

T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ



DERİN ÖĞRENME YÖNTEM VE MODELLERİ
KULLANILARAK SOSYAL MÜHENDİSLİK SALDIRILARININ
TESPİTİNDE YENİ YAKLAŞIMLARIN GELİŞTİRİLMESİ

Zeynep ASLANPENÇESİ

Yüksek Lisans Tezi

YAZILIM MÜHENDİSLİĞİ ANABİLİM DALI

MAYIS 2024

T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Yazılım Mühendisliği Anabilim Dalı

Yüksek Lisans Tezi

DERİN ÖĞRENME YÖNTEM VE MODELLERİ KULLANILARAK
SOSYAL MÜHENDİSLİK SALDIRILARININ TESPİTİNDE YENİ
YAKLAŞIMLARIN GELİŞTİRİLMESİ

Tez Yazarı
Zeynep ASLANPENÇESİ

Danışman
Doç. Dr. Muhammet BAYKARA

MAYIS 2024
ELAZIĞ

T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Yazılım Mühendisliği Anabilim Dalı

Yüksek Lisans Tezi

Başlığı:	Derin Öğrenme Yöntem ve Modelleri Kullanılarak Sosyal Mühendislik Saldırılarının Tespitinde Yeni Yaklaşımların Geliştirilmesi
Yazarı:	Zeynep ASLANPENÇESİ
İlk Teslim Tarihi:	02.01.2024
Savunma Tarihi:	15.05.2024

TEZ ONAYI

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

Danışman:	Doç. Dr. Muhammet BAYKARA Fırat Üniversitesi, Teknoloji Fakültesi	İmza Onayladım
Başkan:	Dr. Öğr. Üyesi Talha Burak ALAKUŞ Kırklareli Üniversitesi, Mühendislik Fakültesi	Onayladım
Üye:	Dr. Öğr. Üyesi Murat AYDOĞAN Fırat Üniversitesi, Teknoloji Fakültesi	Onayladım

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

İmza

Prof. Dr. Burhan ERGEN
Enstitü Müdürü

BEYAN

Fırat Üniversitesi Fen Bilimleri Enstitüsü tez yazım kurallarına uygun olarak hazırladığım “Derin Öğrenme Yöntem ve Modelleri Kullanılarak Sosyal Mühendislik Saldırılarının Tespitinde Yeni Yaklaşımların Geliştirilmesi” Başlıklı Yüksek Lisans Tezimin içindeki bütün bilgilerin doğru olduğunu, bilgilerin üretilmesi ve sunulmasında bilimsel etik kurallarına uygun davrandığımı, kullandığım bütün kaynakları atıf yaparak belirttiğimi, maddi ve manevi desteği olan tüm kurum/kuruluş ve kişileri belirttiğimi, burada sunduğum veri ve bilgileri unvan almak amacıyla daha önce hiçbir şekilde kullanmadığımı beyan ederim.

15.05.2024

Zeynep ASLANPENÇESİ



ÖNSÖZ

Bu tez kapsamında günümüzün vazgeçilmez bir parçası haline gelen sosyal ağlar ve internet aracılığıyla gerçekleştirilen sosyal mühendislik saldırıları incelenmiştir. Günümüzde geline nokta göz önüne alındığı zaman bu saldırıların hızlı ve etkin tespiti oldukça önem kazanmaktadır. Ayrıca bu tespitlerin doğruluk oranı da aynı doğrultuda öneme sahiptir. Bu nedenle sosyal mühendislik saldırılarında literatürde mevcut olan tespit yöntemlerinin yanı sıra Doğal Dil İşleme (DDİ) alanlarını da inceleyerek bu alana katkıda bulunulması amaçlanmıştır. Tez konusu günümüz teknolojisi de dikkate alındığı zaman oldukça önemli bir alanı kapsamaktadır. Bu çalışma için yapılan ön hazırlık aşamasında özellikle DDİ alanında karşılaşılan en büyük sorun, Türkçe kaynak bulma konusunda yaşanan sıkıntılardır. Bu alanda ülkemizde henüz yeterli ve detaylı çalışmalar bulunmamaktadır. Bu nedenle yabancı kaynaklardan yararlanılarak ön hazırlık aşaması tamamlanmıştır.

Çalışmamı hazırlama süreci boyunca her aşamada gösterdiği ilgi ve destekten ötürü değerli hocam ve danışmanım Doç. Dr. Muhammet BAYKARA'ya, yine bu süreçte bana olan destek ve yardımlarından ötürü aileme teşekkür ederim. Bu tez çalışması, Fırat Üniversitesi Bilimsel Araştırma Projeleri Koordinasyon Birimi (FÜBAP) tarafından **TEKF.22.27** protokol numaralı proje ile desteklenmiştir. Projeye verdiği destekten ötürü FÜBAP'a teşekkürlerimizi sunarız.

Zeynep ASLANPENÇESİ

ELAZIĞ, 2024

İÇİNDEKİLER

Sayfa

Önsöz.....	iv
İçindekiler.....	v
Özet	vii
Abstract	viii
Şekiller Listesi.....	ix
Tablolar Listesi.....	x
Simgeler ve Kısaltmalar	xi
1. GİRİŞ	1
2. İLGİLİ ÇALIŞMALAR	4
3. BİLGİ GÜVENLİĞİ.....	10
3.1. Bilgi Güvenliği İlkeleri.....	11
3.1.1. Gizlilik	11
3.1.2. Bütünlük.....	11
3.1.3. Kullanılabilirlik	11
3.2. Bilgi Güvenliğinin Önemi	12
3.3. Bilgi Güvenliği Türleri	13
3.3.1. Uygulama Güvenliği	13
3.3.2. Altyapı Güvenliği.....	14
3.3.3. Bulut Güvenliği.....	14
3.3.4. Uç Nokta Güvenliği	15
3.3.5. Kriptografi.....	16
3.3.6. Güvenlik Açığı Yönetimi	17
3.4. Bilgi Güvenliği Politikası	18
3.5. Bilgi Güvenliği Önlemleri	19
3.6. Bilgi Güvenliği ve Siber Güvenlik	20
4. SOSYAL MÜHENDİSLİK SALDIRI TÜRLERİ	22
4.1. Sosyal Mühendislik Saldırıların Gerçekleştirilmesi	25
4.2. Sosyal Mühendislik Saldırılarından Korunma Yöntemleri.....	26
4.3. Sosyal Mühendislik Saldırıların Tespitindeki Zorluklar	29
4.4. Sosyal Mühendislik Saldırıların Tespitinde Kullanılan Yöntemler	29
4.4.1. Smishing.....	29
4.4.2. Phishing.....	31
4.4.3. Literatürde Yer Alan Diğer Çalışmalar	33
5. DERİN ÖĞRENME	36
5.1. Evrimsel Sinir Ağı	36
5.2. Tekrarlayan Sinir Ağı Mimarileri.....	37
5.2.1. Uzun Kısa Dönemli Hafıza (LSTM)	39
5.2.2. Çift Yönlü Uzun Kısa Dönemli Hafıza (BiLSTM)	40
5.2.3. Gated Recurrent Unit (GRU)	40
6. DOĞAL DİL İŞLEME TEKNİKLERİ	42
6.1. Corpus	42

6.2. Natural Language Toolkit (NLTK) ile Tokenizasyon	43
6.3. NLTK ile StopWords.....	43
6.4. NLTK ile Stemming.....	43
6.5. NLTK ile Lemmatization	44
6.6. Kelime Vektörleri.....	44
6.6.1. Kelime Vektörleri – word2vec	45
6.6.2. Kelime Vektörleri – Glove.....	46
6.7. Doğal Dil İşlemenin Çalışma Mantığı.....	47
6.8. Yapay Zekâ ile Doğal Dil İşleme Arasındaki İlişki	48
7. MATERYAL VE METOTLAR	49
7.1. Doğal Dil İşleme Teknikleri Kullanılarak Hibrit CNN-BiLSTM Yöntemi ile Smishing Saldırıların Tespiti	49
7.1.2. Geliştirilen Yöntem.....	51
7.2. Hibrit CNN – GRU Yöntemi ile SMS Phishing Tespiti	54
7.2.1. Geliştirilen Yöntem.....	54
7.3. Twitter Verileri Kullanılarak Kısa Metinler Üzerinde GRU ile Nefret Söylemi Tespiti	57
7.3.1. Geliştirilen Yöntem	58
8. SONUÇLAR VE TARTIŞMA.....	64
KAYNAKLAR.....	66
ÖZGEÇMİŞ	

ÖZET

Derin Öğrenme Yöntem ve Modelleri Kullanılarak Sosyal Mühendislik Saldırılarının Tespitinde Yeni Yaklaşımların Geliştirilmesi

Zeynep ASLANPENÇESİ

Yüksek Lisans Tezi

FIRAT ÜNİVERSİTESİ
Fen Bilimleri Enstitüsü

Yazılım Mühendisliği Anabilim Dalı

Mayıs 2024, Sayfa: xi + 71

Günümüzün dijital dünyasında meydana gelen gelişmeler neticesinde kullanıcıların sosyal medyada paylaştığı verilerin güvenliğini sağlamak da bilgi güvenliğinin önemini artırmaktadır. Gerekli güvenliğin sağlanmadığı dijital ortamlarda kullanıcıların her türlü verisi saldırgan olarak nitelendirilen kötü niyetli kişilerin açık hedefi haline gelmektedir. Saldırganlar tarafından kullanıcıların başta kişisel verileri ve iş hayatına ilişkin verileri olmak üzere her türlü verisi ele geçirilebilmektedir. Verileri ele geçirmek için kullanılan en yaygın yöntemlerden olan sosyal mühendislik saldırıları bilgi güvenliği ve veri güvenliği açısından büyük tehlike arz etmektedir. Kullanıcıların dijital ortamlarda da en az fiziki ortamlarda olduğu kadar güvende olması günümüzün en önemli sorunları arasında olmakla birlikte bu güvenliğin sağlanması da öncelikle bu saldırıların tespiti ile engellenmesine dayanmaktadır. Bu nedenle öncelikle en yaygın sosyal mühendislik saldırı türleri yakından tanınmalıdır. Akabinde bu saldırıların açıklıkları bulunarak tespiti için yine dijital yöntemlerin geliştirilmesi gerekliliği açıktır.

Tez çalışması kapsamında kullanıcıların veri güvenliğinin korunması için en yaygın sosyal mühendislik saldırısı olan SMS kimlik avı saldırılarının tespitine yönelik yenilikçi bir yaklaşım önerilmiştir. DDİ teknikleri ile bu saldırıların tespiti gerçekleştirilerek kullanıcıların bu saldırılardan korunması amaçlanmıştır. DDİ teknikleri kullanılarak Hibrit CNN-BiLSTM yöntemi ile SMS kimlik avı saldırılarının tespiti gerçekleştirilmiştir. Model başarımı 0.998 olarak elde edilmiştir. Ardından Hibrit CNN-GRU yöntemi kullanılarak SMS phishing tespiti gerçekleştirilmiştir. Bu yöntemle gerçekleştirilen tespitin başarımlı oranı 0.996 olarak elde edilmiştir. Son olarak DDİ teknikleri ile Twitter verileri kullanılarak kısa metinler üzerinde Gated Recurrent Unit (GRU) ile nefret söylemi tespiti gerçekleştirilmiştir. GRU modeli ile 0.998’lik bir başarımlı elde edilmiştir. Tez çalışması ile DDİ teknikleri kullanılarak sosyal mühendislik saldırılarının tespitinin gerçekleştirilmesi ve literatüre bu alanda katkı sağlanması hedeflenmiştir.

Anahtar Kelimeler: Sosyal Mühendislik, Bilgi Güvenliği, Doğal Dil İşleme Teknikleri, Hibrit Modeller, Siber Güvenlik, Derin Öğrenme Yöntemleri

ABSTRACT

Development of New Approaches in Detecting Social Engineering Attacks Using Deep Learning Methods and Models

Zeynep ASLANPENÇESİ

Master's Thesis

FIRAT UNIVERSITY
Graduate School of Natural and Applied Sciences
Department of Software Engineering

May 2024, Pages: xi + 71

As a result of the developments in today's digital world, ensuring the security of the data shared by users on social media increases the importance of information security. In digital environments where the necessary security is not provided, all kinds of data of users become the open target of malicious people who are considered as attackers. All kinds of data of users, especially personal data and business-related data, can be seized by attackers. Social engineering attacks, which are one of the most common methods used to capture data, pose a great danger in terms of information security and data security. While keeping users at least as safe in digital environments as in physical environments is one of the most important problems of today, ensuring this security is primarily based on detecting and preventing these attacks. For this reason, the most common types of social engineering attacks must first be closely recognized. Then digital methods must be developed to detect vulnerabilities in these attacks.

Within the scope of the thesis study, an innovative approach has been proposed to detect smishing attacks, the most common social engineering attack, to protect users' data security. It is aimed to protect users from these attacks by detecting these attacks with Natural Language Processing techniques. Smishing attacks were detected with the Hybrid CNN-BiLSTM method using Natural Language Processing techniques. Model performance was achieved as 0.998. Then, SMS phishing detection was performed using the Hybrid CNN – GRU method. The success rate of the detection performed with this method was obtained as 0.996. Finally, hate speech was detected with GRU on short texts using Natural Language Processing techniques and Twitter data. A performance of 0.998 was achieved with the GRU model. The aim of this thesis is to detect social engineering attacks using Natural Language Processing techniques and to contribute to the literature in this field.

Keywords: Social Engineering, Information Security, Natural Language Processing Techniques, Hybrid Models, Cyber Security, Deep Learning Methods

ŞEKİLLER LİSTESİ

	Sayfa
Şekil 3.1. Bilgi güvenliğinin unsurları.....	10
Şekil 3.2. Uygulama güvenliği	13
Şekil 3.3. Uç nokta güvenliği	16
Şekil 3.4. Güvenlik açığı yönetimi yaşam döngüsü.....	18
Şekil 4.1. Sosyal mühendislik saldırı türleri	22
Şekil 4.2. Phishing saldırısının başlatılması	22
Şekil 4.3. Pretexting saldırısının başlatılması.....	24
Şekil 4.4. Sosyal mühendislik saldırısı yaşam döngüsü	26
Şekil 5.1. CNN mimarisi	37
Şekil 5.2. Tekrarlayan sinir ağı mimarileri	38
Şekil 5.3. RNN fonksiyon yapısı	38
Şekil 5.4. RNN grafiksel gösterimi	39
Şekil 5.5. LSTM çalışma mantığı.....	40
Şekil 5.6. GRU mimarisi	41
Şekil 6.1. Tokenizasyon süreci	43
Şekil 6.2. Stemming ile veri ön işleme	44
Şekil 6.3. Lemmatization ile veri ön işleme	44
Şekil 6.4. Kelimelerin vektörlerle temsil edilmesi	45
Şekil 6.5. Skip-gram algoritması	46
Şekil 6.6. CBOW algoritması	46
Şekil 6.7. Doğal dil işleme süreci	48
Şekil 7.1. Smishing saldırısının başlatılması	50
Şekil 7.2. Spam.csv veri seti dağılım grafiği	51
Şekil 7.3. Spam.csv veri seti.....	51
Şekil 7.4. Spam.csv veri setinde yer alan stopwordsler.....	52
Şekil 7.5. Hibrit CNN-BiLSTM ile Smishing saldırılarının tespiti akış şeması	53
Şekil 7.6. Hibrit CNN-GRU ile SMS Phishing tespiti akış şeması	56
Şekil 7.7. Twitter veri seti	59
Şekil 7.8. Twitter veri seti etiketlerin dağılım grafiği.....	59
Şekil 7.9. GRU ile Twitter veri seti üzerinde nefret söylemi tespiti akış şeması.....	60
Şekil 7.10. Geliştirilen algoritmanın yapısı	62

TABLÖLAR LİSTESİ

	Sayfa
Tablo 4.1. Sosyal mühendislik saldırı türleri ve korunma yöntemleri	28
Tablo 4.2. Literatürde yer alan Smishing saldırılarında kullanılan tespit yöntemlerinin analizi.....	30
Tablo 4.3. Literatürde yer alan Phishing saldırılarında kullanılan tespit yöntemlerinin analizi	31
Tablo 4.4. Literatürde yer alan diğer sosyal mühendislik saldırılarında kullanılan tespit yöntemlerinin analizi.....	33
Tablo 6.1. Makine dilleri ve doğal dillerin karşılaştırılması	42
Tablo 6.2. Kelime vektörlerinin karşılaştırılması.....	47
Tablo 7.1. “spam.csv” veri setine uygulanan algoritmaların doğruluk performansı	54
Tablo 7.2. Veri setinde yer alan stopwordsler.....	55
Tablo 7.3. “spam.csv” veri setine uygulanan algoritmaların doğruluk performansı	57
Tablo 7.4. Geliştirilen modelin Precision, Recall ve F1-Score değerleri	57
Tablo 7.5. Train.csv veri setine uygulanan algoritmaların doğruluk performansı	63

SİMGELER VE KISALTMALAR

Kısaltmalar

MÖ	: Makine Öğrenimi
DÖ	: Derin Öğrenme
LSTM	: Uzun Kısa Dönemli Hafıza
BiLSTM	: Çift Yönlü Uzun Kısa Dönemli Hafıza
GRU	: Gated Recurrent Unit
NLTK	: Natural Language Toolkit
DDoS	: Dağıtılmış Hizmet Reddi
CNN	: Evrimsel Sinir Ağları
RNN	: Tekrarlayan Sinir Ağları
DDİ	: Doğal Dil İşleme
CBOW	: Continuous Bag of Words Model
RO	: Rastgele Orman
KNN	: K-en Yakın Komşu
BERT	: Bidirectional Encoder Representation from Transformers
LR	: Lojistik Regresyon
NB	: Naive Bayes
YSA	: Yapay Sinir Ağı
DVM	: Destek Vektör Makineleri
SGİ	: Stokastik Gradyan İnişi
DNN	: Derin Sinir Ağı

1. GİRİŞ

Dünyanın küresel bir köye dönüşeceği fikri ilk olarak 1962 yılında Marshall McLuhan tarafından ortaya atılmış ve yazar tarafından popülerleştirilmiştir [1]. Yazara göre teknolojinin ve elektronik iletişimin yaygınlaşmasıyla birlikte dünya küresel bir köye dönüşecek ve mesafelerin herhangi bir önemi kalmayacaktır. Günümüzde geline nokta yazarın ileri görüşlülüğünü gösterirken bir yandan da teknoloji ve elektronik iletişimin yaygın hale gelmesinin beraberinde getirdiği sorunları ortaya çıkarmaktadır.

Teknolojinin yaygınlaşması insanların hayatında büyük bir kolaylık sağlamakla birlikte iletişimin yüz yüze gerçekleştiği zamana göre beraberinde oldukça büyük ve tehlikeli sorunlar da getirmiştir. Örneğin iletişimin yüz yüze gerçekleştiği ve insanların genellikle iletişim sağladıkları kişileri tanıdığı dönemlerde kimse kendisine ait ve bir başkası tarafından ele geçirilmesi tehlikeli olan verileri tanımadığı ya da yeteri kadar güvenmediği kişiler ile paylaşmamaktaydı. Ancak günümüzde teknolojinin geldiği noktada telefonun diğer ucundaki kişiyi ve kişinin amacını bilemesek de iyi yazılmış ve cezbedici bir metnin kurbanı olarak kişisel verilerimizi paylaşabilmekteyiz. Bu kapsamda iletişimin evrildiği nokta da önem kazanmaktadır. Örneğin artık bankaların müşterileriyle, işveren şirketlerin çalışanlarıyla mesaj üzerinden belirli kalıplar doğrultusunda kurduğu iletişim çalışmanın devamında “saldırgan” olarak nitelendirilecek kişilerin işini kolaylaştırmaktadır. Çünkü aynı kalıp kullanılarak “hedef” şahsın kişisel verileri talep edilebilmekte söz konusu mesaj kalıbına aşina olan hedef tarafından mevcut güveni nedeniyle sorgulanmadan söz konusu veriler paylaşılabilir.

Sosyal mühendislik saldırıları olarak nitelendirilen bu fiiller ile saldırgan şahıs kendisine avantaj ve menfaat sağlayacak bilgileri hedef şahıstan temin etmektedir. Söz konusu fiiller birden çok şekilde ortaya çıkabilmektedir. Örneğin bir kısa mesaj üzerinden “SMS kimlik avı” saldırısı olarak, mail üzerinden “phishing” saldırısı olarak ya da sesli olarak iletişim kurulmak suretiyle “vishing” olarak gerçekleştirilebilmektedir. Bütün bu saldırı türlerinin ortak özelliği hedefte güven, korku, panik ya da merak duygusu uyandırarak saldırganın kimliğini sorgulamadan hedefin kişisel verilerini paylaşmasını sağlamaktır. Teknolojinin ve iletişimin günümüzdeki kolaylığı göz önüne alındığı zaman bu kolaylığın sağladığı imkânlar saldırgan tarafından kötü amaçlı olarak kullanılabilir ve kendisine menfaat sağlayabilmektedir.

Sosyal mühendislik saldırılarında saldırganlar öncelikle hedefin kimlik ve banka bilgilerine ulaşmayı hedeflemektedir. Çünkü saldırganları hedefine ulaştırabilecek en önemli iki kişisel veri türü bunlardır. Bankacılığın da dijital ortama taşındığı ve bütün hesaplara dört haneli şifreler ile erişim sağlandığı düşünüldüğü zaman saldırganın hedefin kimlik bilgilerine ulaşmasının ardından banka bilgilerine ulaşmasını da kolaylaştıracaktır. Bu kapsamda saldırganlar kullandıkları yöntemlerle verilere ulaşabilmek için hedefin duygu durumunda değişiklik meydana getirecek

senaryolar kullanmaktadır. Örneğin isminin bir terör örgütüyle ilişkilendirildiği son zamanlarda sıkça duyulan bir senaryo niteliğindedir. Bu senaryoda olduğu gibi saldırgan amacına ulaşabilmek için hedefin korku duygusunu harekete geçirerek ani karar almasını ve kararını sorgulamasını engellemektedir. Korku duygusu harekete geçirilen hedef ise saldırganın istediği verileri kendisiyle paylaşarak kendini güvenceye almaya çalışmaktadır. Bir başka durum ise hedefin merak ve heyecan duygusunu harekete geçirecek senaryoların kullanılmasıdır. Örneğin saldırgan tarafından hedefe bir ödül kazandığı, tatil kazandığı ya da ilgili soruları cevaplaması halinde sürpriz ödüller kazanacağı vaat edilmektedir. Merak ve heyecan duygusu harekete geçirilen hedef çoğunlukla ödüle ulaşmak amacıyla kendisinden istenen bilgileri paylaşmakta ya da yöneltilen soruları cevaplayarak ödüle ulaşmayı amaçlamaktadır. Ancak tüm bu yöntemlerin ortak sonucu saldırganı hedefine ulaştırarak kişisel verilerin ele geçirilmesini sağlamaktan başka bir şey değildir.

Sosyal mühendislik saldırılarının yaygınlaşmasının ardından bu saldırılarından korunmak için çözüm üretme yöntemleri de geliştirilmeye başlanmıştır. Söz konusu sosyal mühendislik saldırılarından korunmak için öncelikle kişilerin kendi güvenliklerini sağlamaları için manuel olarak geliştirilen geleneksel yöntemler ortaya çıkmıştır. Bu yöntemde kişilere tanımadığı kişilerden gelen mesajlara, maillere ya da aramalara cevap vermemeleri, kişisel verilerini paylaşmamaları, ödül ya da sürpriz vaat edilen reklamlara güvenmemeleri önerilmektedir. Bu şekilde en azından kişilerin belirli saldırılardan korunabilmesi amaçlanmaktadır. Ancak teknolojinin yaygınlaşması sosyal mühendislik saldırılarının yöntem ve çeşitlerini de artırmaktadır. Hedefin herhangi bir dalgınlıkla ya da yanlışlıkla kendisine iletilen mesajda yer alan bağlantıya tıklaması halinde dahi saldırgan amacına ulaşabilmektedir. Bu da korunma yöntemlerinin dijitalleşmesini, teknolojiyi takip ederek saldırıları hedefe ulaşmadan ortadan kaldırmasını zorunlu hale getirmektedir. Bu korunma yöntemlerinin geliştirilmesinde derin öğrenme yöntemleri de yaygın olarak kullanılmaktadır.

Sosyal mühendislik saldırılarının engellenmesi için literatürde birçok çalışma bulunmaktadır. Çoğunlukla gerçek phishing saldırılarına ait veriler üzerinde yapılan çalışmalarda Makine Öğrenimi (MÖ) ve Derin Öğrenme (DÖ) algoritmaları kullanılarak sosyal mühendislik saldırılarının engellenmesi amaçlanmaktadır. DÖ, büyük veri setlerini sinir ağlarıyla eğitmeye odaklanmaktadır. DÖ'nin ilgilendiği veri türleri ve öğrenme yöntemleri MÖ algoritmalarından farklıdır. Örneğin kedi, köpek gibi bir dizi hayvandan oluşan bir veri setine DÖ algoritmaları uygulandığı zaman MÖ'nden farklı olarak burada DÖ veri setinin ayırt edici özelliklerini belirleyebilmektedir.

DÖ yöntemi ile makineler dünyayı nasıl anlayacakları ve algılayacakları öğretilmektedir. Makinelere sunulan veriler etiketlenerek bu verilerin çalışma alanına ilişkin olarak anlamlı ya da anlamsız olduğu makineler tarafından algılanabilmektedir. Sosyal mühendislik saldırılarında mesaj içerikleri birbirlerine benzerlik taşıdığından söz konusu yöntem ve içerikler etiketlenmek suretiyle

makinelere saldırı olarak öğretilmektedir. Bu yöntemi kullanan cihazlar ise henüz saldırı hedefe ulaşmadan saldırıyı kaynakta engellemektedir.

Bu tez çalışmasında giriş bölümünün ardından Bölüm 2’de detaylı bir literatür taramasına yer verildikten sonra Bölüm 3’te öncelikle bilgi güvenliğinin ne olduğu detaylıca açıklanmış ardından bilgi güvenliğinin içerisinde yer alan temel sorunlardan olan sosyal mühendislik saldırıları Bölüm 4’te açıklanmıştır. Sosyal mühendislik saldırıları ile hedefin duyuu durumunda ani değışiklikler yaratılarak ya da mevcut olan açıklıklardan faydalanılarak (bir kiři ya da řirket olması fark etmeksizin) verilerine ulaşılması amaçlanmaktadır. Bilgi güvenliğinin sağlanabilmesi ise ancak bu saldırıların felsefesi ve yöntemlerinin anlaşılmasından geçmektedir. Bu nedenle çalışmada detaylıca sosyal mühendislik saldırılarından, yaşam döngüsünden, saldırı yöntemlerinden ve korunma yöntemlerinden bahsedilmiştir. Bölüm 5’te DÖ ve Bölüm 6’da DDİ yöntemlerinden bahsedilmiş ardından bu saldırıların tespitine yönelik çalışmalar incelenerek literatür taraması sunulmuştur. Çalışmanın son bölümü olan Bölüm 7’de tez çalışması kapsamında yapılan çalışmalara yer verilmiştir.

2. İLGİLİ ÇALIŞMALAR

Tez çalışmasının bu bölümünde sosyal mühendislik saldırılarının tespit yöntemlerine ilişkin literatürdeki mevcut çalışmalar, çalışmalarda kullanılan yöntemler, veri setleri ve çalışmaların sonucunda elde edilen başarımlar incelenmiştir.

Bustio-Martínez ve diğ. phishing URL'lerini tespit etmek için 52.000 veriden oluşan Lightweight veri setine Rastgele Orman (RO) yöntemini uygulamışlardır. Phishing tespitini gerçekleştirmek için kullandıkları yöntemde 0.995'lik bir başarımla elde etmişlerdir [2]. Bhardwaj ve diğ. kimlik avı saldırılarını tespit etmek ve azaltmak için kimlik avı taksonomisi önermişlerdir. Gmail, Yahoo ve son kullanıcıya ait 500'den fazla özel postallardan oluşan veri setine geliştirdikleri 3 aşamalı ortalama tespit yöntemini uygulamışlardır. Çalışmaları sonucunda 0.954'lük bir başarımla elde ettiklerini belirtmişlerdir [3]. Minocha ve Singh çalışmalarında kimlik avı tespiti için AV-şekil transfer fonksiyonu (AV-BMEO) ve k-en yakın komşu (KNN) sınıflandırıcı ile Binary Modified Equilibrium Optimizer (BMEO) kullanan yeni bir sistem önermişlerdir. Geliştirdikleri yöntemde iki ayrı veri seti kullanmışlardır. Bu veri setlerinden ilki 10.000 veriden oluşan Mendeley3 veri seti ve ikincisi 11.055 veriden oluşan UCI deposundan indirilen 18 (PenglungEW, Zoo, Lymphography, Wine, Sonar, SpectEW, HeartEW, Vote, Ionosphere, CongressEW, BreastEW, BreastCancer, Tic-tac-toe, Exactly2, Exactly, M_of_n, KrVsKpEW, WaveformEW) veri setidir. Eğitimler sonucunda 0.963 ve 0.974'lük başarımlar elde etmişlerdir. Önerdikleri yöntemin 17 algoritmaya kıyasla daha iyi sonuçlar verdiğini belirtmişlerdir [4]. Chiew ve diğ. internet kullanıcılarını kimlik avı saldırılarından korumak için bir web sitesinin gerçek kimliği ile tasvir edilen kimliği arasındaki tutarlılığı belirlemek için logo görüntüsünü kullanmayı önermişlerdir. Önerdikleri yöntemde web sitesinin logosunun çıkarılması ve kimlik doğrulamasının yapılması için iki aşama bulunmaktadır. MÖ yöntemleri ile doğru logo görüntüsünü çıkardıktan sonra Google görsel aramayı kullanarak logo ve alan adı arasındaki ilişki tespit edilerek tespit işlemi gerçekleştirilmiştir. Önerilen yöntem için 3894 veriden oluşan Veri Seti 1, 3000 veriden oluşan Veri Seti 2 ve 500 veriden oluşan Veri Seti 3 kullanılmıştır. Önerilen yöntem ile 0.998'lik bir başarımla elde edilmiştir [5]. Barraclough ve diğ. kimlik avı tespitini gerçekleştirmek için 20.000 veriden oluşan Phistank, 500 veriden oluşan Miilersmiles ve 10.000 veriden oluşan Relbanks veri setlerine ANFIS, Naive Bayes, PART, J48 ve JRIP yöntemlerini uygulamışlardır. Uygulanan yöntemlerin 0.990 ile 0.993 arasında başarımlara sahip olduğunu belirtmişlerdir. Ancak en iyi performansın 0,006 saniye hızla 0.993'lük başarımla PART yöntemine ait olduğunu saptamışlardır [6]. Azeez ve diğ. çalışmalarında kimlik avı saldırılarını tespit etmek için "White-list (Beyaz Liste)" yaklaşımı sunmuşlardır. Bilinen güvenilir sitenin benzerliklerini ve alan adını bir beyaz liste içerisinde bir araya getirmişlerdir. Ardından tespit yapılmadan önce İnternet Protokolü (IP) adresleri ile eşleştirme yaparak tespiti gerçekleştirmişlerdir. Deneysel çalışmalarını Phistank ve

Alexa veri setleri üzerinde gerçekleştirerek 0.967'lik bir başarıma ulaşmışlardır [7]. Alhogail ve Alsabih kimlik avı e-postalarını tespit etmek için DÖ yöntemlerini kullanarak DÖ algoritmalarını uygulayan grafik evrişimli bir ağ yöntemi (GCN) geliştirmişlerdir. Geliştirdikleri yöntemi 8579 veriden oluşan Fraud veri seti üzerinde eğittikten sonra 0.982'lik bir başarımla elde ettiklerini belirtmişlerdir [8]. Oram ve diğ. kimlik avcısı web sitelerinin URL'leri taklit eden özelliklerini kullanan makine tabanlı kimlik avı e-posta algılama modelini arttıran hafif gradyan yöntem önermişlerdir. LightGBM olarak isimlendirdikleri bu yöntemi geliştirmek için 5000 veriden oluşan Phishing veri setini kullanmışlardır. Çalışmaları sonucunda 0.969'luk bir başarımla elde etmişlerdir [9]. Yang ve diğ. çalışmalarında kimlik avının tespiti için ters olmayan matris çevrimiçi sıralı aşırı öğrenme makinesine (NIOSELM) dayalı yeni bir yaklaşım sunmuşlardır. Yaklaşımlarında 65.000 veriden oluşan Phishank veri setini kullanmış ve 0.969'luk bir başarımla elde etmişlerdir [10]. Wei Wei ve diğ. çalışmalarında evrişimli sinir ağını kullanarak kimlik avı tespitini gerçekleştirmişlerdir. Çalışmalarında 10.604 veriden oluşan Phishtank veri setini kullanmışlardır. Deneysel çalışmaları sonucunda 0.999'luk bir başarımla elde etmişlerdir [11].

Xia ve Chen çalışmalarında smishing tespiti için ayırık bir Gizli Markov Modeli (HMM) önermişlerdir. 5547 veriden oluşan UCI veri havuzunu kullanmışlardır. Çalışmaları sonucunda 0.969'luk başarımla elde etmişlerdir [12]. Akande ve diğ. çalışmalarında spam SMS'lerin nasıl tespit edileceğini ve amatör mesajlardan nasıl ayırt edileceğine odaklanmışlardır. SMS kimlik avı saldırılarını tespit etmek ve önlemek için kural tabanlı bir SMS hizmeti sunan mobil uygulama geliştirmişlerdir. Spam SMS'lerin tespitini gerçekleştirmek için 5547 veriden oluşan UCL SMS spam veri seti kullanılmıştır. Önerilen yöntemde model 0.979'luk bir başarımla sağlamıştır [13]. Waja ve diğ. çalışmalarında spam SMS'leri tespit etmek için çok kanallı bir Evrişimsel Sinir Ağı (CNN) mimarisi önermişlerdir. Önerilen modeli geliştirmek için 8751 veriden oluşan UCI SMS toplama veri seti ve kendi kendine toplanan mesajlar birleştirilmiştir. Çalışmalar sonucunda 0.961'lik bir başarımla elde edilmiştir [14]. Hameed ve Ali çalışmalarında SMS spam mesajlarını tespit etmek için ikili parçacık sürü optimizasyonuna dayalı bulanık kural seçimine odaklanan yeni bir yaklaşım önermişlerdir. Çalışmalarında kullandıkları 5574 veriden oluşan UCI veri setinden ilk olarak önemli özellikleri çıkarmışlardır. Daha sonra çıkardıkları özelliklere dayalı bir dizi bulanık kurallar oluşturmuşlardır. Ardından karmaşıklığı azaltan ve modelin performansını iyileştiren bulanık kuralları seçmek için ikili parçacık sürüsünü kullanmışlardır. Önerilen modelden 0.985'lik bir başarımla elde etmişlerdir [15]. Saeed çalışmasında spam SMS'leri filtrelemek için J48, KNN ve Karar Ağacı (KA) makine öğrenimi yöntemlerini kullanmıştır. 5574 veriden oluşan UCI veri seti üzerinde modelini eğitmiştir. Deneysel sonuçlarında KNN algoritması için 0.911'lik ve KA algoritması için 0.970'lik bir başarımla elde etmiştir [16]. Hossain ve Sen çalışmalarında öz dikkat mekanizmasını kullanarak SMS spam tespiti gerçekleştirmişlerdir. Önerdikleri yaklaşımda önceden eğitilmiş Bidirectional Encoder Representations from Transformers (BERT) yöntemini ve

ek katmanları kullanmışlardır. Yöntemi 5572 veriden oluşan Kaggle “SMS Spam collection” veri seti ile eğitmişlerdir. Eğitim sonucunda 0.988’lik bir başarımla elde etmişlerdir [17]. Ojugo ve Oyemade çalışmalarında SMS’leri meşru ve istenmeyene göre sınıflandırmak için hibrit bir spam filtreleme tekniği üzerinde DÖ topluluğu olarak kullanılan bir dizi eşleştirme algoritması önermişlerdir. Çalışmada ağ yapısını seçmek ve eğitmek için gizlenmiş Markov ağı kullanılmıştır. Modeli 5574 veriden oluşan UCI veri seti üzerinde eğitmişler ve 0.970’lik bir başarımla elde etmişlerdir [18]. Ghourabi ve diğ. çalışmalarında Arapça ve İngilizce mesajlar üzerinde spam SMS’lerin tespitini gerçekleştirebilmek için Hibrit CNN - LSTM yöntemini önermişlerdir. Modeli eğitmek için Suudi Arabistan’daki birkaç yerel akıllı telefonda toplanan mesaj içeriklerini ve İngilizce mesajlar olmak üzere 8304 veriden oluşan veri setini kullanmışlardır. Modelden 0.981’lik başarımla elde etmişlerdir [19]. Ojugo ve Eboka çalışmalarında gürültülü başarımları normalleştirmeyi, altta yatan öğrenme sezgisellerini eğitmek için kelime anlamı belirsizliğini giderme tekniğini kullanan sözlük bilimsel ve semantik sözlükleri kullanarak metni genişletmeyi amaçlayan Genetik Algoritma Eğitimli Bayes Ağını önermişlerdir. Önerilen bu hibrit model metin ön işleme, özellik seçimi, eğitim ve sınıflandırma aşamalarından oluşmuştur. Çalışmalarında özellik seçimini gerçekleştirmek için genetik algoritmanın kullanıldığı bir hibrit bir genetik algoritma ile eğitilmiş bayes modeli kullanılmıştır. Önerilen modeli eğitmek için KDD-CUP ’99 veri setini kullanmışlardır [20]. Srinivasarao ve Sharaff çalışmalarında spam SMS’lerin tespitini gerçekleştirmek için Makine Zekâsı Tabanlı Hibrit Sınıflandırıcı yöntemini önermişlerdir. Modeli eğitmek için üç ayrı veri seti kullanmışlardır. İlk olarak 5574 veriden oluşan UCI’nin SMS veri seti, ikinci olarak 5172 veriden oluşan Email veri setini ve son olarak 3252 veriden oluşan Spam Assassin veri setini kullanmışlardır. Çalışmaları sonucunda önerdikleri hibrit yöntemden 0.998’lik başarımla elde etmişlerdir [21].

Basiri ve diğ. çalışmalarında duygu analizini gerçekleştirmek için Dikkat Tabanlı Çift Yönlü CNN - RNN Derin Modeli (ABCDM) yöntemini önermişlerdir. Modeli eğitmek için App, Kindle, Movies, Electronics, CDs, Airline Twitter, Sentiment140 ve T4SA veri setlerini kullanmışlardır. Veri setlerinden 0.818 ile 0.968 arasında doğruluk elde etmişlerdir. En iyi performansı 0.968’lik doğrulukla T4SA veri setinden elde etmişlerdir [22]. Lopez-Vizcaino ve diğ. çalışmalarında sosyal medya ağlarında siber zorbalık tespitini erken gerçekleştirmek için eşik ve ikili olarak adlandırılan iki farklı spesifik erken tespit modeli ile denetimli bir öğrenme yöntemi önermişlerdir. İlk basit bir yaklaşım izlerken ikincisinin iki makine öğrenimi modeli gerektirdiğini belirtmişlerdir. Vine sosyal medya ağından toplanan verilerden oluşan veri setini kullanmışlar ve 0.950’lik bir başarımla elde etmişlerdir [23]. Chia ve diğ. çalışmalarında MÖ ve Öznitelik Mühendisliği tekniklerini kullanarak Twitter’daki alay ve ironiyi keşfetmeyi amaçlamışlardır. Geliştirdikleri MÖ ve Öznitelik Mühendisliği Tabanlı İğneleme çalışması için Semantik Evaluation 2017 Task 3 isimli İngilizce Tweetlerde İroni Tespiti veri seti kullanılmıştır. Çalışmalarında 0.896’lık bir başarımla elde

etmişlerdir [24]. Eronen ve diğ. siber zorbalık sorununu çözebilmek için Özellik Yoğunluğu yöntemini önermişlerdir. Önerilen yöntemi English Cyberbullying veri seti, Japanese Cyberbullying veri seti ve Ploish Cyberbullying veri setleri üzerinde geliştirmişlerdir [25]. Ayo ve diğ. çalışmalarında Twitter’da nefret söylemi sınıflandırması için olasılıksal bir kümeleme modeli önermişlerdir. Önerilen yöntemi eğitmek için Twitter Akış API’si ile toplanan veri setini kullanmışlardır. Önerdikleri yaklaşım ile 0.925’lik başarımlar elde etmişlerdir [26]. Bozyiğit ve diğ. çalışmalarında siber zorbalık tespitinde sosyal medya özniteliklerinin önemini ortaya koymayı amaçlamışlardır. Birçok sosyal medya özelliğine sahip Turkish veri setini hazırlayıp kullanmışlardır. Veri setine AdaBoost yöntemi uygulanarak 0.901’lik başarımlar elde etmişlerdir [27]. Rajesh ve Sharanya çalışmalarında sosyal medyada siber tacizin tanınması ve önlenmesi için MÖ sınıflandırma algoritmalarını kullanmışlardır. Irkçılık ve cinsiyetçilik içeren yaklaşık 17.000 Tweet kimliği içeren sosyal medya veri setine MÖ yöntemlerini uygulayarak 0.928’lik başarımlar elde etmişlerdir [28]. Yazgılı ve Baykara çalışmalarında Türkçe metinlerde MÖ yöntemlerini kullanarak siber zorbalık tespiti gerçekleştirmişlerdir. Çalışmalarında Kaggle’da yer alan veri setine Bagging, Boosting, C4.5, GBM, K-Means, KNN, Lojistik Regresyon (LR), Naive Bayes (NB), Yapay Sinir Ağı (YSA), RO, Destek Vektör Makineleri (DVM), Stokastik Gradyan İnişi (SGİ) ve Aşırı Gradyan Artırma (XGBoost) yöntemlerini uygulamışlardır. Veri setlerinden 0.510 ile 0.883 arasında başarımlar elde etmişlerdir. En yüksek başarımları LR yöntemi ile elde etmişlerdir [29]. Toğaçar çalışmasında web sitelerinde gerçekleştirilen ortalama saldırılarının tespiti için RO, DVM, KNN Algoritması ve KA’yı kullanmıştır. Çalışmasında 11.055 veriden oluşan web sitesinin içeriği veri seti olarak kullanılmıştır. RO algoritması 0.965, DVM 0.932, KNN 0.931 ve KA 0.961’lik başarımlar sağlamıştır [30]. Toğaçar ve diğ. çalışmalarında internet ortamında yayınlanan sahte haberlerin tespitini gerçekleştirmek için Yapay Zekâ Tabanlı DDİ yaklaşımını önermişlerdir. Çalışmalarında 6335 haber başlığı ve içeriğinden oluşan real_or_fake isimli veri setini kullanmışlardır. Önerdikleri yöntemden 0.914’lük bir başarımlar elde etmişlerdir [31].

Yüce, Sabuncu ve Atay tez çalışmalarında siber zorbalık içeren tweetlerin yapay zekâ yardımıyla tespit edilip bu tarz tweetleri paylaşan hesapların otomatik olarak tespitini gerçekleştiren sistem geliştirmeyi amaçlamışlardır. Çalışmalarında kamu ve siyaset, kültür-sanat, basın-yayın, spor, iş dünyası ve fenomenlerden belirledikleri yüksek takipçili hesaplara sahip 378 kişiye ait tweetleri RapidMiner Studio aracılığıyla 3 ayrı parçada çekerek veri seti hazırlamışlardır. Veri setinin yaklaşık olarak 30.000 veriden oluştuğunu ve yapılan ön işlemler sonrasında çalışmada bu verilerin 29.198’ini kullandıklarını belirtmişlerdir. Veri setini manuel olarak etiketleyip ardından bu veri seti üzerinde NB, Genelleştirilmiş Doğrusal Model, DÖ Modeli, KA, RO Algoritması, Gradient Boosted Trees Algoritması ve DVM yöntemlerini uygulamışlardır. Uyguladıkları yöntemlerden 0.807 ile 0.911 arasında başarımlar elde etmişlerdir. En yüksek doğruluk başarımlarını 0.911 ile DVM’den elde ettiklerini belirtmişlerdir [32]. Avaroğlu ve Nergiz

yüksek lisans tez çalışmalarında sosyal medya platformlarında yer alan Türkçe yorumların siber zorbalık davranışları içerip içermediğini DÖ yöntemlerini kullanarak tespit etmeyi amaçlamışlardır. Veri seti oluşturmak için Twitter'dan 84.356, Instagram'dan 42.680 ve Youtube'dan 52.964 yorumu bir araya getirerek toplamda 180.000 adet yorumdan oluşan verileri manuel olarak etiketleme işlemini yapmışlardır. Veri setine Word2Vec, Fasttext ve Doc2Vec kelime gömme yöntemleri ile geliştirdikleri 3 farklı modeli uygulamışlardır. Çalışmaları sonucunda Word2Vec ile oluşturdukları modelden 0.913, Fasttext ile oluşturdukları modelden 0.931 ve Doc2Vec ile oluşturdukları modelden 0.895 oranında başarımlar elde ettiklerini belirtmişlerdir [33]. Yüksel yüksek lisans tez çalışmasında Nefret Söylemi tespitini gerçekleştirmek için hedef odaklı dilsel özelliklerle desteklenen BERT tabanlı bir model geliştirmiştir. Veri seti olarak çeşitli kaynaklardan Türk medyasında yer alan basılı haberleri kullanmıştır. Oluşturduğu veri setine geliştirdiği RuleBert yöntemini uygulayarak 0.913'lük bir başarımlar elde etmiştir [34]. Diri, Şahingöz ve Korkmaz doktora tezi çalışmalarında ortalama saldırılarının tespiti için "İki aşamalı hibrit ortalama saldırısı tespit sistemi" adını verdikleri yöntem ile hem URL hem de içerik analizi tespiti yapan sistem geliştirmişlerdir. Çalışmada 2021 yılında PhishTank internet sitesi verilerinden oluşturulmuş YRU veri setini, 2022 yılında PhishTank, DMOZ ve Open PageRank olmak üzere 3 farklı kaynaktan elde edilmiş X_URL veri setini ve içerik analizini gerçekleştirmek için YRİ veri setlerini kullanmışlardır. Veri setleri toplamda 14.782,355 adet veriden oluşmaktadır. URL tabanlı derin öğrenme modeli olarak geliştirdikleri GDNN - CNN modelini eğitmek için X_URL ve YRU veri setlerini kullanmışlardır. Bu modeli X_URL veri setinde test etmiş ve URL tespiti için 0.976'lık başarımlar elde etmişlerdir. Aynı şekilde X_URL+YRU veri setini kullanarak içerik analizini tespit etmek için Deep Neural Network (DNN) - CNN modelini eğitmişlerdir. Bu modelde ise yine X_URL veri seti üzerinde test yaparak 0.973'lük başarımlar elde etmişlerdir. URL ve içerik analizlerini YRİ veri seti üzerinde test ederek GDNN - CNN modeli için 0.960 ve DNN - CNN modeli için 0.967'lik başarımlar elde etmişlerdir [35]. Bulut, Onan ve Eğin geliştirdikleri yüksek lisans tez çalışmasında sosyal ağlar üzerinde mültecilere yönelik nefret söylemini tespit etmeyi amaçlamışlardır. Çalışmalarında Twitter'da yer alan içeriklerden "mülteci, mülteciler, göçmenler" etiketlerine sahip paylaşımları SNScrape aracılığıyla çekerek veri seti oluşturmuşlardır. Toplamda 9778 tweetten oluşan veri setine LR, DVM, KA, RO Algoritması ve YSA modellerini ve BERTurk yöntemini uygulamışlardır. Çalışmaları sonucunda LR 0.69, DVM 0.63, KA 0.65, RO 0.67 ve YSA 0.67'lik başarımlara sahip olmuştur. Tekrar bu algoritmaları TF-IDF kelime gömme yöntemi ile kullanarak LR 0.81, DVM 0.81, KA 0.74, Ro 0.81 ve Ysa 0.78'lik başarımlar elde edilmiştir. Ayrıca veri setine farklı öğrenme oranları kullanılarak BERTurk yöntemini de uygulamışlardır. BERTurk modelinden "3e-5" öğrenme oranıyla 0.86'lık başarımlar elde ettiklerini belirtmişlerdir [36]. Albayrak ve Alarbi doktora tezi çalışmalarında phishing URL'lerinin tespitini gerçekleştirmek için KNN algoritmasından türetilen ve denetimsiz K-means

algoritması ile hibridize edilen çekirdek sınıflandırma algoritması (CCA) adı verilen yeni bir yöntem geliştirmişlerdir. Ardından geliştirilmiş çekirdek sınıflandırma algoritmasını (ICCA) tanıtmışlardır. Çalışmalarında Mendeley ve Kaggle dahil olmak üzere çeşitli platformlardan alınmış 549.346 veriden oluşan veri setini kullanmışlardır. Çalışmaları sonucunda geliştirdikleri CCA yönteminden 0.701 ve ICCA yönteminden ise 0.757'lik başarımlar elde ettiklerini belirtmişlerdir [37].



3. BİLGİ GÜVENLİĞİ

Bilgi güvenliği, her türden tüzel ve gerçek kişinin bilgilerini gizlilik, kullanılabilirlik ve bütünlükleri açısından sürekli olarak korunmasını sağlamaktadır [38]. Özellikle şirketler açısından bilgi güvenliğinin sağlanması büyük önem taşımaktadır.

Teknolojinin yaygınlaşması ile birlikte insan yaşamında meydana gelen kolaylıklar bazı şartların yerine getirilmesiyle sağlanmaktadır. Bu şartlar herhangi bir sipariş verdiğimizde ya da internet üzerinden herhangi bir başvuruda bulunduğumuzda kendimize ait verileri paylaşmamızdan ibarettir. Söz konusu verileri bünyesinde tutmak ve bu verilerin güvenliğini sağlamak ise paylaşım sağlanan şirketlerin sorumluluğundadır. Bu sorumluluğun yerine getirilmemesi ya da ihlali şirketlerin hukuki ve cezai sorumluluklarını gündeme getirmektedir. Bu nedenle şirketler veri güvenliğini sağlamak konusunda oldukça hassas hareket etmek zorundadır.

Bilgi güvenliği ile ağ sistemlerinden teknik sistemlere kadar elde edilen bütün verilerin korunması sağlanmaktadır. Temel olarak bilgi güvenliği sağlanırken verilerin sınıflandırılması yapılmaktadır. Bu veri sınıflandırması, ifşası halinde şirketi/ kişileri en az zarara uğratacak veriden en çok zarara uğratacak veriye doğru yapılır. Bunun sonucunda en önemli veriler ayıklanmış olur. Geniş çerçevede ise bilgi güvenliği ile çalışanların ya da veriyle ilgili kişilerin söz konusu verilere erişimi sağlanırken yetkisiz kişilerin erişimi engellenmektedir. Bilgi güvenliği unsurlarına ait görsel Şekil 3.1’de verilmektedir.



Şekil 3.1. Bilgi güvenliğinin unsurları

3.1. Bilgi Güvenliđi İlkeleri

Bilgi güvenliđi gizlilik, bütünlük ve kullanılabilirlik olmak üzere üç temel ilkeden oluşmaktadır [39]. Bu üç temel ilkenin herhangi birinin eksik olması bilgi güvenliđinin tam olarak sağlanamadığını göstermektedir.

3.1.1. Gizlilik

Gizlilik elde edilen veriye erişim izni olmayan kişilerin erişememesini ve bu verilerin ifşa edilememesini ifade etmektedir. Gizlilik, yetkisiz kişinin erişimini engellerken yetkili kişinin de bu veriye kolaylıkla erişebilmesini çift yönlü olarak sağlayan bir ilkedir. Kimi zaman bilgi güvenliđi sağlanırken gizlilik için ek önlemler de alınabilmektedir. Ek önlemler çift adımlı doğrulamalar, güvenlik sorusu doğrulamaları ya da kimlik doğrulamaları olarak çeşitli şekillerde karşımıza çıkmaktadır [39].

3.1.2. Bütünlük

Bütünlük ilkesi verinin değıştirilmesini ya da silinmesini engellemeyi ifade etmektedir. Temel olarak verinin ilk hali ile korunmasını esas almaktadır. Verinin bütünlüğünü sağlamak için çeşitli yöntemler geliştirilmiştir. Bu yöntemlerle veri üzerinde değışiklik yapılmasına izin verilmemekte, değışiklik yapılması halinde ise veri sahibinin bu değışiklikten haberdar olması garanti edilmektedir [39]. Bu kapsamda avukatların kullandığı e-imza protokolleri güncel bir örnektir. Bu protokolda birden fazla e-imza ile imzalanması gereken bir evrakta veri sahibi olan avukattan habersiz olarak gerçekleştirilen değışiklikler ya da veri üzerinden yapılan güncellemeler olması halinde protokol geređi veri sahibi avukatın e-imzası silinerek verinin ilk halinin değıştirildiđi bildirilmektedir. Yine bu kapsamda veride üçüncü şahıslar tarafından yapılan değışiklikler e-imzayı geçersiz kılarak avukata ait olan e-imza verisinin bütünlüğünü sağlamaktadır.

3.1.3. Kullanılabilirlik

Kullanılabilirlik ilkesi veriye erişim izni olanların istedikleri her anda veriye erişim sağlayabilmelerini ifade etmektedir. Aynı zamanda verinin kaybolmasını ya da erişimde herhangi bir sorun yaşanmamasını amaçlamaktadır. Bu amaç doğrultusunda verinin kopyalarını ana depo haricinde başka bir yerde daha yedeklenmesinin ya da geliştirilen sistemler sayesinde herhangi bir saldırıya karşı söz konusu verilere erişimin kesintiye uğramaması için gerekli tedbirlerin alınmasını kapsamaktadır [39].

3.2. Bilgi Güvenliğinin Önemi

Kişiler ve şirketler bünyelerinde bulundurdıkları verileri uygun şekilde korumakla yükümlüdürler. Söz konusu yükümlülük piramidin en başında devletlerin de üzerindedir. Özellikle devlet güvenliği ve günümüzdeki casuslukların teknoloji ile paralel gelişmesi dijital olarak saklanan verileri açık bir hedef haline getirmekte ve korunmalarını devlet güvenliği seviyesinde önem arz etmektedir.

Verilerin gerektiği gibi korunamaması birden çok soruna yol açabilmektedir. En temelde verinin uygun şekilde korunmaması gizli bilgilerin açığa çıkmasını sağlamaktadır. Ortaya çıkan gizli bilginin bir kişiye ait olması halinde kişisel hakların ihlalinin, şirkete ait olması halinde itibar kaybı ve ekonomik zarardan, devletlere ait olması halinde ise güvenlik tehlikesinden bahsedilebilecektir. Görüldüğü üzere bilgi güvenliğinin sağlanması toplumun en küçük yapıtaşı olan bireylerden devlete kadar herkes için önem taşımaktadır. Meydana gelen zararların yanı sıra açığa çıkan gizli bilgiler yetkisiz kişilerin erişimini de kolaylaştırmakla birlikte veriye erişimi de imkânsız hale getirebilmektedir [40].

Kullanıcı hataları ya da sosyal mühendislik saldırıları neticesinde bilgi güvenliğinin sağlanamaması halinde birden çok zarar meydana gelebilir. Bunlardan başlıcaları:

- 1- Kişilere ve şirketlere ait gizli ya da hassas bilgiler
- 2- Şirketlerin iş sürekliliği
- 3- Şirketlerin itibarı, ekonomik durumları
- 4- Ticari sırlar, adli veriler olarak sıralanabilmektedir.

Söz konusu başlıklar ise şirketlerin ya da kişilerin ciddi zarara uğramasına yol açmaktadır. Geniş bakış açısıyla bu, devletlerin güvenlik politikalarının, devlet sırrı niteliğindeki bilgilerin açığa çıkması anlamına gelecektir. Bu kapsamda teknolojik devirde devletlerin güvenlikleri de bilgi güvenliğinin tam ve eksiksiz sağlanmasından geçmektedir.

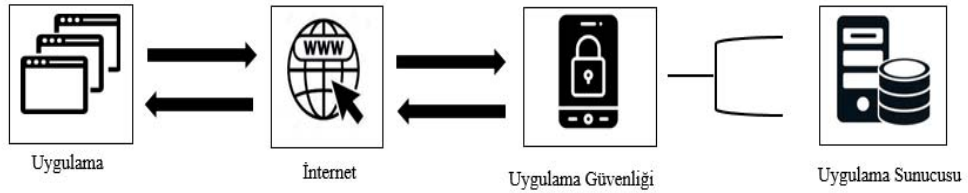
Yukarıda bahsedildiği üzere bilgi güvenliğinin sağlanması oldukça önem taşımakla birlikte bilgi güvenliğinin sağlanmasından ise belirli gruplar sorumludur. Hiç şüphesiz bu sorumluluk veri sahibine aittir. Veri sahibi, söz konusu veriyi kimlerle ve ne şartlar altında paylaştığına dikkat etmelidir. Örneğin henüz tanımadığı kişilerle ya da güvenilirliğinden ve bilinirliğinden emin olmadığı kuruluşlarla verilerini paylaşmaması alınabilecek en temel tedbirdir. Ancak söz konusu verinin paylaşılması halinde ise bu sorumluluk artık birden fazla grubun üzerinde olmaktadır. Bu aşamadan sonra artık sorumluluk hem veri sahibinde hem de söz konusu veriyi bünyesine alarak onu kullanacak olan kişi ya da şirkettir. Veriyi bünyesine katan eğer bir şirket ise bilginin güvenliğinden veri sistemini yöneten gerçek kişilerin de sorumluluğu meydana gelecektir. Esasen söz konusu durumda veri güvenliğinden önemi ile doğru orantılı olarak veriye erişim sağlayan herkesin sorumlu olduğu söylenebilecektir.

3.3. Bilgi Güvenliđi Türleri

Bilgi güvenliđi verilerin izinsiz kullanılmasını, deđiştirilmesini ya da kaybedilmesini engellemek gibi her türlü tehdit karşısında verinin güvenliđini sađlamayı amaçlamaktadır. Bu nedenle tehdit türlerine ve korunma yöntemlerine bađlı olarak birden fazla bilgi güvenliđi çeşidi bulunmaktadır. Bu kısımda bilgi güvenliđi türleri açıklanmıştır.

3.3.1. Uygulama Güvenliđi

Kullanıcılar tarafından akıllı telefonlarda, tablet ve bilgisayarlarda birden çok uygulama kullanılmaktadır. Bu uygulamaları tam anlamıyla etkin kullanmak için ise bazı veri erişimlerine yine kullanıcılar tarafından izin verilmektedir. Bu kapsamda ortaya çıkan en büyük sorun uygulamanın herhangi bir güvenlik açığına sahip olması halinde veri güvenliđinin tehlikeye düşme ihtimalidir. Uygulamanın bir güvenlik açığına sahip olması ise saldırganın o açıktan içeriye sızarak verilerin ele geçirilmesi anlamına gelmektedir. Bu nedenle uygulama kullanımı sırasında erişimine izin verilen verilerin güvenliđinin sađlanması önem taşımaktadır. Uygulama güvenliđi ise hata ve özellikle izinsiz girişleri durdurmak, tespit etmek için uygulama ve API'leri korumaya odaklanmaktadır [39]. Uygulama güvenliđi söz konusu korumayı sađlarken çeşitli araçlar kullanılmaktadır. Bu araçlar uygulama koruma, test ve tarama için özel olarak geliştirilmiştir. Söz konusu araçlar kullanılarak uygulamalardaki güvenlik açıklıkları, izinsiz girişler tespit edilerek öncelikle uygulama geliştiricilerinin güvenlik açıklıklarından haberdar olmalarını sađlamaktadır. Yine aynı şekilde izinsiz girişlerin de tespiti sađlanarak engellenmesi amaçlanmaktadır. Bu kapsamda uygulama güvenliđi için sürekli olarak uygulama açıklarını bulmak izlemek ve azaltmak için güvenlik duvarları ve tarayıcılar başlıca kullanılan tekniklerdendir. Uygulama güvenliđine ait görsel Şekil 3.2'de verilmektedir.



Şekil 3.2. Uygulama güvenliđi

3.3.2. Altyapı Güvenliği

Altyapı; bilgisayarları, bulut sistemleri ve yazılımları içeren bileşenleri ifade etmektedir. Teknolojinin gelişimiyle orantılı olarak ortaya çıkan güvenlik açıkları kişileri ve özellikle şirketleri öncelikle altyapı sistemlerini korumaya yönlendirmiştir. Altyapı, dış dünyada ve dijital dünyada varlığını aynı anda sürdürdüğü için hem fiziksel hem de siber saldırılardan oldukça kolay bir şekilde etkilenebilmektedir. Bu nedenle altyapı güvenliğini sağlamak daha kapsamlı, geleneksel ve modern yöntemlerin bir arada kullanılmasını gerektirmektedir.

Altyapı güvenliğine yönelik birden çok tehdit bulunmaktadır. Bunlar, dağıtılmış hizmet reddi (DDoS), sosyal mühendislik saldırıları ve siber saldırılar gibi dijital tehditler olabileceği gibi hırsızlık, yangın, deprem ve sel gibi doğal ve dış dünyaya ilişkin fiziksel tehditler de olabilmektedir. Birden çok ve birbirinden farklı potansiyel tehditlere maruz kalan altyapı güvenliği alanı özellikle fikri mülkiyet ve rekabet hususları açısından verilerin hayati önem taşıması nedeniyle şirketler nezdinde korunması önem taşıyan bir alandır. Bu nedenle şirketler altyapı güvenliği hususunda oldukça hassas davranmaktadırlar. Ancak altyapı güvenliği, kişisel verilerin hızla dijitalleştiği dünyada bireyler açısından da önem taşımaktadır. Bu kapsamda gerekli tedbirlerin alınmaması halinde şirketlerin kurumsal ve kamusal itibarlarının zarar görmesinin yanında kullanıcıların da kişisel verilerinin ifşa olması gibi sorunlar gündeme gelecektir.

Altyapı güvenliğini, dolayısıyla şirketlerin itibarlarını ve kullanıcıların kişisel verilerini korumanın birden çok yolu vardır. Bu yöntemler geleneksel olabileceği gibi dijital/ teknolojik yöntemler de olabilmektedir. Uzmanlar, tam bir korumanın sağlanabilmesi her ne kadar zor olsa da hem geleneksel hem de modern korunma yöntemlerinin bir arada kullanılması halinde maksimum korunma düzeyine ulaşılabileceğinden bahsetmektedir. Geleneksel olarak altyapıyı oluşturan bilgisayar, hard disk gibi donanımların korunması için en yaygın yöntemler erişim kontrolü, güvenlik kamerası, güvenlik görevlisi ve çalışanların bu alanda bilinçli hareket etmesini sağlayacak eğitimlerin uygulanması şeklinde karşımıza çıkmaktadır. Ancak altyapıyı oluşturan bulut sistemler, yazılımlar ve uygulamalar için modern yöntemlerle korunmaya da ihtiyaç bulunmaktadır. Bu kapsamda da modern korunma yöntemleri olarak karşımıza sızma testleri, ağ izleme ve şifreleme teknolojileri çıkmaktadır.

3.3.3. Bulut Güvenliği

Bulut bilişim sistemleri temelde verinin fiziksel olarak değil dijital olarak saklanması, bu nedenle de veriye ulaşmak için altyapı maliyetini azaltma fikrine dayanmaktadır. Ancak bir verinin dijital olarak saklanması, kolay erişim sağlamak ve altyapı maliyetini azaltmak konusunda ne kadar avantajlı görünse de beraberinde ciddi güvenlik sorunlarını getirmektedir. Özellikle büyük verileri depolayan şirketler bakımından müşteri memnuniyetinin ve müşterinin şirkete güven duymasının

önemi göz önüne alındığı zaman kullanımı hızla artan bulut bilişim sistemlerinin güvenliği de o kadar önem taşımaktadır. Çünkü temelde her şirketin itibarı, kendisine iletilen verilerin güvenliğini ne kadar sağlayabildiğinden geçmektedir.

Bulut güvenliği bulut bilişim sistemlerinin güvenliğinin sağlanmasını amaçlamaktadır. Bu kapsamda bulut güvenliği bulut ortamlarını, bulutta tutulan verileri ve uygulamaları koruyan teknoloji, protokol ve güvenlik uygulamalarından oluşan bir bütün niteliğindedir. Bulut güvenliği ile en temelde veriler korunmakta olup marka itibarı, kişisel hakların korunmasından söz edilebilmektedir. Her ne türde bulut bilişim sistemi korunursa korunsun temel olarak bulut güvenliği ile fiziksel ağlar, veri depolama, veri sunucuları, bilgisayar sanallaştırma çerçeveleri, işletim sistemleri, ara yazılım, çalışma zamanı ortamları, veri, uygulamalar ve son kullanıcı donanımları korunmaktadır. Bahsi geçen hususların korunması verilerin ve dolayısıyla bulut bilişim sistemlerinin korunmasını sağlamaktadır.

Yukarıda bahsedilen hususların korunması yani bulut bilişim sistemlerinin korunmasının özünde yatan fikir herhangi bir veri kaybı halinde veriyi kurtarmaya, saldırıdan önce veriyi korumaya, saldırı halinde saldırının etkisini azaltmaya ve saldırı ihtimalini azaltmak için caydırıcılığa dayanmaktadır. Bu nedenle bulut güvenliğini sağlamak birden çok yönüyle ele alınması gereken bir husustur. Bulut güvenliğinin sağlanmasında hiç şüphesiz en önemli olan ve en öne çıkan husus saldırıyı henüz başlamadan bertaraf etmektir. Bu nedenle bulut bilişim sistemlerini korumak için güvenlik duvarını güçlendirme, birden fazla aşamalı kimlik doğrulama sistemlerini kullanma, bulut sistemine başka bir cihazdan erişim sağlanması halinde bilgilendirme maili gönderilerek erişimin engellenmesini sağlama yöntemleri kullanılmaktadır. Ancak her nasılsa saldırının meydana gelmesi halinde kaybedilen ya da zarara uğrayan verilerin kurtarılması önem kazanmaktadır. Bu nedenle bulut bilişim sistemleri kurtarma protokolleri başta olmak üzere verinin en son haline erişim sağlayarak veriyi kurtarmak için yöntemler kullanılmaktadır. Ancak yine de bulut güvenliğini sağlamak dijital önlemleri almak kadar veri sahipleri olan kullanıcıların ve çalışanların bilinçli hareket etmesine teknik ve davranışsal açıdan kullanıcı güvenliği eğitimlerine de bağlıdır.

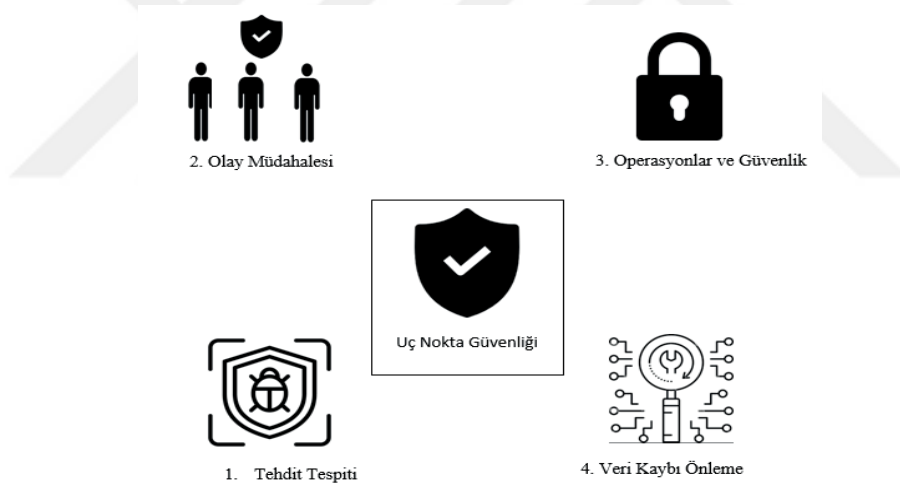
3.3.4. Uç Nokta Güvenliği

Uç nokta, kurumsal ağlara güvenlik duvarının dışından bağlanan cihazları ifade etmektedir. Bu cihazlar, dizüstü bilgisayarlar, masaüstü bilgisayarlar, akıllı telefonlar, anahtarlar, yazıcılar ve diğer iletişim cihazlarıdır. Uç nokta güvenliği de en basit tabiriyle bu cihazların korunmasını ifade etmektedir.

2020 yılının başlarından itibaren bütün dünyanın hayatına pandeminin girmesiyle birlikte temasın azaltılması için şirketler çalışanlarıyla yeni bir çalışma şekli olarak uzaktan çalışma

yöntemini denedi. Bu nedenle yaygınlık kazanan uzaktan çalışma yöntemi ise verilerin güvenliği açısından yeni bir tartışma alanı yarattı: kurumsal ağlara güvenlik duvarı dışından bağlanan cihazlar. Bu çalışma şeklinde verilerin saldırıya daha açık hale gelmesi, şirketlerin sağladığı güvenlik imkânlarının bireysel olarak çalışanlar tarafından karşılanamayacak olması temel bir çözüm fikri olan dışarıdan bağlantı sağlayan cihazların korunmasını gündeme getirdi. Bu nedenle uç nokta güvenliği her zaman önemli olmakla birlikte pandeminin başlamasıyla birlikte daha fazla tartışılan ve daha görünür hale gelen bir alan oldu.

Uç nokta güvenliği, her türlü uç noktayı siber tehditlerden korumak için kullanılan ve merkezi olarak yönetilen güvenlik çözümlerinden oluşmaktadır. Bu çözümler, yöneticilerin çalışanların çalışma şekillerini izlemek, yönetmek ve müdahale etmek için bulut veya hibrit yöntemlerden yararlanan konsollar sunmaktadır. Bu konsollar aracılığıyla yöneticiler güvenlik duvarı dışındaki uç noktaların faaliyetlerini izleyebilmekte bu faaliyetleri yönetebilmekte ve gerek gördüğü takdirde müdahalede bulunabilmektedir. Yukarıda bahsedilen uç nokta güvenliğine ait yapı Şekil 3.3'te verilmektedir.



Şekil 3.3. Uç nokta güvenliği

3.3.5. Kriptografi

Kriptografi ya da şifreleme, bir verinin içerisinde bulunan bilginin üçüncü kişiler tarafından okunabilmesinin engellenmesi için veriyi şifreleyen ve üçüncü kişiler tarafından anlaşılamayacak hale dönüştüren bir yöntemdir [39]. Kriptografi yöntemi simetrik ve asimetrik olarak iki farklı yöntemden oluşmaktadır. Simetrik yöntemde veriyi şifreleyen anahtar ile veriyi şifresiz haline geri döndüren anahtar aynı iken asimetrik yöntemde genel ve özel anahtarlar bulunmaktadır. Asimetrik yöntemde genel anahtarlar herkese kolaylıkla dağıtılabilirken verinin şifresini çözen özel anahtarlar ise sadece yetkili kişilerde bulunmaktadır [41].

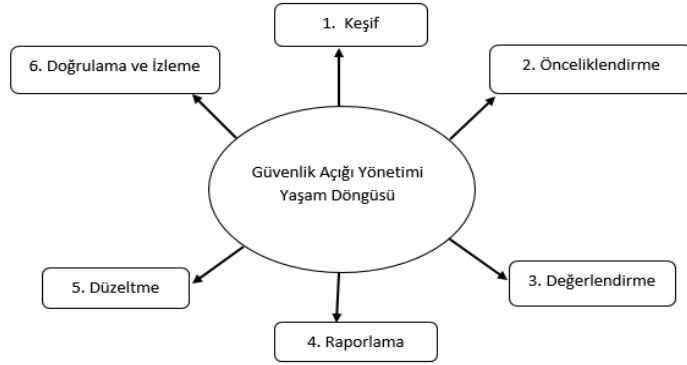
Üçüncü kişilerin veriyi okuyamayacağı şekilde bir kriptografi oluşturacak kodun dört parametresi vardır. Bunlar gizlilik, doğrulama, inkar edilememe ve kimlik doğrulamadır. Gizlilik parametresi kriptografinin en temel parametresi olup verilere sadece belirlenen kişilerin erişmesini üçüncü kişilerin erişiminin engellenmesini sağlamaktadır. Eğer bu parametre gerektiği gibi kullanılmazsa şifreli bir veriden bahsedilemeyecektir. Doğrulama parametresi ise verinin gönderilme ve görüntüleme aşamasında değiştirilmesini içermektedir. İnkâr edilememe parametresi ise göndericinin gönderdiği bilgiyi inkar edememesini içermekte olup kimlik doğrulama parametresi ise gönderici ile alıcının kimliklerini ve verinin olduğu yeri doğrulamaktadır. Bu dört parametre birlikte ve eksiksiz çalıştığı zaman şifrelenmiş bir veriden bahsedilebilecektir.

3.3.6. Güvenlik Açığı Yönetimi

Güvenlik açığı yönetimi sistemlerdeki ve yazılımlardaki güvenlik açıklıklarını bularak onları değerlendirme, raporlama ve iyileştirmeye yönelik süreklilik içeren bir süreci ifade etmektedir. Bu yönetim ile amaçlanan şey kritik öneme sahip iş süreçlerinde araçların kullanımının siber saldırılardan korunarak etkinliği kaybetmesinin önüne geçmektir.

Güvenlik açığı yönetimi bir şirketin çalışması için araçların etkinliği korumakla birlikte birden çok bileşenden oluşmaktadır. Bu bileşenler güvenlik açığı taraması, güvenlik açığı değerlendirmesi, yama yönetimi ve güvenlik açığı düzeltme olmak üzere dört adettir. Her bir bileşenin yönetim aşamasında belirlenen sıklıklarla kullanılması gerekmekte olup bu yönetim için mühendislerden oluşan bir güvenlik ekibine ihtiyaç duyulmaktadır.

Güvenlik açığı taraması, kuruluş tarafından yönetilen içerideki ya da dışarıdaki adresleri, varlıkların incelenmesini ifade etmektedir. Bu tarama ile temel olarak web siteleri, uygulamalar, cihazlar, ağlar ve bulut hizmetleri taranarak güvenlik açıklıkları keşfedilmektedir. Güvenlik açığı değerlendirmesi ile tespit edilen bu güvenlik açıklıkları değerlendirilerek gözden geçirilmektedir. Yama yönetimi ile yazılıma yapılacak “yama”lar daha organize bir şekilde ele alınarak olası saldırıların önüne geçmek hedeflenirken güvenlik açığı düzeltme bileşeni ile mevcut güvenlik açıklıkları giderilmeye çalışılmaktadır. Güvenlik açığı yönetimi yaşam döngüsüne ait görsel Şekil 3.4’te verilmektedir.



Şekil 3.4. Güvenlik açığı yönetimi yaşam döngüsü

3.4. Bilgi Güvenliği Politikası

Bilgi güvenliğini sağlamak bir şirket için ekonomik ve prestij açılarından oldukça önem taşımaktadır. Bu nedenle şirketler bilgi güvenliğini sağlamak amacıyla gerekli önlemleri almak için danışmanlık hizmetleri, uygulama alımları gibi birden çok yönetime başvurmaktadır. Bu kapsamda bir şirketin hangi güvenlik açıklıklarına sahip olduğunu, hangi tehditlerle daha fazla muhatap olduğunu ve bu tehditlerden hem geleneksel olarak hem de dijital olarak nasıl korunacağını bilmesi önem taşımaktadır. Bu doğrultuda bir şirketin bilgi güvenliğini, siber saldırılardan korunarak nasıl sağlayacağını düzenleyen politika ise bilgi güvenliği politikası olarak tanımlanmaktadır [42]. Eksiksiz bir bilgi güvenliği politikası oluşturmanın bazı önemli sonuçları bulunmaktadır. Bunlar, veri gizliliğini sağlama, güvenlik riskini azaltma, güvenlik önlemlerinin bir bütün halinde yürütülmesini sağlama ve yasal gereksinimlerin sağlanmasıdır.

Bilgi güvenliği politikası ile şirketin genel siber güvenlik amaçları, hedefleri ve kapsamı tanımlanmaktadır. Bu tanımlama ile politikanın temeli oluşturulmakta ve özellikle çalışanlar nezdinde temel davranış biçimlerinin şekillenmesi sağlanmış olmaktadır. Oluşturulan bu bilgi güvenlik politikası ile şirkette bulunan herkesin öncelikle alınan güvenlik önlemlerinin ve bu önlemlere uygun davranış sergilemenin önemini anlaması sağlanır. Bu kavrayış şirketin insan unsuru tarafından gerçekleştirildikten sonra büyük oranda geleneksel koruma yöntemlerinin de geniş alanda uygulanabilir olması sağlanmaktadır. Dolaylı olarak bilgi güvenliği politikası ile en üst katmanda şirket itibarı da korunmuş olmaktadır.

Bilgi güvenliği politikasının tüm güvenlik hususlarını kapsayacak şekilde geniş olması gerekmektedir. Her ne kadar her alanın kendine özgü politika oluşturması gerekse de ortak paydada bilgi güvenliği politikasının belirli unsurları vardır. Bunlar, amaç, kapsam, zaman çizelgesi, yetki, hedef, uyumluluk, politikayı oluşturan gövde, uygulama, eğitim, kişiler ve sürüm geçmişi olarak on bir ana başlıkta toplanabilmektedir. Eksiksiz bir şekilde tüm güvenlik hususlarını kapsayacak

bir bilgi güvenliği politikası bahsedilen bu unsurları içererek kapsayıcı bir güvenlik önlemi sistemi sunmaktadır.

3.5. Bilgi Güvenliği Önlemleri

Teknolojinin gelişmesi ve hemen her alanda kullanılabilir hale gelmesiyle günümüzde bilgi güvenliğini sağlayabilmek de önemli hale gelmiştir. Bu nedenle özellikle büyük verileri depolayan ticari amaç güden şirketler için bilgi güvenliğini sağlayabilmek hayati bir önem taşımaktadır. Çünkü bilgi güvenliğini sağlayabilmek söz konusu şirketlerin ticari itibarı ve ekonomik anlamda sürdürülebilirliği ile doğrudan ilgili haldedir [40]. Bu kapsamda her an çeşitli saldırılara maruz kalma ihtimaline karşı bilgi güvenliğini sağlamak için bazı yöntemlere başvurulmaktadır.

- **Güçlü şifre ve çift adımlı kimlik doğrulama kullanımı:** Şirketler ve kişiler verilerini genellikle belirli bir depolama alanında saklamaktadır. Bu nedenle söz konusu depolama alanına erişim belirli bir şifre ve çift adımlı kimlik doğrulama ile gerçekleşmektedir. Veriye erişim sağlanmasının ön koşulu niteliğinde olan şifrelerin kolay tahmin edilemeyen, büyük ve küçük harfler içeren, en az sekiz karakterden oluşan, rastgele sayılar içeren şekilde düzenlenmesi gerekmektedir. Ayrıca bu kapsamda belirlenen şifre uzun bir süre için kullanılmamalı belirli aralıklarla güncellenmelidir. Kolay tahmin edilebilir ve kısa şifrelerin kullanılması veriye erişimi kolaylaştıracak ve bilgi güvenliğinin sağlanması için büyük bir tehdit haline gelecektir. Bu kapsamda güçlü bir şifre kullanımının yanında çift adımlı kimlik doğrulama yöntemi de ekstra bir güvenlik duvarı oluşturacaktır.

- **Verilerin yedeklenmesi:** Bilgi güvenliğinin temel koruma alanı olan veriler, her an saldırı tehlikesi ile karşı karşıya olup herhangi bir saldırının başarılı olması halinde bir daha erişilemeyecek şekilde kaybedilebilmektedir. Bu nedenle verilerin hem dijital hem de yangın, sel gibi doğal tehlikelere karşı korunması daha geleneksel bir yöntem olan yedekleme ihtiyacını doğurmaktadır. Kişisel veya şirketlere ait olan bu veriler birden fazla yerde ve birden fazla şekilde yedeklenerek verinin herhangi bir saldırıya uğraması halinde geri dönülmez şekilde kaybolmasını engellemektedir. Son zamanlarda özellikle bulut yedekleme sistemleri ile verilere her yerden erişim sağlanabilmekte ve saldırıya uğraması halinde verinin en son haline erişme imkânı sağlamaktadır. Bu nedenle özellikle büyük verileri depolayan şirketler basılı olarak saklamak gibi en eski yöntemlerden bulut yedekleme sistemlerini kullanmak gibi en yeni yöntemleri birlikte kullanmaktadır.

- **Güvenlik politikası oluşturma:** Yukarıda bir başlık altında detaylı olarak anlatılan güvenlik politikası, şirketlerin herhangi bir saldırıdan nasıl korunacağını, saldırı anında nasıl hareket edeceğini belirleyen bir belgedir. Bu belge sayesinde çalışanlar tarafından güvenlik politikası daha anlaşılabilir hale gelmekte ve saldırıya karşı kolektif bir koruma sağlamaktadır.

- **Yazılımı güncel tutma:** Şirketler ve kişiler tarafından kullanılan uç noktaların yazılımlarının güncel tutulması bilgi güvenliğinin sağlanması açısından önem taşımaktadır. Belirlenen güvenlik açıklıklarına karşı yenilenen yazılımların güncellenmesi depolanan verilerin korunmasında önemli rol oynamaktadır.

- **Eğitim:** Verilerin güvenliğinin sağlanmasında en önemli faktör insan faktörüdür. Bu nedenle saldırılara ve bu saldırılara karşı korunma yöntemlerine ilişkin şirket çalışanları ve kişilere gerekli eğitimlerin sağlanması gerekmektedir. Bu eğitimlerin sağlanması halinde dijital korunma yöntemlerinin yanında insan faktörü ile geleneksel korunma yöntemleri de bir arada gerçekleştirilmiş olmaktadır.

- **Antivirüs yazılımı kullanımı:** Herhangi bir indirme ya da sosyal mühendislik saldırısı ile uç noktalara virüs bulaşması ve verilerin zarar görmesi oldukça muhtemel bir tehlikedir. Bu nedenle doğru bir antivirüs yazılımı kullanmak ve bu yazılımı güncel tutmak bu tehlikeye karşı en temel korunma yöntemidir. Antivirüs yazılımı ile bilgisayarlar ve diğer akıllı cihazlar periyodik olarak herhangi bir kötücül yazılımın var olup olmadığı hususunda tarama ve var olması halinde bu yazılım etkisiz hale getirilir [39]. Bu kapsamda güncel bir antivirüs yazılımı kullanmak bilgi güvenliğini sağlamanın en temel ve kolay yöntemleri arasında yer almaktadır.

- **Güvenli Wi-Fi ağlarına bağlanma:** Uzaktan çalışma yönteminin hayatımıza girmesiyle birlikte çalışanlar internet bağlantısı olan herhangi bir yerde iş tanımlarını yerine getirebilmektedir. Genellikle şirketler çalışanlarına güvenli bir ağ bağlantısı sağlamakta ve bu bağlantıyı kontrol altında tutmaktadır. Ancak kurumsal bağlantı dışında herhangi bir bağlantıya bağlanarak internet erişimi sağlanması halinde veriler saldırıya açık hale gelebilmektedir. Bu nedenle herkese açık herhangi bir Wi-Fi bağlantısına bağlanması halinde öncelikle bu bağlantının güvenliğinden emin olunması gerekmektedir. Güvenliğinden şüphe duyulan herhangi bir Wi-Fi bağlantısına bağlanmak veri güvenliği açısından büyük bir tehlike oluşturmaktadır.

- **Verilere erişimin kısıtlanması:** Özellikle şirketler nezdinde birden çok veri türü depolanmaktadır. Ancak her departmanda çalışanlar farklı türden verilere ihtiyaç duymaktadır. Bu nedenle şirket içerisinde hangi departmanın hangi tür veri ile çalıştığı ve hangi tür verilere ihtiyaç duymadığı titizlikle belirlenerek erişimlerinin sadece ihtiyaç duyulan veriler ile kısıtlanması gerekmektedir [40]. Ayrıca şirketten ayrılan çalışanların ve bu çalışanlara ait cihazların verilere erişimi askıya alınarak engellenmelidir.

3.6. Bilgi Güvenliği ve Siber Güvenlik

Siber güvenlik ve bilgi güvenliği sıklıkla birbiri yerine kullanılan ancak belirli yönleriyle birbirinden ayrılan, temel farklılıklara sahip olan iki farklı terimi ifade etmektedir. Bilgi güvenliği temelde verinin dijital ve fiziksel olarak korunmasını amaçlayan ve bütüncül bir yaklaşım içerirken

siber güvenlik verinin dijital olarak korunmasını amaçlamaktadır [38]. Siber güvenlik ve bilgi güvenliği arasındaki en temel farklılık burada ortaya çıkmaktadır.

Teknolojinin gelişmesiyle birlikte veriler dijital olarak da depolanmakla birlikte dijital ortamda güvenliğin nasıl sağlanacağı da önemli bir konu haline gelmeye başlamıştır. Bu sebeple siber güvenlik ve bilgi güvenliği alanları öne çıkmaya ve dikkate değer araştırma konuları olmaktadır [43]. Bu kapsamda öncelikle her iki terimin de aynı şeyi korumayı amaçlayıp amaçlamadığı cevaplanması gereken en temel sorulardan biridir. Bilgi ve veri kavramları çoğunlukla aynı anlamda kullanılsa da teknik olarak farklı anlamlar ifade edebilmektedir. Veri, bir bilginin ham parçacığı iken bilgi, anlamlandırılmış bir veridir. Dolayısıyla bilgi veriyi de kapsayan daha geniş bir kümeyi ifade etmektedir. Bu doğrultuda siber güvenlik ve bilgi güvenliğinin korumayı amaçladıkları alanların da daha farklı alanlar olduğu açıktır.

Bilgi güvenliği, içerisinde siber güvenliği de barındıran daha geniş bir yapı sunmaktadır. Bu nedenle bilgi güvenliği hem fiziksel hem de dijital olarak bilgiyi her türlü tehdide karşı korumayı amaç edinirken siber güvenlik ile verinin dijital olarak korunması amaçlanmaktadır. Teknolojinin gelişmesiyle paralel olarak verilerin ve bilgilerin dijital ortamda depolanması bilgi güvenliği ve siber güvenliği de birbiri içerisinde karışan iki alan haline gelmiştir. Bu nedenle sıklıkla karıştırılmakta ve birbiri yerine kullanılmaktadır.

Her iki terimin birbiri ile örtüştüğü belirli alanlar olmakla birlikte bunlardan en önemlisi kullanılan modeldir. Bilgi güvenliği gibi siber güvenlik de bilginin gizliliği, bütünlüğü ve kullanılabilirliği olan üçlü modeli kullanmaktadır. Yukarıda detaylıca bahsedildiği üzere gizlilik, bilginin yalnızca yetkili kişiler tarafından erişilebilir olmasını, bütünlük bilginin değiştirilememesini ve güvenilir olmasını, kullanılabilirlik ise bilginin ihtiyaç halinde kullanılabilir olmasını ifade etmektedir. Siber güvenlik alanı da veri için aynı üçlü modeli kullanarak güvenlik politikasını geliştirmektedir. Aynı zamanda bilgi güvenliğinin sağlanması için kullanılan birden çok yöntem siber güvenliğin sağlanması için de elverişli yöntemler olmaktadır.

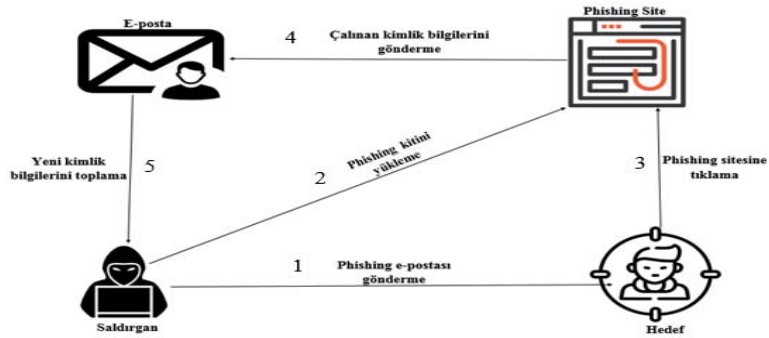
4. SOSYAL MÜHENDİSLİK SALDIRI TÜRLERİ

Saldırgan çeşitli popüler yollara başvurarak sosyal mühendislik saldırılarını gerçekleştirmektedir. Saldırı türleri Şekil 4.1’de gösterildiği gibidir. Bu kısımda saldırı türleri detaylıca açıklanmıştır.



Şekil 4.1. Sosyal mühendislik saldırı türleri

• **Phishing:** Sosyal mühendislik saldırıları ile hedef manipüle edilerek istenilen bilgilere erişmek amaçlanmaktadır. Saldırıların en çok tercih edilen türü phishing (kimlik avı) saldırıdır. Phishing saldırılarında genellikle güvenilir kurumlardan ya da kişilerden gelmiş gibi görünen aldatıcı mailler ile hedeften kişisel bilgileri ve kredi kartı bilgileri gibi üçüncü şahıslarla paylaşılması güvenlik ihlali teşkil edecek verilerin alınması sağlanmaktadır. Şekil 4.2’de bir phishing saldırısının nasıl başlatıldığı tüm adımlarıyla birlikte gösterilmiştir.



Şekil 4.2. Phishing saldırısının başlatılması

• **E-posta kimlik avı (Email phishing):** Kimlik avı saldırıları içerisinde yaygın olarak kullanılan türdür. Bilinen kurum ya da kuruluş tarafından gönderilmiş gibi görünen bir e-posta ile hedefin kişisel bilgilerine ulaşmak istenmektedir. E-posta kimlik avı saldırısı belirli bir hedef özelinde hazırlanmamıştır. Aksine toplu bir şekilde tasarlanmış bir saldırı türüdür [44]. E-posta kimlik avının “spear phishing, whaling (balina avcılığı) ve CEO fraud (CEO dolandırıcılığı) ve clone phishing” olmak üzere çeşitli türleri bulunmaktadır.

Spear phishing saldırılarında temel phishing saldırılarının aksine hedef gözetilir. Gönderilecek olan yanıltıcı e-posta hedef için özellikle düzenlenmiştir. Bu ise spear phishing yapacak saldırganın öncesinde hedefe ilişkin bir ön araştırma yapması anlamına gelmektedir [44].

Whaling ve CEO Fraud yöntemleri birbirinin zıttı bir şekilde çalışmaktadırlar. Whaling saldırısında saldırgan üst düzey yöneticileri hedef alarak resmi bir devlet otoritesinden gelmiş gibi görünen ya da bir müşteri şikâyeti gibi görünen bir mail ile üst düzey yöneticinin dikkatini çekmeyi amaçlamaktadır [44]. CEO fraud saldırısında ise çalışanlara yönelik bir üst düzey yöneticiden gelmiş gibi görünen bir e-posta ile saldırı gerçekleştirilir. CEO fraud yönteminin başarı olasılığı çok düşük olsa da genellikle finans bölümü çalışanları hedeflendiği için başarılı olma durumunda saldırganın büyük meblağlar kazanması mümkündür [45].

Clone phishing saldırısı geleneksel phishing saldırılarından daha farklı bir yol izlemektedir. Bu saldırıda hedefe daha önce gönderilmiş meşru bir e-posta kullanılır. Bu e-postanın birebir kopyası saldırgan tarafından inşa edilir ve e-postanın ek ve bağlantıları zararlı yazılımları ve virüsleri içeren ek ve bağlantılar ile değiştirilir. Hedef bu ek ve bağlantıları görüntülemek için e-postaya tıkladığı zaman aynı e-posta hedefin gelen kutusundaki diğer kişilere de iletilmektedir [46].

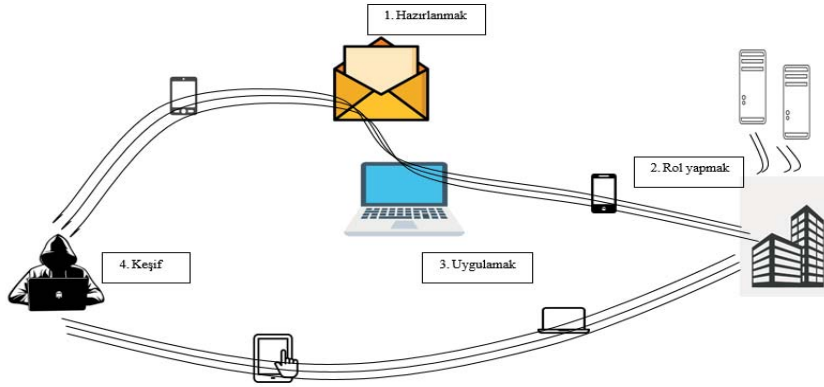
• **Sesli kimlik avı (Voice phishing):** Sesli kimlik avı saldırılarında diğer phishing saldırılarından farklı olarak e-posta değil telefon kullanılmaktadır. Saldırgan kendisini bir müşteri temsilcisi gibi tanıtarak hedeften kişisel bilgilerini paylaşmasını istemektedir. Telefon numarası çoğunlukla kurumsal telefon numaralarına benzer şekilde düzenlenir. Bu şekilde henüz hedefle iletişime geçilmeden belirli bir güven seviyesine ulaşılmış olur. Bu aşamadan sonra sosyal mühendislik yöntemleri kullanılarak veriler saldırgan tarafından ele geçirilir [44].

• **SMS kimlik avı (Smishing):** Akıllı telefonların yaygınlaşmasıyla birlikte artık geleneksel e-posta kimlik avı, telefonlara gönderilen metin mesajları ile de yapılabilir. Bu yöntemde saldırgan tarafından bir bağlantı, telefon numarası ya da saldırganın kontrol ettiği bir web adresi içeren metin mesajı hedefe gönderilir. Bağlantıya tıklanması durumunda artık hedef saldırganın tuzağına düşmüş olur [44].

• **Sayfa ele geçirme (Page hijacking):** Bu phishing saldırı türünde saldırgan tarafından meşru bir web sitesinin güvenliği tehlikeye atılır ve bu siteyi ziyaret edenler saldırganın düzenlediği kötü amaçlı yazılımlar içeren web sitesine yönlendirilir. Oldukça yaygın olarak kullanılan bir phishing türüdür.

• **Typosquatting with phishing (Alan adı benzerliği ile kimlik avı):** Bu saldırı türünde bilinen ve popüler olan web sitelerinin alan adlarında karıştırılabilen harflerde yapılan küçük değişikliklerle hedefler saldırganın düzenlediği kötücül yazılımlar içeren ya da satın alma işlemi gerçekleştirilen web sitelerine yönlendirilirler. Saldırganlar, hedeflerin dikkatsizliğinden faydalanmaktadırlar.

• **Pretexting:** Bu sosyal mühendislik saldırı türünde saldırgan hedefin güven duygusu üzerinde hâkimiyet kurarak başka şartlar altında paylaşmayacağı kişisel bilgilerini paylaşmasını sağlamaktadır. Saldırgan tarafından yapılan hedefe ilişkin ön araştırma neticesinde hedefle ilgili elde edilen doğru bilgiler içerisinde tuzakların yerleştirildiği senaryo uygulanırken kullanılmaktadır. Saldırgan tarafından aktarılan doğru bilgiler hedefte güven duygusu oluşturur ve senaryoya inanan hedef saldırgana istediği bilgileri vererek saldırganı amacına ulaştırmış olmaktadır. Şekil 4.3'te pretexting saldırısına ait görsel verilmiştir.



Şekil 4.3. Pretexting saldırısının başlatılması

• **Quid Pro Quo:** Sözlük anlamı “bir iyilik için iyilik olan” bu saldırı türünde saldırgan hedefe bir hizmet ya da kolaylık sağlamak için hedefin kişisel bilgilerine veya hesap bilgilerine erişim izni talep etmektedir. Kendisine sağlanan hizmetin sağlıklı ve sorunsuz bir şekilde ilerleyebilmesi için söz konusu erişim izinlerinin verilmesi gerektiğini düşünen ve bunda bir sakınca görmeyen hedef tarafından ilgili izinler verilerek saldırı amacına ulaştırılmış olmaktadır [47].

• **Baiting:** Bu saldırı türünde saldırgan tarafından hedefe ulaşmak için fiziksel ya da dijital dünyada belirli yemler bırakılmaktadır. Bu yemler halka açık bir alanda bırakılan USB ve CD gibi fiziksel materyaller olabileceği gibi web sitelerine yönlendirilen cazip teklifler içeren reklamlarda olabilmektedir. Söz konusu yemleri gören hedefin merak duygusu uyandırıldığı zaman hedef

tarafından kişisel bilgisayarına ya da telefonuna enjekte edilen USB veya CD neticesinde kötücül yazılımlar cihazlarına yüklenmiş ve kişisel bilgileri çalınmış olmaktadır [47].

• **Tailgating:** Diğer sosyal mühendislik saldırı türlerinden farklı olarak bu saldırı türünde saldırgan hedef ile gerçek dünya üzerinde fiziksel bir temas kurmaktadır. Saldırgan bir senaryo kurarak aslında erişim izni olmayan ya da girmeye yetkisinin bulunmadığı kurumlara girmek için hedeften yardım istemektedir. Hedef tarafından senaryoya uygun bir şekilde söz konusu kuruma girme izninin bulunduğu düşünülen saldırgan içeriye alınarak saldırganın kurumla ilgili bilgi ve belgelere erişimi sağlanmış olmaktadır.

• **Scareware:** Saldırgan tarafından hedefe bir sosyal ağ üzerinden ani ve korkutucu bildirimler gönderilmektedir. Söz konusu korkutucu durumdan kurtulmak için hedef saldırganın kendisine sunduğu yazılımı almaya yönlendirilmektedir. Korku duygusu harekete geçen hedef tarafından saldırgan istenilen ödeme yapılarak gerçekte var olmayan korkutucu durumdan kurtulmak için herhangi bir işlevi olmayan yazılım satın alınarak saldırgan amacına ulaştırılmaktadır.

• **Dumpster Diving:** Çöp kutusu dalışı olarak bilinen bu yöntemde saldırgan tarafından hedefe yönelik önemli bilgilerin elde edilmesi için hedefe ait çöplerin karıştırılmasıdır. Çöp karıştırma işlemi fiziksel ya da dijital olarak gerçekleştirilmektedir. Hedef tarafından önemsiz olduğu düşünülen bilgilerin mevcut olduğu kağıtların çöpe atılması ve ardından saldırgan tarafından bu kağıda ulaşılması neticesinde söz konusu kağıttaki bilgiler kullanılarak başka tür sosyal mühendislik saldırıları gerçekleştirilmektedir.

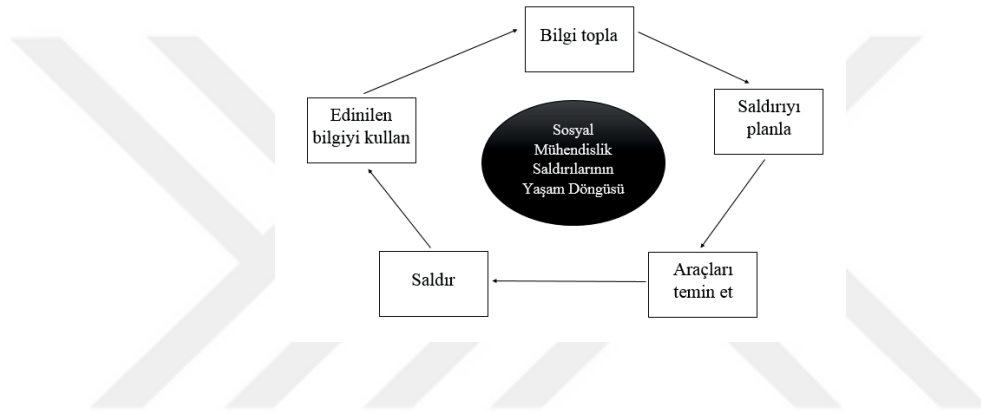
4.1. Sosyal Mühendislik Saldırılarının Gerçekleştirilmesi

Sosyal mühendislik saldırıları, temel olarak kişilerin zafiyetlerinden ve duygu durumlarından faydalanarak onların kişisel ve başkaları tarafından ulaşılması sakıncalı olan bilgilerine ulaşılması prensibine dayanır. Saldırganlar çoğu zaman hedeflerinin merak, korku, istek gibi insani duygularını harekete geçirerek saldırılarını başlatmaktadır. Birden çok yöneme ve türe sahip olan sosyal mühendislik saldırılarının iki aktörü vardır: Saldırgan ve hedef. Saldırgan, sosyal mühendislik yöntemlerini kullanarak kişisel bilgiye erişmeyi amaçlayan tarafı, hedef ise duyguları harekete geçirilerek kişisel bilgileri ele geçirilmek istenen tarafı teşkil etmektedir.

Sosyal mühendislik saldırılarında saldırgan öncelikle hedefini belirleyerek hedef hakkında onda güven duygusunu oluşturacak ya da merak, korku, endişe gibi duygularını harekete geçirecek senaryolar için araştırma yapmaktadır. Ön hazırlık niteliğindeki bu aşamada hedef ve senaryo belirlendikten sonra saldırgan senaryoya ve hedefe en uygun sosyal mühendislik saldırı yöntemini belirleyerek hedefe ulaşma aşamasına geçmektedir. Saldırgan tarafından ani bir bildirimle hedefte korku yaratılması, cazip bir teklifle hedefin istek duygusunun harekete geçirilmesi, çeşitli sorular

ile hedefin merak duygusunun harekete geçirilmesi gibi yöntemler kullanılarak kişisel bilgilere erişim sağlanabileceği gibi fiziksel olarak temasa geçilerek de saldırı gerçekleştirilerek istenilen bilgilere erişim sağlanabilmektedir.

Sosyal mühendislik saldırılarının yaşam döngüsünün üç aşamadan ibaret olduğu söylenebilir; araştırma aşaması, saldırının gerçekleştiği aşama ve saldırının sona erdiği aşama. Söz konusu aşamalar tamamlandığı zaman saldırgan tarafından tekrar araştırma aşamasına dönülerek yeni hedefler belirlenerek yeni saldırılar başlatılmaktadır. Bahsedilen sosyal mühendislik saldırılarının yaşam döngüsü Şekil 4.4'te gösterilmiştir.



Şekil 4.4. Sosyal mühendislik saldırısı yaşam döngüsü

Araştırma Aşaması: Bu aşamada saldırgan, hedefini harekete geçirecek tetikleyicileri araştırmaktadır. Salırgan istediği bilgilere erişmek için zor kullanmak yerine hedefini harekete geçirecek tetikleyicileri bularak onları manipüle etmeyi tercih etmektedir. Bu aşamada saldırgan temel olarak hedefini nasıl manipüle edeceğini araştırmaktadır [47].

Saldırı Aşaması: Bu aşamada hedefinin nasıl manipüle edeceğini bulan saldırgan bu manipülasyonu yapabilmek için gereken yemleri hedefinin görebileceği yerlere yerleştirmektedir. Korku, merak, endişe ya da istek duygusu harekete geçirilen hedef tarafından saldırganın erişmek istediği bilgiler kolayca saldırganı teslim edilmiş olmaktadır [47].

Saldırının Sonlanma Aşaması: Bu aşamada ise istediği bilgilere kolaylıkla erişen saldırgan hedefi ile bir önceki aşamada kurduğu iletişimi sonlandırarak tekrar araştırma aşamasına geçerek yeni bir saldırıya hazırlanmaktadır [47].

4.2. Sosyal Mühendislik Saldırılarından Korunma Yöntemleri

Sosyal mühendislik saldırılarına karşı kişiler bir başka kurum, kuruluş veya herhangi bir programdan güvenliklerinin sağlanmasını beklemek yerine kendi güvenliklerini sağlayacak

yöntemler hakkında da bilgi sahibi olmalılardır. Bu yöntemler oldukça basit ve uygulanması kolay yöntemlerdir.

Sosyal mühendislik saldırılarına karşı korunmak için her şeyden önce sosyal mühendislik kavramının ne anlama geldiğini ve sosyal mühendislerin nasıl davrandıkları konusunda kişilerin genel bir bilgi çerçevesine sahip olması gerekmektedir.

Hedef, kendisine ulaşan sosyal mühendisten korunmak amacıyla ilk olarak kendisine ulaşılma yöntemini sorgulamalıdır. Bu yöntem sonucunda kendisinde meydana gelen duygu değişimlerini iyi gözlemleyebilmeli ve ani duygu artışı yaşadığı durumlarda bunun normal olmadığını ve bir sosyal mühendislik saldırısıyla karşı karşıya olabileme ihtimalini anlayabilmek için acele etmeden sakince düşünmelidir.

Saldırgan, çoğu saldırıda hedeflerin kişisel bilgilerini, banka bilgilerini ve sosyal medya hesap bilgilerini elde etmeyi amaçlar. Bu nedenle kendisine karşı böyle bir istekte bulunanlara karşı şüpheci davranılmalı ve bu tür özel bilgiler kimseyle kolayca paylaşılmamalıdır. Kendisini yetkili bir kişi olarak tanıtan saldırgandan bu bilgileri doğrulaması istenilmeli ve tatmin edici bir doğrulama sağlanmazsa bilgiler karşı tarafla paylaşılmamalıdır.

Kişiler e-posta kutularına düşen veya kendilerine başka bir yolla gönderilen bağlantılara gitmeden önce bağlantıların güvenilirliğinden emin olmalılardır. Bunu yapmak için URL'ler kontrol edilmeli ya da bağlantı genişletici programlar kullanılmalıdır [48].

Kişiler herhangi bir program, belge, doküman veya uygulama indirecekleri zaman indirme yapacakları sitelerin ve indirecekleri kaynakların güvenilir olduğundan emin olmalıdırlar. Bunun için sık kullandıkları siteleri tercih etmeli ya da URL'sinde "https" olan siteleri kullanmalıdır. URL'sinde "http" olması yeterli olmayıp "s" ekinin de bulunması gereklidir çünkü bu ek sayesinde web sitelerinin güvenlik sertifikasının olduğu çıkarımında bulunulur [48].

Her gün e-posta kullanıcıları onlarca istenmeyen mail almaktadırlar. Sosyal mühendisler de istenmeyen e-postalar aracılığıyla saldırılarını gerçekleştirebilmektedir. Bu nedenle iyi bir istenmeyen e-posta filtresi kullanmak saldırılara karşı iyi bir önlem yolu olabilmektedir.

Sosyal mühendislik saldırıları ile karşı karşıya kalındığı zaman bunun gerçekten bir saldırı olup olmadığını anlamak oldukça zordur. Bu nedenle öncelikle sakin kalarak sonrasında karşılaşılan bu durumun gerçeğe uygunluğu değerlendirilmelidir. Gerçeğe uygunluk konusunda tereddüt yaşandığı zaman karşı tarafın talep ettiği bilgi, belge vb. şeylerin teslimi gerçekleştirilmemelidir.

Sosyal mühendislik saldırıları çoğunlukla teknolojik cihazlar üzerinden gerçekleştirilir. Bu nedenle cihazları güvence altına almak ve başkalarının cihaza erişim sağlamaları halinde güvenlik sorunlarına neden olacak özel bilgileri bu cihazlarda depolamaktan imtina etmek konusunda özen gösterilmelidir. Teknolojik cihazlar üzerinden değil de fiziksel olarak gerçekleştirilen "Dumpster Diving" gibi sosyal mühendislik saldırılarına karşı alınacak en etkili önlem çöplerin kağıt

parçalayıcısından geçirilerek birleştirilmesi engellenecek şekilde küçük parçalara ayrıştırılmasıdır. Tablo 4.1’de sosyal mühendislik saldırı türlerine ait korunma yöntemleri verilmiştir.

Tablo 4.1. Sosyal mühendislik saldırı türleri ve korunma yöntemleri

Saldırı Türü	Korunma Yöntemleri
Vishing	➤ Saldırganı hedefe ait olmayan başka bir cihazla aranması, ➤ Telefon görüşmeleri sırasında kişisel bilgilerin paylaşılmaması, ➤ Kişisel bilgileri talep eden saldırgan ile yüz yüze görüşme istenmesi, ➤ Tanınmayan numaraların engellenmesi, ➤ Saldırganın olduğu iddia ettiği kişi olup olmadığını sadece o kişinin cevaplayabileceği sorularla tespit edilmesi [49].
Phishing/ Smishing	➤ Mesaj ve mail içeriklerinde yer alan bağlantılara tıklamadan önce gönderi sahibinin güvenilirliğinden emin olmak, ➤ Spam mail ve sms engelleyici eklentiler kullanmak, ➤ Kişisel bilgi talep eden metinlere mesaj içeriğinde cevap vermeme, ➤ Sosyal medya hesaplarında, banka hesaplarında farklı şifreler kullanmak, ➤ Saldırı tespit uygulamaları kullanmak, ➤ Güvenilirliğinden emin olunmayan web siteleri üzerinden indirme işlemi gerçekleştirilmemesi, ➤ Yetki sorgulama işleminin yapılması, ➤ Penetrasyon testi yapılması.
Pretexting	➤ Saldırganın kimliğinin doğrulanması, ➤ Mail analizinin yapılması, ➤ Saldırıların tespit edilmesi için geliştirilen uygulamaların kullanılması.
Dumpster Diving	➤ Kişisel ya da kurumsal bilgilerin yazılı olduğu kağıtların, kağıt parçalayıcısından geçirilmesi, ➤ Bilgilerin yer aldığı harddisklerin çöpe atılmaması, ➤ Kişisel bilgisayarların halka açık alanlarda açık bırakılmaması, ➤ Kişisel bilgilerin yer aldığı klasörlerin şifrelenmesi.
Scareware	➤ Güvenilirliğinden emin olunmayan web sitelerinin kullanılmaması, ➤ Web sitelerinde karşımıza çıkan ani bildirimlere tıklanmaması, ➤ Güvenli bir virüs tespit sisteminin kullanılması, ➤ Güvenilirliğinden emin olunmayan web siteleri üzerinden herhangi bir yazılım indirilmemesi, ➤ Güncel tarayıcılar kullanılması.

4.3. Sosyal Mühendislik Saldırılarının Tespitindeki Zorluklar

Sosyal mühendislik saldırıları her ne kadar yaygın olsa da tespit yöntemleri de bir o kadar yaygındır. Saldırıların tespiti modern ve geleneksel yöntemlerle olabilmektedir. Geleneksel olarak kişilerin tanımadığı mailleri açmaması, bağlantılara tıklamaması gibi yöntemlerle korunma sağlanırken spam önleyicilerin kullanılması ya da saldırıların tespiti için geliştirilen uygulamalar modern olarak kullanılmaktadır. Ancak söz konusu yöntemler her zaman tam bir koruma sağlamayıp yeni geliştirilen sosyal mühendislik saldırılarının farkına varılana kadar bazı zorlukları da beraberinde getirmektedir.

Sosyal mühendislik saldırılarının tespitinde kullanılan sistemler temelde varoluşları ile doğru orantılı olarak bazı sorunları beraberinde getirmektedir. Geliştirilen yöntemler sınırsız veri setlerine ya da kendi kendine yeni geliştirilen tanıma özelliğine sahip değildir. Bu nedenle sosyal mühendislik saldırılarının tespiti için DÖ ve MÖ yöntemleri ile geliştirilen uygulamalarda kullanılan veri setlerinin temel mantığı dışında başka bir türden veri ile karşılaştığı zaman uygulama bu durumu saldırı olarak tespit etmekte ve engellemekte yetersiz kalmaktadır. Aynı şekilde mevcut saldırı türleri üzerinde geliştirilen uygulamaların yeni ortaya çıkmış saldırı türlerini tanımakta güçlük çekmesi de sosyal mühendislik saldırılarının tespitinde yaşanan zorluklardandır.

Geliştirilen uygulamalarda kullanılan DÖ ve MÖ yöntemleri sınırlı bir tanıma ve hareket alanına sahip olmaktadır. Dolayısıyla saldırının bir insan tarafından tespitinin ise bir uygulama tarafından gerçekleştirilmesi temel olarak saldırının tespitinde önemli zorlukları beraberinde getirmektedir.

4.4. Sosyal Mühendislik Saldırılarının Tespitinde Kullanılan Yöntemler

Sosyal mühendislik saldırılarından smishing ve phishing başta olmak üzere diğer sosyal mühendislik saldırılarının tespitinde kullanılan yöntemler, veri setleri ve başarımları oranları bu kısımda detaylı bir şekilde incelenmiştir.

4.4.1. Smishing

Saldırgan tarafından hedefe bir kısa mesaj gönderilerek hedefin kişisel bilgilerine erişilen Smishing yöntemine ilişkin literatürde yer alan çalışmalar Tablo 4.2’de gösterilmiştir.

Tablo 4.2. Literatürde yer alan Smishing saldırılarında kullanılan tespit yöntemlerinin analizi

Referans	Kullanılan Yöntem	Veri Seti	Veri Sayısı	Doğruluk Performansı
[12]	Hidden Markov Modeli	UCI Veri Havuzu	5547	0.969
[19]	Hibrit CNN – LSTM Modeli	İngilizce ve Arapça mesajlardan oluşan veri seti(Suudi Arabistan’daki birkaç yerel akıllı telefonda toplanan mesaj içerikleri)	8304	0.981
[14]	Çok Kanallı Evrişimli Sinir Ağı	UCI SMS Spam Toplama veri seti ve kendi kendine toplanan mesajların birleştirilmesi	8751	0.961
[15]	Bulanık kurala dayalı sınıflandırma	UCI Veri Seti	5574	0.985
[50]	OC-SVM modelinin uygulanması	Kaggle “SMS Spam collection dataset”	5574	0.980
[16]	J48, KNN, KA	UCI Veri Seti	5574	KNN = 0.911 KA = 0.970
[17]	Önceden eğitilmiş BERT + ek katmanlar	Kaggle “SMS Spam collection dataset”	5572	0.988
[18]	Hibrit kısa mesaj hizmeti spam filtreleme tekniği için Boyer Moore dize eşleştirme çerçevesi	UCI Veri Seti	5574	0.970
[20]	Genetik Algoritma Eğitimli Bayes Ağı çözümü	KDD-CUP '99 Veri Seti	-	-
[21]	Makine zekası tabanlı hibrit sınıflandırıcı	SMS Dataset (UCI) Email Dataset Spam assassin	5574 5172 3252	0.998
[51]	Multinomial Naive Bayes KNN) DVM RO Ekstra Ağaçlar	Kaggle “SMS Spam collection dataset”	5572	0.993
[52]	Malay SMS Spam algılama çerçevesi	SMS spam in Malay language	-	-
[13]	Kural Tabanlı SMS Hizmeti	UCL SMS spam dataset	5547	0.979

4.4.2. Phishing

Genel olarak kullanıcılara güvenilir ve bilinen bir kurum, kuruluş ve bankaları kullanarak meşru bir e-posta aracılığıyla çeşitli bağlantılar gönderilerek gerçekleştirilen phishing saldırılarına ilişkin çalışma Tablo 4.3'te gösterilmektedir.

Tablo 4.3. Literatürde yer alan Phishing saldırılarında kullanılan tespit yöntemlerinin analizi

Referans	Kullanılan Yöntem	Veri Seti	Veri Sayısı	Doğruluk Performansı
[53]	RO	Lightweight dataset	52.000	0.995
[54]	ECSPAD (Kurumsal Kimlik Bilgisi Sperar-phishing Saldırısı Kimlik Bilgisi)	-	-	-
[3]	3 aşamalı kimlik avı tespit çerçevesi	Gmail, Yahoo ve son kullanıcıya ait özel posta kutuları	500'den fazla e-posta	0.954
[5]	Kimlik avı tespiti için web sitesi logosunun kullanılması	Veri Kümesi 1 Veri Kümesi 2 Veri Kümesi 3	3894 3000 500	0.998
[6]	ANFIS NB PART J48 JRIP	Phistank Millersmiles Relbanks	20.000 500 10.000	ANFIS = 0.990 NB = 0.993 PART = 0.993 J48 = 0.993 JRIP = 0.993
[9]	Hafif Gradyan Artırma Makinesi Tabanlı (LightGBM)	Makine öğrenimi için kimlik avı veri kümesi	5000	1.0
[10]	Ters olmayan matris çevrimiçi sıralı aşırı öğrenme makinesine (NIOSELM) dayalı yeni yaklaşım	Phishtank	65.000	0.969
[55]	FS-CNN	5000 BEST WEBSITES(https://5000best.com/websites) https://www.phishtank.com	51.200	0.995

Tablo 4.3. (Devamı) Literatürde yer alan Phishing saldırılarında kullanılan tespit yöntemlerinin analizi

[56]	ANN	Spambase	4061	0.994
		CSDMC2010	4327	
		Phishing_corpus(SpamAssassin + Nazario)	5850	
[57]	Metasezgisel öğrenme odaklı kimlik avı tespiti (MDLPD-SSE)	Benign Phishing	7053 5995	0.957
[58]	KNN RO DVM	Phishtank	1000	KNN = 0.933 RO = 0.983 DVM = 0.917
[59]	LIGHTGBM	MillerSmiles	-	0.855
	XGBOOST	Phishtank		0.853
	Çok katmanlı algılayıcılar			0.851
	Catboost sınıflandırıcı			0.850
	RO			0.808
	KA			0.808
	DVM			0.796
[4]	İkili Modifiye Denge Optimize Edici (BMEO), AV-şekli transfer fonksiyonu (AV-BMEO) ve k-en yakın komşu sınıflandırıcı.	Dataset 1 : Mendeley3 Dataset 2: UCI deposundan indirilen 18 veri seti kullanılmıştır. (PenglunEW, Zoo, Lymphography, Wine, Sonar, SpectEW, HeartEW, Vote, Ionosphere, CongressEW, BreastEW, BreastCancer, Tic-tac-toe, Exactly2, Exactly, M_of_n, KrVsKpEW, WaveformEW)	Dataset1: 10.00 Dataset2: 11.055	0.963 0.974
[7]	Beyaz liste yaklaşımı	Phishtank Alexa	-	0.967
[8]	DDİ yöntemlerini kullanarak derin öğrenme algoritmalarını uygulayan bir grafik evrişimli ağ (GCN).	Fraud dataset	8579	0.982
[11]	Evrişimsel sınır ağı yaklaşımı	Phishtank	10.604	0.999

Tablo 4.3. (Devamı) Literatürde yer alan Phishing saldırılarında kullanılan tespit yöntemlerinin analizi

[60]	Makine öğrenimine (PhishNot) dayalı kimlik avı URL'ni algılama sistemi.	Phishtank Alexa	88.646	0.974
[61]	Hibrit DNN-BiLSTM Modeli	Dataset 1: Ebbu2017 Dataset 2: Phishtank2 + Marchal2014	36.400 26.000	0.992

4.4.3. Literatürde Yer Alan Diğer Çalışmalar

Önceki kısımlarda detaylıca anlatılan ve yaygın olarak kullanılan sosyal mühendislik yöntemleri haricinde kalan diğer sosyal mühendislik saldırı şekillerine ilişkin literatür analizi Tablo 4.4'te gösterilmiştir.

Tablo 4.4. Literatürde yer alan diğer sosyal mühendislik saldırılarında kullanılan tespit yöntemlerinin analizi

Referans	Çalışmanın Konusu	Kullanılan Yöntem	Veri Seti	Doğruluk Performansı
[22]	Duygu Analizi	Dikkat Temelli Çift Yönlü CNN-RNN Derin Modeli (ABCDM)	App Kindle Movies Electronics CDs Airline Twitter Sentiment140 T4SA	0.921 0.934 0.905 0.906 0.887 0.927 0.818 0.968
[23]	Sosyal medya ağlarında siber zorbalık tespiti	Eşik ve ikili olarak adlandırılan iki farklı spesifik erken tespit modeline sahip denetimli öğrenme yöntemi.	Vine sosyal medya ağından toplanan verilerden oluşan veri seti	0.950
[62]	Siber zorbalık tespiti	Derin karar ağacı sınıflandırmasına sahip çok özellikli tabanlı yapay zeka modeli	Twitter dataset	0.935
[24]	Siber zorbalık tespiti	Makine Öğrenimi ve alay konusuna yönelik özellik mühendisliği tabanlı çalışma	Semantic Evaluation 2018 Task 3: İngilizce Tweetlerde İroni Tespiti (Van Hee et al., 2018) tarafından sağlanan veri seti kullanılmıştır.	0.896

Tablo 4.4. (Devamı) Literatürde yer alan diğer sosyal mühendislik saldırılarında kullanılan tespit yöntemlerinin analizi

[24]	Siber zorbalık tespiti	Özellik Yoğunluğu	İngilizce siber zorbalık veri kümesi Japon siber zorbalık veri kümesi Ploish siber zorbalık veri kümesi	-
[63]	Zorbalık söylemi	Denetimli makine öğrenimi	Çalışmada 7 Ağustos 2019 ile 31 Mart 2020 arasında toplanan 847.548 tweet veri seti olarak kullanılmıştır.	0.780
[64]	Siber saldırganlık	M-BERT, distil-BERT, Bangla-BERT, XLM-R kullanan ağırlıklı topluluk modeli	BAD-Bangla-Aggressive Text Dataset (https://github.com/omar-sharif03/BAD-Bangla-Aggressive-Text-Dataset)	0.934
[65]	Nefret söylemi sınıflandırması	Olasılıksal kümeleme modeli	Twitter Akış API'si ile toplanan veri seti kullanılmıştır.	0.925
[27]	Siber zorbalık tespiti	AdaBoost	Turkish Twitter dataset	0.901
[65]	Nefret söylemi sınıflandırması	Kural tabanlı kümeleme	Twitter dataset	0.950
[28]	Siber tacizin ve tanınması önlenmesi	MÖ Sınıflandırma Algoritmaları	Social Media dataset	0.928
[29]	Siber Zorbalık Tespiti	Bagging Boosting C4.5 GBM K-Means KNN LR NB YSA RO DVM SGD XGBoost	Kaggle'da halka açık şekilde yer alan veri seti kullanılmıştır. (https://www.kaggle.com/code/tbrknt/detection-of-cyberbullying-in-turkish/notebook)	0.825 0.865 0.852 0.695 0.510 0.652 0.883 0.753 0.862 0.853 0.780 0.869 0.767

Tablo 4.4. (Devamı) Literatürde yer alan diğer sosyal mühendislik saldırılarında kullanılan tespit yöntemlerinin analizi

[30]	Oltalama	RO	11.055 web sitesinin	0.965
	Saldırıların	DVM	içeriği veri seti olarak	0.932
	Tespiti	KNN	kullanılmıştır.	0.931
		KA		0.961
[66]	Web	Multilayer Perceptron	1000 web sitesi içeriği	0.740
	uygulamalarında	DVM	veri seti olarak	1.0
	güvenlik	KA	kullanılmıştır.	0.695
	açıklıklarına neden	NB		1.0
	olan kötü amaçlı uygulamaların tespiti	RO		
[67]	Nefret söylemi tespiti	Sıralı Minimal Optimizasyon (SMO)	Turkish hate speech dataset	En yüksek doğruluk oranının
		NB	(https://github.com/ima-yda/turkish-hate-speech-dataset-1)	SMO yönteminde
		KA		0.799 olduğu
		RO		belirlenmiştir.
[31]	Sahte haberlerin tespiti	LSTM ile	6335 haber başlığı ve içeriğinden oluşan real_or_fake	0.914
			(https://www.kaggle.com/datasets/rchitic17/real-or-fake) veri seti kullanılmıştır.	
[68]	E-posta spam tespiti	DDİ tabanlı rastgele orman yöntemi (NLP-RF)	Spam e-posta dataset	-

5. DERİN ÖĞRENME

Derin öğrenme, en basit tabiriyle insan beyninin işlevlerini bir bilgisayarın yapmasını sağlayan yapay zekâ yöntemidir. Makine öğreniminin bir alt kümesi olan DÖ, yapay zekânın yaygınlaşması ve popüler hale gelmesiyle geniş bir çalışma alanı bulmakla birlikte bilgisayarların insan beynine benzer şekilde davranabileceklerini de göstermiş olmaktadır. Bu kapsamda derin öğrenme teknolojisi günlük yaşantıyı kolaylaştıran ve sürekli olarak kullanılan dijital asistanlar, sesle yazma özellikleri ve otomatik yüz algılama uygulamalarından tıp ve havacılık alanlarına kadar geniş bir alanda kullanılmaktadır.

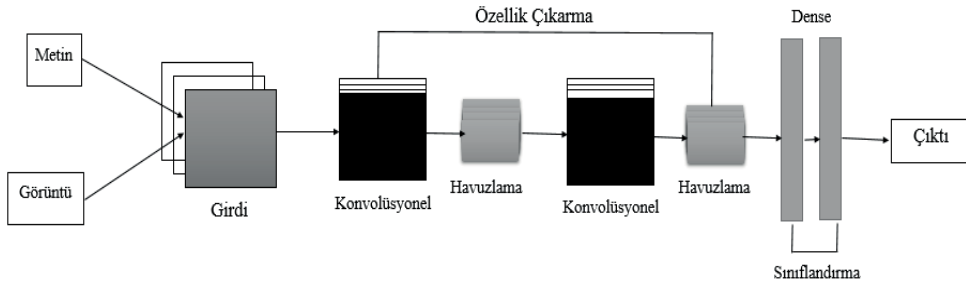
İnsan beyni birbirine bağlı milyonlarca nöronları olan ve herhangi bir şeyi öğrenmek için bu nöronların birbirleri ile etkileşime geçtiği karmaşık bir yapıyı ihtiva eder. DÖ algoritmaları da verinin öğrenilmesi için insan beyninde olduğu gibi birden çok yapay sinir ağının birbiri ile etkileşimini kullanmaktadır. Bu yapay sinir ağları da verileri işlemek için matematiksel hesaplamalar yapan birden çok yapay nöronlardan oluşmaktadır. Büyük çerçevede ise bir DÖ algoritmasının girdi katmanı, gizli katman ve çıktı katmanı olmak üzere üç katmanı bulunmaktadır. Yukarıda açıklanan yapay sinir ağları ve yapay nöronlardan meydana gelen bu katmanların her birinin verinin öğrenilmesinde farklı görevleri bulunmaktadır. Girdi katmanı dış dünyadan verileri alarak derin öğrenme alanına veriyi dahil eden ilk katmandır. Bu katmanda dış dünyadan alınan veriler bir diğer katman olan gizli katmana iletilir. Gizli katman girdi katmanından kendisine iletilen verilere farklı dönüşümler uygulayan nöronlara sahiptir. Gizli katman ve nöron sayısı arttıkça derin öğrenmenin verimi de aynı oranda artmaktadır. Bu katmanda farklı dönüşümler uygulanarak işlenen veriler en son gizli katmandan çıktı katmanı olarak adlandırılan son katmana iletilmektedir. Bu katman veriden istenilen tahmini oluşturan ve sonuç olarak kullanıcıya sunan son katmandır. Bu kapsamda özetlemek gerekirse bir verinin bilgisayara öğretilmesi için dış dünyadan alınan veri sırasıyla girdi, gizli ve çıktı katmanından geçerek öğrenilmekte ve kullanıcıya istenilen sonucu sunmaktadır.

Yukarıda bahsedildiği üzere DÖ ile etiketsiz verilerde sonuç elde etmek her ne kadar daha avantajlı olsa da bu avantajın beraberinde getirdiği bazı zorluklar da mevcuttur. Bu kapsamda bir DÖ teknolojisinden verim alınmak isteniyorsa büyük miktarda yüksek kaliteli veri ve bu verileri işleyebilecek büyük işlem gücüne sahip altyapının hazırlanması gerekmektedir.

5.1. Evrişimsel Sinir Ağı

Evrişimsel Sinir Ağları (CNN), görüntü işleme ve doğal dil işleme alanlarında özellik çıkarmak amacıyla kullanılan bir derin öğrenme algoritmasıdır. ESA'da giriş verilerinin belirlenmiş olan filtre boyutları ile örtüşmesine göre çıkış bilgileri elde edilmektedir.

Konvolüsyonel katman, havuzlama katmanı ve dense (Tam Bağlantı) katmanlarından oluşmaktadır. Mimariye verilen görüntü ya da metin şeklinde olan girdi konvolüsyonel katmandan geçirilerek özellikler ve bilgiler çıkarılır. Bu katmanda özellikler çıkarıldığı için girdi verisinin boyutu artmaktadır. Verinin boyutunu küçültmek amacıyla veri havuzlama katmanından geçirilir. Genellikle maksimum ve ortalama havuzlama katmanları kullanılmaktadır. Maksimum havuzlama katmanında belirlenen filtre boyutunun örtüştüğü alandaki en büyük değeri, ortalama havuzlama katmanında ise belirlenen filtre boyutundaki değerlerin ortalaması alınmaktadır. Böylece verinin boyutu azaltılır ve önemli özellikler elde edilmiş olunur. Veri seti, mimaride belirlenen sayıdaki konvolüsyonel ve havuzlama katmanlarından geçirildikten sonra elde edilen matrisi vektör haline dönüştürerek çıktıyı elde etmek için dense katmanından geçirilir. Dense katmanında bir katmandaki düğümler bir sonraki katmandaki bütün düğümlere bağlıdır ve her birinin ayrı bir ağırlığı bulunmaktadır. Şekil 5.1’de Evrimsel Sinir Ağı mimarisi verilmektedir.



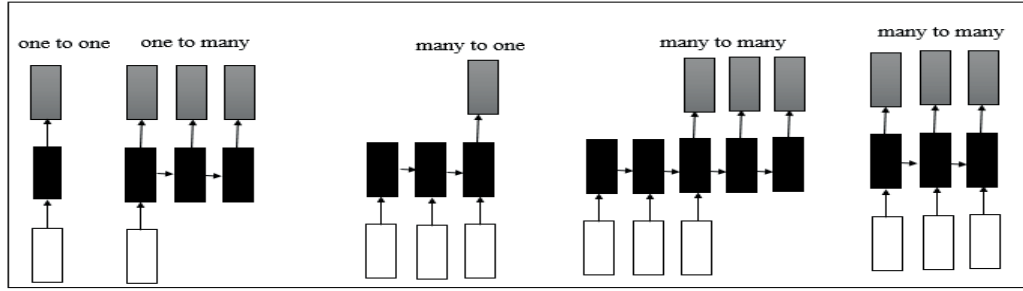
Şekil 5.1. CNN mimarisi

5.2. Tekrarlayan Sinir Ağı Mimarileri

Bizler konuşurken ya da bir yazıyı okurken kelimelerin cümle içerisindeki sıralaması ve cümlelerin metin içerisindeki sıralaması çok önemlidir. Eğer metinde yer alan tüm kelimeler karıştırılırsa metin bütün anlamını yitirebilir. Normal sinir ağıları kullanıldığı zaman tam olarak bu işlem yapılmaktadır. Hangi verinin hangi veriden önce ya da sonra geldiği dikkate alınmamaktadır. Sadece o anda gelen veri dikkate alınarak işlem yapılmaktadır. Bir input geldiği zaman o inputa bağlı output verilmektedir. Sinir ağıları bir önceki veriyi unutup o andaki veri girişine göre çıktı vermektedir. Ancak DDİ ile çalışma noktasında girdi oldukça önem kazanmaktadır. Sinir ağının bir hafızası olmalıdır ki önceki girdi verileri hakkında aldığı bilgileri saklayabilsin ve kullanılabilsin. Tekrarlayan Sinir Ağları (RNN) ile yapay sinir ağına bir hafıza eklenmektedir. Böylece sinir ağı daha önce aldığı girdileri göz önünde bulundurarak bir çıktı üretecektir. RNN içerisinde kendine beslenen bir döngü bulunmaktadır ve burada ağın içerisindeki birçok nörona bakılmaktadır. Nöronlar zaman içerisinde bir önceki nöronla bağlandığı için şu anda çalışan sinir

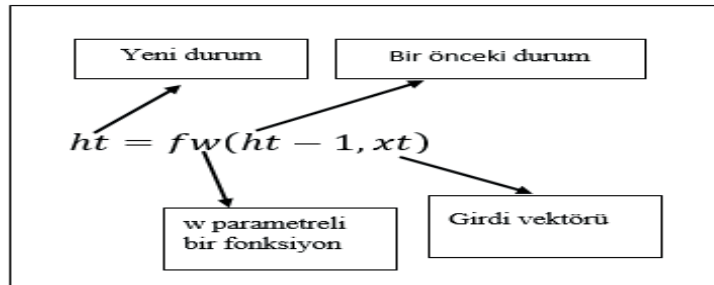
ağı bir önceki zamanda çalışan sinir ağından bilgi alabilmektedir. Bu sayede bilgi akışı sağlamaktadır.

RNN’ler farklı mimarilerde görülebilir ve kullanılabilir. Şekil 5.2’de de verildiği gibi öncelikle “one-to-one” mimarisi bir tekrarlayan sinir ağı değil standart sinir ağıdır. Belirli boyutta bir girdi alarak belirli bir boyutta çıktı vermektedir. İkinci sırada yer alan “one-to-many” ise bir tekrarlayan sinir ağıdır. Tek bir girdi alıp birden çok çıktı üretmektedir. Tam tersi olarak “many-to-one” ile birden çok girdi alıp tek bir çıktı elde edilebilir. Ardından “many-to-many” ağlarında Şekil 5.2’de de görüldüğü üzere farklı sayıda girdiler verilerek farklı sayıda çıktılar elde edilmektedir. Ancak ikisinin birbirinden farkı bir önceki many to many mimarisinde önce girdi işlenip ardından çıktı elde edilmekte diğer many to many mimarisinde ise her işlenen girdi için bir çıktı üretilmektedir.



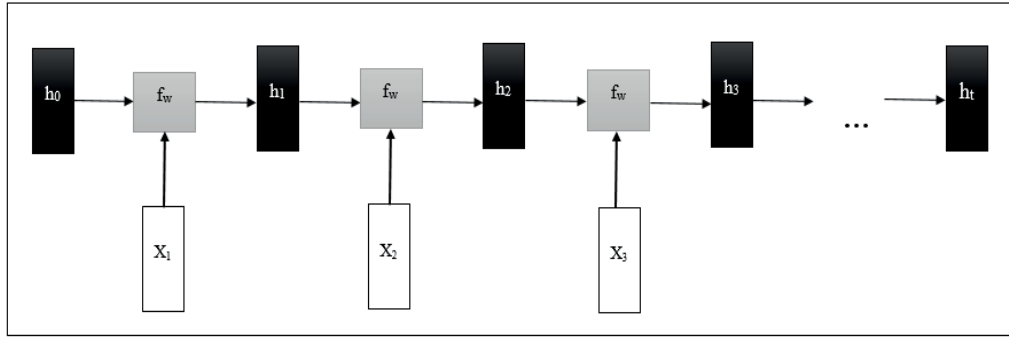
Şekil 5.2. Tekrarlayan sinir ağı mimarileri

Şekil 5.3’te verildiği üzere RNN bir girdi alır ve ardından bu girdi üzerinde işlemler gerçekleştirir. Aynı zamanda “Hidden State” denilen bir hafızası bulunmaktadır. Yani RNN her girdi aldığı anda “Hidden State” güncellenmektedir. Model yeniden farklı bir girdi aldığı anda ise önceki “Hidden State” bilgisini de alarak daha önceki işlemlerde ne olduğu hakkında fikir sahibi olmasını sağlamaktadır.



Şekil 5.3. RNN fonksiyon yapısı

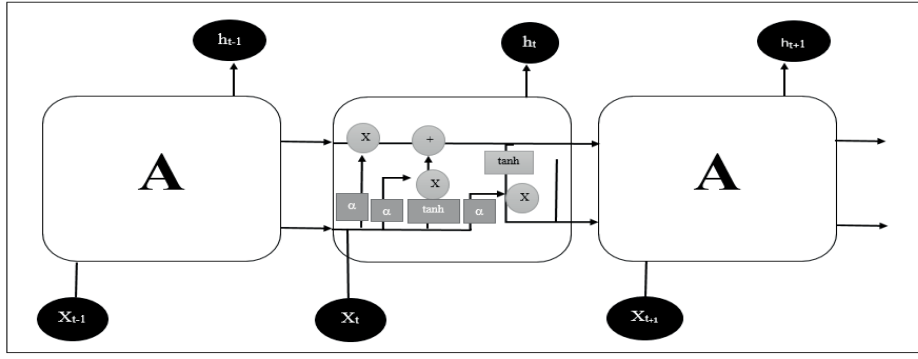
Şekil 5.4'te yapısı görülen RNN çalışmaya başladığında başlangıçta tamamı sıfırlardan oluşan bir h_0 durumu bulunmaktadır. RNN, X_1 isimli girdi verisi f_w fonksiyonuna yani tekrarlayan sinir ağına verilmektedir. Sonuç olarak h_1 isimli yeni bir durum oluşturulur. Girdi olarak X_2 verisi verildiğinde aynı zamanda bir önceki adımda üretilen h_1 durumu da verilmektedir. Sonuçta f_w aldığı h_1 durumunu güncelleyerek h_2 durumunu üretmektedir. Bu işlem bu şekilde tekrar tekrar uygulanmaktadır. Her adımda weight parametresi sabit kalmaktadır.



Şekil 5.4. RNN grafiksel gösterimi

5.2.1. Uzun Kısa Dönemli Hafıza (LSTM)

Uzun Kısa Dönemli Hafıza(LSTM) “Cell State” ile bilgi akışını bir hücreden diğer hücreye sağlıklı bir şekilde gerçekleştirmektedir. LSTM, Cell State’den gerekli görmediği bilgileri çıkarabilir ya da bazı yeni bilgiler ekleyebilir. Böylece kendisine gelen bilgiyi Cell State’yi düzenleyerek bir sonraki hücrelere aktarabilir. LSTM, Cell State’i güncelleyebilmek için bazı geçitlere ihtiyaç duymaktadır. Bu geçitlerden ilki olan “Forget Gate” LSTM, kendisine gelen bilgilerden hangisini unutmaya gerektiğine karar vermektedir. Bu işlemi ise 0 ile 1 arasında sıkıştırma yapan sigmoid fonksiyonunu kullanarak gerçekleştirmektedir. Yani 0 geldiyse unut, 1 geldiyse geç demektir. Input Gate ise Cell State’de hangi bilgilerin saklanacağına karar vermektedir. Bütün işlemler tamamlandıktan sonra Cell State güncellenir. Toplama yapılarak yeni bilgiler Cell State eklenir ve bir sonraki hücreye aktarmak için hazır hale getirilir. Son olarak Output Gate ise çıktının ne olması gerektiğine karar verilir ve bu çıktı Cell State bağlı olur. Cell State filtre uygulanarak sigmoid katmanında Cell State neleri çıktı olarak verecek ise belirlenir. Ardından “tanh” geçirilerek Cell State’de yer alan değerler -1 ile 1 arasında sıkıştırılır ve sigmoidden çıkan değerlerle çarpılarak sadece Cell State’in istenilen parçaları çıktı olarak verilmektedir. Şekil 5.5’te LSTM mimarisinin çalışma şekli görülmektedir.



Şekil 5.5. LSTM çalışma mantığı

5.2.2. Çift Yönlü Uzun Kısa Dönemli Hafıza (BiLSTM)

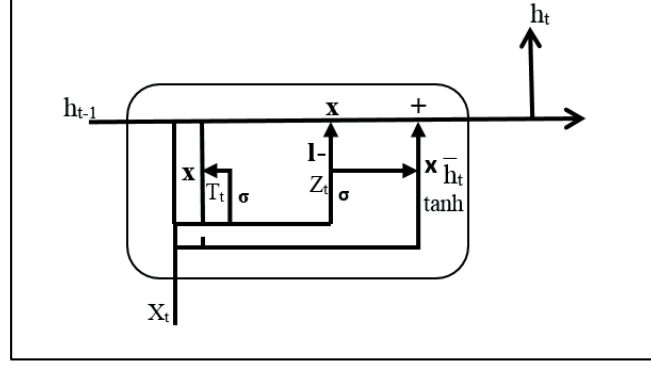
Bidirectional LSTM temelde iki bağımsız RNN'i bir araya getirmektedir. BiLSTM yapısıyla ağların zaman adımı dizi hakkında ileri ve geri yönlü olarak bilgiye sahip olunması sağlanmaktadır. Çalışma prensibi verilerin birinin geçmişten geleceğe diğerinin de gelecekte geçmişe işlenmesine dayanmaktadır. Bu kapsamda BiLSTM'i, LSTM'den ayıran en önemli fark iki gizli durumun birleştirilerek kullanılmasıdır. BiLSTM tarafından geçmişe ya da geleceğe ait bu bilgiler herhangi bir zamanda korunmaktadır. Genellikle DDİ alanında kullanılan BiLSTM'ler bağlamı daha iyi anladıkları için oldukça iyi sonuçlar vermektedir. Örneğin bir cümledeki sonraki kelime tahmin edilmek isteniyorsa LSTM tek yönlü olduğu için sadece elindeki veriye bağlı olarak tahmin etmeye çalışmaktadır. Ancak BiLSTM gelecekte gelen bilgileri de kullanarak sonraki kelimenin ne olduğunu tahmin etmeye çalışmaktadır. Özetlemek gerekirse BiLSTM ile bir kelimeyi anlamak için önceki ve sonraki kelimeler kullanılarak cümle bağlamına en uygun sonuç elde edilmeye çalışılmaktadır.

5.2.3. Gated Recurrent Unit (GRU)

GRU, bir tür tekrarlayan sinir ağıdır. Ancak GRU'nun bazı durumlarda LSTM'e göre daha avantajlı olduğu kabul edilir. Örneğin LSTM'den daha az bellek kullanır ve daha hızlıdır. Tekrarlayan Sinir Ağlarının kaybolan gradyan sorunu GRU'lar tarafından ele alınmaktadır. Bu kapsamda sinir ağındaki katmanın öğrenememesi durumunda RNN'lerin unutma sorununu çözmek için Update Gate ve Reset Gate olmak üzere iki kapı kullanılmaktadır. Bu kapılarla hangi verinin çıktıya geçeceğine karar verilir ve sorun çözülmüş olur.

Reset gate ve update gate adı verilen iki geçit mekanizması ağı içindeki ve dışındaki bilgi akışını kontrol etmek için kullanılmaktadır. Bu geçit mekanizmalarından Reset Gate kendisine gelen verinin ne kadarının unutulması gerektiğine karar verirken Update Gate ise bu verinin ne

kadarının kullanılması gerektiğine karar vermektedir. Çıktı ise Update Gate'den çıkan veri durumuna göre hesaplanmaktadır. Şekil 5.6'da GRU mimarisine ait şekil görülmektedir.



Şekil 5.6. GRU mimarisi

6. DOĞAL DİL İŞLEME TEKNİKLERİ

Doğal Dil İşleme (DDİ), yapay zekâ, MÖ ve dil biliminin ortak olarak ilgilendiği konu olmakla birlikte aynı zamanda yapay zekânın da alt dalıdır. DDİ, gerçek hayatta insanların konuştuğu dilleri kapsamaktadır. DDİ ile insanların konuştuğu dil genelde metin halinde bilgisayara verilerek bilgisayarın anlaması ve metin üzerinde istenilen işlemleri yapması amaçlanmaktadır. DÖ'nün son zamanlarda yaygınlaşması ile DDİ alanında da gelişmeler kaydedilerek DDİ'nin çalışma alanı gün geçtikçe genişlemektedir.

Dünya üzerinde kaç dil konuşulduğu kesin olarak bilinmemekle birlikte bu dillerin tamamı insanların ömürleri boyunca edindiği görsel, durumsal ve işitsel tecrübeler ile anlam kazanmaktadır. Öyle ki farklı vurgularla söylenen kelimeler karşı taraf için farklı anlamlar ifade edebilirken bazen de cümlenin tamamlanması karşı tarafın deneyimlerine bırakılmaktadır. Bu nedenlerle doğal diller makine dillerine göre oldukça fazla kelimeye, esnekliğe ve kuralla sahiptir. Makine dillerini kullanırken her değişken tanımlanıp belirlenen kurallar çerçevesinde kodlamalar yapılırken doğal dillerde kelime sayısının oldukça fazla olması, bir kelimenin birden fazla anlamının olması ve kelime oyunları gibi bilgisayara öğretilmesi oldukça zor olan durumlar mevcuttur. Bu gibi durumlar sebebiyle DDİ oldukça kaotik ve zorlu bir süreç içermektedir. Tablo 6.1'de makine dilleri ve doğal dillerin ayrıntılı karşılaştırılması verilmektedir.

Tablo 6.1. Makine dilleri ve doğal dillerin karşılaştırılması

Makine Dilleri	Doğal Diller
C#, Ruby, Python vs.	Türkçe, İngilizce, Rusça vs.
Az sayıda kelime bulunur.	Çok sayıda kelime bulunur.
Belirli kurallar çerçevesinde oluşturulmuştur ve bu kuralların dışına çıkılamaz.	Kesin kuralları bulunmamaktadır.
	Sınırsız sayıda kelime kombinasyonları olabilir.

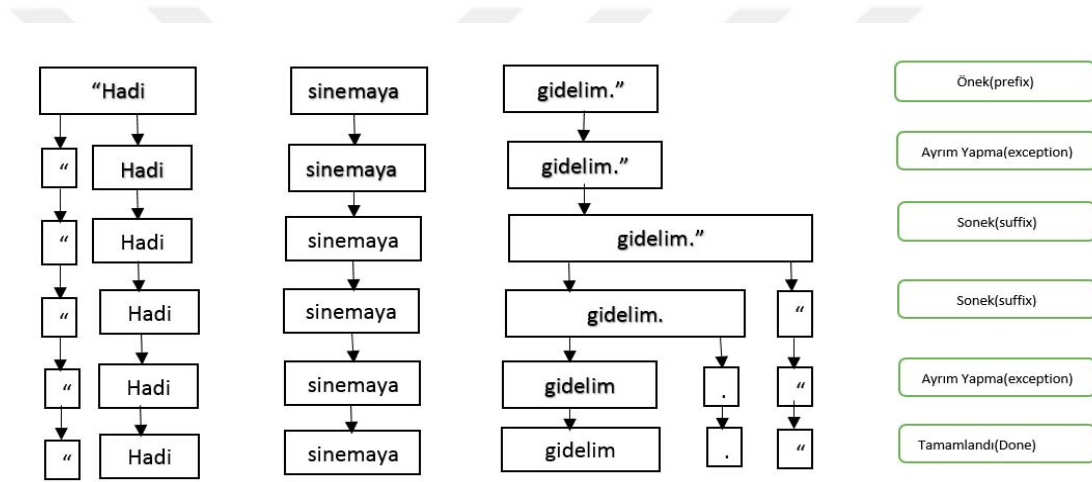
6.1. Corpus

DDİ'de corpus (derlem) veri seti olarak verilen bir metnin koleksiyonudur. Corpus DDİ içerisinde her şey olabilir. Örneğin Web Scraping (Web Kazıma) ile "sıkayetvar.com"dan çekilen yorumlar bir excel dosyası içerisinde veri seti haline getirildikten sonra DDİ'de corpus olarak kullanılabilir. Corpus bir metin belgesini içerirken birden çok metin belgesi içeren yapı ise "Corpora" olarak adlandırılır.

6.2. Natural Language Toolkit (NLTK) ile Tokenizasyon

DDİ alanında çalışmalar gerçekleştirilirken NLTK kütüphanesi kullanılır. Bu kütüphane ile DDİ yapılırken ilk olarak corpus içerisinde yer alan paragraflar cümlelere ya da cümleler kelimelere ayrılır. Ayrılan her bir cümle ya da kelime “token” adını alır. Bir sonraki adım olan tokenizasyon işleminde ise token haline dönüştürülen metin belgesi integer karşılıklarına dönüştürülür. Bu aşamada metin belgesi integer değerlere dönüştürülerek aslında metin, bilgisayarın anlayacağı dile dönüştürülmüş olmaktadır.

Şekil 6.1’de verildiği gibi üzerinde DDİ yapılacak olan veri bir cümleden oluşmaktadır. Bu cümle kelimelere ayrılmış ve cümlede bulunan noktalama işaretleri kelimelerden ayrıştırılmış böylece tokenizasyon süreci tamamlanmıştır.



Şekil 6.1. Tokenizasyon süreci

6.3. NLTK ile StopWords

Veri analizi yapılırken analize başlamadan önce veriler üzerinde ön işleme gerçekleştirilir. NLTK kütüphanesi, verilerin ön işlemlerden geçirilmesine yardımcı olmaktadır. Veri ön işleme aşamalarından birisi ise “StopWords” olarak bilinen Corpus içerisinde yer alan öznelilik çıkarımı yapılırken gereksiz kelimelerin (ben, sen, -de, -da, -ki, ile, ...) temizlenmesi aşamasıdır. Bu kelimeler veri setinde yer kaplamakta ve eğitim aşamasında daha fazla bellek kapladığı için eğitim süresini uzatmaktadır.

6.4. NLTK ile Stemming

Veri analizi yapılırken gerçekleştirilen ön işlemlerde bir diğeri ise stemming (kök alma) işlemidir. Stemming ile aldığı eklerle gövde haline gelmiş olan bir kelime kök haline getirilir.

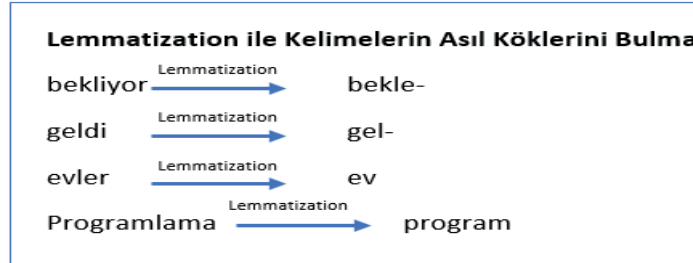


Şekil 6.2. Stemming ile veri ön işleme

Şekil 6.2’de görüldüğü gibi Stemming işlemi yapılırken kelimenin asıl anlamı dikkate alınmadığı için sadece kelimedeki bulunan ekleri bulup bu ekleri kelimedenden çıkarılmaktadır. Bu işlemin ön işleme gerçekleştirilen verilerle yapılan çalışmalarda karmaşıklıklara neden olmaktadır.

6.5. NLTK ile Lemmatization

Veri analizi gerçekleştirilirken kullanılan ön işleme işlemlerinden bir diğeri ise lemmatization işlemidir. Bu işlemde kelimenin kökü bulunur. Ancak Stemming işleminden farklı olarak bu yöntemde kelimedeki var olan ekler çıkarılırken kelimenin asıl anlamı esas alınır. Dolayısıyla ekler atıldığı zaman ortaya kelimenin asıl kökü çıkmış olmaktadır. Şekil 6.3’te açıklanan aşamaların uygulaması görülmektedir.



Şekil 6.3. Lemmatization ile veri ön işleme

6.6. Kelime Vektörleri

DDİ’nin temel sorunu kelimelerin anlamlarını makinelerin anlayabileceği şekilde sembolize etmektir. Makinelerin kelimeleri anlayabilmesi için kelimeler sayısal olarak ifade edilmelidir. Ancak sayısal olarak ifade edildikten sonra DDİ ile kelimeler anlaşılabilir hale gelmektedir. Kelimeleri sayısal olarak ifade edebilmek için ise kelime vektörlerine ihtiyaç duyulmaktadır.

Tek elemanı 1, geriye kalan tüm elemanları 0 olan vektörler “one-hot” vektörler olarak adlandırılır. Kelime sayısı ile vektörlerin boyutları aynı sayıda olmalıdır. Örneğin on kelime var ise vektörlerin uzunluğu da on olmalıdır.

Gül	=	[1, 0, 0, 0, 0, 0, 0, 0, 0, 0]
Lale	=	[0, 1, 0, 0, 0, 0, 0, 0, 0, 0]
Karanfil	=	[0, 0, 1, 0, 0, 0, 0, 0, 0, 0]
Zambak	=	[0, 0, 0, 1, 0, 0, 0, 0, 0, 0]
Nergis	=	[0, 0, 0, 0, 1, 0, 0, 0, 0, 0]
Sümbül	=	[0, 0, 0, 0, 0, 1, 0, 0, 0, 0]
Nilüfer	=	[0, 0, 0, 0, 0, 0, 1, 0, 0, 0]
Kasımpatı	=	[0, 0, 0, 0, 0, 0, 0, 1, 0, 0]
Lilium	=	[0, 0, 0, 0, 0, 0, 0, 0, 1, 0]
Papatya	=	[0, 0, 0, 0, 0, 0, 0, 0, 0, 1]

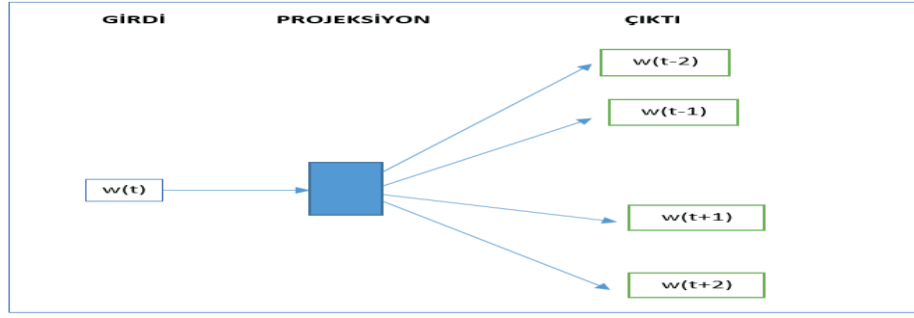
Şekil 6.4. Kelimelerin vektörlerle temsil edilmesi

Şekil 6.4’te on adet çiçek ismi mevcuttur. Her indeks bir çiçek ismine karşılık gelmektedir ve kelime sayısı kadar vektör oluşturulmaktadır. Her kelimenin karşılık geldiği indeks değeri one-hot vektör kullanılarak 1 ile temsil edilip makinenin anlayabileceği sayısal karşılıklara dönüştürüldü.

6.6.1. Kelime Vektörleri – word2vec

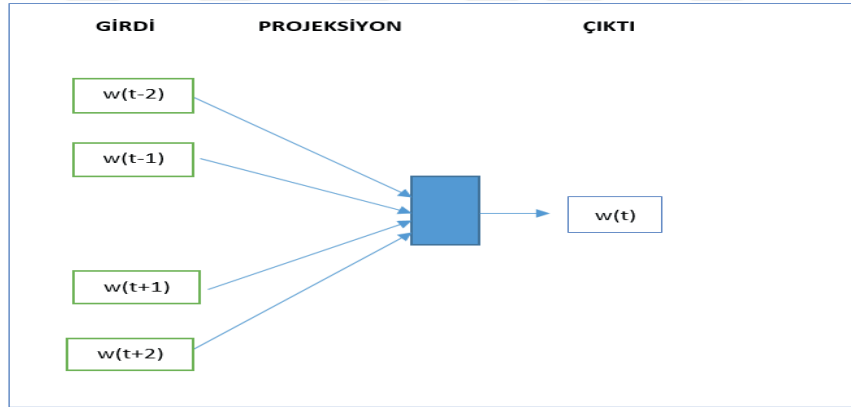
İnsanlar cümle içerisinde geçen herhangi bir kelimenin anlamını bilmediği zaman anlam çıkarabilmek için o kelimenin cümle içerisindeki kullanımına ve yerine bakarak yorumlamaktadır. Word2vec ile de bir kelimenin etrafındaki kelimeler kullanılarak asıl kelimenin anlamının bulunması sağlanmaktadır. Bir corpus içerisinde bulunan cümlelerdeki herhangi bir kelimenin anlaşılabilmesi için o kelime ve etrafındaki kelimeler birer vektöre dönüştürülür. Ardından bu vektörler makinelere öğretilmek için “Gensim” kütüphanesi kullanılarak eğitilir. Eğitim sonucunda loss kayıpları sıfıra yakın ise sonuç oldukça isabetli olur. Ancak loss kayıpları sıfırdan uzaklaştıkça modelin kelimeyi doğru tahmin etme oranı da aynı doğrultuda düşmektedir. Word2vec ile kelime vektörleri oluşturulmak için iki farklı algoritma geliştirilmiştir:

- **Skip-gram:** Başlangıçta algoritmaya hedef kelime ($w(t)$) verilmektedir. Ağırlık matrisi ile hedef kelime arasında gizli bir katman bulunmaktadır. Gizli katman nokta çarpımı sonucunu çıktı katmanına iletmektedir. Burada çıktı katmanı, gizli katmanın çıkış vektörü ile çıkış katmanının ağırlık matrisi arasındaki nokta çarpımını hesaplamaktadır. Şekil 6.5’te de görüldüğü üzere bu algoritma ile ortada yer alan kelimeye bakılarak etrafındaki kelimeler tahmin edilmektedir.



Şekil 6.5. Skip-gram algoritması

• **Continuous Bag of Words Model (CBOW):** CBOW mimarisinde bir giriş katmanı, bir gizli katman ve bir çıkış katmanı bulunmaktadır. Giriş katmanı vektörlerle ifade edilmektedir ve her bir vektör bir kelimeye karşılık gelmektedir. Gizli katman ise öğrenme olayını gerçekleştirilmektedir. Son olarak hedef kelimenin tahmin edilebilmesi için çıktı katmanı kullanılmaktadır. Şekil 6.6’da görüldüğü gibi Skip-gram algoritmasının tam tersine CBOW kenardaki kelimeleri kullanarak ortadaki kelimeyi tahmin etmektedir.



Şekil 6.6. CBOW algoritması

6.6.2. Kelime Vektörleri – Glove

Açılımı Global Vectors olan Glove tüm corpus içerisinde bir matris çıkarmaktadır. Ardından matris üzerindeki her kelime gezilmektedir. Her aşamada tek tek vektörler optimize edilmektedir. Corpus içerisinde yer alan en çok tekrar eden kelimeler için üst limit konulmaktadır. Bu şekilde stopwords belirli noktaya kadar sayılmaktadır. Belirlenen sınırdan fazla tekrar eden kelimeler varsa yok sayılmaktadır. Böylece hızlı bir eğitim gerçekleştirilmektedir. Corpus üzerinde tek tek

gezilip güncelleme yapmak yerine kelimelerin corpusta kaç defa geçtiği hesaplanmaktadır. Ardından optimizasyon yapılmaktadır. Glove ile,

- Hızlı eğitim gerçekleştirilmektedir.
- Çok büyük corpuslarda çalışılabilmektedir.
- Küçük corpus ve vektörlerde iyi performans elde edilmektedir.

Tablo 6.2’de skip-gram, word2vec ve glove ile ilgili karşılaştırma verilmektedir.

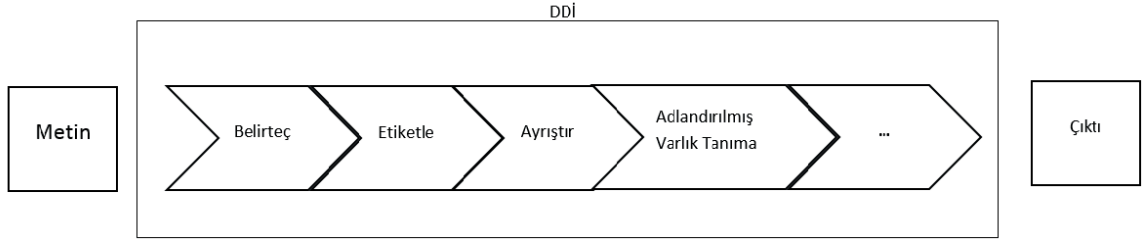
Tablo 6.2. Kelime vektörlerinin karşılaştırılması

Skip-gram	Glove	Word2vec
Metni kelime kelime gezerek iki kelimenin birbiriyle ne kadar sık yan yana bulunduğu hesaplanmaktadır.	Glove ise iki kelimenin ne kadar sık bulunduğu dair istatistik çıkarmaktadır.	-
-	Glove bahsedilen kelimelere ait istatistikleri elde ettikten sonra optimizasyonu gerçekleştirmektedir.	Word2vec kelime kelime gezerken aynı zamanda optimizasyon yapmaktadır

6.7. Doğal Dil İşlemenin Çalışma Mantığı

DDİ ile doğal diller bir makinenin anlayabileceği hale getirilir. Makinenin anlamasını sağlamak için veri belirli işlemlerden geçirilir ve çıktıya ulaşılır. Bu işlemler DDİ’nin çalışma mantığını oluşturur.

DDİ’de çalışılmak istenen dile ait metin sisteme girdi olarak verilir ancak sistem bu dili anlayamayacağı için metin bazı işlemlere tabi tutulur. Öncelikle makineye okutulan metin cümleler ya da kelimeler haline parçalanarak her bir cümleye ya da kelimeye karşılık gelen tam sayı değerleri atanır. Ardından “embedding” işlemi uygulanır. Bu işlem ile tam sayı halinde bulunan tokenler için vektörler oluşturulur ve bu vektörlerin boyutlarının her kelime için eşit olması sağlanır. Vektörler oluşturulduktan sonra verinin eğitileceği “RNN” ya da “CNN” belirlenip veriden elde edilmek istenen çıktıya uygun konuşma parçası etiketleme, duygu analizi ya da adlandırılmış varlık tanıma gibi DDİ çıktıları elde edilir. Şekil 6.7’de görüldüğü üzere DDİ sürecine tabi tutulacak metin girdi olarak verildikten sonra sayısallaştırma işlemi gerçekleştirilir. Sayısallaştırılan veriye çeşitli DDİ algoritmaları uygulanarak istenilen çıktılar elde edilmektedir.



Şekil 6.7. Doğal dil işleme süreci

6.8. Yapay Zekâ ile Doğal Dil İşleme Arasındaki İlişki

DDİ, yapay zekânın bir alt dalı niteliğindedir. DDİ insanların konuştuğu doğal dilleri makinelerin anlayacağı şekle dönüştürerek insan ve makine arasındaki iletişimi sağlamak amacıyla yapay zekâ yöntemlerine başvurmaktadır.

Yapay zekâ, insan zekâsı gibi davranabilen, bilgi toplayabilen, verilen verilerden istenilen durumlara göre çıkarımlar yapabilen bir teknolojidir. DDİ ise makine ve insan arasındaki iletişimi sağlamayı amaçlayan bir teknolojidir. DDİ'nin amacını gerçekleştirebilmesi için ilk adım olarak DDİ araçlarına bir veri seti verilir. Ancak burada DDİ'nin verilen veri setini anlayabilmesi ve çıkarımlar yapabilmesi için yapay zekâyı ihtiyaç duyulmaktadır. Yapay zekâ ve DDİ'nin temel ilişkisi bu noktada başlamaktadır. DDİ ile metin özetleme, duygu analizi, cümle işaretleme, konuşma parçası etiketleme ve metin madenciliği gibi işlemlere başlanmadan önce dili yöneten yapı ve kurallar yapay zekâ aracılığı ile anlaşılır hale getirilir. Bu işlem tamamlandıktan sonra DDİ'nin adımları uygulanmaya başlanır. Buradan da anlaşılacağı üzere yapay zekâ alanı MÖ, DÖ ve DDİ algoritmalarını kapsamaktadır. DÖ, MÖ'nin içerisinde yer alsa da aynı zamanda DDİ içerisinde de yer aldığı görülmektedir.

7. MATERYAL VE METOTLAR

Bu bölümde tez kapsamında yapılan çalışmalara yer verilmiştir. Tez kapsamında Doğal Dil İşleme Teknikleri Kullanılarak Hibrit CNN-BiLSTM Yöntemi ile Smishing Saldırılarının Tespiti, Hibrit CNN-GRU modeli ile SMS Phishing saldırılarının tespiti ve Doğal Dil İşleme Teknikleri Kullanılarak Twitter Verileri Kullanılarak Kısa Metinler Üzerinde GRU ile Nefret Söylemi Tespiti gerçekleştirilmiştir.

7.1. Doğal Dil İşleme Teknikleri Kullanılarak Hibrit CNN-BiLSTM Yöntemi ile Smishing Saldırılarının Tespiti

SMS mesajları, kullanıcının göndermesi için genellikle çok ucuz olup SMS spam bir mobil ağ üzerinden iletilir. Bu sayede SMS mesajları sosyal mühendisler açısından cazip hale gelmektedir. SMS spam, e-posta spam kadar yaygın olmasa da akıllı telefon kullanımının yaygınlığı düşünüldüğü zaman yine de akıllı telefon kullanıcılarını rahatsız eder. SMS mesajları oldukça yaygın ve geniş bir kullanım alanına sahip olması nedeniyle spammerların ilgi odağı halindedir. Aynı zamanda SMS spamleri kullanılarak telefon kullanıcılarının kişisel bilgilerine ulaşmaları oldukça kolaydır. Çünkü spammerlar bir mesaja yerleştirdikleri bağlantı ile akıllı telefon kullanıcılarını istedikleri alana doğru özellikle merak duygularını harekete geçirerek yönlendirmektedirler. Geniş kullanıcı ağına ve kullanım alanına sahip olması nedeniyle SMS spamlerini sınıflandırmak ve tespit etmek özellikle iletişimin ulaştığı boyut ile SMS ilişkisi düşünüldüğü zaman önem arz etmektedir.

Akıllı telefon kullanıcıları kendilerine gelen smishing niteliğindeki bir mesajda yer alan linke tıkladıkları zaman linke yerleştirilen kötücül yazılımı telefonlarına indirmiş olurlar. Kötücül yazılımı başarıyla hedefinin telefonuna yerleştiren sosyal mühendis için artık yapılması gereken tek şey ulaşabildiği bilgileri ele geçirmek ve kendine fayda sağlamak olmaktadır. Ancak her zaman smishing saldırıları bir bağlantı üzerinden gerçekleşmemekle birlikte geleneksel olarak merak uyandırıcı sorular, cezbedici teklifler üzerinden de yapılabilmektedir. Smishing saldırısına maruz kalan kullanıcılar için spam SMS'lerin tehlikeleri oldukça çoktur:

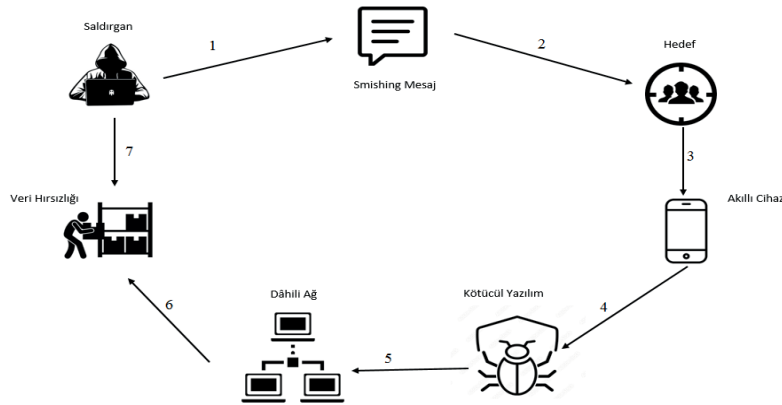
- İstenmeyen reklam
- Özel bilgilerin açığa çıkması
- Dolandırıcılık veya finansal planın kurbanı olma
- Kötü amaçlı yazılım ve kimlik avı web sitelerine girme
- Uygunsuz içeriklere istemsiz maruz kalma

Kullanıcılar bahsi geçen bu tehlikelerden bir tespit yazılımı içeren uygulamalar aracılığıyla korunabileceği gibi herhangi bir yazılıma ihtiyaç duymadan geleneksel yöntemlerle de korunabilir. Sıkça başvuru geleneksel korunma yöntemleri şunlardır:

- Göndericisi şüpheli olan ya da tanınmayan mesaj içeriklerinde yer alan linklere tıklamamak,
- Göndericisi şüpheli olan ya da tanınmayan mesaj içeriklerinde yer alan kişisel bilgilere yönelik soruları cevaplamamak,
- Göndericisi şüpheli olan ya da tanınmayan mesaj içeriklerinde yer alan banka hesap bilgilerine yönelik soruları cevaplamamak,
- Herhangi bir şekilde banka hesap bilgileri ya da kişisel bilgilere erişim sağlayan “ödüllere” inanmamak,
- Mesaj yoluyla gönderilen ve tavsiye edilen yazılımları yüklememek.

Yukarıda bahsedilen yöntemler ile her ne kadar smishing saldırılarından belirli oranda korunmak mümkünse de bu yöntemler bir tespit mekanizması içeren yazılım kadar iyi koruma sağlamamaktadır.

Geliştirilen Hibrit CNN-BiLSTM yapısı aracılığıyla bu tür smishing saldırısı içeren SMS’ler tespit edilerek kullanıcılara bilgi güvenliği sağlanmaktadır. Geliştirilen bu yapı ile sosyal mühendislerin kullanıcılara gönderdikleri SMS’lerin içeriklerinde spam olup olmadığı yüksek başarımları ile tespit edilmiştir. Şekil 7.1’de bir smishing saldırısının nasıl başlatıldığına ait yapı verilmektedir.



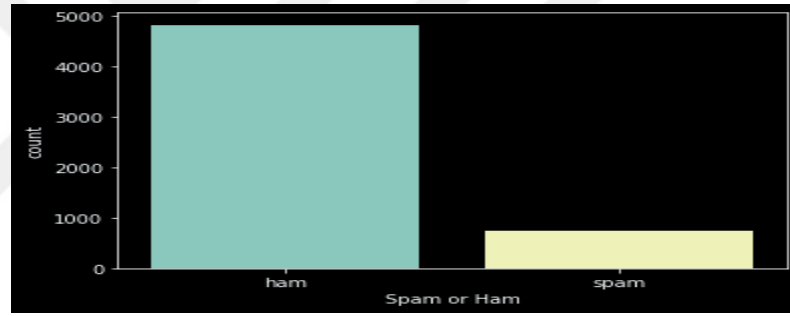
Şekil 7.1. Smishing saldırısının başlatılması

7.1.2. Geliştirilen Yöntem

Çalışmanın bu bölümünde kullanılan veri setine, veri setine uygulanan ön işlemlere, eğitim için geliştirilen modelin yapısına ve eğitim sonunda modelin doğruluk performansına ilişkin bilgilere detaylıca yer verilmiştir.

A. Veri Seti

Hibrit CNN-BiLSTM yöntemi kullanılarak smishing saldırılarının tespiti için geliştirilen çalışmada veri bilim platformu olan Kaggle’da [69] erişime açık olarak sunulan “spam.csv” isimli veri seti kullanılmıştır. Veri seti 5572 satır ve 2 sütundan oluşmaktadır. Sütunlar “v1” ve “v2” olarak iki özneliğe sahiptir. V1 olarak belirlenen sütunda SMS mesajların “spam” ya da “ham” olarak iki sınıfı yer almaktadır. V2 olarak belirtilen sütunda ise SMS mesajların içerikleri yer almaktadır. Veri seti içerisinde “Ham” olarak sınıflandırılan 4825 mesaj ve “Spam” olarak sınıflandırılan 747 mesaj bulunmaktadır. Veri setinin dağılım grafiği Şekil 7.2’de verilmiştir.



Şekil 7.2. Spam.csv veri seti dağılım grafiği

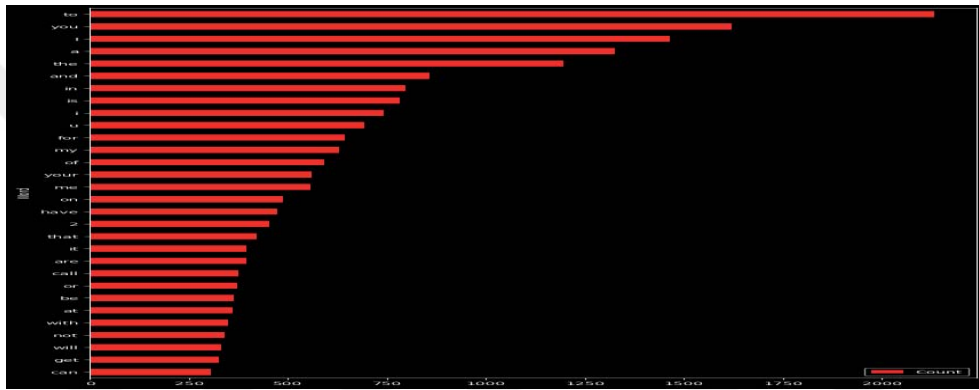
Veri setinin içeriğine ait belirli bir kısım ise Şekil 7.3’te verilmiştir.

```
v1,v2...
ham,"Go until Jurong point, crazy.. Available only in bugis n great world la e buffet... Cine there got amore wat..."
ham,Ok lar... Joking wif u oni....
spam,Free entry in 2 a wkly comp to win FA Cup final tkts 21st May 2005. Text FA to 87121 to receive entry question(std txt rate)T&C's apply 08452810075over18's...
ham,U dun say so early hor... U c already then say...
ham,"Nah I don't think he goes to usf, he lives around here though"
spam,"FreeMsg Hey there darling it's been 3 week's now and no word back! I'd like some fun you up for it still? Tb ok! XxX std chgs to send, $£1.50 to rcv"
ham,Even my brother is not like to speak with me. They treat me like aids patient..
ham,As per your request 'Melle Melle (Onu Minnaminungente Nuruugu Vettam)' has been set as your callertune for all Callers. Press *9 to copy your friends Callertune...
spam,WINNER!! As a valued network customer you have been selected to receive a £900 prize reward! To claim call 09061701461. Claim code KL341. Valid 12 hours only
spam,I'd your mobile 11 months or more? U R entitled to Update to the latest colour mobiles with camera for Free! Call The Mobile Update Co FREE on 08002986030...
ham,"I'm gonna be home soon and i don't want to talk about this stuff anymore tonight, k? I've cried enough today."
spam,"SIX chances to win CASH! From 100 to 20,000 pounds txt> CSH11 and send to 87575. Cost 150p/day, 6days, 16+ TsandCs apply Reply HL 4 info"
spam,"URGENT! You have won a 1 week FREE membership in our £100,000 Prize Jackpot! Txt the word: CLAIM to No: 81010 T&C www.dbuk.net LCCLTD POBOX 4403LU
ham,I've been searching for the right words to thank you for this breather. I promise i wont take your help for granted and will fulfil my promise. You have been wonderful an
spam,"XXXMobileMovieClub: To use your credit, click the WAP link in the next txt message or click here>> http://wap. xxxmobilemovieclub.com?n=QJKGIGHJJCBL"
ham,Oh k..i'm watching here:)
ham,Eh u remember how 2 spell his name... Yes i did. He v naughty make until i v wet...
ham,Fine if that's the way u feel. That's the way its gota be...
spam,"England v Macedonia - dont miss the goals/team news. Txt ur national team to 87077 eg ENGLAND to 87077 Try:WALES, SCOTLAND 4txt/141.20 PBOXox36504V
ham,Is that seriously how you spell his name?
ham,"I'm going to try for 2 months ha ha only joking...
ham,So L.. pay first lar... Then when is da stock comin....
ham,M's finish my lunch then i go str down lor. Ard 3 smth lor. U finish ur lunch already?
ham,!!!!!! Alright no way I can meet up with you sooner?
ham,Just forced myself to eat a slice. I'm really not hungry tho. This sucks. Mark is getting worried. He knows I'm sick when I turn down pizza. Lol,,
ham,Lol your always so convincing...
ham,Did you catch the bus ? Are you frying an egg ? Did you make a tea? Are you eating your mom's left over dinner ? Do you feel my Love ?...
ham,"I'm back &amp; we're packing the car now, I'll let you know if there's room"
ham,Take care...
```

Şekil 7.3. Spam.csv veri seti

B. Ön İşlemler

SMS mesajları içeren 5572 satır veriden oluşan veri kümesinde mesajlar “Spam” ve “Ham” olarak sınıflandırılmıştır. DDİ ile verilere önce anlamlı köklerini elde etmek için “Lemmatization” işlemi uygulanmıştır. Daha sonra SMS mesajları içerisinde yer alan ve stopwords kelimeler olarak adlandırılan etkisiz kelimeler veri setinden çıkarılmıştır. Veri seti içerisinde yer alan stopwords kelimelere ait dağılım grafiği Şekil 7.4’e verilmiştir. Aynı zamanda Şekil 7.4’e ait stopwordsler Tablo 7.2’de detaylıca gösterilmiştir. Stopword kelimeleri veri setinden temizlendikten sonra veri setinde yer alan sınıflar (Spam ve Ham) kategorik veriler olduğu için tokenize (sayısallaştırma) işlemi gerçekleştirilmiştir.



Şekil 7.4. Spam.csv veri setinde yer alan stopwordsler

Veri setinin %75'i eğitim ve %25'i test verisi olacak şekilde parçalanmıştır. Tüm veriye tokenleştirme işlemi uygulanmıştır. Modeli eğitmek için girdi olarak verilecek veri setinin tamamının kelime boyutunun aynı olması gerekmektedir. Veri setindeki her bir cümle kelimelere parçalanmış ve her bir kelimeye sayısal değer verilmiştir. Bu sayısal değerler içerisinde maksimum tokene sahip olan cümle dikkate alınarak tüm veri seti embedding vektörlere dönüştürülmüştür.

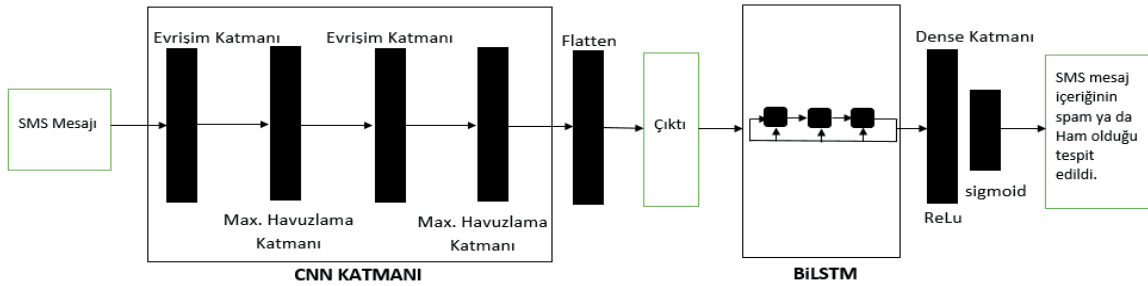
Veri seti üzerinde sayısallaştırma işlemi tamamlandıktan sonra keras kütüphanesi kullanılarak eğitim ve test veri setine padding eklenmiştir. Kelime ve kelimelerin tokenlerinin yer aldığı bir liste oluşturularak geri döndürülmüştür.

C. Hibrit CNN-BiLSTM Modelinin Oluşturulması

Veri seti geliştirilen hibrit sistemlerde eğitim için hazır hale getirildikten sonra modelin yapısı kurulmuştur. İlk olarak model sıralı bir şekilde oluşturulmuş ve girdi veri setini daha dar bir alanda sıkıştırmak için “Embedding” katmanı oluşturulmuştur. Verilerden uzamsal özellikleri çıkarabilmek için ikinci katman olarak CNN katmanı oluşturulmuştur. CNN katmanında

aktivasyon fonksiyonu olarak “ReLU” belirlenmiştir. Ardında maksimum havuzlama katmanı oluşturulmuştur. Modelin hem geriye hem de ileriye doğru bilgi sahibi olarak daha iyi sonuçlar vermesi için son olarak BiLSTM katmanı oluşturulmuştur. Bu katmanda aktivasyon fonksiyonu yine “ReLU” olarak tercih edilmiştir. Veri setinde yer alan veriler “Spam” ve “Ham” olmak üzere iki sınıfa ayrıldığı için “sigmoid” fonksiyonu kullanılmıştır. Optimizasyon algoritması olarak kerasta yer alan “Adam” algoritması ve performans metriği olarak “accuracy” belirlenmiştir. Tüm bu işlemlerden sonra 5, 10 ve 20 epochta olmak üzere veri seti 3 farklı epochta eğitilmiştir.

Spam.csv veri seti kullanılarak geliştirilen Hibrit CNN-BiLSTM modeli ile SMS mesajların spam olup olmadığını tespit eden uygulamaya ait akış şeması Şekil 7.5’te verilmiştir. Şekil 7.5’te görüldüğü üzere sisteme eğitmek üzere bir girdi verilmiştir. İlk olarak CNN katmanına tabi tutulan girdinin bu katmandaki çıktısı BiLSTM katmanına girdi olarak verilmiştir. Bu katmanda eğitimi tamamlanmış olan girdi Dense katmanında yer alan aktivasyon fonksiyonlarından geçirilerek sınıflandırılma işlemi tamamlanmıştır.



Şekil 7.5. Hibrit CNN-BiLSTM ile Smishing saldırılarının tespiti akış şeması

D. Modelin Başarımı ve Literatürdeki Çalışmalarla Karşılaştırılması

Geliştirilen uygulama ile Hibrit CNN-BiLSTM yöntemi kullanılarak SMS içeriğinin spam olup olmadığını tespit edilmesi amaçlanmıştır. Çalışma kapsamında Kaggle’da [70] yer alan açık kaynaklı “spam.csv” isimli 5572 satır ve 2 sütundan oluşan veri seti kullanılmıştır. Veri setine ilk olarak lemmatization ve stopwords işlemleri uygulanmıştır. Ardından veri seti tokenize (sayısallaştırma) edilmiştir. Ön işlemler tamamlanıp veri seti eğitime hazır hale getirildikten sonra Hibrit CNN-BiLSTM modeli ile eğitilmiştir. Eğitim sonucunda SMS içeriğinin spam olup olmadığını tespitindeki başarımı %98.87 olarak saptanmıştır. Tablo 7.1’de ise literatürde yer alan aynı veri setine uygulanan farklı yöntemlerin başarımlarına ait karşılaştırma çalışması verilmiştir.

Tablo 7.1. “spam.csv” veri setine uygulanan algoritmaların doğruluk performansı

Literatürde Kullanılan Algoritma	Doğruluk Ölçütü
LogisticRegression [71]	0.97
MultinomialNB [71]	0.97
SVC [71]	0.98
RandomForestClassifier [71]	0.97
SMS Spam Detection using LSTM [70]	0.94
Geliştirilen Hibrit CNN-BiLSTM çalışması	0.9987

7.2. Hibrit CNN – GRU Yöntemi ile SMS Phishing Tespiti

Smishing bir SMS metninin içerisine yerleştirilen bağlantı aracılığıyla mobil cihaz kullanıcılarının başta merak duygusu olmak üzere çeşitli duygularını harekete geçirerek kişisel verilere, hesap verilerine ulaşmayı sağlayan bir saldırı türüdür. Akıllı telefonların yaygınlaşmasıyla sosyal mühendisler kişisel verileri ele geçirebilmenin çeşitli yollarını üretmişlerdir. Smishing ile de akıllı telefon kullanıcılarının telefonlarına yerleştirilen kötücül yazılımlarla fotoğraflarına, banka hesap bilgilerine, e-posta içeriklerine üçüncü kişilerin erişmesini sağlamaktadır. SMS metin verileri doğrusal olmayan karmaşık bir yapıya sahiptir. Bu yapısı yüzünden smishing tespitini gerçekleştirmek zorlu bir iş haline gelmektedir. Bu çalışma kapsamında smishing tespitini gerçekleştirebilmek için “spam.csv” veri setine DDİ teknikleri kullanılarak çeşitli ön işlemler uygulanmıştır. Ardından veri setine uygulanacak olan Hibrit CNN-GRU modeli oluşturularak veri seti eğitilmiştir. Hibrit CNN-GRU yöntemi ile normal kullanıcı davranışları ile saldırganlara ait içeriklerin anormallikleri arasındaki farklılıklar tespit edilerek saldırı girişimi engellenmesi amaçlanmıştır. Önerilen yöntemle 0.996’lık başarımlar elde edilmiştir. Aynı zamanda çalışmamız “2024 3rd International Conference on Computer Engineering, Technologies and Applications” konferansında bildiri olarak sunulmuştur.

7.2.1. Geliştirilen Yöntem

Çalışmanın bu bölümünde kullanılan veri setine, veri setine uygulanan ön işlemlere, eğitim için geliştirilen modelin yapısına ve eğitim sonunda modelin doğruluk performansına ilişkin bilgilere detaylıca yer verilmiştir.

A. Veri Seti

Çalışmada, veri bilim platformu olan Kaggle’da [69] erişime açık olarak sunulan “spam.csv” isimli veri seti kullanılmıştır. Veri seti 5572 satır ve 2 sütundan oluşmaktadır. Sütunlar “v1” ve “v2” olarak iki özneliğe sahiptir. V1 olarak belirlenen sütunda SMS mesajların “spam” ya da “ham” olarak iki sınıfı yer almaktadır. V2 olarak belirtilen sütunda ise SMS mesajların içerikleri yer almaktadır. Veri setinin %75’i eğitim verisi, %25’i ise test verisi olarak kullanılmıştır. Veri seti içerisinde “Ham” olarak sınıflandırılan 4825 mesaj ve “Spam” olarak sınıflandırılan 747 mesaj bulunmaktadır.

B. Ön İşlemler

DDİ teknikleri kullanılarak, veri kümesi ilk olarak anlamlı kelime köklerinin çıkarılması için "Lemmatizasyon" işlemine tabi tutulmuştur. Bu işlemle kelimelerin kökleri ve ekleri birbirinden ayrılarak kelimenin anlamsal analizi sağlanır. Bu analizin amacı, benzer kökleri paylaşan kelimelerle anlamsal bağlantılar oluşturarak tespit ve analiz çalışmalarının hassasiyetini arttırmaktır. Daha sonra veri setini gereksiz yere kalabalıklaştırmamak ve veri setinin eğitildiği aşamada öğrenmeyi olumsuz etkilememek için stopwordsler veri setinden çıkarılmıştır. Tablo 7.2’de modelin eğitimini olumsuz etkileyecek gereksiz kelimeler olarak değerlendirilen veri setindeki stopwordsler sayılarıyla birlikte gösterilmiştir.

Tablo 7.2. Veri setinde yer alan stopwordsler

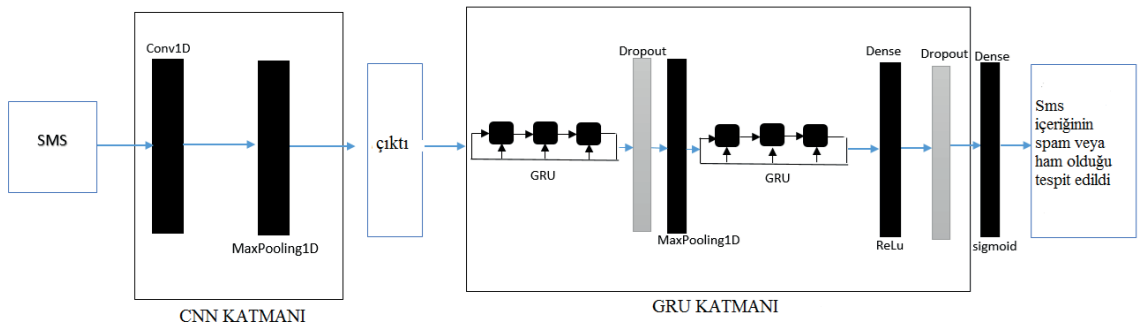
Kelime	Adet	Kelime	Adet	Kelime	Adet	Kelime	Adet
can	304	at	359	u	692	have	474
get	326	call	376	i	742	on	487
will	330	me	557	is	781	the	1197
not	340	your	560	in	798	a	1327
with	349	of	591	it	396	I	1466
be	362	my	628	that	421	you	1622
or	373	for	644	2	453	are	396

Veri seti kategorik verilerden oluştuğu için tokenizasyon (dijitalleştirme) yapılmıştır. Tokenizasyon ile algoritmanın algılayamadığı ve dolayısıyla sonuç üretemediği metin verileri, algoritmanın anlayabileceği sayısal verilere dönüştürülür. Bu işlem sayesinde metin, her kelimeye belirli sayısal değerler atanarak sayısal verilere dönüştürülür. Ayrıca, tokenizasyon yoluyla yinelenen kelimeler veri kümesinden izole edildi ve bu da modelin eğitimi için daha akıcı bir veri kümesi elde edilmesini sağladı. Veri seti eğitim verileri için %75 ve test verileri için %25'e

bölünmüştür. Modeli eğitmek için tüm giriş verilerinin aynı kelime boyutuna sahip olması çok önemlidir. Bunu başarmak için tüm veri seti, maksimum sayıda belirteç içeren cümleye dayalı olarak gömme vektörlerine dönüştürülmüştür. Daha sonra kelimeleri ve bunlara karşılık gelen simgeleri içeren bir liste oluşturulmuş ve döndürülmüştür.

C. Hibrit CNN – GRU Modelinin Oluşturulması

Veri kümesinin ön işlenmesinin ardından model, girdi veri kümesini daha kompakt bir alana sıkıştırmak için tasarlanmış bir gömme katmanının oluşturulmasıyla başlayarak sırayla oluşturulmuştur. Daha sonra verilerden mekânsal özellikleri çıkarmak için ikinci katman olarak bir CNN katmanı eklenmiştir. Aktivasyon fonksiyonu ReLU (Düzeltilmiş Doğrusal Birim), bu CNN katmanı için ağı doğrusal olmamayı sağlamak amacıyla seçilmiştir. Daha sonra maksimum havuzlama katmanı eklenmiştir. Bellek kaynaklarını korurken hesaplamaları hızlandırmak için modele iki GRU katmanı dahil edilmiştir. GRU katmanlarının ardından aşırı uyumu azaltmak için bir Dropout katmanı dahil edilmiştir. Dense katmanında aktivasyon fonksiyonu bir kez daha ReLU olarak seçilmiştir. Veri setinin "Spam" ve "Ham" olmak üzere iki sınıftan oluştuğu göz önüne alındığında, ikili sınıflandırma için sınıflandırma katmanında bir "sigmoid" aktivasyon fonksiyonu kullanılmıştır. Adam algoritması optimizasyon algoritması olarak seçilmiş ve doğruluk performans ölçütü olarak kullanılmıştır. Tüm bu adımların ardından veri seti 10 epochs boyunca eğitilmiştir. Şekil 7.6'da verilen akış şeması, veri seti ile geliştirilen Hibrit CNN-GRU modelini kullanarak SMS mesajlarını spam veya ham olarak sınıflandırmak için tasarlanan uygulamayı göstermektedir. Gösterilen süreçte, sisteme başlangıçta eğitim için giriş verileri sağlanır. CNN katmanı aracılığıyla işlenen girişin çıktısı daha sonra GRU katmanına beslenir. Bu katmanda, eğitim sonrası girdi, Dense katmanda aktivasyon fonksiyonlarına tabi tutulur ve sonuçta sınıflandırma süreci tamamlanır.



Şekil 7.6. Hibrit CNN-GRU ile SMS Phishing tespiti akış şeması

D. Modelin Başarımı ve Literatürdeki Çalışmalarla Karşılaştırılması

Akıllı telefonların yaygın olarak benimsenmesi ve SMS'e olan bağımlılığın artması, sosyal mühendislik saldırganlarının dikkatini bu araca çekti. Saldırganlar, hedefin merakını uyandıran mesajlardan yararlanarak hedefin kişisel bilgilerine ve özel verilerine erişebilir. Geleneksel koruyucu önlemlerin smishing saldırılarına karşı kapsamlı ve etkili savunma sağlamada yetersiz kaldığı bu sorun, bilgi güvenliği alanında önemli bir zorluk olarak ortaya çıkıyor. Geleneksel yöntemlerin yetersizliği göz önüne alındığında araştırmacılar odaklarını yapay zekâ yaklaşımlarına kaydırdılar. Bu çalışmada yapay zekânın bir alt kümesi olan derin öğrenme kullanılmış ve hibrit bir model tasarlanmıştır.

Önerilen yöntem %99,96 doğruluk puanına ulaşmış ve bu sonuç literatürde aynı veri setini kullanan diğer çalışmalardan elde edilen bulgularla karşılaştırılmıştır. Ayrıntılı karşılaştırmalar ve sonuçlar Tablo 7.3'te sunulmuştur.

Tablo 7.3. “spam.csv” veri setine uygulanan algoritmaların doğruluk performansı

Literatürde Kullanılan Algoritma	Doğruluk Ölçütü
RO [50]	0.97
Hibrit KNN-DVM [72]	0.99
Hibrit CNN-GRU [73]	0.98
Geliştirilen Hibrit CNN-GRU Yöntemi	0.996

Tablo 7.3'te başarımları verilen yöntemin smishing olarak tespit edilen tüm mesajlar arasında doğru şekilde tanımlanan smishing mesajlarının oranını gösteren kesinlik değerine ilişkin ayrıntılar Tablo 7.4'te verilmiştir.

Tablo 7.4. Geliştirilen modelin Precision, Recall ve F1-Score değerleri

Sınıf	Precision	Recall	F1-Score
0	0.99	0.99	0.99
1	0.95	0.94	0.94
Macro avg	0.97	0.97	0.97
Weighted avg	0.99	0.99	0.99

7.3. Twitter Verileri Kullanılarak Kısa Metinler Üzerinde GRU ile Nefret Söylemi Tespiti

Gelişen teknoloji ile akıllı cihazların kullanımı yaygınlaşmıştır. İnsanlar akıllı cihazlar üzerinden kolayca erişebildiği birçok sosyal medya platformunu kullanmaktadır. Bu mecralardan önde gelen uygulamalar arasında Twitter yer almaktadır. Özellikle Twitter üzerinden insanlar düşüncelerini kolayca paylaşma imkânı bulmaktadırlar. Tüm bunlara bağlı olarak Twitter

üzerinden nefret söylemi oldukça yaygın olarak gerçekleşmektedir. Nefret söylemi, çeşitli konularda nefret ifadeleri içeren kişilere açıkça yapılan şiddet uygulamaları olarak tanımlanır.

Kullanıcılar hedef edindikleri kişilerle yüz yüze olmama avantajından faydalanarak bu eylemleri gerçekleştirirler. Bu eylemlerin sonucunda mağdurların yaşantıları olumsuz etkilenir. Bu çalışmada toplumsal bir sorun haline gelen nefret söylemi tespitinin gerçekleştirilmesi amaçlanmıştır. Böylece sosyal medya üzerinden paylaşılan bu tarz içerikler tespit edilecektir. Ardından bu içerikleri paylaşan kişi ya da gruplara çeşitli yaptırımlar uygulanarak bu sorunla mücadele edilebilmesi hedeflenmektedir.

Bu çalışmada kullanılan veri seti cinsiyetçi ve ırkçı tweetlerden oluşmaktadır. Veri setinin etiketleri nefret söyleminin olup olmadığına göre iki sınıfta yapılmıştır. Veri setine doğal dil işleme teknikleri kullanılarak ön işlemler uygulanmıştır. Eğitim ve test veri seti olarak parçalanmıştır. Ardından RNN algoritmalarından GRU mimarisi ile model eğitimi tamamlanmıştır. Eğitim sonunda model %99.80'lik başarıma ulaşmıştır. Aynı zamanda çalışmamız IDAP 2022 International Artificial Intelligence and Data Processing Sempozyumunda <http://idap.inonu.edu.tr/resources/proceeding/IDAP-2022-Procedding.pdf> linkinde yer alan sayfa 78 – 85 sayfalarında bildiri olarak yayınlanmıştır.

7.3.1. Geliştirilen Yöntem

Çalışmanın bu bölümünde kullanılan veri setine, veri setine uygulanan ön işlemlere, eğitim için geliştirilen modelin yapısına ve eğitim sonunda modelin doğruluk performansına ilişkin bilgilere detaylıca yer verilmiştir.

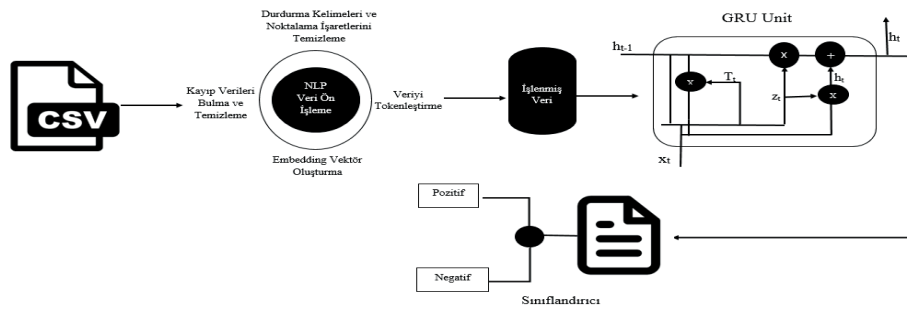
A. Veri Seti

Kısa metinler üzerinden nefret söyleminin tespitini gerçekleştirmek amacıyla yapılan çalışmada Kaggle'da [71] erişime açık olarak sunulan “train.csv” isimli veri seti kullanılmıştır. Veri seti 31.951 satır ve 3 sütundan oluşmaktadır. Sütunlar “id”, “label” ve “tweet” özniteliklerine sahiptir. Veri setinde “1” etiketi tweet'in ırkçı ya da cinsiyetçi bir söyleme sahip olduğunu “0” etiketi ise böyle bir söylemin olmadığını ifade etmektedir. Veri setinin bir bölümü Şekil 7.7'de, etiketli verilere ait grafik Şekil 7.8'de verilmiştir.

kullanılarak eğitim ve test verilerine padding eklenmiştir. Kelime ve kelimelerin tokenlerinin yer aldığı bir liste oluşturularak geri döndürülmüştür.

C. GRU Modelinin Oluşturulması

Ön işlemler tamamlandıktan sonra veri setini eğitmek için Tekrarlayan Sinir Ağı algoritmalarından olan GRU mimarisi kurulmuştur. Modelde olması gereken her şeyi modele katman katman eklemek için başlangıçta sıralı bir model oluşturulmuştur. Her kelimeye karşılık gelecek vektör uzunluğu belirlenmiştir. Girdi veri setini daha dar bir alanda sıkıştırmak için Kerasta “Embedding” katmanı oluşturulmuştur. Embedding katmanına parametre olarak kelime sayıları, kelime vektörlerinin uzunluğu, gelen girdinin boyutu ve bu katmandaki ağırlıkların boyutunu daha sonra çağırmak için isim verilmiştir. Bu katman girdiyi aldıktan sonra girdi içerisindeki kelimelerin vektörlerini bir sonraki katmana gönderecektir. Embedding katmanının çıktısı bir sonraki katmanın girdisi olacaktır. Ardından Yinelenen Sinir Ağını oluşturmak için üç adet GRU katmanı kullanılmıştır. Üç adet GRU katmanı kullanılmasının sebebi ise Yinelenen Sinir Ağlarında çok derin oluşturulan katmanlardan iyi sonuçlar alınamamasıdır. Keras kütüphanesinin “.add()” metodu ile GRU katmanları sırasıyla eklenmiştir. İlk GRU katmanının nöron sayısı 16 olarak girilmiştir. Bu katmanda modele 16 tane çıktı verilecektir. İkinci GRU katmanının nöron sayısı 8 olarak girilmiştir. Son GRU katmanında ise nöron sayısı 4 olarak girilmiştir. Bir sonraki katman tek bir nörondan oluşacağı için son GRU katmanından uygun bir çıktı üretilmesi sağlanmıştır. Bir katmandan alınan nöronun bir sonraki katmana girdi olarak verilmesini sağlamak için tek bir nörondan oluşan “Dense” katmanı modele eklenmiştir. Eğitim sonunda elde edilecek çıktı iki sınıftan oluştuğu için aktivasyon fonksiyonuna “sigmoid” verilerek çıktı “0” ile “1” arasında sıkıştırılmıştır. Optimizasyon algoritması “Adam” olarak belirlenmiştir. İki sınıf bulunduğu için Loss Fonksiyonu olarak “binary_crossentropy” ve modelin başarı oranını görebilmek için metrik olarak “accuracy” kullanılmıştır. Model 256’lık batch_size’ler halinde 10 epochs olacak şekilde eğitilmiştir. Model eğitime ait akış şeması Şekil 7.9’da verilmiştir.



Şekil 7.9. GRU ile Twitter veri seti üzerinde nefret söylemi tespiti akış şeması

Geliştirilen modele ait tüm adımlar aşağıda maddeler halinde verilmiştir.

Doğal Dil İşleme ile Ön İşlem Adımları

1. train.csv veri seti pandas kütüphanesi ile okundu.
2. train.csv veri setinin kayıp satır ve sütunları taranıp kayıp veri olmadığı tespit edilmiştir.
3. Veri seti durdurma kelimelerinden “stopwords” kütüphanesi ile temizlenmiştir.
4. Veri setinin %75’i eğitim, %25’i test verisi olacak şekilde parçalanmıştır.
5. Veri seti içerisinde yer alan cümlelerin kelime boyutu belirlenerek “Tokenizer” kütüphanesi ile tüm veri seti sayısallaştırılmıştır.
6. Kelime vektör yöntemi olarak “Bag of Words” kullanılarak 50.000 kelimedenden oluşan embedding vektörler oluşturulmuştur.
7. Tokenleştirme işlemi tamamlanan eğitim ve test veri setine Keras kütüphanesi ile padding eklenmiştir.

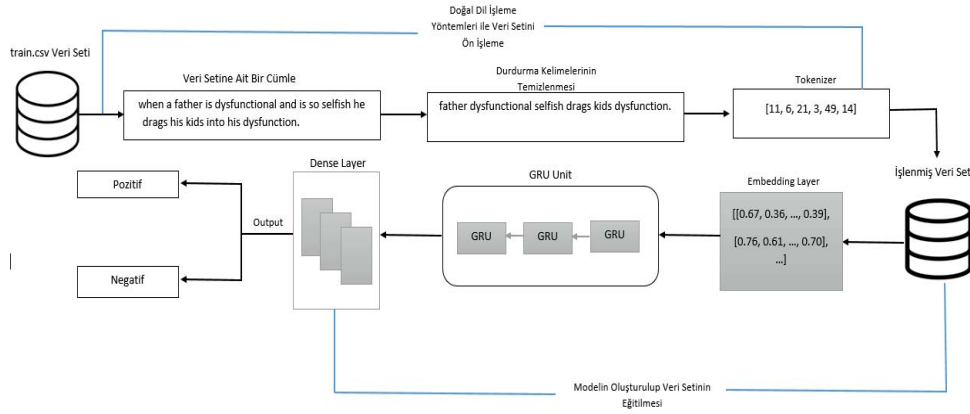
Eğitim Modelinin Oluşturulması

1. Sıralı bir model oluşturulmuştur.
2. Girdi veri setini daha dar bir alanda sıkıştırmak için Kerasta “Embedding” katmanı oluşturulmuştur. Embedding katmanına parametre olarak kelime sayıları, kelime vektörlerinin uzunluğu, gelen girdinin boyutu ve bu katmandaki ağırlıkların boyutunu daha sonra çağırmak için isim verilmiştir.
3. Ardından 16, 8 ve 4 nöronlu oluşan üç adet GRU katmanı oluşturulmuştur.
4. Dense katmanı oluşturulup aktivasyon fonksiyonu belirlenerek bu katmanda parametre olarak girilmiştir.
5. Optimizasyon algoritması belirlenmiştir.
6. Modeli derlemek için “model.compile” oluşturulmuştur. Parametre olarak loss fonksiyonu, optimizasyon algoritması ve başarımleri metriği girilmiştir.
7. Model 10 epochs’ta 256’lık batch_size’ler ile eğitilmiştir.

Eğitilen Modelin Test Edilmesi

1. Algoritmanın başarımleri test etmek için kullanıcıdan bir metin girmesi istenmiştir.
2. Ardından girilen metin “sent_tokenize” kütüphanesi ile cümlelere ayrılmış ve listeye içerisinde eklenmiştir.
3. Cümleleri içeren liste tokenleştirildikten sonra eğitilen modele girdi olarak verilip metin içerisinde nefret söylemi olup olmadığı başarılı bir şekilde tespit edilmiştir.

Yukarıda adım adım verilen algoritmanın adımları Şekil 7.10’da detaylı olarak verilmiştir.



Şekil 7.10. Geliştirilen algoritmanın yapısı

D. Modelin Başarımı ve Literatürdeki Çalışmalarla Karşılaştırılması

Bu çalışmada, tweetlerden oluşturulan veri seti doğal dil işleme teknikleri kullanılarak temizlenmiş, vektör uzunlukları belirlenmiş ve sayısallaştırma işlemleri gerçekleştirilmiştir. Model eğitimi için GRU mimarisi kullanılmıştır. Eğitim tamamlandıktan sonra modelin %99.80'lik bir başarıma sahip olduğu belirlenmiştir. Veri bilim platformu olan Kaggle'da yer alan veri setine [74] literatürdeki çeşitli algoritmalar kullanılarak model eğitimi gerçekleştirilmiştir. Bu algoritmalarından elde edilen başarımlar ve bu çalışmada kullanılan yaklaşımın başarımları Tablo 7.5'te verilmiştir.

Tablo 7.5. Train.csv veri setine uygulanan algoritmaların doğruluk performansı

Literatürde Kullanılan Algoritma	Doğruluk Ölçütü
LR [75]	0.487
KNN [75]	0.942
DVM [75]	0.480
Kernel DVM [75]	0.93
NB [75]	0.913
ROt Sınıflandırıcı [75]	0.951
XG Boost Sınıflandırıcı [75]	0.953
KA [75]	0.925
RO Sınıflandırıcı [76]	0.90
Linear SVC [77]	0.956
Multinomial NB [78]	0.94
LSTM [79]	0.952
Multinomial NB [80]	0.937
LR [80]	0.945
Linear SVC [80]	0.962
BiLSTM [80]	0.952
LR [81]	0.931
Bu çalışma/GRU	0.998

8. SONUÇLAR VE TARTIŞMA

Bilgi güvenliği bilgiyi her türlü fiziksel ve dijital tehdide karşı korumayı amaç edinen bir alandır. Bu alanda bilgi güvenliği sağlanırken birden çok yöntem kullanılmakla birlikte bu yöntemler hem geleneksel hem de modern yöntemler olabilmektedir. Bilginin depolandığı alanı fiziksel olarak korumak, uç noktaların fiziksel güvenliğini sağlamak ve insan faktörünü eğitmek gibi geleneksel yöntemlerin yanı sıra antivirüs programları kullanma, güncel yazılım kullanma, kriptografi yöntemleri gibi dijital yöntemler bu korunma yöntemlerinden bazılarıdır. Bu korunma yöntemleri kullanılarak bilginin ilk haliyle kalması ve sadece yetkili kişilerce erişilmesi sağlanmaktadır. Bilgi güvenliğinin içerisinde yer alan siber güvenlik alanı ise verinin siber alanda dijital olarak korunmasını amaçlamaktadır.

Günümüzde artan internet kullanımı beraberinde bazı güvenlik açıklıklarını da getirmektedir. Bu güvenlik açıklıklarının en yaygın olanı ise modern bir bilgi hırsızlığı olan sosyal mühendislik saldırılarıdır. Sosyal mühendislik saldırılarında saldırganlar birbirinden farklı yöntemler kullanarak hedefinin kişisel bilgileri, banka bilgileri, işyerine ait bilgileri ve bunun gibi üçüncü kişilerin ulaşması tehlikeli olan bilgilerini ele geçirmektedir. Bu saldırılardan korunmak, tanınmayan kişilerle hassas kişisel bilgilerin paylaşılmaması gibi geleneksel yöntemlerle olabileceği gibi bu saldırıları tespit etmek için geliştirilen yöntemleri kullanmak gibi modern çözümlerle de olmaktadır.

Bu çalışmada öncelikle bilgi güvenliği açıklanarak ardından sosyal mühendislik saldırılarının ne olduğu, phishing, smishing gibi çeşitli sosyal mühendislik saldırıları detaylı bir biçimde açıklanarak okuyucu saldırı türleri ile ilgili bilgilendirilmiş daha sonrasında bir sosyal mühendislik saldırısının yaşam döngüsünün nasıl olduğu ve bu saldırılardan nasıl korunulacağına ilişkin genel bir çerçeve sunulmuştur. Ayrıca sosyal mühendislik saldırılarının tespitine yönelik olarak literatürde yer alan makaleler kapsamında elde edilen doğruluk performansı, kullanılan veri seti ve yöntemler geniş bir yelpazede incelenmiştir. Çalışmanın devam eden bölümlerinde DDİ ve DDİ teknikleri hakkında önemli noktalara değinilmiştir.

Çalışmanın son bölümünde ilk olarak DDİ teknikleri kullanılarak Hibrit CNN-BiLSTM Yöntemleri ile smishing saldırılarının tespiti gerçekleştirilerek modelin doğruluk performansı 0.998 olarak, Hibrit CNN – GRU yöntemiyle Smishing saldırılarının tespiti gerçekleştirilerek doğruluk performansı 0.996 olarak, Twitter verileri kullanılarak kısa metinler üzerinde GRU ile nefret söylemi tespiti gerçekleştirilerek doğruluk performansı 0.998 elde edilmiştir.

Tez çalışması için yapılan ön hazırlık aşamasında özellikle DDİ alanında karşılaşılan en büyük sorun Türkçe kaynakların yetersiz olmasıdır. Bu alanda ülkemizde henüz yeterli ve detaylı çalışmalar bulunmamaktadır. Bu kapsamda Türkçe veri setleri oluşturularak geliştirilen

yöntemlerin elde edilen Türkçe veri setlerine uygulanmasıyla bilgi güvenliği açısından ülkemizde yaşanan sorunlarında önüne geçilmesi sağlanabilir.



KAYNAKLAR

- [1] M.Ç. Varol, E. Varol, Kavram ve Kuramlarıyla Marshall McLuhan 'a Bakış : Günümüzün Egemen Medya Araçları Ekseninde Bir Değerlendirme, *Int. J. Cult. Soc. Stud.* 5 (2019) 137–158. <https://dergipark.org.tr/tr/download/article-file/744648#:~:text=Medya%0AEkolojisi%0ATEorisi%0A1960%0Ayılında,bir%0Agüç%0Adışında%0Aolduğuna%0Ainanmaktadır.>
- [2] K. Haynes, H. Shirazi, I. Ray, Lightweight URL-based phishing detection using natural language processing transformers for mobile devices, *Procedia Comput. Sci.* 191 (2021) 127–134. <https://doi.org/10.1016/j.procs.2021.07.040>.
- [3] A. Bhardwaj, F. Al-Turjman, V. Sapra, M. Kumar, T. Stephan, Privacy-aware detection framework to mitigate new-age phishing attacks, *Comput. Electr. Eng.* 96 (2021) 107546. <https://doi.org/10.1016/j.compeleceng.2021.107546>.
- [4] S. Minocha, B. Singh, A novel phishing detection system using binary modified equilibrium optimizer for feature selection ☆, *Comput. Electr. Eng.* 98 (2022) 107689. <https://doi.org/10.1016/j.compeleceng.2022.107689>.
- [5] K.L. Chiew, E.H. Chang, S.N. Sze, W.K. Tiong, Utilisation of website logo for phishing detection, *Comput. Secur.* 54 (2015) 16–26. <https://doi.org/10.1016/j.cose.2015.07.006>.
- [6] P.A. Barraclough, G. Fehringer, J. Woodward, Intelligent cyber-phishing detection for online, *Comput. Secur.* 104 (2021) 102123. <https://doi.org/10.1016/j.cose.2020.102123>.
- [7] N.A. Azeez, S. Misra, I.A. Margaret, L. Fernandez-Sanz, S.M. Abdulhamid, Adopting automated whitelist approach for detecting phishing attacks, *Comput. Secur.* 108 (2021) 102328. <https://doi.org/10.1016/j.cose.2021.102328>.
- [8] A. Alhogail, A. Alsabih, Applying machine learning and natural language processing to detect phishing email, *Comput. Secur.* 110 (2021) 102414. <https://doi.org/10.1016/j.cose.2021.102414>.
- [9] E. Oram, P.B. Dash, B. Naik, J. Nayak, S. Vimal, S.K. Nataraj, Light gradient boosting machine-based phishing webpage detection model using phisher website features of mimic URLs, *Pattern Recognit. Lett.* 152 (2021) 100–106. <https://doi.org/10.1016/j.patrec.2021.09.018>.
- [10] L. Yang, J. Zhang, X. Wang, Z. Li, Z. Li, Y. He, An improved ELM-based and data preprocessing integrated approach for phishing detection considering comprehensive features, *Expert Syst. Appl.* 165 (2021) 113863. <https://doi.org/10.1016/j.eswa.2020.113863>.
- [11] W. Wei, Q. Ke, J. Nowak, M. Korytkowski, R. Scherer, M. Woźniak, Accurate and fast URL phishing detector: A convolutional neural network approach, *Comput. Networks.* 178 (2020). <https://doi.org/10.1016/j.comnet.2020.107275>.
- [12] T. Xia, X. Chen, A weighted feature enhanced Hidden Markov Model for spam SMS filtering, *Neurocomputing.* 444 (2021) 48–58. <https://doi.org/10.1016/j.neucom.2021.02.075>.
- [13] O.N. Akande, O. Gbenle, O.C. Abikoye, R.G. Jimoh, H.B. Akande, A.O. Balogun, A. Fatokun, SMSPROTECT: An automatic smishing detection mobile application, *ICT Express.* 9 (2022) 168–176. <https://doi.org/10.1016/j.ict.2022.05.009>.

- [14] G. Waja, G. Patil, C. Mehta, S. Patil, How AI Can be Used for Governance of Messaging Services: A Study on Spam Classification Leveraging Multi-Channel Convolutional Neural Network, *Int. J. Inf. Manag. Data Insights.* 3 (2023) 100147. <https://doi.org/10.1016/j.jjime.2022.100147>.
- [15] S.M. Hameed, Z.H. Ali, SMS Spam Detection Based on Fuzzy Rules and Binary Particle Swarm Optimization, *Int. J. Intell. Eng. Syst.* 14 (2021) 314–322. <https://doi.org/10.22266/ijies2021.0430.28>.
- [16] V.A. Saeed, A Method for SMS Spam Message Detection Using Machine Learning, *Artif. Intell. Robot. Dev. J.* 3 (2023) 214–228. <https://doi.org/10.52098/airdj.202366>.
- [17] S.M.M. Hossain, A. Sen, K. Deb, Detecting Spam SMS Using Self Attention Mechanism, *Lect. Notes Networks Syst.* 569 LNNS (2023) 175–184. https://doi.org/10.1007/978-3-031-19958-5_17.
- [18] A.A. Ojugo, D.A. Oyemade, Boyer moore string-match framework for a hybrid short message service spam filtering technique, *IAES Int. J. Artif. Intell.* 10 (2021) 519–527. <https://doi.org/10.11591/ijai.v10.i3.pp519-527>.
- [19] A. Ghourabi, M.A. Mahmood, Q.M. Alzubi, A hybrid CNN-LSTM model for SMS spam detection in arabic and english messages, *Futur. Internet.* 12 (2020) 1–16. <https://doi.org/10.3390/FI12090156>.
- [20] A.A. Ojugo, A.O. Eboka, Memetic algorithm for short messaging service spam filter using text normalization and semantic approach, *Int. J. Informatics Commun. Technol.* 9 (2020) 9. <https://doi.org/10.11591/ijict.v9i1.pp9-18>.
- [21] U. Srinivasarao, A. Sharaff, Machine intelligence based hybrid classifier for spam detection and sentiment analysis of SMS messages Content courtesy of Springer Nature , terms of use apply . Rights reserved . Content courtesy of Springer Nature , terms of use apply . Rights reserved , (2023).
- [22] M.E. Basiri, S. Nemati, M. Abdar, E. Cambria, U.R. Acharya, ABCDM: An Attention-based Bidirectional CNN-RNN Deep Model for sentiment analysis, *Futur. Gener. Comput. Syst.* 115 (2021) 279–294. <https://doi.org/10.1016/j.future.2020.08.005>.
- [23] M.F. López-Vizcaíno, F.J. Nóvoa, V. Carneiro, F. Cacheda, Early detection of cyberbullying on social media networks, *Futur. Gener. Comput. Syst.* 118 (2021) 219–229. <https://doi.org/10.1016/j.future.2021.01.006>.
- [24] Z.L. Chia, M. Ptaszynski, F. Masui, G. Leliwa, M. Wroczynski, Machine Learning and feature engineering-based study into sarcasm and irony classification with application to cyberbullying detection, *Inf. Process. Manag.* 58 (2021) 102600. <https://doi.org/10.1016/j.ipm.2021.102600>.
- [25] J. Eronen, M. Ptaszynski, F. Masui, A. Pohl, G. Leliwa, M. Wroczynski, Improving classifier training efficiency for automatic cyberbullying detection with Feature Density, *Inf. Process. Manag.* 58 (2021) 102616. <https://doi.org/10.1016/j.ipm.2021.102616>.
- [26] F.E. Ayo, O. Folorunso, F.T. Ibharaolu, I.A. Osinuga, A. Abayomi-Alli, A probabilistic clustering model for hate speech classification in twitter, *Expert Syst. Appl.* 173 (2021) 114762. <https://doi.org/10.1016/j.eswa.2021.114762>.
- [27] A. Bozyigit, S. Utku, E. Nasibov, Cyberbullying detection: Utilizing social media features, *Expert Syst. Appl.* 179 (2021). <https://doi.org/10.1016/j.eswa.2021.115001>.
- [28] S. Rajesh, B. Sharanya, Recognition and prevention of cyberharassment in social media using

- classification algorithms, Mater. Today Proc. (2021). <https://doi.org/10.1016/j.matpr.2020.10.502>.
- [29] E. Yazgılı, M. Baykara, Türkçe metinlerde makine öğrenmesi yöntemleri ile siber zorbalık tespiti, Gümüşhane Üniversitesi Fen Bilim. Enstitüsü Derg. 12 (2022) 443–453. <https://doi.org/10.17714/gumusfenbil.935448>.
- [30] M. Toğaçar, Web Sitelerinde Gerçekleştirilen Oltalama Saldırıların Yapay Zekâ Yaklaşımı ile Tespiti, Bitlis Eren Üniversitesi Fen Bilim. Derg. 10 (2021) 1603–1614. <https://doi.org/10.17798/bitlisfen.988001>.
- [31] M. Toğaçar, K.A. Eşidir, B. Ergen, Yapay Zekâ Tabanlı Doğal Dil İşleme Yaklaşımını Kullanarak İnternet Ortamında Yayınlanmış Sahte Haberlerin Tespiti, J. Intell. Syst. Theory Appl. 5 (2021) 1–8. <https://doi.org/10.38016/jista.950713>.
- [32] F.B. Enstit, Sosyal medya analitiği ile siber zorbalık tespiti, (n.d.).
- [33] Z.D.E. Nerg, D. Mers, Derin öğrenme teknikleri kullanılarak sosyal ağlar üzerinde siber zorbalık tespiti yüksek lisans tezi, (2021).
- [34] Sheila Maria Belgis Putri Affiza, Hate speech detection in turkish news using a transformer-based model enhanced with linguistic features, הָאָרָץ. (2022) 2003–2005.
- [35] B. Mühendisliği Anabilim Dalı Bilgisayar Mühendisliği Programı Danışman Banu Diri, Eş Danışman Özgür Koray Şahingöz, Oltalama saldırılarının derin öğrenme tabanlı url ve içerik analizi ile hibrit tespiti mehmet korkmaz doktora tezi, (2023).
- [36] Y. Mühendisliği, A. Bilim, Sosyal Ağlar Üzerinde Mültecilere Yönelik Nefret Söyleminin Metin Madenciliğine Dayalı Tespiti, (2022).
- [37] A new machine learning classification algorithm for phishing urls detection Abdalraouf Almahdi Mohammed Alarbi thesis advisor, (2023).
- [38] A. Sangging, Kurumsal bilgi güvenliği yönetişimi ve bilgi güvenliği için insan faktörünün önemi, Block Caving – A Viable Altern. 21 (2017) 1–9.
- [39] S. Zulaikha, Bilişim sistemleri için saldırı tespit ve engelleme yaklaşımlarının tasarımı ve gerçekleştirilmesi, Rev. Bras. Ergon. 9 (2016) 10. <https://www.infodesign.org.br/infodesign/article/view/355%0Ahttp://www.abergo.org.br/revista/index.php/ae/article/view/731%0Ahttp://www.abergo.org.br/revista/index.php/ae/article/view/269%0Ahttp://www.abergo.org.br/revista/index.php/ae/article/view/106>.
- [40] R. putri Indahningrum, lia dwi jayanti, Çalışanların bilgi güvenliği farkındalığına ilişkin algıları: bankacılık sektöründe bir araştırma, 2507 (2020) 1–9. <http://journal.um-surabaya.ac.id/index.php/JKM/article/view/2203>.
- [41] A.M. and E.A. Haile G, Kuantum Sonrası Kriptografi, 4 (2023) 88–100.
- [42] bidin A, Kamu kurumlarının bilgi güvenliği politikalarının kurumsal bilgi güvenliğinin sağlanması açısından etkinliğinin analiz edilmesi, 4 (2017) 9–15.
- [43] W.S. Admass, Y.Y. Munaye, A.A. Diro, Cyber security: State of the art, challenges and future directions, Cyber Secur. Appl. 2 (2024) 100031. <https://doi.org/10.1016/j.csa.2023.100031>.
- [44] R. Alabdan, Phishing attacks survey: Types, vectors, and technical approaches, Futur. Internet. 12 (2020) 1–39. <https://doi.org/10.3390/fi12100168>.
- [45] E. Business, M. To, T. Fighting, F. This, CEO Fraud CEO, (n.d.).

- [46] Z. Alkhalil, C. Hewage, L. Nawaf, I. Khan, Phishing Attacks: A Recent Comprehensive Study and a New Anatomy, *Front. Comput. Sci.* 3 (2021). <https://doi.org/10.3389/fcomp.2021.563060>.
- [47] K. Chetoui, B. Bah, A.O. Alami, A. Bahnasse, Overview of Social Engineering Attacks on Social Networks, *Procedia Comput. Sci.* 198 (2021) 656–661. <https://doi.org/10.1016/j.procs.2021.12.302>.
- [48] J.W. Bullée, M. Junger, Social engineering, *Palgrave Handb. Int. Cybercrime Cyberdeviance*. (2020) 849–875. https://doi.org/10.1007/978-3-319-78440-3_38.
- [49] A. Yasin, R. Fatima, L. Liu, J. Wanga, R. Ali, Z. Wei, Counteracting social engineering attacks, *Comput. Fraud Secur.* 2021 (2021) 15–19. [https://doi.org/10.1016/S1361-3723\(21\)00108-1](https://doi.org/10.1016/S1361-3723(21)00108-1).
- [50] S.Y. Yerima, A. Bashar, Semi-supervised novelty detection with one class SVM for SMS spam detection, *Int. Conf. Syst. Signals, Image Process.* 2022-June (2022). <https://doi.org/10.1109/IWSSIP55020.2022.9854496>.
- [51] P. Joseph, S.Y. Yerima, A comparative study of word embedding techniques for SMS spam detection, *Proc. - 2022 14th IEEE Int. Conf. Comput. Intell. Commun. Networks, CICN 2022*. (2022) 149–155. <https://doi.org/10.1109/CICN56167.2022.10008245>.
- [52] E.I. Hanif, C.F.M. Foozy, I.R.A. Hamid, R. Mohamed, M.M. Yusof, R.A. Aziz, P. Shamala, Malay SMS Spam Detection Tool Using Keyword Filtering Technique, *J. Phys. Conf. Ser.* 1793 (2021). <https://doi.org/10.1088/1742-6596/1793/1/012064>.
- [53] L. Bustio-Martínez, M.A. Álvarez-Carmona, V. Herrera-Semenets, C. Feregrino-Urbe, R. Cumplido, A lightweight data representation for phishing URLs detection in IoT environments, *Inf. Sci. (Ny)*. 603 (2022) 42–59. <https://doi.org/10.1016/j.ins.2022.04.059>.
- [54] Y. Al-Hamar, H. Kolivand, M. Tajdini, T. Saba, V. Ramachandran, Enterprise Credential Spear-phishing attack detection, *Comput. Electr. Eng.* 94 (2021) 107363. <https://doi.org/10.1016/j.compeleceng.2021.107363>.
- [55] R. Wazirali, R. Ahmad, A.A.K. Abu-Ein, Sustaining accurate detection of phishing URLs using SDN and feature selection approaches, *Comput. Networks*. 201 (2021) 108591. <https://doi.org/10.1016/j.comnet.2021.108591>.
- [56] S. Magdy, Y. Abouelseoud, M. Mikhail, Efficient spam and phishing emails filtering based on deep learning, *Comput. Networks*. 206 (2022) 108826. <https://doi.org/10.1016/j.comnet.2022.108826>.
- [57] M. Abdullah Alohal, N. Alasmari, M. Maashi, A.M. Nouri, M. Rizwanullah, I. Yaseen, A. Elneil Osman, A.A. Alneil, Metaheuristics with deep learning driven phishing detection for sustainable and secure environment, *Sustain. Energy Technol. Assessments*. 56 (2023) 103114. <https://doi.org/10.1016/j.seta.2023.103114>.
- [58] T.O. Ojewumi, G.O. Ogunleye, B.O. Oguntunde, O. Folorunsho, S.G. Fashoto, N. Ogbu, Performance evaluation of machine learning tools for detection of phishing attacks on web pages, *Sci. African*. 16 (2022) e01165. <https://doi.org/10.1016/j.sciaf.2022.e01165>.
- [59] S. Pandiyan S, P. Selvaraj, V.K. Burugari, J. Benadit P, K. P, Phishing attack detection using Machine Learning, *Meas. Sensors*. 24 (2022) 100476. <https://doi.org/10.1016/j.measen.2022.100476>.
- [60] M.M. Alani, H. Tawfik, PhishNot: A Cloud-Based Machine-Learning Approach to Phishing URL Detection, *Comput. Networks*. 218 (2022) 109407. <https://doi.org/10.1016/j.comnet.2022.109407>.

- [61] A. Ozcan, C. Catal, E. Donmez, B. Senturk, A hybrid DNN–LSTM model for detecting phishing URLs, *Neural Comput. Appl.* 0123456789 (2021) 17. <https://doi.org/10.1007/s00521-021-06401-z>.
- [62] N. Yuvaraj, V. Chang, B. Gobinathan, A. Pinagapani, S. Kannan, G. Dhiman, A.R. Rajan, Automatic detection of cyberbullying using multi-feature based artificial intelligence with deep decision tree classification, *Comput. Electr. Eng.* 92 (2021). <https://doi.org/10.1016/j.compeleceng.2021.107186>.
- [63] K. Dhungana Sainju, N. Mishra, A. Kuffour, L. Young, Bullying discourse on Twitter: An examination of bully-related tweets using supervised machine learning, *Comput. Human Behav.* 120 (2021) 106735. <https://doi.org/10.1016/j.chb.2021.106735>.
- [64] O. Sharif, M.M. Hoque, Tackling cyber-aggression: Identification and fine-grained categorization of aggressive texts on social media using weighted ensemble of transformers, *Neurocomputing*. (2021). <https://doi.org/10.1016/j.neucom.2021.12.022>.
- [65] F.E. Ayo, O. Folorunso, F.T. Ibharalu, I.A. Osinuga, Machine learning techniques for hate speech classification of twitter data: State-of-The-Art, future challenges and research directions, *Comput. Sci. Rev.* 38 (2020) 100311. <https://doi.org/10.1016/j.cosrev.2020.100311>.
- [66] M. Toğaçar, Siber Saldırlara Karşı Kullanılan Makine Öğrenme Yöntemlerinin Web Uygulamalarında Güvenlik Etkinliğinin Ölçümü, *Gazi Üniversitesi Fen Bilim. Derg. Part C Tasarım ve Teknol.* 9 (2021) 608–620. <https://doi.org/10.29109/gujsc.950639>.
- [67] İ. Mayda, B. Diri, T. Yıldız, Türkçe Tweetler üzerinde Makine Öğrenmesi ile Nefret Söylemi Tespiti, *Eur. J. Sci. Technol.* (2021) 328–334. <https://doi.org/10.31590/ejosat.903854>.
- [68] A. Rayan, A.I. Taloba, A. Rayan, A.I. Taloba, Detection of Email Spam using Natural Language Processing Based Random Forest Approach Detection of Email Spam using Natural Language Processing Based Random Forest Approach, (2021) 0–12.
- [69] SMS Spam Detection using LSTM | Kaggle, (n.d.). <https://www.kaggle.com/code/basilb2s/sms-spam-detection-using-lstm/input> (accessed April 1, 2023).
- [70] SMS Spam Detection using LSTM | Kaggle, (n.d.). <https://www.kaggle.com/code/basilb2s/sms-spam-detection-using-lstm/notebook> (accessed April 2, 2023).
- [71] SMS Spam Detection Analysis using Python (NLP) | Machine Learning Project Tutorial, (n.d.). <https://www.hackersrealm.net/post/sms-spam-detection-analysis-using-python> (accessed April 2, 2023).
- [72] U. Srinivasarao, A. Sharaff, Machine intelligence based hybrid classifier for spam detection and sentiment analysis of SMS messages Content courtesy of Springer Nature , terms of use apply . Rights reserved . Content courtesy of Springer Nature , terms of use apply . Rights reserved , (2023) 31069–31099.
- [73] R.E. Ulfath, H. Alqahtani, M. Hammoudeh, I.H. Sarker, Hybrid CNN-GRU Framework with Integrated Pre-trained Language Transformer for SMS Phishing Detection, *ACM Int. Conf. Proceeding Ser.* (2021) 244–251. <https://doi.org/10.1145/3508072.3508109>.
- [74] Twitter Sentiment Analysis | Kaggle, (n.d.). <https://www.kaggle.com/datasets/arkhoshghalb/twitter-sentiment-analysis-hatred-speech> (accessed August 17, 2023).
- [75] twitter-sentiment-analysis | Kaggle, (n.d.). <https://www.kaggle.com/code/aayush349/twitter-sentiment-analysis>

- sentiment-analysis (accessed August 17, 2023).
- [76] Twitter Duygu/Konuşma Analizi ve Modelleme | kaggle, (n.d.).
<https://www.kaggle.com/code/danielbeltsazar/twitter-sentiment-speech-analysis-and-modelling>
(accessed May 6, 2022).
- [77] Clasificación sentimientos | Kaggle, (n.d.).
<https://www.kaggle.com/code/joseangelroyolopez/clasificaci-n-sentimientos> (accessed May 6, 2022).
- [78] Twitter Sentiment Analysis | Kaggle, (n.d.). <https://www.kaggle.com/code/tenzintsega/twitter-sentiment-analysis> (accessed May 6, 2022).
- [79] Hate Speech - Tweet Classification | Kaggle, (n.d.).
<https://www.kaggle.com/code/isidronavarroporto/hate-speech-tweet-classification> (accessed May 6, 2022).
- [80] Twitter Sentiment Analysis - Bidirectional LSTM | Kaggle, (n.d.).
<https://www.kaggle.com/code/rashmimanwani/twitter-sentiment-analysis-bidirectional-lstm>
(accessed May 6, 2022).
- [81] Twitter Sentiment Analysis-sklearn | Kaggle, (n.d.).
<https://www.kaggle.com/code/tracyporter/twitter-sentiment-analysis-sklearn> (accessed May 6, 2022).