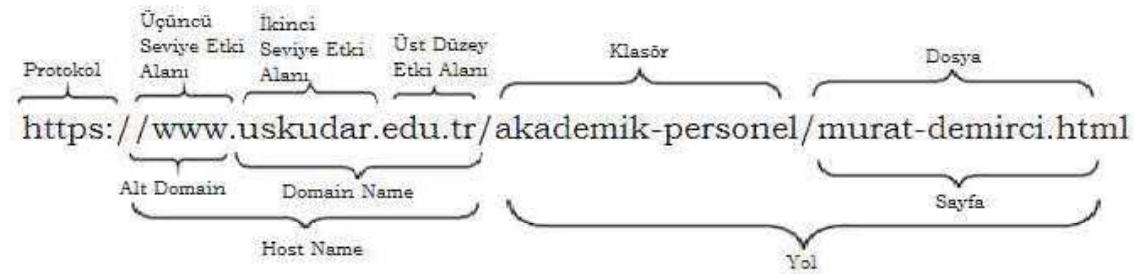
- Sahte E-Mail Adreslerinin Tespiti
- Siber Farkındalık Eğitimleri
- Devlet Politikası
- Güvenilen Liste – Güvenilmeyen Liste
- Makine Öğrenmesi

Siber farkındalık eğitimi, oltalama saldırılarına karşı kullanıcılarda farkındalık oluşturmak amaçlı düzenlenen ve kısa zaman aralıkları ile tekrarlanması gereken eğitim türüdür. Kullanıcıların, bu suçlara ve faaliyetlere karşı ve bu saldırıların yöntemlerine karşı eğitilmesi gereklidir. Kullanıcılar güvenli web sitelerindeki işaretleri nasıl izleyebileceğini öğretilmeli. Kullanıcılar, oltalama saldırısı ile alakalı temel bilgilerden yoksundurlar ve saldırılara aldanma ihtimalleri yüksektir bu yüzden de saldırılara ve

yöntemlere karşı farkındalık kazanmaları ve eğitilmeleri, oltalama ile mücadelede çok faydalıdır.

URL, isimlerinin analizi, URL (Uniform Resource Loader) internette buluna web adreslerine ulaşmak için kullanılan terimdir. İçeriginde web sitesinin protokolü, domain adı ve adres yolu barındırmaktadır.

**Şekil 8: URL yapısı (<https://seouzmanim.net>)**



Şekil 8'de görüldüğü gibi URL'nin ayırt edici özellikleri vardır. Bunlardan bazlarını aşağıda sıralanmıştır.

- URL'nin uzunluğu
- Gerçek domain isimlerinin küçük değişiklerle benzetilmesi (Typosquatted google.com → goggle.com)
- Bilinen bir markanın adının olup olmadığı (apple-icloud-login.com)
- URL'nin alt alan domain sayısı
- Top Level Domain yaygın olarak kullanılan bir uzantı olup olmadığı (.clup)

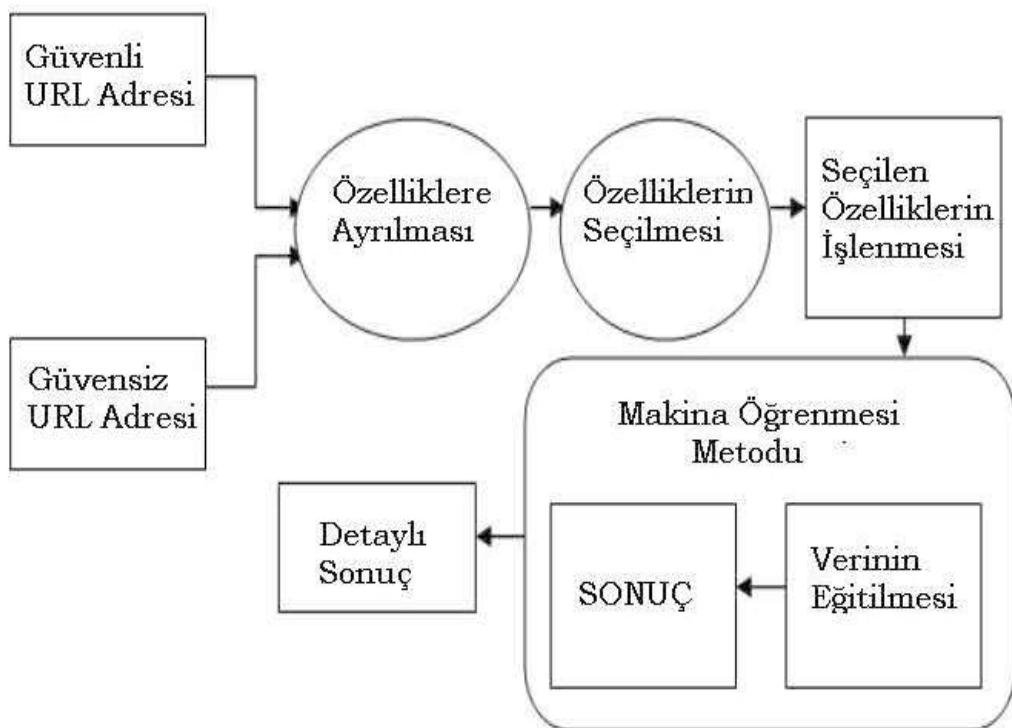
URL analizi sonucunda yukarıda belirtilen özelliklere göre engelleme yapılarak oltalama ile mücadele edilebilir.

Devlet Politikası olarak, Dünya genelinde oltalama saldırının siber suç olarak tanınması Amerika Birleşik Devleti ile başlamıştır. Ülkemizde de birçok siber suç yasal olarak kanunlaşmış ve yaptırım gücüne sahiptir. 2007 yılında çıkartılan 5651 yasası ile internet ortamında yapılan yayınların ve içeriklerinin kontrolleri sağlanarak siber suçlarla mücadele edilmeye başlanmıştır. Günümüzde BTK, Emniyet Müdürlüğü Siber Suçlar, İçişleri Bakanlığının farkındalık oluşturması için geliştirdiği siberay.org projesi ve CDDO (Cumhurbaşkanlığı Dijital Dönüşüm Ofisi) gibi kurumların hepsi siber suçlarla mücadele ve siber suçların önlenmesi için kurulmuş kurumlardır (KILINÇ 2016).

Makine Öğrenmesi, URL adreslerinden oluşan veri setleri kullanılarak otomatik olarak adresin güvenli mi güvensiz mi sonucunu oluşturacak algoritmalar ile kullanılan

uygulamalardır. Sistem sahte mail adresini veya web adresini yakalayıp kullanıcıya ulaşmadan engelleyebilmektedir.

Şekil 9: URL özellik çıkartma (<https://www.researchgate.net>)



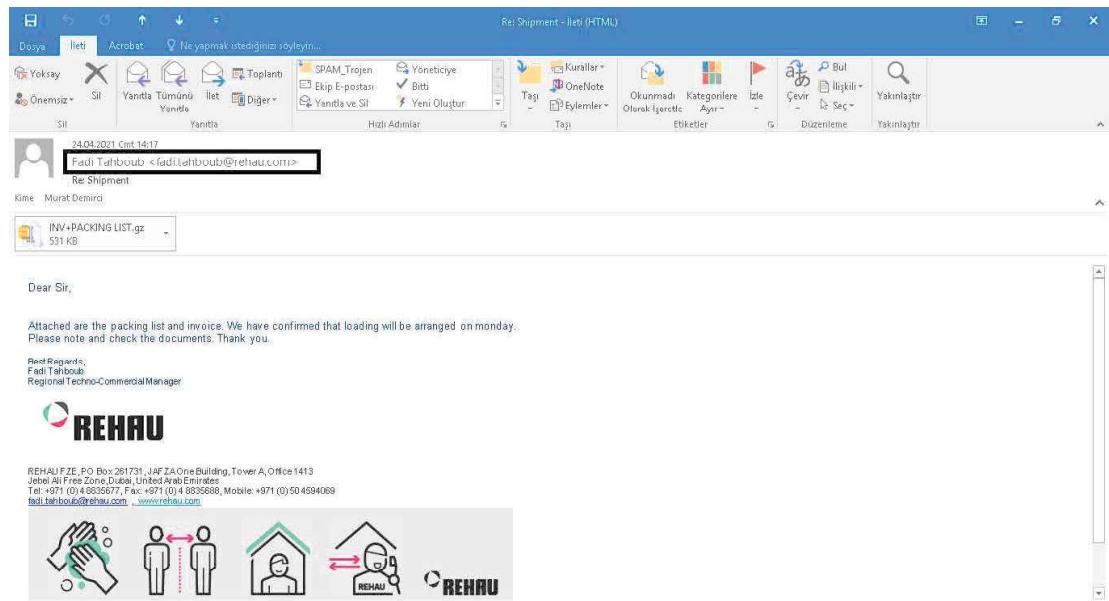
Güvenilen Liste – Güvenilmeyen Liste, bu listelerin en önemli özelliği kullanıcı tarayıcıya URL adresini yazdığında bu adresin güvenli mi güvensiz mi olduğunu sorgular ve güvensiz listede ise kullanıcıya uyarı mesajı vererek engellemeye çalışmasıdır. Güvensiz zararlı olarak görülen web sitelerine ait URL adreslerini veya mail adreslerine ait gönderici IP veya DNS adreslerine göre sınıflandırarak kayıt eder ve ağ düzeyinde koruma sağlar. Bu listeler firewall ’da veya DNS sunucularında olabilir. Bu liste ne kadar güncel tutulursa o kadar güvenlik artırılmış olacaktır. Türkiye’de USOM (Ulusal Siber Olaylara Müdahale Merkezi ) bünyesinde güncel zararlı bağlantı linkleri verilmektedir (USOM 2014).

Kullanıcıların birçoğu yeni bir mail aldığında bu mailin kimden geldiğine bakmazlar. Gelen kutusunda "BT Yönetimi" gibi bir ad ve konu satırında benzer bir yazı olması yeterlidir. Bundan dolayı direk mailin içeriğine bakılmaktadır. Saldırganlar sahte mail adreslerini oluşturduklarında, genellikle e-mail adresiyle hiçbir ilgisi olmayan sahte mail adresleri kullanarak gönderirler. Bunun yanında farklı smtp kullanarak gerçek mail

adresinden geliyormuş gibi de mail gönderebilirler. Şekil 20'de orijinal mailmiş gibi gelen sahte mail örneği gösterilmiştir.

Bundan dolayı, gelen kutunuzda görünen Google adıyla sahte bir e-posta adresi kullanabilirler. Sahte e-posta adresleri, adresin yerel kısmında sahte kuruluşun adını kullanacaklardır.

**Şekil 10: Sahte mail örneği ([www.globalsign.com](http://www.globalsign.com))**



### **3. GEREÇ VE YÖNTEM**

Bu kısımda, tez sürecinde kullanılan teknik ve bilimsel bilgilerle birlikte kullanılan yöntemler hakkında detaylı bilgi verilmiştir. Tezin temelini oluşturan veri setinden ve makine öğrenmesinde kullanılan algoritmalarla ve kullanılan programlama dilinden bahsedilmiştir.

#### **3.1. Veri Madenciliği**

Veri madenciliği bilimi Big Data olarak bilinen büyük verilerin arasından algoritmalar kullanarak, bu verilerden gizli olarak tanımlanan bilgilerin ve modellerin ortaya çıkartılması sürecidir. Veri madenciliği, teknolojinin hızlı bir şekilde büyümesi sonucu devlet işlerinde, işletmelerde ve bilim adamları tarafından birçok bilimsel alanda yaygın olarak kullanılmaya başlanmıştır. Bu bilim dalı genel olarak, yazılım kullanarak eldeki mevcut verideki bilgilerin keşif edilmesi için kullanılan algoritmalar olarak tanımlanabilir. Bu tanımlamadaki “bilginin keşfi” sözcüğü oldukça önemlidir. Veri madenciliği, veri tabanlarında elde edilen bilgileri kullanmaktadır. Veri madenciliği fonksiyonlarını yerine getirmek için geliştirilen algoritmaların kullanımı veri madenciliğinde temelini oluşturmaktadır. Bu algoritmalar, makine öğrenmesi kavramını esas almaktadır. Bu noktada, büyük veri kümelerinden sonuçların makul zamanda elde edilmesi için yüksek donanım kaynağına sahip hızlı hesaplama yapabilen sunucu gereksinimi söz konusudur. Hesaplama sonuçların kullanıcılar tarafından kolaylıkla yorumlanabilmesi ve anlaşılabilmesini sağlayan görselleştirme araçları veri madenciliği projelerinde faydalankılmaktadır (TEKEREK 2010).

Gizli bilginin ortaya çıkartılması, veri tabanlarından ve veri ambarlarından tipik olarak bilginin elde edilmesini içermektedir. Veri madenciliği, bilgi keşfi sürecinde önemli rol oynamaktadır. Bilgi keşfi süreci, veri tabanlarından bilginin elde edilmesi ile başlamaktadır. Veri tabanlarından elde edilen bilginin sağlıksız ve hatalı olan kısmının temizlenmesi ve veri tabanının farklı kaynaklarından yer alan verinin entegrasyonu bu aşamada kritik önem taşımaktadır. Veri tabanlarından veri madenciliği projeleri için kullanılacak veri ambarının oluşturulması ve veri ambarlarından ilişki verinin alınması bilgi keşfi sürecinin bir sonraki adımındır. Bilgi keşfi sürecinin diğer admında, problem

alanına bağlı olarak veri madenciliği algoritmalarının uygulanması gerçekleştirilmektedir.

Son olarak, algoritma sonucunda elde edilen verilerin değerlendirilmesi sonucunda bilgiye ulaşılmaktadır.

### **3.2. Kullanılan Alanlar**

Çalışma hayatında veri madenciliği birçok şekilde kullanılmakta ve birçok sanayii alanında da çok etkili olmaktadır. Bunların yanında sağlık alanında, hasta tanı kayıtlarından baz alınarak veri madenciliği algoritmalarından faydalılmaktadır. Araç üreticileri, daha etkin taşıma rotalarını belirlemek ve müşterilere araç sevk zamanlarını azaltmak için veri madenciliği kullanmaktadır. Bankalarda, kredi kartı müşterilerinin teşvikinde ve risk değerlendirmesinde, telekomünikasyon şirketlerinde dolandırıcılık tespitinde, üretim firmalarında kalite kontrolde veri madenciliği algoritmaları önemli fırsatlar sunmaktadır. Veri madenciliğindeki ilk uygulamalar, perakendecilikte pazar-sepet analizi alanında gerçekleştirilmiştir. Örneğin; soğuk algınlığı ilacı ile ilgilenen bir müşterinin kâğıt mendil ile de ilgileneceği tahminlersek her iki ürünü birbirlerine kolaylıkla ulaşabilecek şekilde konumlandırma gereksinimi veri madenciliği süreci ile elde edilir. İki tip veri madenciliği çalışması vardır. İlkî, hipotez testidir. Aksiyonlar ve sonuçlar arasındaki ilişki analiz edilir. Örneğin; reklamın daha fazla karlılık sağlayacağı hipotezinin doğruluğu bu yönde bir çalışma ile belirlenebilir. İkincisi ise bilgi keşfidir, bir önyargı yoktur, veri ile ilişkiler belirlenebilir (ÖZCAN 2015).

### **3.3. Siber Güvenlik Alanında Kullanım Alanları**

Pandemi ile birlikte artan internet kullanım oranı ve bununla eş zamanlı artan saldırılardan engellemesi ve önlenmesi açısından güvenlik sistemlerin oluşturduğu log kayıtlarını da büyük veri olarak değerlendirebiliriz. Son kullanıcıdan tutun büyük küçük tüm firmalar siber saldırırlara maruz kalmaktadırlar.

Ticari veriler ve ticari sırlar herhangi bir kurum ve kuruluşun sahip olduğu en değerli varlığıdır. Yenilikçi ve girişimci firmalar ise bu verilerin piyasadaki değerinin giderek arttığını ve önem kazandığının farkındalardır. Artık internette bağlı cihaz sayısı ve bunların kişiselleşmesi sonucu mobil cihazlar ve giyilebilen teknolojiler, girişimci işletmelerin kullanıcılar hakkında büyük oranda verilerin toplamasına yardımcı oluyor.

Bu durumda bu bilgilerin güvenliği ve gizliliğinin korkutacak şekele geldiği görülmektedir. Bu işletmeler, hassas iş verilerini siber saldırganlara maruz kalmamak adına savunma ve koruyucu önlemler almak ve sürekli güncel tutmak içim destek etmek zorundalar.

Ticari kuruluşların ve Devlet kurumlarının çevrimiçi yapılan işlemler sonucunda anlık artan veri boyutlarından kaynaklanan sürekli artan siber tehditlerle mücadele etmek için büyük veri analitiğini kullanmaya başladılar. Büyük veri analitiği ve makine öğrenmesi kullanımı, işletmelerde toplanan bilgilerin eksiksiz bir şekilde analizini gerçekleştirmelerini sağlar. Yapılan bu çalışmalar sonucu işletmeye gelebilecek olası siber tehditler hakkında tahmin ve ipuçlarını verebilecektir.

Bu anlamda, büyük veri hem bir tehdit hem de bir fırsat olarak görülmektedir. Nitekim veri hacminin her geçen zamanda daha da artması siber saldırganların iştahını kabartırken büyük veri analitiği, büyük ölçüde veri depolayarak, analistlerin bir ağ içinde olan usulsüzlükleri incelemelerine, gözlemlenmelerine ve tespit etmelerine yardımcı olacak veri analizi yapmayı kolaylaştırıyor. Diğer taraftan, büyük veri analizinden elde edilen güvenlikle ilgili bilgiler, bir sorunu tespit edebilmek ve çözebilmek için gereken zamanı kısaltıyor. Bu çerçevede, siber güvenlik analistlerinin izinsiz giriş ve saldırısı ihtimallerini tahmin edebilmelerine katkı sağlarken, saldırılara karşı korunmak kolaylaşıyor (AKTAN 2018).

Güvenlik ve risk yönetimi ile ilgili araştırmalar yapan Central Statistics Office araştırmasına göre kurum ve kuruluşların %84'ü gelebilecek ya da oluşabilecek siber saldırılara karşı büyük veriyi analitiği çalışmalarında kullanıyor. Ayrıca kurum ve kuruluşların büyük veri çalışmaları kullanmaya başlamalarının ardından siber güvenlik olaylarında önemli bir oranda düşüş kaydedildiği görülmüştür. Büyük veri analiz çalışmalarında gelen veriler, kötü niyetli yazılım, fidye yazılımı saldırıları ve kullanıcı hatalarından dolayı oluşabilecek siber güvenlik tehditlerini tespit edebilmek için de kullanılmaktadır. Araştırmacılara göre, büyük veri analitiğinin çalışmaları siber güvenlik alanında çözümler üretebilecek ve iyileştirmelerde bulunabileceği ön görülmektedir.

Büyük verinin analizi için kullanılan uygulamalar gerçek zamanlı olarak çalışıyor ve kritiklik seviyesine göre güvenlik alarmları oluşturabiliyor. Bu alarmlar, siber ihlallerin hızlı tespiti ve atakların azaltılması için daha ayrıntılı detaylar ile genişletilebiliyor. Veri uzmanları sizin siber güvenlik tehditlerini azaltmak için büyük veri analizinden nasıl

faydalabileceğinize dair önerilerde bulunurlar. Bu öneriler, eski verilerin analizi, iş akışlarını izleme ve otomatize saldırısı tespit sistemlerini kapsamaktadır.

Saldırganların hedefinde olan şirketler ise güçlü şifreleme algoritmaları kullanmayan, siber güvenlik önlemlerinden firewall, anti virus gibi en temel sistemleri kullanmayan bununla birlikte bulut sisteme taşınmayı düşünen kurumlardır. Diğer taraftan yapısının karmaşıklığı ve sayıları giderek artan siber saldırılarının çoğalması ve artık her türlü şirketin hedef haline gelmesi ile bilinen yöntemlerin yeterli korumayı sağlayamayan bir duruma gelebiliyor. Tam bu durumda ise büyük veri biliminin önemi devreye girerek ve siber güvenlik alanında akıllı çözümlerin geliştirilmesinin hızı artıyor (AKTAN 2018).

İnternetteki her hareketin kayıt altına alınması ile ağ trafigi, günlük olaylar, sistem kayıtları gibi düzensiz ve şüpheli hareketler izlenebilir. Çoğalan siber saldırılarının karmaşıklığı düşünüldüğünde ağ tabanlı izinsiz girişlerin tespit edilmesi gibi tehditleri tespit etmede son derece önemli olduğunu vurgulamak gereklidir.

Diğer taraftan, siber güvenlik alanı, büyük veri analizi tarafından sağlanan risk değerlendirmeden geçirilebilir bir yapay zekâ ile de doğrudan ilişkilendirilebilir. Bu verilerinin analiz edebilecek yöntemlere sahip olmak bile, bu uygulamaların görevlerini otomatik hale getirmemesi ve önemli verileri daha hızlı bir şekilde yetkili kişilere ulaştıramaması önemli bir zafiyet oluşturmaktadır.

Elde edilen büyük verilerin, işletmeniz için çok önemli olsa da, verilerin yetersiz şekilde işlenmiş ya da hiç işlenmemişse bu metodların, tehdit analizi için çok da bir etkisi olmayabilir. Böyle olmasına rağmen yapılan araştırmalar da, yapay zekâ ve makine öğrenimi ile desteklenen büyük veri analiz çözümlerinin, işletmelere bilgisayar korsanlığı veya siber güvenlik ihlali karşısında güvende kalmayı vadettiğini belirtiyor.

Bu yöntemler ile büyük veri analizinin gücünü kullanarak, yöneticilerin siber tehdit algılama yöntemlerini geliştirmek ve veri analizi tekniklerini öğrenerek saldırılara karşı koruma gücünü artırmasını mümkündür.

Güvenlik açıkları bazen bir altyapıdaki, güvenlik analistlerin ve uzmanlarının gözlerinin önündedir ama saldırganlar fark edilemeden sistemlerde kalmayı sürdürürler. İşletim sistemlerindeki zafiyetler, uygulamalardaki hataların, doğru yapılandırılmayan ve riskli son kullanıcı hareketleri siber güvenlik açıklarının en yaygın yerlerden bazılıdır.

### **3.4. Veri Madenciliği Adımları**

#### **3.4.1. Veri toplama**

Veri madenciliğinin ilk aşaması veri toplamadır. Veriler birçok farklı ortamda depolanmaktadır. Örneğin, Microsoft'ta tutulan veriler yüzlerce veri tabanında ve 70'in üzerinde veri ambarında saklanmaktadır( TANG ve MACKENNAN, 2005). Burada ilk adım veri tabanlarından veya veri ambarlarından yapılacak uygulama için uygun verileri çekmektir Veri toplama işlemi tamamlandıktan sonra, veriler test ve analiz veri seti olarak iki gruba ayrılır. Genellikle yapılan örneklerde elde edine verilerin %80'i analiz için %20'si ise test verisi için kullanılır (TEKEREK, 2010).

#### **3.4.2. Veri temizleme ve dönüştürme**

Verilerin dönüştürülmesinin gerekçe, elden edilen veri setindeki kaynak veriyi farklı formata veya değerlere dönüştürmek içindir. Veri setindeki mantıksal bir alan sayısal bir veriye dönüştürülebilir. Bunun sebebi de kullanılan bazı veri madenciliği hesaplamalarının Integer veri tipiyle Boolean veri tipine göre daha başarılı sonuçlar elde edilmesidir. Verilerin temizlemesinde amaç ise, veriler içinde uygunsuz olsan veya giriş yapıtlarken hatalı girilen verilerin ayrılmasıdır. Bu işlem sırasında boşalan alanlara uygun veriler eklenir. Bu işlem sonucunda çok fazla veri eksik ise bu kayıtların silinmesi gereklidir (TEKEREK, 2010).

#### **3.4.3. Model kurma**

Veri madenciliğinin en önemli adımlarından biri model kurma aşamasıdır. Modeli doğru bir şekilde kurabilmek ve doğru sonuç verebilmesi için yapılacak projenin amacını çok iyi bir şekilde biliyor olmamız gerekmektir. Her istenilen sonuç için birden fazla algoritma hesaplamaları kullanılabilir. Böylelikle elimizdeki veri setinde birçok algoritmalar kullanılarak en doğru çıktıyı veren algoritma ulaşılmaya çalışılmaktadır.

#### **3.4.4. Model değerlendirme**

Toplanan verilerin uygun algoritmalar kullanarak çalıştırıldıktan sonra en doğru sonuca hangisinin varacağını bulmak için çeşitli yöntemler vardır. Mesela, tahmine yönelik sayısal veriler varsa ve kullanılan yöntemin doğruluğunu test etmek isteniliyorsa MAPE (Mean Absolute Percentage Error) yöntemini kullanabilir.

### **3.4.5. Raporlama**

Veril madenciliği algoritmaların hesaplanması sonucunda elde edilen çıktılarının yayınlanması olarak hazırlanan sonuçlardır. Raporlama veri madenciliği sonuçlarını yaynlamak için önemli bir kanaldır.

### **3.4.6. Değerlendirme**

Veri madenciliği projelerinde, doğru sonuçlara ulaşmak için örüntüleri elde etmek çalışmanın yarısını oluşturur. Bu aşamadaki amaç, değerlendirme için modeli kullanmaktadır. Değerlendirme veri madenciliği biliminde puanlama olarak da adlandırılır. Değerlendirme yapabilmek için kullanılan model ve yeni durumları içeren veri setinin olması gereklidir. Böylece, eğitilen model kullanılarak yeni oluşan/olacak durumlar için tahminlerde bulunulabilir (TEKEREK, 2010).

### **3.4.7. Uygulama entegrasyonu**

Bu adımda hazırlanan veri madenciliği modeli gerçek zamanlı olarak çalıştmak üzere geliştirilen uygulamaya eklenir.

### **3.4.8. Model yönetimi**

Tüm veri madenciliği modellerinin bir yaşam döngüsü vardır. Bazı hesaplamalarda işler, durağandır ve modelin yeniden eğitilmesine gerek yoktur. Fakat birçok uygulamada işlerin özellikleri sık sık değişir. Yeni veriler geldikçe uygulanan modelin yeniden derlenerek eğitilmesine gerekmektedir. Özette kurulan bir modelin veri setinde çok değişiklik oluyorsa model sık sık eğitilerek güncellenmelidir.

## **3.5. Veri Madenciliği Uygulamaları**

Weka, uygulaması Yeni Zelanda Waikato Üniversitesi'nin geliştirdiği açık kaynaklı java dilinde yazılmış GPL lisansı ile kullanılmaktadır. Weka uygulamasında algoritmaları doğrudan bir veri seti kümesine uygulayabilir veya kendi geliştirdiğimiz Java kodlarınızdan çağrılabilirsiniz. Weka, veri görselleştirme, ön işleme, sınıflandırma, kümeleme, gerileme ve ilişkilendirme araçlarını içerir (TURNA 2011).

Özellikleri:

1. Kullanımı grafiksel ara yüz olduğu için kolay ve fare ile işlemler çok kolay yapılabilir.
2. Genel Kamu Lisansı (GNU ) altında lisanslandığı için ücretsiz kullanılabılır.
3. WEKA' da yüklü olmayan algoritmalar, daha sonradan yüklenebilir ve Java ile geliştirme için özel kütüphaneler ve yazım kuralları yine aynı üniversite tarafından hazırlanmaktadır.
4. Weka kütüphaneleri açık kaynaktır ve Java projelerinden direk çağrılabılır.

Rapid Miner, Alman Teknoloji Dortmund Üniversitesi'nin Yapay Zekâ bölümünde geliştirilmiş bir veri madenciliği uygulamasıdır. Java ile geliştirilmiştir. Bu uygulama üzerinde metin madenciliği, veri madenciliği, makine öğrenmesi, tahmin edici analiz ve iş analizine yönelik çalışmalar yapılmaktadır. Yazılım Rapid Miner firması tarafından satın alınarak lisanslı ücretli olan profesyonel versiyon hazırlanmış ama bunun yanında ücretsiz topluluk destekli yazılım lisansı ile de daha kısıtlı bir versiyonu kullandırmaktadır. Yazılım büyük oranda iş ve ticari projelerde kullanılmasının yanı sıra eğitim-öğretim, araştırma-geliştirme gibi farklı amaçlarda da kullanılır. Ayrıca veri madenciliği sürecinin bütün aşamalar Rapid Miner firması tarafından desteklenmektedir, bundan dolayı verilerin hazırlama, sonuçları görsel hale dönüştürülmesi, doğrulama ve sorunları gidermek gibi amaçlarla kullanılabilir (KAYA ve ÖZEL 2013).

Özellikleri;

1. Ara yüzünün görsel olmasından kaynaklı diğer alternatif yazılımlara göre avantajlar sağlar.
2. Kısıtlı versiyon olan topluluk lisansı altında gelen versiyonu, AGPL açık kaynak lisansı altında ücretsiz kullanılabılır.
3. Profesyonel versiyonu ile üst düzey teknik destek ücretli olarak sağlanmaktadır.

### **3.5.1 Python ile veri madenciliği yöntemi**

Python programlama dili son dönemlerde en fazla tercih edilen programlama dillerinden biri haline gelmiştir. Bunun en büyük sebebi, Python'da veri bilimi ile ilgili ve birçok bilimsel çalışmalarda kullanılabilecek kütüphanelerin çok olmasın

kaynaklanmaktadır. Birçok büyük firmalar başta Facebook, Twitter ve Google olmak üzere, bu dile kullanmaları piyasada bu dile gösterilen ilginin artmasına sebep olmuştur. İnternet üzerinde üretilen büyük verilerin ve bunların analizinin yapılması için geliştirilen veri bilimi bu uygulamaların büyümesine ve geliştirilmesine sebep olmuştur (ARSLAN 2020).

Veri madenciliği uygulamaları, bilgisayarların verilerle nasıl karar vereceğini öğrenmesi için bir yol gösterir. Bu karar örneğin, yarının hava durumunu tahmin etmek, bir spam e-postanın gelen kutunuza girmesini engellemek, bir web sitesinin hangi dilde geliştirildiğini tespit etmek için kullanılabilir. Verilerin sürekli büyümesi sonucu keşfedilen yeni uygulamalarla birlikte birçok farklı veri madenciliği uygulaması olmuştur. Veri madenciliği bilimi algoritmalar, kara verme modelleri, istatistik, mühendislik, optimizasyon ve bilgisayar biliminin bir parçasıdır. Elde edilen sonuçların etkili olabilmesi için bu alana özgü bilginin algoritmalarla entegre edilmesi gerekmektedir. Birçok veri madenciliği uygulamalarında ayrıntılar genellikle oldukça önemli ölçüde değişse de, aynı üst düzey görünümle çalışır. Veri madenciliği süreci gerçek dünyadan bir yönünü tanımlayan bir veri kümesi oluşturarak başlıyor.

Çok karmaşık veri analizleri günümüzde bilişim teknolojileri için en önemli konulardan biri haline gelmiştir. Python dili ise bu analizlerin yapılabilmesi için geliştirilmiş programlama dildir. Python programlama dilinde kütüphanelerin birçoğu makine öğrenimi ve veri bilimi üzerine geliştirilmiştir. Bu bilim dalındaki kütüphaneler üzerinden dağıtılan algoritma ve scriptler sürekli geliştirilerek bu alanın sürekli güncel kalması sağlanmaktadır.

## **4. BULGULAR**

### **4.1. Makina Öğrenmesi ve Oltalama Sayfalarının Tespiti**

Makine öğrenmesi, bilgisayarların direkt programlanmadan, kendilerine verilen bilgiler sayesinde insanlar gibi davranış yaparak öğrenme bilimidir. Bu öğrenme süreci insanlardan esinlenerek geliştirilmiştir. Bu öğrenme sürecinde algoritmaların faydalanyılmasından yararlanmaktadır.

Oltalama saldıruları, URL adreslerini ve web sayfalarını taklit eden yaygın bir sosyal mühendislik yöntemidir. Bu web sitelerini tahmin etmek ve URL adreslerini tespit etmek için oluşturulan veri kümesi üzerinde makine öğrenimi modellerini ve derin sinir ağlarını eğitmektir. Web sitelerinin hem kimlik avı hem de zararsız URL'leri bir veri kümesi oluşturmak için toplanır ve bunlardan gerekli URL ve web sitesi içeriğine dayalı özellikler çıkarılır. Her modelin performans düzeyi ölçülür ve karşılaştırılır.

Kimlik avı URL'leri kümesi, PhishTank adlı açık kaynak hizmetinden alınmıştır. Bu hizmet sürekli güncellenen csv, json vb. gibi birden çok biçimde bir dizi kimlik avı URL'si sağlamamaktadır. Verileri indirmek için: <https://github.com/muraddemirci/oltalamasaldiritespit> adresinde veri setlerine ulaşılabilir. Bu veri kümesinden, makine öğrenmesi modellerini eğitmek için 5000'er rastgele kimlik avı URL'si alınmıştır.

Legal URL'ler, University of New Brunswick'in açık veri kümelerinden elde edilmiş olup <https://www.unb.ca/cic/datasets/url-2016.html> adresinden alınmıştır. Bu veri kümesi, legal, spam, kimlik avı, kötü amaçlı yazılım ve zararlı URL'lerinden oluşan bir veri setine sahiptir. Tüm bu türlerin dışında, bu proje için legal URL veri kümesi düşünülmüştür. Bu veri kümesinden, makine öğrenmesi modellerini eğitmek için 5.000 rastgele legal URL toplanmıştır.

### **4.2. Özellik çıkartma**

URL verilerinin özellikleri aşağıda çıkartılmıştır. Adres çubuğu dayalı özellikler, bu kategoride 9 özellik çıkartılmıştır. Etki alanına dayalı özellikler, bu kategoride 4 özellik çıkartılmıştır. HTML ve Javascript tabanlı 4 özellik çıkartılmıştır.

### **4.3. Modeller ve eğitim**

ML model eğitiminin belirtmeden önce, veriler 8.000 eğitim örneğine ve 2.000 test örneğine bölünür. Veri kümelerinden, bunun denetimli bir makine öğrenimi görevi olduğu açıklır. Sınıflandırma ve regresyon adı verilen iki ana tür denetimli makine öğrenimi çalışılmıştır.

Bu veri seti, giriş URL'si illegal veya legal olarak sınıflandırmaya girer. Bu çalışmada veri kümelerini eğittiği düşünülen denetimli makine öğrenimi modelleri (sınıflandırma):

**Karar ağaçları:** Karar ağaçları, sınıflandırma ve regresyon görevleri için yaygın olarak kullanılan modellerdir. Esasen, bir karara yol açan if/else sorularından oluşan bir hiyerarşi öğrenirler. Bir karar ağını öğrenmek, bizi doğru cevaba en hızlı şekilde ulaştıran if/else sorularının sırasını öğrenmek demektir. Makine öğrenimi ortamında bu sorulara testler denir (modelimizin ne kadar genelleştirilebilir olduğunu görmek için test etmek için kullandığımız veriler olan test seti ile karıştırılmamalıdır). Bir ağaç oluşturmak için, algoritma tüm olası testleri araştırır ve hedef değişken hakkında en bilgilendirici olanı bulur.

**Rastgele Orman:** Rastgele ormanlar, regresyon ve sınıflandırma için şu anda en yaygın kullanılan makine öğrenimi yöntemleri arasındadır. Rastgele orman, esasen, her ağaçın diğerlerinden biraz farklı olduğu bir karar ağaçları topluluğudur. Rastgele ormanların ardındaki fikir, her ağaçın nispeten iyi bir tahminde bulunabileceği, ancak verilerin bir kısmına büyük olasılıkla fazla sığacağıdır. Hepsi iyi çalışan ve farklı şekillerde fazla takılan çok sayıda ağaç inşa edersek, sonuçlarının ortalamasını alarak fazla yerleştirme miktarını azaltabiliriz. Rastgele bir orman modeli oluşturmak için, oluşturulacak ağaç sayısına karar vermeniz gereklidir (Random Forest Regressor veya Random Forest Classifier'in `n_estimators` parametresi). Çok güçlündürler, genellikle parametrelerin yoğun bir şekilde ayarlanması olmadan iyi çalışırlar ve verilerin ölçeklendirilmesini gerektirmezler.

**Çok Katmanlı Algılayıcılar:** Çok katmanlı algılayıcılar (MLP'ler), (vanilya) ileri beslemeli sinir ağları veya bazen sadece sinir ağları olarak da bilinir. Çok katmanlı algılayıcılar hem sınıflandırma hem de regresyon problemleri için uygulanabilir.

MLP'ler, bir karara varmak için birden çok işlem aşamasını gerçekleştiren doğrusal modellerin genellemeleri olarak görülebilir.

XGBoost: Son dönemin en popüler makine öğrenme algoritmalarından biridir. XGBoost, eXtreme Gradient Boosting anlamına gelir. Eldeki tahmin görevinin türü ne olursa olsun; regresyon veya sınıflandırma. XGBoost, hız ve performans için tasarlanmış gradyan destekli karar ağaçlarının bir uygulamasıdır.

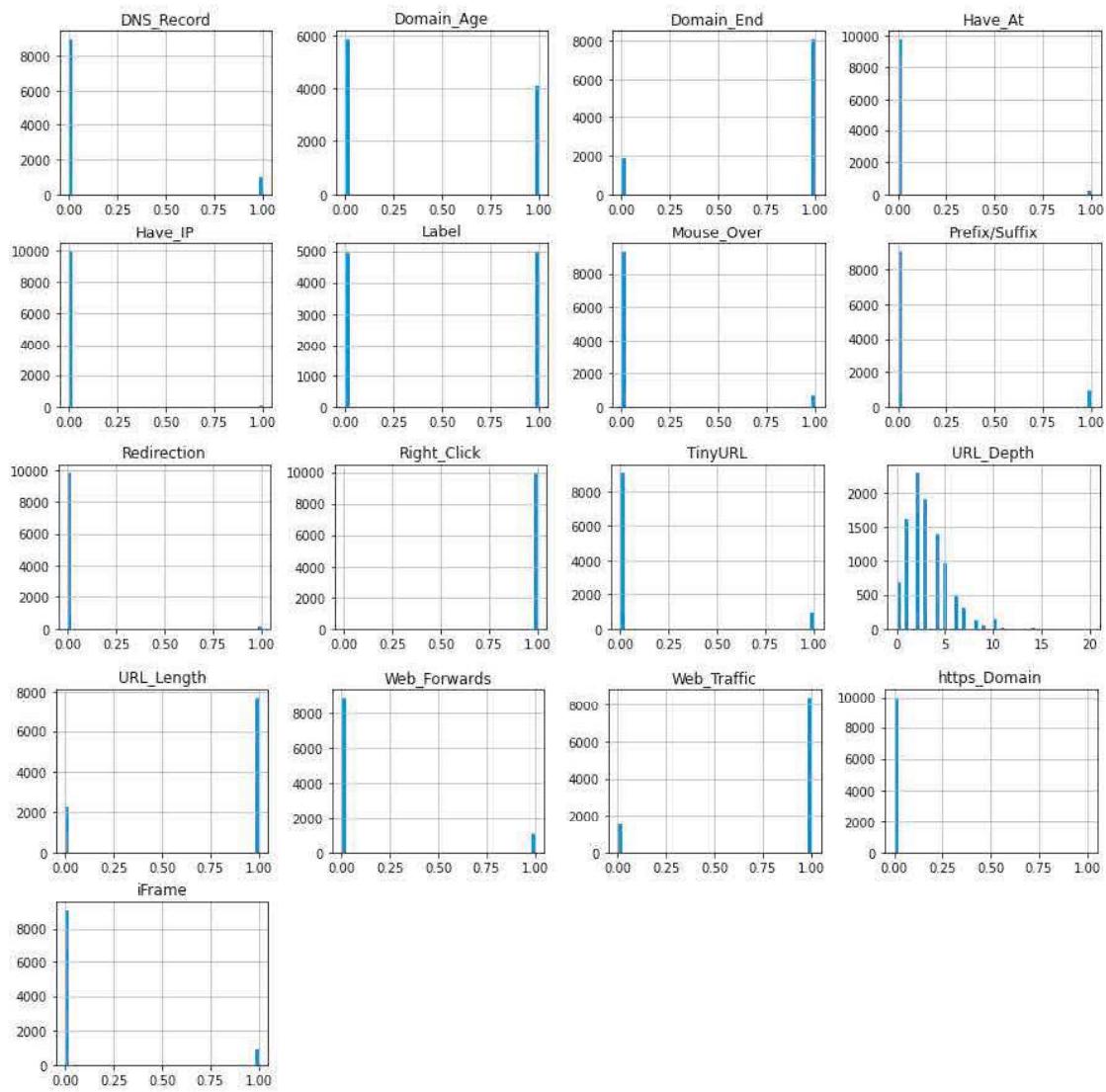
Otomatik Kodlayıcı Sinir Ağı: Otomatik kodlayıcı, çıktılarla aynı sayıda giriş nöronuna sahip bir sinir ağıdır. Sinir ağının gizli katmanları, giriş/çıkış nöronlarından daha az nörona sahip olacaktır. Daha az nöron olduğu için, otomatik kodlayıcı, girişi daha az gizli nörona kodlamayı öğrenmelidir. Tahminciler ( $x$ ) ve çıktı ( $y$ ) bir otomatik kodlayıcıda tamamen aynıdır.

Vektör Makineleri Desteklemek: Makine öğreniminde, destek vektör makineleri (SVM'ler, ayrıca destek vektör ağları), sınıflandırma ve regresyon analizi için kullanılan verileri analiz eden ilişkili öğrenme algoritmaları ile denetimli öğrenme modelleridir. Her biri iki kategoriden birine veya diğerine ait olarak işaretlenmiş bir dizi eğitim örneği verildiğinde, bir SVM eğitim algoritması, bir kategoriye veya diğerine yeni örnekler atayan ve onu olasılıksal olmayan bir ikili doğrusal sınıflandırıcı yapan bir model oluştururlar.

**Tablo 1:** Veri seti örneği

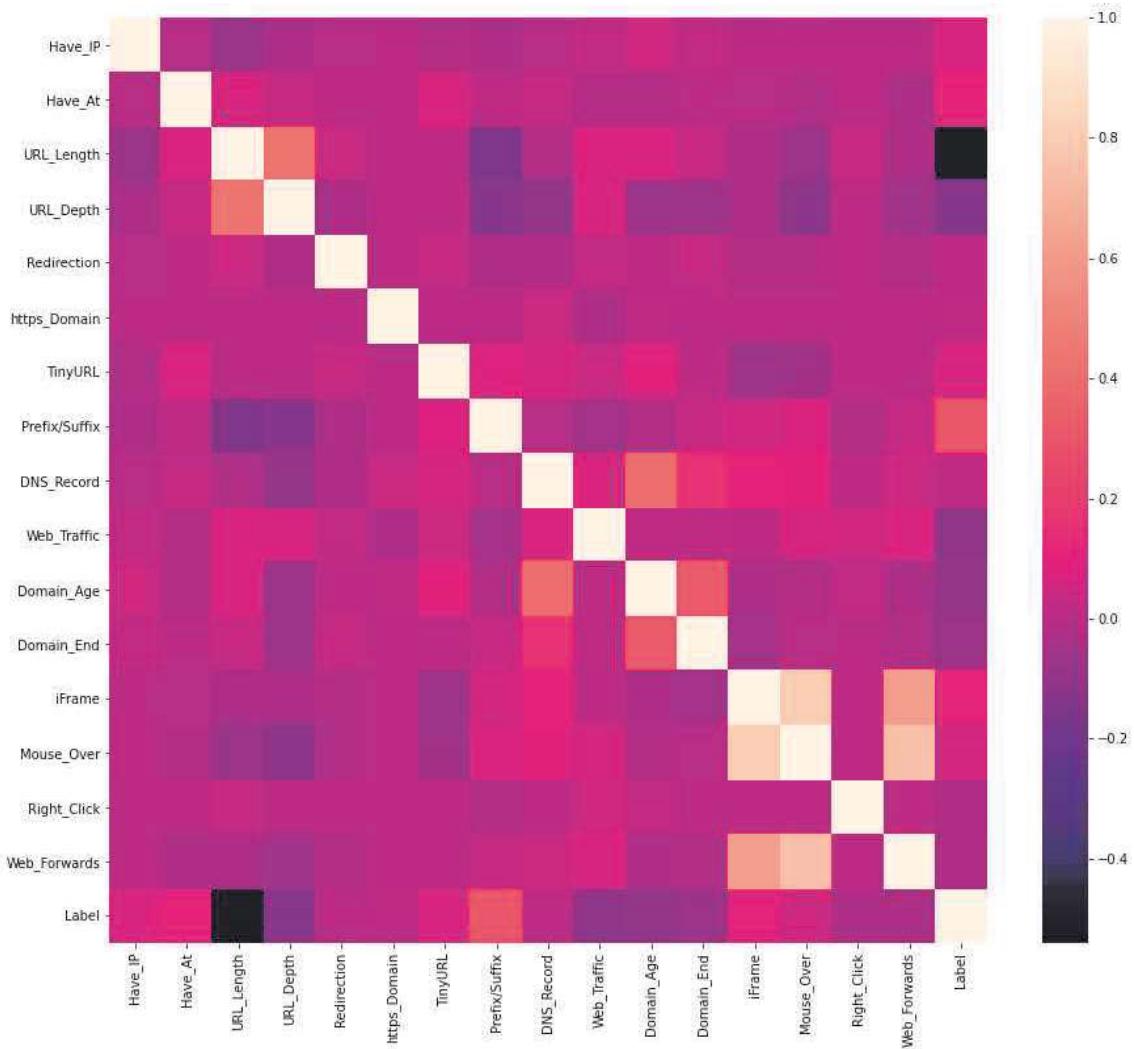
DNS	Have_IP	Have_Art	URL_Length	URL_Depth	Redirection	https_Domain	TinyURL	Prefix_Suffix	DNS_Record	Web_Traffic	Domain_Age	Domain_End	iFrame	Mouse_Over	Right_Click	Web_Forward	Label	
0	graphicriver.net	0	0	1	1	0	0	0	0	0	0	1	1	1	0	0	1	0
1	ecnavi.jp	0	0	1	1	1	0	0	0	0	0	1	1	1	0	0	1	0
2	hubpages.com	0	0	1	1	0	0	0	0	0	0	1	0	1	0	0	1	0
3	extratorrent.cc	0	0	1	3	0	0	0	0	0	0	1	0	1	0	0	1	0
4	icicibanlk.com	0	0	1	3	0	0	0	0	0	0	1	0	1	0	0	1	0

**Şekil 11: Veri dağılımı**



Verilerin grafiksel olarak dağılımına baktığımızda özelliklerin birbiri ile nasıl ilişkide olduğunu görmekteyiz.

Şekil 12: Korelasyon haritası



Veri setindeki değişkenler (öznitelikler) arasındaki korelasyon katsayısı değerlerini gösterirken en çok kullanılan yöntemlerden birisi de ısı haritası ile görüntüleme yapmaktadır. Heatmap() fonksiyonu yukarıdaki gibi kullanılarak değişkenler arasındaki korelasyon değerleri ısı haritası oluşturulmuştur. İşı haritasında negatif yönlü ilişkilerin şiddeti artıkça rengi koyulacaktır. Pozitif yönlü olduğunda renk açılmasına başlayacaktır.

**Tablo 2: Veri ön işleme**

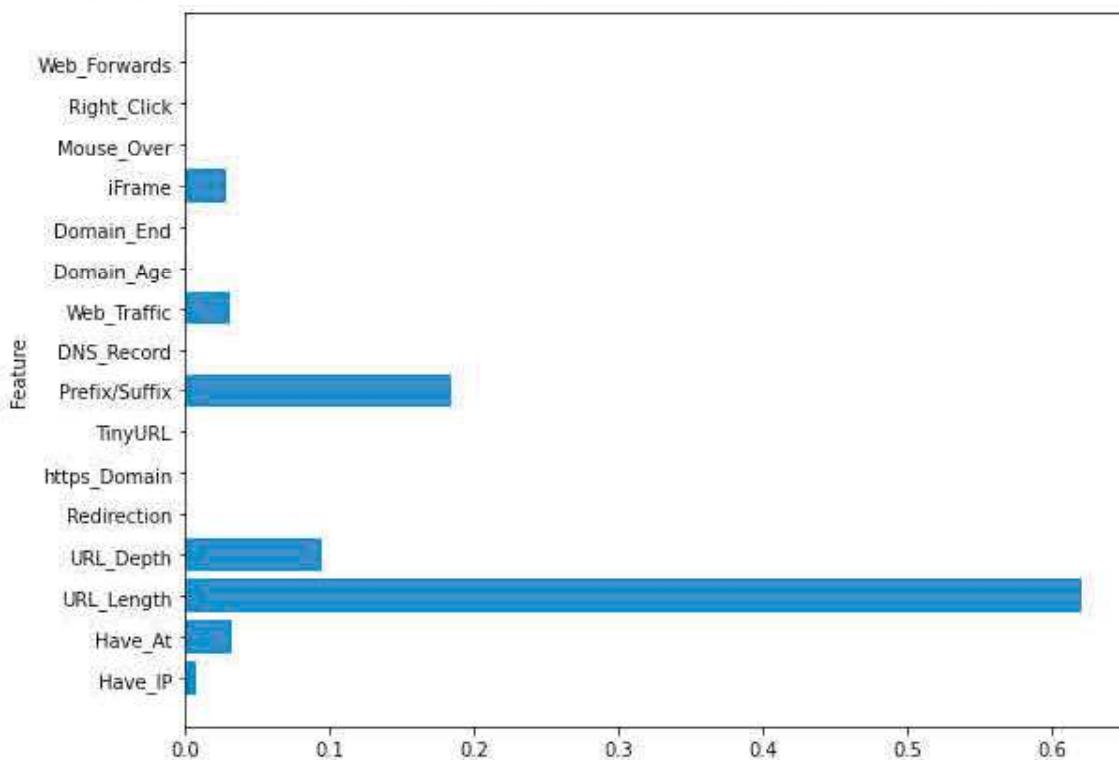
	Have_IP	Have_At	URL_Length	URL_Depth	Redirection	https_Domain	TinyURL	Prefix/Suffix	DNS_Record
count	10.000.000.000	10.000.000.000	10.000.000.000	10.000.000.000	10.000.000.000	10.000.000.000	10.000.000.000	10.000.000.000	10.000.000.000
mean	0.005500	0.022600	0.773400	3.072.000	0.013500	0.000200	0.090300	0.093200	
std	0.073961	0.148632	0.418653	2.128.631	0.115408	0.014141	0.286625	0.290727	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.000000	0.000000	1.000.000	2.000.000	0.000000	0.000000	0.000000	0.000000	
50%	0.000000	0.000000	1.000.000	3.000.000	0.000000	0.000000	0.000000	0.000000	
75%	0.000000	0.000000	1.000.000	4.000.000	0.000000	0.000000	0.000000	0.000000	
max	1.000.000	1.000.000	1.000.000	20.000.000	1.000.000	1.000.000	1.000.000	1.000.000	1.000.000

**Tablo 3: Veri ön işleme**

	Have_IP	Web_Traffic	Domain_Age	Domain_End	iFrame	Mouse_Over	Right_Click	Web_Forwards	Label
count	10.000.000.000	10.000.000.000	10.000.000.000	100.000.000	10.000.000.000	1.000.000.000	1.000.000.000	1.000.000.000	10.000.000.000
mean	0.100800	0.845700	0.413700	0.8099	0.090900	0.06660	0.99930	0.105300	
std	0.301079	0.361254	0.492521	0.3924	0.287481	0.24934	0.02645	0.306955	
min	0.000000	0.000000	0.000000	0.0000	0.000000	0.00000	0.00000	0.000000	
25%	0.000000	1.000.000	0.000000	10.000	0.000000	0.00000	100.000	0.000000	
50%	0.000000	1.000.000	0.000000	10.000	0.000000	0.00000	100.000	0.000000	
75%	0.000000	1.000.000	1.000.000	10.000	0.000000	0.00000	100.000	0.000000	
max	1.000.000	1.000.000	1.000.000	10.000	1.000.000	100.000	100.000	100.000	1.000.000

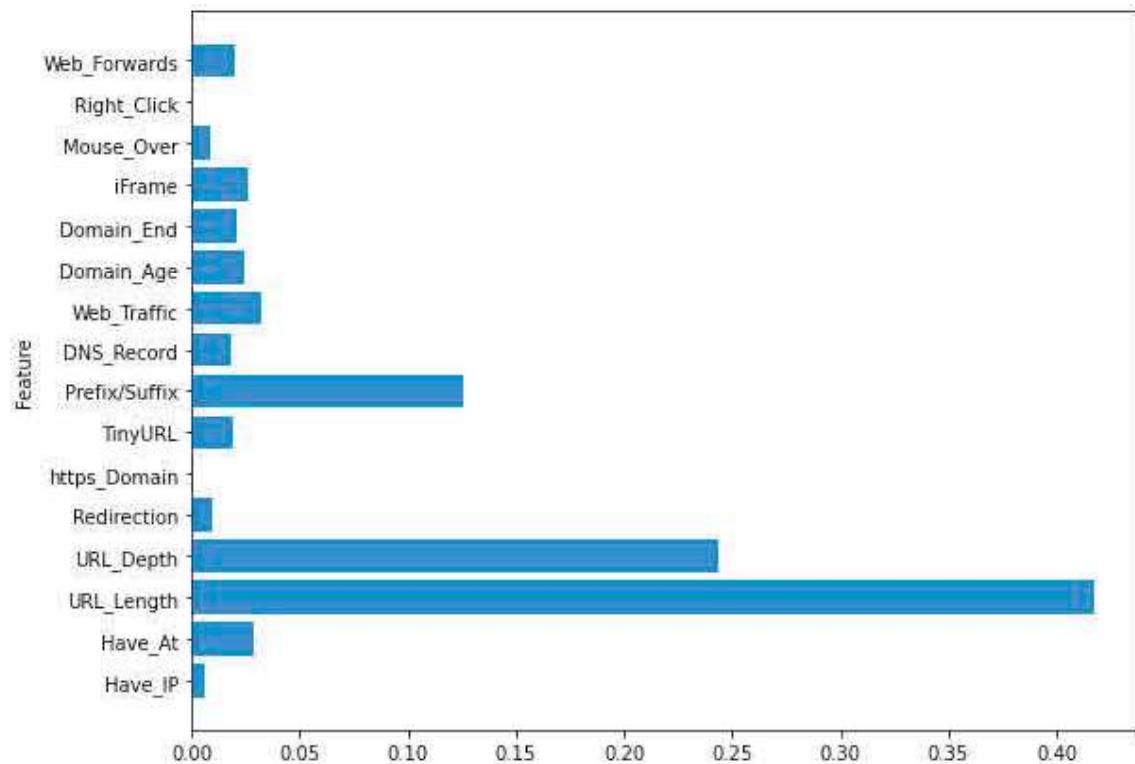
Yukarıda elde edilen sonuçlar, verilerin çoğunun 'Domain' ve 'URL\_Depth' sütunları dışında 0'lar ve 1'lerden oluştuğunu göstermektedir. DNS sütununun, makine öğrenimi modeli eğitimi için herhangi bir önemi yoktur.

**Şekil 13: Karar ağacı sınıflandırıcısı**



10.000 veri setinden 8.000'i algoritmada kullanılan 2.000'ide testte kullanmak için ayırmıştık Karar Ağacı yöntemi ile 8.000 veri seti üzerinde hesaplama yapıldığında URL adresleri ve Prefix özelliğinin derinliği ve uzunluğunun daha fazla olduğu görülmekte olup bu özellikler niteliklerine göre legal illegal URL ayrimı yapılabilecektir.

**Şekil 14: Rastgele orman sınıflandırıcısı**



**Tablo 4: Sonuç listesi**

ML Model	Train Accuracy	Test Accuracy	Result
0	Karar ağacı	0.810	0.826
1	Rastgele Orman	0.814	0.834
2	Çok Katmanlı Algılayıcılar	0.858	0.863
3	XGBoost	0.866	0.864
4	Otomatik Kodlayıcı Sinir Ağrı	0.819	0.818
5	Vektör Makineleri	0.798	0.818

Yukarıdaki modellerin elde edilen sonuçlarından XGBoost makine öğrenmesi, %86,4 ile en yüksek model performansına sahip olduğu görüşmüştür. Dikkat edilmesi gereken bir husus da diğer hesaplamaların da %80'nin üzerinde başarılı olduğu görülmektedir. Bu da makine öğrenmesi ve veri madenciliğinin bu saldırıları engelleme uygulamalarında başarı sonuçlar elde etmeyi sağlamaktadır.



## **5. TARTIŞMA**

Siber saldırıların çok yoğun bir şekilde yaşandığımız günümüzde internet tabanlı işlemlerin her geçen gün artması sonucu ve interneti herkesin kullanması gerektiğini ve yapılan işlemlerin kritiklik seviyesini artamamaktadır. E-Devlet işlemlerinden tutun banka şubesine gitmeden görüntülü ve yeni kimlik kartlarının NFC özellikleri ile hesap açtırmaya kadar hayatımız her alanına girmiştir. Yapılan bu işlemlerden anlaşılıyor ki bunların hepsi kişisel bilgilerimizi ve ticari bilgilerimizi içermektedir. Bu da işlemlerin ne kadar kritik ve güvenliğin zorunlu olması gerektiğini göstermektedir.

Oltalama saldırıları büyük oranda insan odaklı olduğundan dolayı birçok çalışma bu saldırıları otomatik engellemeye ve farkındalık eğitimlerinin artması yönünde güvenlik seviyeleri artırmaya çalışılmaktadır. Şirketler, kurum ve kuruluşlar siber güvenlik altyapılarını korumak için büyük miktarlarda paralar harcanmış olsalar dahi saldırganlar, bu saldırının türünü kullanarak hedef sistemler insanlar üzerinden kolaylıkla erişim elde edebilirler.

Bu saldırının yöntemi saldırganlar tarafından en fazla tercih edilen bir yöntemdir. Bunun sebebi insan odaklı ve aldatma sanatı olarak da ifade edilen ve kullanıcılar üzerinde heyecan, korku ve stres oluşturabilecek psikolojik baskı uygulaması da bu saldırının ayrı bir yönüdür. Kullanıcılar mail yoluyla veya sms ile gelen linklere tıkladıklarında sitenin saldırganlar tarafından hazırlanmış sahte bir site olduğunu fark etmeden istenilen bilgilerini bu siteye girerek tüm bilgilerini kendi eliyle saldırganlara vermiş olacaktır. Son dönemde artan sosyal paylaşım sitelerinde takipçi sayısının çok olduğu ünlülerin veya siyasetçilerin hesaplarını ele geçirmek için bu bahsedilen yöntemler kullanılarak ele geçirildiği görülmüştür.

Saldırıların tespitine yönelik yapılan çalışmalar ve Literatür araştırmalarından da görüldüğü üzerinde oltalama saldırıları ile mücadelenin zor olduğu, bundan dolayı makine öğrenmesi metotları ile çalışmalar yapılarak engellenmeye ve kullanıcılarla farkındalık eğitimleri verilerek saldırının akılda kalmasına ve yılda en az 2 defa sosyal mühendislik testleri yapılarak güvenlik sevileri artırılmalıdır.

Bu saldırısı yöntemlerinin engellenmesi veya tespit edilmesine yönelik yapılan çalışmalarında URL adreslerinin yapısının incelenmesinin dışında e-posta veri kümesi oluşturularak metin sınıflandırma yöntemi olan Ki Kare yöntemi ile diğer yöntemler olan öz nitelik seçme yöntemleri ile karşılaştırıldığında Ki Kare yönteminin daha başarılı olduğu görülmüştür.

Bazı çalışmalar sosyal mühendislik saldırısı simülasyonu yaparak ve farkındalık eğitimleri düzenleyerek bu saldırılardan gerçek boyutta verdikleri zararları uygulamalı olarak göstermeleri sonucundan farkındalığın daha çok arttığını ortaya koymuştur.

Diğer bir çalışmada kimlik avı saldırılara karşı geliştirilen kural tabanlı sistem kullanarak mail içerisindeki linklerin şüpheli sayfalara yönlendirildiğinin tespiti yönünde olup legal bağlantılardan ayırmak için geliştirilen genetik algoritma kullanılır. Oluşturulan veri seti bir veri tabanından saklanır ve kural tabanlı sistemdeki kurallar ile karşılaştırılarak bağlantının illegal mi yoksa legal mi olduğunu tespit edilmeye çalışmıştır.

Ülkelerinde siber saldırılara karşı hem yasal hem de teknoloji anlamında ürünler ve yasalar çıkartarak bu saldırılara yapan Hacker veya Hacker gruplarına karşı yaptırımlar söz konusu olmuştur. Bizim ülkemizde de 2007 yılında çıkartılan 5651 yasal kanun ile amaç ve kapsam belirlenmiştir. Kanun, “*İçerik sağlayıcı, yer sağlayıcı, erişim sağlayıcı ve toplu kullanım sağlayıcıların yükümlülük ve sorumlulukları ile internet ortamında işlenen belirli suçlarla içerik, yer ve erişim sağlayıcıları üzerinden mücadeleye ilişkin esas ve usulleri düzenlemektir.*” (Resmi Gazete Sayı: 25530, 2007). Şeklindedir.

Saldırılarla yasal mücadele yapmak adına 5651 kanunun dışından Kişisel Verilerin Korunması Kanunu 2016 yılında 6698 sayılı karar ile Mecliste kabul edilmiştir. Bu Kanunun amacı, kişisel verilerin işlenmesinde başta özel hayatın gizliliği olmak üzere kişilerin temel hak ve özgürlüklerini korumak ve kişisel verileri işleyen gerçek ve tüzel kişilerin yükümlülükleri ile uyacakları usul ve esasları düzenlemektir (Resmi Gazete Sayı: 26977, 20016).

Bu tez çalışmasında sosyal mühendislik yöntemi olan oltalama saldırılarda kullanılan URL’lerin tespit edilmesine yönelik bir sistem geliştirilmeye çalışılmıştır. Tespit edilen URL adresleri pasif veri türündendir. URL adresinden yola çıkılarak saldırıların hızlı bir şekilde tespit edilmesi yönünde çalışma yapılmıştır. Bu sayede yapılan saldırılar, hedefe ulaşmadan önce tespit edilerek bloklanması amaçlanmıştır. Oluşturulan bir veri seti üzerinde makine öğrenmesi modellemeleri ile verilerin eğitilmesi sağlanmış olup, URL ve web sitelerinin içeriğindeki legal bağlantılar ve illegal bağlantıları karşılaştırılarak korumaya çalışmaktadır.

Özet olarak bu çalışmada ve yapılan diğer çalışmalarдан görülmüþ ki oltalama saldırıların tespit edilmesi ve bu saldırıların en aza indirgenmesi açısından kullanıcıların eğitilmesi gereği ve veri madenciliği ile yapay zekâ ile geliştirilen teknolojiler ile saldırının hedef kullanıcıya ulaşmadan önce tespit edilmesi ve yüksek oranda engellenmesi amaçlanmaktadır.

## 6. SONUÇ VE ÖNERİLER

Sosyal Mühendislik saldırılardan en fazla saldırganlar tarafından tercih edilen insan odaklı saldırı yöntemi olan oltalama saldırılara karşı savunma sağlamak adına geliştirmek ve mağduriyet yaratmadan engellenmesini amaçlanan veri madenciliği yöntemleri ve makine öğrenmesi algoritmaları kullanarak geliştirilmeye çalışılmıştır.

Legal ve illegal veri setlerinden rastgele 5000 adet URL adresleri rastgele seçilerek çalışma yapılacak veri seti (legalveillegal.csv) oluşturulmuştur. Veri setleri ile alakalı bilgi aşağıdaki belirtilmiştir.

1. legal.csv: Legal URL adreslerinin olduğu veri seti.
2. illegal.csv: Açık kaynaklı PhishTank üzerinden alınan veri seti.
3. random\_legalurl.csv: Legal URL adreslerin 5000 adet rasgele seçilmiş veri seti.
4. random\_ilegalurl.csv: Illegal(PhishTank) adresler arasından 5000 adet rasgele seçilmiş veri seti.
5. legalveillegal.csv: Legal ve Illegal URL adreslerinin bulunduğu veri seti.

legalveillegal.csv veri setinde 10.000 adet URL adresi kayıt edilmiştir. Veri setleri ile alakalı içerikleri <https://github.com/muraddemirci/oltalamasaldiritespit> adresinden detayları bulunabilir.

Tüm bu modeller veri seti üzerinde eğitilir ve test veri seti ile modelin değerlendirilmesi yapılır. Bu çalışmada kullanılan veri setleri ve kaynak kodlar aşağıdaki Google Colab linkinden ulaşılabilir durumdadır. Erişim tarihi, 11.01.2022

<https://colab.research.google.com/drive/157d3RuN0x6NHP2OYvaDvlcP0l4whuoDS?usp=sharing>

Tüm bu çalışmalar ve araştırmalar sonucu gösteriyor ki siber güvenlik alanında sosyal mühendislik yöntemlerinde insan odaklı yapılan saldırının engellenmesine karşı tüm Dünyada hem Ülkeler bazında olsun hem Üniversite bazındaki çalışmalar olsun hem de özel ticari kuruluşların bu sadırılarla mücadele etmeleri anlamında zorluk geçmektedir.

Bu saldırılara karşı bilinçli ve farkında olarak özellikle üçüncü şahıslara, aniden ortaya çıkan ‘hayırsevere’ güvenmemek, en başta gelen yöntemlerdendir biridir. Yine de sosyal mühendislik saldırularına karşı hazırlıkta olmanın son derece zor olduğunu unutmayın. Çünkü insanı değerler noktasında sevdiklerimize yardım etme güdüsü hemen hepimizde vardır veya aniden gelişen bir olay karşısında hızlı ve heyecanla hareket etme zafiyetimiz vardır. Bu durumların farkında olan saldırganlar da bu özellikleri bildikleri için sürekli saldırıda bulunuyorlar.

Bu saldırılara maruz kalmamak için hem kişisel hem de tüzel farklılıkların olarak almamız gereken birçok önlemler bulunmaktadır. Kullandığımız cihazları güvenliğini artırmak için anti virüs yazılımı kesinlikle kullanmalıyız. Sosyal mühendislik saldırısı veya başka bir saldırıyla maruz kalsak bile erişilebilecek alanlar kısıtlı olacaktır. Kullanılan cihazlar ister akıllı telefon, ister standart bir ev kullanıcısı ağı, isterse büyük bir kurumsal sistem olsun, temel alınması gereken adımlar aynıdır. Başlıca alabileceğimiz önlemler şu sıralayabiliriz.

1. Kullanacağınız anti virüs yazılımları lisanslı ve muhakkak güncel olmalıdır. Mail yoluyla gelecek veya kötü amaçlı yazılımların yüklenmesini önlemeye yardımcı olabilir.
2. Telefonunuza tam yetkili erişim hakkı tanımak için root yapmayın yaptırmayı veya ağınızı ya da bilgisayarınızı yönetici modunda çalıştırmayın. Bir sosyal mühendislik saldırısı kullanıcı hesabınızın kullanıcı parolasını alsa bile, yetkisiz bir hesap olacağı için sisteminizi yeniden yapılandırmamasına veya üzerine yazılım yüklemesine izin vermez.
3. Birçok hesaplarınız için aynı parolayı kullanmayın. Bir sosyal mühendislik saldırısını sonucunda ele geçirilen parola ile diğer hesabınızı korumuş olursunuz.
4. Önemli hesaplarınız için iki aşamalı kimlik doğrulama kullanın, böylece bu hesaplara erişmek için yalnızca parolanız yeterli olmaz. Buna ses tanıma, bir güvenlik cihazı kullanımı, parmak izi veya SMS onay kodları dahil olabilir.

5. Kullanıcı hesap adınızı ve parolasını geçici de olsa birine verdiyseniz saldırıyla maruz kalabilecek düşüncesiyle parolayı hemen değiştirin.
6. Siber güvenlikle alakalı farkındalık eğitimleri alarak risklerin farkında olmanızda fayda vardır. Yeni saldırısı yöntemleri ortaya çıktıktan sonra onlar hakkında bilgi sahibi olarak mağdur olma ihtimalinizi azaltabilirsiniz.
7. Size gelen bir mail içeriğinde veya tıklamanızı isteyen bir soru ile karşılaşığınızda URL/Adres kontrolü yani tarayıcınızda bulunan adresi kontrol etmenizdir. Veya gelen maile cevap yazarken gönderilecek olan adresin doğruluğunu teyit etmeniz son derece önemlidir.

Ülkemizde Emniyet Genel Müdürlüğü'nün Siber Suçlarla Mücadele kapsamında hazırladıkları SİBERAY projesi Ülkemizdeki vatandaşların güvenli internet kullanımı için kullanıcılarına ve vatandaşlarına yol gösteren rehberler hazırlamaktadır. Bu kapsamında aşağıdaki linkten rehbere ulaşılabilir (<https://www.siberay.com/guvende-kalma-rehberi>).  
Erişim tarihi 11.01.2022

## 7. KAYNAKLAR

<https://data.tuik.gov.tr>, <https://data.tuik.gov.tr>, [ [https://data.tuik.gov.tr/Bulten/Index?p=Hanehalki-Bilisim-Teknolojileri-\(BT\)-Kullanim-Arastirmasi-2020-33679](https://data.tuik.gov.tr/Bulten/Index?p=Hanehalki-Bilisim-Teknolojileri-(BT)-Kullanim-Arastirmasi-2020-33679) ]

Erişim tarihi: 01.10.2021

<https://docs.apwg.org>, [https://docs.apwg.org/reports/apwg\\_trends\\_report\\_q3\\_2020.pdf](https://docs.apwg.org/reports/apwg_trends_report_q3_2020.pdf)  
[https://docs.apwg.org/reports/apwg\_trends\_report\_q3\_2020.pdf]

Erişim tarihi: 05.10.2021

www.secureworld.io, State of the Phish Report 2021: 4 Key Metrics,  
[https://www.secureworld.io/industry-news/state-of-the-phish-report-2021]

Erişim tarihi: 09.10.2021

<https://www.tessian.com>, Must-Know Phishing Statistics: Updated 2021,  
[https://www.tessian.com/blog/phishing-statistics-2020/]

Erişim tarihi: 10.10.2021

<https://www.researchgate.net>, Example of an email based phishing attack2021,  
[https://www.researchgate.net/figure/Example-of-an-email-based-phishing-attack\_fig1\_343543963]

Erişim tarihi: 11.10.2021

<https://not2phish.co.uk>, Life Cycle of a Phishing Attack [https://not2phish.co.uk/knowledge-base/]  
Erişim tarihi: 15.10.2021

<https://not2phish.co.uk>, Phishing Activity Trends Report, 3rd Quarter 2019 [https://not2phish.co.uk],  
Life Cycle of a Phishing Attack [https://not2phish.co.uk/knowledge-base/]  
Erişim tarihi: 15.10.2021

<https://www.researchgate.net>, Phishing URL detection system based on URL features using SVM,  
[https://www.researchgate.net/profile/BireswarBanik/publication/333166694/figure/fig1/AS:75953484090113@1558098612378/Block-diagram-of-our-proposed-system\_W640.jpg]  
Erişim tarihi: 15.10.2021

Majed, R. (2019). Visualisation Model Based on Phishing Features. Journal of Information & Knowledge Management Vol. 18, No. 1, 1-17.

James, J., L., S., & Thomas, C. (2013). Detection of Phishing URLs Using Machine Learning Techniques. International Conference on Control Communication and Computing (ICCC), (s. 304-309). Mindeb.

Anti-PhishingWorking Group. Global Phishing Report 2H 2014. [[http://docs.apwg.org/reports/APWG\\_Global\\_Phishing\\_Report\\_2H\\_2014.pdf](http://docs.apwg.org/reports/APWG_Global_Phishing_Report_2H_2014.pdf)]

Anti-Phishing Working Group. Global Phishing Survey: Trends and Domain Name Use in 2016. [http://docs.apwg.org/reports/APWG\\_Global\\_Phishing\\_Report\\_2015-2016.pdf](http://docs.apwg.org/reports/APWG_Global_Phishing_Report_2015-2016.pdf), 2017.

Erişim tarihi: 04.10.2021

Learning based Malicious Web Sites Detection using Suspicious URLs 2017  
<http://users.eecs.northwestern.edu/~hlc720/349/index.html>

Erişim tarihi: 20.10.2021

Must-Know Phishing Statistics: Updated 2021 Suspicious URLs 2020

<https://www.tessian.com/blog/phishing-statistics-2020/>

Erişim tarihi: 04.11.2021

Veri madenciliği 2015. [http://auzefkitap.istanbul.edu.tr/kitap/endustirimuhlt\\_ue/verimadenciligi.pdf](http://auzefkitap.istanbul.edu.tr/kitap/endustirimuhlt_ue/verimadenciligi.pdf).

Erişim tarihi: 10.11.2021

Yang, X., Yan, L., Yang, B., & Li, Y.-f. (2017). Phishing Website Detection Using C4.5 Decision Tree. 2nd International Conference on Information Technology and Management Engineering (ITME 2017).

Jain A. K. and Gupta B. B. (2018) PHISH-SAFE: URL features-based phishing detection system using machine learning, Advances in Intelligent Systems and Computing, 2018; 729: 467-474.

Google,

Google

Colab,

[<https://colab.research.google.com/drive/157d3RuN0x6NHP2OYvaDvlcPOl4whuoDS#scrollTo=C297HhYulXcb>]

Erişim tarihi: 17.12.2021

Githup, muraddemirci /oltalamasaldiritespit, [ <https://github.com/muraddemirci/oltalamasaldiritespit/> ]  
Erişim tarihi: 11.01.2022

Detection of phishing emails using data mining algorithms 2015  
<https://ieeexplore.ieee.org/abstract/document/7399985>  
Erişim tarihi:25.12.2021

The Effect of SMiShing Attack on Security of Demand Response Programs 2020  
[https://www.researchgate.net/publication/344147847\\_The\\_Effect\\_of\\_SMiShing\\_Attack\\_on\\_Security\\_of\\_Demand\\_Response\\_Programs](https://www.researchgate.net/publication/344147847_The_Effect_of_SMiShing_Attack_on_Security_of_Demand_Response_Programs)  
Erişim tarihi: 10.11.2021

Bilgisayar Sistemlerine Yapılan Saldırılar Ve Türleri 2007 <https://dergipark.org.tr/tr/download/article-file/252301>  
Erişim tarihi:25.12.2021

Junaid Ahsenali Chaudhry, Shafique Ahmad Chaudhry, and Robert G Rittenhouse. Phishing attacks and defenses. International Journal of Security and Its Applications, 10(1):247–256, 2016.

F. Zhou, “Phishing Sites and Prevention Measures”, Int. J. Secur. Its Appl.vol. 9, 1–10, 2015

Ali W., (2017) Phishing Website Detection based on Supervised Machine Learning with Wrapper Features Selection, International Journal of Advanced Computer Science and Applications, 2017.

Patil V., Thakkar P., Shah C., Bhat T. and Godse S.P. (2018) Detection and Prevention of Phishing Websites Using Machine Learning Approach, in Proceedings - 2018 4th International Conference on Computing, Communication Control and Automation, ICCUBEA 2018.

Hossain S., Sarma D. and Chakma R. J. (2020) Machine learning-based phishing attack detection, International Journal of Advanced Computer Science and Applications, 2020; 11: 378-388.

Siber Suçlar ve Türkiye'nin Siber Güvenlik Politikaları 2013  
<https://www.acarindex.com/dosyalar/makale/acarindex-1423936102.pdf>  
Erişim tarihi:10.11.2021

Tang, Z., MacLennan, J., "Data Mining with Sql Server 2005", Wiley, 2005