

DETECTING TURKISH PHISHING ATTACKS WITH MACHINE LEARNING
CLASSIFIERS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS OF
THE MIDDLE EAST TECHNICAL UNIVERSITY
BY

MELİH TURHANLAR

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE
OF MASTER OF SCIENCE
IN
THE DEPARTMENT OF CYBERSECURITY

NOVEMBER 2019

DETECTING TURKISH PHISHING ATTACKS WITH MACHINE LEARNING
CLASSIFIERS

Submitted by MELİH TURHANLAR in partial fulfillment of the requirements for the degree of **Master of Science in Cyber Security Department, Middle East Technical University** by,

Prof. Dr. Deniz Zeyrek Bozşahin
Dean, **Graduate School of Informatics**

Assoc. Prof. Dr. Aysu Betin Can
Head of Department, **Cyber Security**

Assoc. Prof. Dr. Cengiz Acartürk
Supervisor, **Cognitive Science Dept., METU**

Examining Committee Members:

Assoc. Prof. Dr. Aysu Betin Can
Cyber Security Dept., METU

Assoc. Prof. Dr. Cengiz Acartürk
Cognitive Science Dept., METU

Assoc. Prof. Dr. Burcu Can
Computer Engineering Dept., Hacettepe University

Date: **06 November 2019**

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name: Melih TURHANLAR

Signature : _____

ABSTRACT

DETECTING TURKISH PHISHING ATTACKS WITH MACHINE LEARNING CLASSIFIERS

TURHANLAR, MELİH

MSc., Department of Cyber Security

Supervisor: Assoc. Prof. Dr. Cengiz Acartürk

November 2019, 82 pages

Phishing Attacks are social engineering attacks that aim at stealing victim's credit card numbers, credentials, and personal information by exploiting victim's emotions, such as curiosity and fear. The attacker usually sends a webpage link in embodied in textual content. If the victim clicks the link, they usually connect to a mock webpage that imitates a real, institutional webpage. Filling the HTML forms in the mock webpage, the victim sends their credentials unwittingly to the attacker. In our day, phishing is a global issue. This study presents a framework for detecting phishing text in Turkish by running machine learning classifiers on an imbalanced phishing data set. The training dataset covers e-mails, SMS text and tweets. The results show that Logistic Regression Synthetic Minority Over-Sampling Technique achieves high performance, as indicated by F-measures, compared to a set of 32 machine learning models in our study.

Keywords: Turkish Phishing Attacks, Machine Learning, Imbalanced Dataset

ÖZ

TÜRKÇE OLTALAMA SALDIRILARININ MAKİNE ÖĞRENMESİ ALGORİTMALARI İLE TESPİTİ

TURHANLAR, MELİH

Yüksek Lisans, Siber Güvenlik Bölümü

Tez Yöneticisi: Doç. Dr. Cengiz Acartürk

Kasım 2019, 82 sayfa

Oltalama Saldırıları, kurbanın merak ve korku gibi duygularını kullanarak, kredi kartı numaralarını, kimlik bilgilerini ve kişisel bilgilerini çalmayı amaçlayan sosyal mühendislik saldırılarıdır. Saldırgan genellikle metin içerisinde bir web sayfası bağlantısı gönderir. Kurban bağlantıyı tıklarsa, genellikle gerçek ve kurumsal bir web sayfasını taklit eden sahte bir web sayfasına bağlanır. Sahte web sayfasındaki HTML formlarını dolduran kurban, kimlik bilgilerini istemeden saldırgana gönderir. Günümüzde, phishing küresel bir konudur. Bu çalışma, dengesiz bir oltalama veri setinde makine öğrenme sınıflandırıcıları vasıtasıyla Türkçe oltalama metnini tespit etmek için bir çerçeve sunmaktadır. Eğitim veri setinde e-postalar, SMS metinleri ve tweet'ler bulunmaktadır. Çalışmamızda 32 makine öğrenme modelinin F-puan sonuçları karşılaştırıldığında, Lojistik Regresyon Sentetik Azınlık Aşırı Örnekleme Tekniğinin yüksek performans elde ettiği görülmektedir.

Anahtar Sözcükler: Türkçe Oltalama Saldırıları, Makine Öğrenimi, Eşit dağılımı olmayan Veri Seti

To My Family

ACKNOWLEDGMENTS

First of all, I would like to express my gratitude and my respect to my thesis advisor Assoc.Prof.Dr. CENGİZ ACARTÜRK for his priceless guidance, encouragement and continuous support to make this research possible.

I also thank my thesis jury members, Assoc. Prof. Dr. Aysu Betin Can, Assoc. Prof. Dr. Burcu Can for their suggestions and reviewing my work.

Finally, I would like to express my deepest gratitude to my wife, Anna for her love, encouragement, and support, my son Hikmet Can and my daughter Yasemin for their patients during my study. This thesis would not have been written without them.

TABLE OF CONTENTS

(1) ABSTRACT	iv
(2) ÖZ	v
(3) ACKNOWLEDGMENTS.....	vii
(4) TABLE OF CONTENTS.....	viii
(5) LIST OF TABLES.....	x
(6) LIST OF FIGURES	xi
(7) LIST OF ABBREVIATIONS	xii
1. INTRODUCTION	1
1.1. Motivation.....	1
1.2. Research Question.....	2
1.3. Scope of the Thesis.....	3
1.4. Thesis organization.....	3
2. BACKGROUND INFORMATION AND LITERATURE REVIEW	5
2.1. Social Engineering	5
2.2. Information Retrieval Methods	10
2.3 Machine Learning Applications on Phishing Attacks.....	12
2.4 Machine Learning Classification Metrics.....	14
3. METHODOLOGY	17
3.1. Dataset	18
3.2. Preprocessing	20
3.2.1. Clearing Syntax.....	21
3.2.2. Tokenization	21
3.2.3. Removing Stop Words	22
3.2.4. Lemmatization	22
3.3. Sampling Methods.....	23
3.3.1 Non-Sampling.....	23
3.3.2 Under Sampling	24
3.3.3 Over Sampling	24
3.3.4 Synthetic Minority Over Sampling Technique (SMOTE)	24
3.4. Information Retrieval and Feature Selection	25

3.5.	Model Creation	29
3.5.1	Random Forest	29
3.5.2	Logistic Regression	29
3.5.3	AdaBoost.....	30
3.5.4	Support Vector Machine	30
4.	RESULTS	31
4.1.	Experiment Setup	31
4.2.	Results	31
5.	DISCUSSION AND CONCLUSION	35
5.1.	Discussion	35
5.2.	Conclusion.....	38
5.3.	Future Work	38
5.4.	Limitations of Thesis	39
(8)	REFERENCES	41
(9)	APPENDICES	49
(10)	APPENDIX A.....	49
(11)	APPENDIX B	59

LIST OF TABLES

Table 1 2018 Cyber Threats and Status Change according to ENISA Threat Landscape Report (European Union. Agency for Network and Information Security, 2019)	1
Table 2 Bag of Words (Term Frequency – Inverse Document Frequency) Example	11
Table 3 Machine Learning Applications on Phishing Attacks	13
Table 4 Keywords for searching phishing attacks in social media platform	18
Table 5 Most Phishing Tweets	19
Table 6 Dataset	20
Table 7 Under Sampling Data Size	24
Table 8 Over Sampling Data Size	24
Table 9 SMOTE Dataset Size	25
Table 10 Top Ten words of Dataset	25
Table 11 Binary Search of Feature Count and F1 scores.	27
Table 12 Results of F1-Scores and Means	28
Table 13 View of BoW (TF-IDF) Vector	28
Table 14 View of BoW (Frequency) Vector	28
Table 15 F1 and AUC Scores of Classifiers	31
Table 16 Precision and Recall Values of Classifiers	32
Table 17 Confusion Matrix of Classifiers	33
Table 18 Phishing Attack Detections	36
Table 19 Results of Other Studies Authors	37

LIST OF FIGURES

Figure 1 A genuine internet bank web site.	6
Figure 2 A bogus internet bank web site	7
Figure 3 Phishing mail with Linkedin attack vector and translation	8
Figure 4 Phishing SMS and translation	8
Figure 5 Phishing tweet and translation	9
Figure 6 Academic phishing	9
Figure 7 Bag of Words (Frequency)	11
Figure 8 Methodology	17
Figure 9 Implementation of Preprocessing	21
Figure 10 Stop Words in Turkish	22
Figure 11 Lemmatization Example	23
Figure 12 Finding Feature Count	26

LIST OF ABBREVIATIONS

BOW	Bag of Words
CWLC	Confidence-Weighted Linear Classifiers
ENISA	European Union Agency for Network and Information Security
FN	False Negative
FP	False Positive
HTML	Hypertext Markup Language
kNN	k-Nearest Neighbor
LR	Logistic Regression
ML	Machine Learning
NLP	Natural Language Processing
NLTK	Natural Language Toolkit
OCR	Optical Character Recognition
RAM	Random Access Memory
RBF	Radial Basic Function
RF	Random Forest
SMOTE	Synthetic Minority Over-Sampling Technique
SMS	Short Message Service
SVM	Support Vector Machine
TF-IDF	Term Frequency – Inverse Document Frequency
TN	True Negative
TP	True Positive
URL	Uniform Resource Locator
XML	Extensible Markup Language

CHAPTER 1

INTRODUCTION

1.1. Motivation

Cyber-attacks in our day consist of social engineering aspects, as well as technically complicated methodologies. Phishing attack is an illegal activity, in which the attacker mimics the genuine web site with a fake one. Using social engineering techniques, attackers deceive computer users in order to obtain their personal information like credit card number, username, password, etc. (Goodrich & Tamassia, 2014; Kishan Gajera, 2018)

Recently, mobile technologies find a broader use by the society in large compared to the previous years. The user of social media follows the spread of mobile devices, thus increasing the coverage in the population. Accordingly, cyber-attacks are getting broader in terms of their social engineering aspects. (European Union. Agency for Network and Information Security, 2019) reports that social media attacks through SMS and messaging applications have increased by 85% since 2011.

Human users are the main target of phishing attacks. In particular, attackers aim at creating attacks that can exploit humanly vulnerabilities, such as fear, anxiety, rather than vulnerabilities due to technical system components (Wu, Miller, & Little, 2006). In order to reach his goal, an attacker can send bulk emails or can use social media platforms by sending phishing tweets and SMS messages (Goodrich & Tamassia, 2014; Kishan Gajera, 2018; Toolan & Carthy, 2010; Varol et al., 2018).

According to a report by the European Union Agency for Network and Information Security (ENISA), both in 2017 and in 2018, phishing attacks were ranked as the fourth in the top threats ranking (Table 1).

Table 1 2018 Cyber Threats and Status Change according to ENISA Threat Landscape Report (European Union. Agency for Network and Information Security, 2019)

Top Threats of 2017	Top Threats of 2018
1. Malware	1. Malware
2. Web-based attacks	2. Web-Based Attacks
3. Web application attacks	3. Web Application Attacks

4. Phishing	4. Phishing
5. Spam	5. Denial of Service
6. Denial of service	6. Spam
7. Ransomware	7. Botnets
8. Botnets	8. Data Breaches
9. Insider threat	9. Insider Threat
10. Physical manipulation / damage / theft / loss	10. Physical Manipulation / damage / theft / loss
11. Data breaches	11. Information Leakage
12. Identity theft	12. Identify theft
13. Information leakage	13. Cryptojacking
14. Exploit kits	14. Ransomware
15. Cyber espionage	15. Cyber Espionage

An investigation of the USOM National Cert of Turkey web page¹ reveals that 56% of malicious URLs are Turkish phishing attack URLs (2014-2019). The main motivation of this thesis is to study Turkish phishing attacks given their increasing importance in our daily life.

1.2. Research Question

In literature there, existing methodologies about detecting phishing attacks with machine learning classifiers.

The aims of the present study are:

- Analyzing Turkish texts in phishing attack vectors,
- Detecting Turkish Phishing Attacks texts with machine learning classifiers,
- Comparing machine learning classifiers results on our dataset,
- Comparing BoW (TF-IDF) and BoW (Frequency) information retrieval methods.
- Comparing under sampling, over sampling and SMOTE sampling method results on our dataset,
- Creating and supplying a Turkish Phishing Attack dataset.

¹ <https://www.usom.gov.tr/zararli-baglantililar/1.html>

1.3. Scope of the Thesis

In this thesis, information retrieval methods have been used with supervised machine learning algorithms. Bag of Words (Frequency) and Bag of Words (TF-IDF) methods have been used for feature selection. The URLs of the phishing attacks are not in the scope of this thesis. The proposed method has been applied to only the text sections of phishing attacks.

Since there is no publicly available dataset, for both training and testing purposes the dataset of our research was created from Turkish phishing emails which came to ODTU Computer Center official email address, short messages (SMS) and social media Turkish phishing attacks which are collected during the study of this thesis on social media.

1.4. Thesis organization

This thesis is organized in five chapters. The motivation of our work and research question stated in this chapter. The researches on phishing attacks with background terminology and machine learning classifiers on phishing attacks explained in Chapter 2. Chapter 3 shows the methodology of our work. The results of our methodology with machine learning classifiers given in Chapter 4. Chapter 5 concludes the discussion and conclusion of the thesis with future work proposals.

CHAPTER 2

BACKGROUND INFORMATION AND LITERATURE REVIEW

In this chapter, background information of phishing attacks and countermeasures of these attacks with information retrieval methods which includes machine learning applications are given.

2.1. Social Engineering

Humans when compared to computers, more likely believe to other humans. Therefore, cyber criminals choose humans in order to bypass computer security systems, applications, etc. In social engineering, attackers by using some tricks and manipulating computer users, aim to steal personal information for infiltrating into information systems. Since humans are the most fragile ring in security chain, preventing these types of attacks are hard. Social engineering can be done in many ways. Most popular tactics are,

- *Pretexting*, inventing a history and pretending to be someone,
- *Baiting*, promising to give or giving some kind of gift to the victim,
- *Quid pro Quo*, pretending to help to the victim in order to obtain confidential information (Goodrich & Tamassia, 2014; Jain & Gupta, 2016; Salahdine & Kaabouch, 2019).

Phishing Attacks are a type of social engineering attack in which attackers use baiting tactic and aims to steal victim's credit card numbers, credentials, and personal information by utilizing victim's curiosity and fear emotions (Goodrich & Tamassia, 2014). On these attacks, the attacker sends a link with a text which triggers the victim's curiosity and fear emotions. If the victim clicks the link in the text, then the victim connects to a so-called web server or a site that looks like the real web site but which is indeed under the control of the attacker. By filling the HTML forms on the fake webpage, the victim may send their credentials unwittingly not to the real web site but actually to the attacker.

A phishing attack mostly involves three steps; firstly, there must be a company or an entity which is reliable and known by the victims, secondly the entity or company which is the

main part of the attack vector must have a real web site, thirdly the web site must have some HTML forms which take inputs from users (Goodrich & Tamassia, 2014; Jakobsson & Myers, 2006; Ramzan & Wuest, 2007; Verma, Shashidhar, & Hossain, 2012).

Phishing attacks are mostly executed in three phases, these are (Aleroud & Zhou, 2017):

- **Preparation of a phishing attack:** In that phase, attackers prepare their attack vector. In order to perform that they can use some penetration test tools or can write code for preparing their attack vector (Sponchioni, 2015). Attackers first copy the real internet page before starting a phishing campaign (Figure 1). After that, they decide on a URL that will be meaningful with the text in the attack vector (Figure 2). In addition, the bogus web site is served on a web server which is totally under control of the attacker. The attacker can also take some third-party https certificate in order to affect their victims more quickly by showing bogus web site like the real web site.

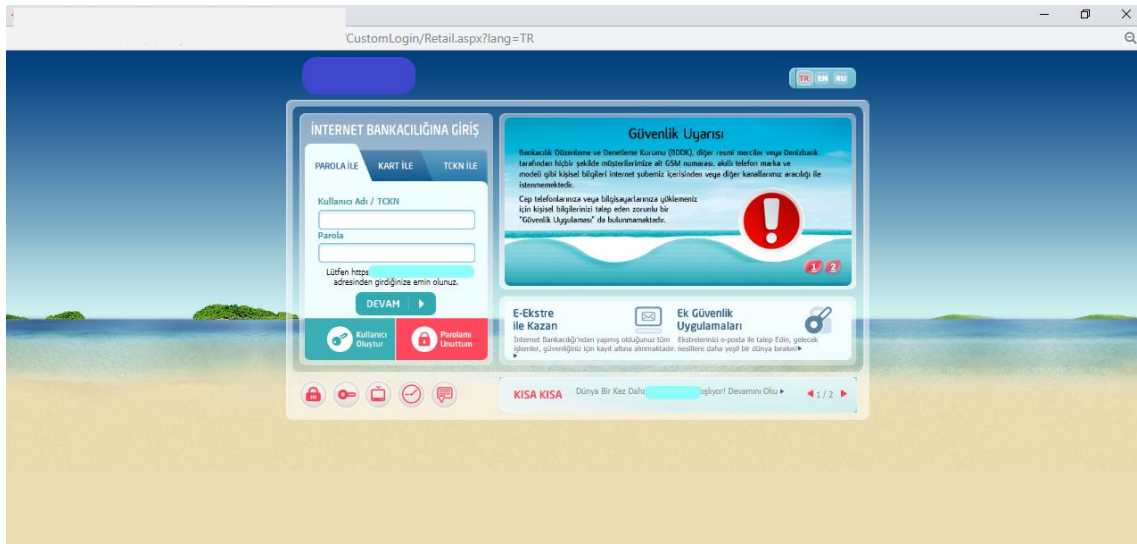


Figure 1 A genuine internet bank web site.

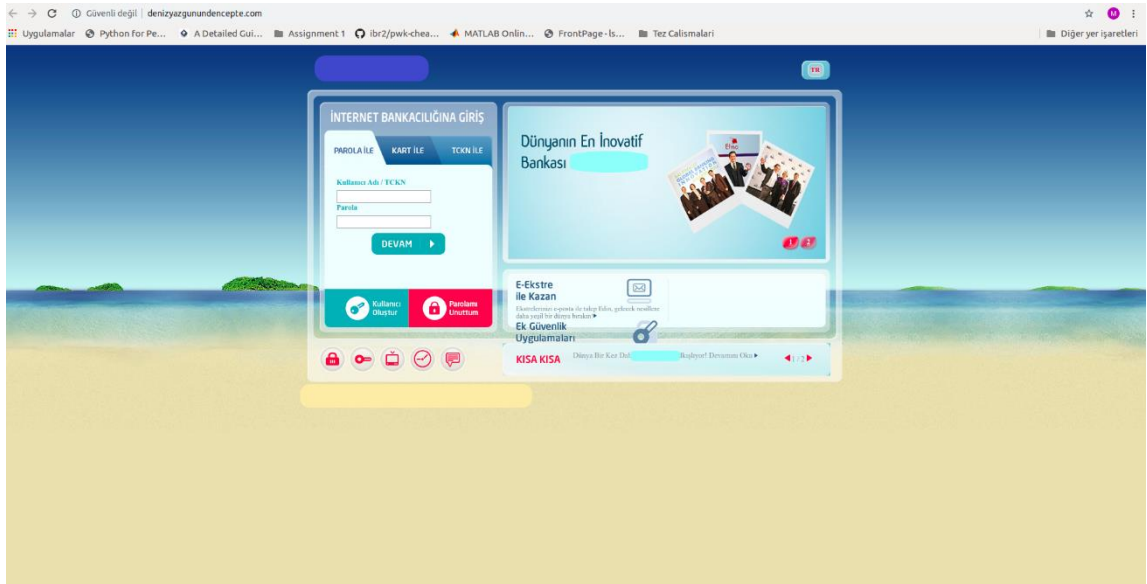


Figure 2 A bogus internet bank web site

- Execution of a phishing attack:** In this phase of the attack, the attacker chooses a media type. These can be emails (Figure 3), SMS (Figure 4) or social media platforms (Figure 5), etc. The text content is adapted according to the type of social media. In these texts, the attackers mostly use news on media of the country or a company, email, cloud systems, credit card or online payment topics and by doing that they aim to trigger victims' fear and curiosity emotions. The topic of the mail can also be academic as seen in Figure 6. Lastly, they send this attack vector to the bulk of users, also on social media platforms, attackers use advertisements as a media to prepare an attack vector opportunity.

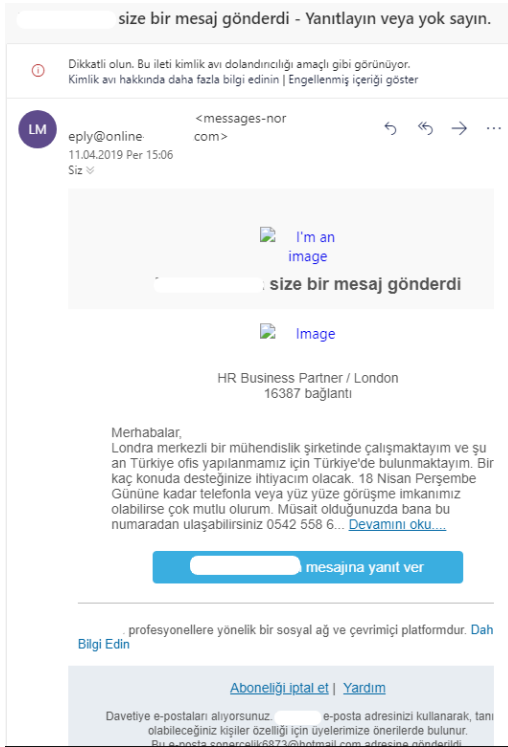


Figure 3 Translation:

..... has sent you a message.
Hello,
I am working in an engineering company based in London, and now in Turkey for our office structuring. I will need your support on a few issues Thursday, April 18 I would be very happy if we could have a phone call or a face-to-face meeting by the day. You can contact me at this number when you are available. 0542 558 6 ...
Continue Reading

Link (Give Answer to message)

Figure 3 Phishing mail with LinkedIn attack vector and translation



Figure 4 Translation:

Bonus Campaign from bank. The first 10000 people special to individual users 300 bonus points. For participation

Link

Figure 4 Phishing SMS and translation



Figure 5 Translation:
Official Information

From now on, our customers who log in from the link above and accept our active internet banking terms will receive a bonus of 2000 TL for all purchases. Participations are provided only from the entrance section above the image.

Figure 5 Phishing tweet and translation

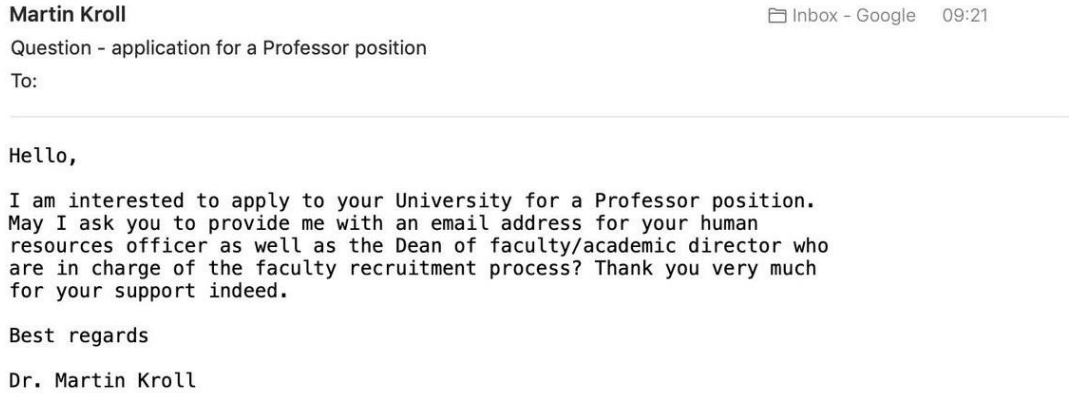


Figure 6 Academic phishing

- **The exploitation of the phishing attack:** This is the last step of the phishing attack. In this step, the attackers use credentials, credit card numbers, personal information of victims which they obtained during their campaign.

In this section, we have defined social engineering term and phishing attacks. In literature, there is research methods for studying phishing attacks. These are presented in the following sections.

2.2. Information Retrieval Methods

In text categorization, the need for representation of texts in computer brings us to information retrieval methods. By using these methods, we can represent our text data in the computer for using text categorization. Since each machine learning algorithm works with features, our words in the text are features (Thorsten Joachims, 1998). In this thesis two methods will be used, these are;

- Bag of Words (Term Frequency-Inverse Document Frequency),
- Bag of Words (Frequency);

explained respectively.

2.2.1 Bag of Words (Term Frequency-Inverse Document Frequency)

Term Frequency-Inverse Document Frequency (TF-IDF) is a method for extracting the weight of a word in a text. In a text, a word TF-IDF is a product of $f_{w,t}$, where word frequency in a text, and $\log(|T|/f_{w,T})$ where $|T|$ is a count of text in a text corpus, $f_{w,T}$ is the frequency of the word in text corpus (Ramos, 2003).

$$w_d = f_{w,t} * \log\left(\frac{|T|}{f_{w,T}}\right) \quad (2,1)$$

In our mail corpus let's say there is two mail Mail-1 = {*This is a phishing attack*} and Mail-2 = {*This is not a phishing attack*}. Since the mail corpus consists of two mail the results of BoW (TF-IDF) values mails can be seen. In this example, we can see the importance of {*not*}.

Table 2 Bag of Words (Term Frequency – Inverse Document Frequency) Example

Example	Mail-1	Mail-2	IDF	BoW (TF-IDF)	
				Mail-1	Mail-2
This	1	1	$\log\left(\frac{2}{2}\right) = 0$	0	0
is	1	1	$\log\left(\frac{2}{2}\right) = 0$	0	0
Not	0	1	$\log\left(\frac{2}{1}\right) = 1$	0	1
a	1	1	$\log\left(\frac{2}{2}\right) = 0$	0	0
Phishing	1	1	$\log\left(\frac{2}{2}\right) = 0$	0	0
Attack	1	1	$\log\left(\frac{2}{2}\right) = 0$	0	0

2.2.2 Bag of Words (Frequency)

Sentences are constructed from words and all together they have a meaning. Bag of word is a simple method that is used in text classification for transforming sentences into vectors (Thorsten Joachims, 1998; Zhang, Jin, & Zhou, 2010).

[0, 0]

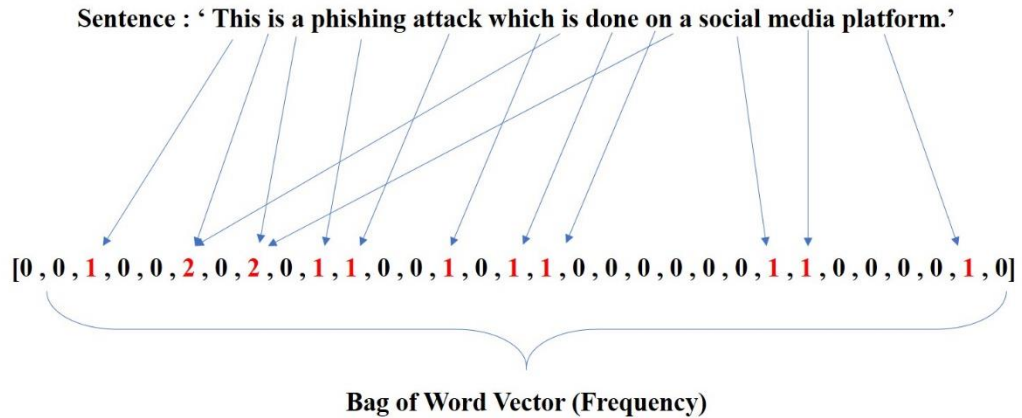


Figure 7 Bag of Words (Frequency)

All of the text in our corpus transformed into a sentence with their frequency in text. As it is seen in Figure 7 the sentence can be shown in vectors. All of the words that we are using in our language have their place in Bag of Words (Frequency). Therefore, the frequency of words in text will be written in their place. During our study we used Bag of Words (Frequency) information retrieval method.

2.3 Machine Learning Applications on Phishing Attacks

As it is stated in Table 1 (European Union. Agency for Network and Information Security, 2019) phishing attacks are forth in the world cyber-attacks. Since it is so popular there are lots of diverse scientific research in literature. (Alsharnouby, Alaca, & Chiasson, 2015) have tried to identify users' behaviors on phishing emails and sites by using eye trackers. (Birk, Gajek, Gröbert, & Sadeghi, 2007; Chandrasekaran, Chinchani, & Upadhyaya, 2006; Ferolin, 2012; Knickerbocker, Yu, & Li, 2009; Li & Schmitz, 2009; McRae & Vaughn, 2007) have worked on more proactive ways like phishing the phisher by sending fake entries to phishing sites. (Dhamija & Tygar, 2005; Medvet & Kruegel, 2008) have studied on the visuality of real web pages and phishing copies of them. One time password also has been studied as the prevention of phishing attacks (Khan, 2013). While (Jain & Gupta, 2016) have worked on the white list method, (Prakash, Kumar, Rao Kompella, & Gupta, 2010) have made research on blacklisting methods of phishing URLs.

Machine Learning Algorithms also have been used widely in researches (Table 3). (Garera, Provos, Chew, & Rubin, 2007) have worked on malicious phishing URLs and with their proposed methods by using Logistic Regression tried to detect phishing URLs. They classified phishing URLs in 4 types. Dataset they have used consists of 1,245 malicious and 1,263 non-malicious URL. The accuracy rate of the trained model was %95,8 true positive and %1.2 false positive rate.

(Zareapoor & Seeja, 2015) have classified phishing attacks in two categories as deceptive phishing and malware phishing. Their aim was to detect deceptive phishing attacks. They proposed a methodology with 4 steps. These are respectively, Preparing Dataset, Feature Extraction Using Text Mining, Feature Selection Using Selection Techniques and lastly Classification. The dataset they have used in their experiment is English and consists of totally 1900 emails with 500 phishing emails and 1400 legitimate e-mails. They have used 1067 features and their success above %99,5 with SVM, Ada Boost, Random Forest. In this thesis studied not only emails but also social media phishing attacks in Turkish.

(Fette, Sadeh, & Tomasic, 2007) have used not only text in their study but also some other features like emails contain javascript or not, number of dots, number of reference links, etc. They have used 860 phishing and 6950 non-phishing English emails. Information retrieval methods they have just used BoW (TF-IDF) and not social media sharing in their dataset.

(Abu-Nimeh, Nappa, Wang, & Nair, 2009) have used 71 features and 60 of them is BoW (TF-IDF) value. Without using social media phishing attacks, they have just used 5,152 raw English emails with 1,409 phishing email and in their dataset, they have found RF algorithm outperformed all of the other algorithms.

(Miyamoto, Hazeyama, & Kadobayashi, 2009) employ 9 machine learning algorithm techniques on 1,500 phishing and 1,500 legitimate URL, By using heuristic models that have been employed in CANTINA which includes BoW (TF-IDF) (Hong & Cranor, 2007). They have evaluated results. The highest result of their experiment was AdaBoost with 0.9342 AUC score. The dataset which they have been used in their experiment is English.

(Basnet & Sung, 2010) without using heuristic features they have used the contents of 5,152 ham and 1,409 phishing English emails. By using the English phishing mail corpus (Jose Nazario, 2019), they have deployed Confidence-Weighted linear classifiers. Firstly, they have implemented preprocessing methods like removing stop words, tokenization of words and stemming. After that, by using bag of word methods they have extracted features of the dataset. After the implementation of these, they have reached 99.77% accuracy rate.

(Toolan & Carthy, 2010) have used (Jose Nazario, 2019) 4,116 phishing email with 4,202 ham emails. In their study, they have used 5 features. These are IP address, HTML codes in email content, script in email content, number of links in email body, period numbers of link. In their experiment, they have studied C5.0, which is a decision tree algorithm, k-NN, SVM, Naïve Bayes, Linear Regression. SVM and C5.0 show the best accuracy rates over other machine learning algorithms.

Table 3 Machine Learning Applications on Phishing Attacks

Authors	Features	Algorithms	Dataset	
			Phishing	Non-Phishing
(Garera et al., 2007)	URLs	Logistic Regression	1,245	1,263
(Zareapoor & Seeja, 2015)	Email Content	SVM Ada Boost Random Forest	500	1,400
(Fette, Sadeh, & Tomasic, 2007)	URLs and Email Content	Random Forest	860	6,950

(Abu-Nimeh, Nappa, Wang, & Nair, 2009)	URLs and Email Content	Bayesian Additive Regression Trees Random Forest SVM Logistic Regression	1,409	5,152
(Miyamoto, Hazeyama, & Kadobayashi, 2009)	URLs	Ada Boost	1,500	1,500
(Basnet & Sung, 2010)	Email Content	Linear Classifiers	1,409	5,152
(Toolan & Carthy, 2010)	URLs, IP, Html Codes	k-NN Naïve Bayes Linear Regression SVM C5.0	4,116	4,202

2.4 Machine Learning Classification Metrics

During the conclusion of machine learning classifications Accuracy, Precision, Recall, , F1-Score and Area Under the ROC Curve (AUC) are used. These values can give an idea about our model which we have constructed during our research, also can enlighten us about our model have wrongs or not. *True Positive* (TP), i.e., phishing attacks which are detected as a positive; *True Negative* (TN), i.e., non-phishing instances that are classified as a negative; *False Positive* (FP), i.e., non-phishing instances which are detected as phishing attacks; *False Negative* (FN), i.e., phishing attacks which are detected as non-phishing instances.

Accuracy is the percentage of correctly detected data and is calculated as

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} . \quad (2.2)$$

Precision is the percentage of phishing attacks which are correctly detected and calculated as

$$Precision = \frac{TP}{TP+FP} . \quad (2.3)$$

Recall (Detection Rate) is the percentage of actual phishing attacks and calculated as

$$Recall = \frac{TP}{TP+FN} . \quad (2.4)$$

F1-Score is the weighted average of precision and recall:

$$F1 - Score = 2 \times \frac{(precision \times recall)}{(precision + recall)} . \quad (2.5)$$

The Area Under the Receiver Operating Characteristic Curve (ROC) is another performance metric used in machine learning studies. Receiver Operating Characteristic curve shows balance between FP (False Positive) and TP (True Positive). The AUC curve is a comprehensive assessment of the machine learning classifiers which performs across all decision thresholds and gives a single value between 0.5 and 1.

CHAPTER 3

METHODOLOGY

The implementation of methodology can be seen in Figure 8.

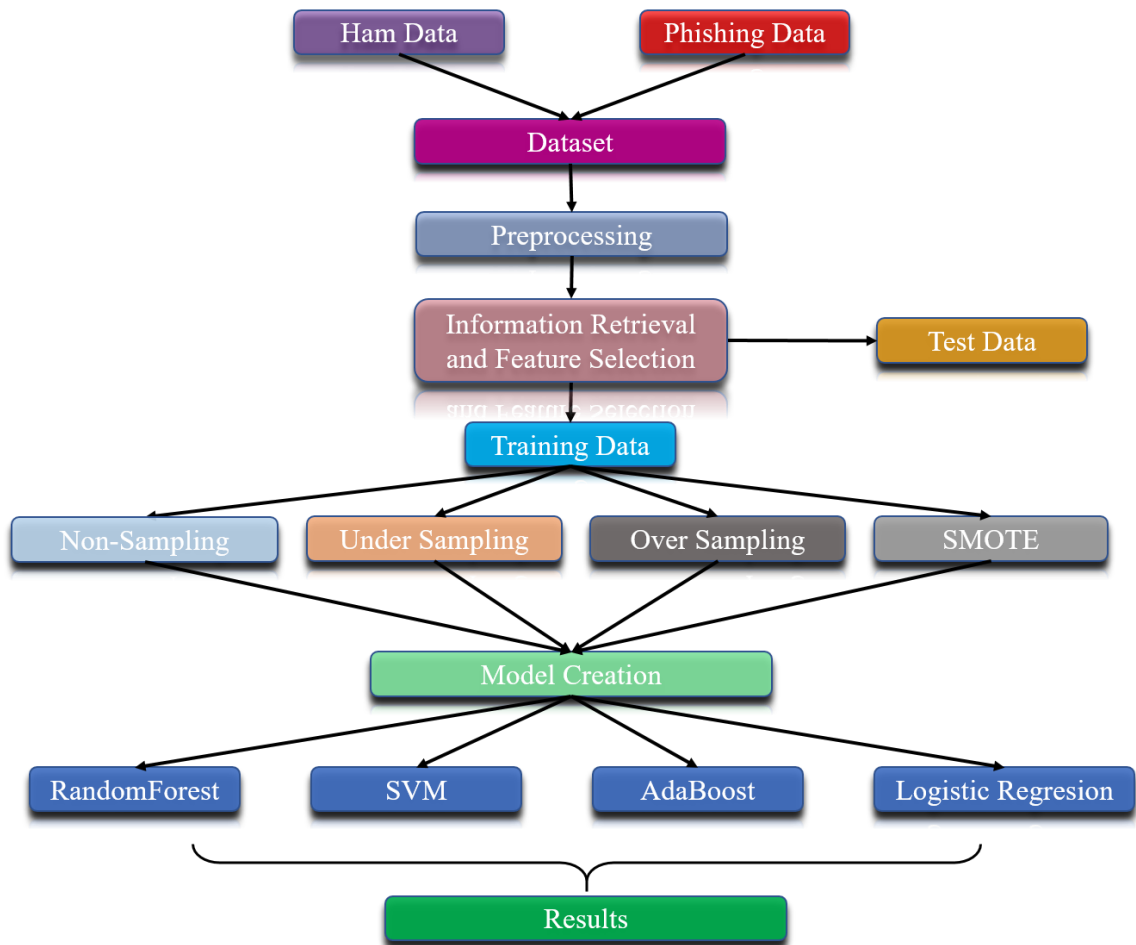


Figure 8 Methodology

The following sections described the details of the methodology employed in this thesis.

3.1. Dataset

The dataset which was collected for this thesis had two categories, the first category was ham emails and tweets category, the second one was phishing data. Ham emails obtained from METU Computer Center call center, namely the Hotline Service. The ham emails obtained in mbox² format and were checked manually before the implementation of the methodology. All non-Turkish emails were left out of the analyses. During the control of ham emails, Thunderbird Mail (version 60.8.0) application was used and with the extension of this application, all checked ham emails were saved in CSV format without email header. On the other hand, ham tweets were obtained from Twitter³. Since most of the phishing attacks were in the domain of banking, ham tweets were selected from Turkish Bank Tweets. 1,200 ham tweets were added to datasets that were tweeted by Turkish Banks on Tweeter.

The phishing data contains phishing emails that were obtained from the Hotline Service. Phishing SMS and social media posts were obtained from Twitter. All emails were checked manually to make sure that they include phishing attack texts. Sender and receiver parts of Turkish emails, links of emails were examined and after that, the emails were tagged as a phishing email. The Tweetdeck⁴ application was employed to collect phishing text from Twitter. Tweetdeck is an application that gives the opportunity to follow up multiple keywords at the same time on Twitter. The Tweetdeck keywords included Named Entities, as well as Turkish common words that are related to phishing. The keywords are presented in Table 4.

Table 4 Keywords for searching phishing attacks in social media platform

	Keywords	Translation
1	Ten Turkish Banks	Explicit bank names hidden for security
2	BDDK	Banking Supervision and Assessment Agency
3	Gelir İdaresi	Revenue Administration
4	Emniyet Genel Müdürlüğü	Turkish Police Agency
5	Oltalama	Phishing
6	Dolandırıcı	Swindler

All attacks were identified manually and checked by looking at the links inside the attack vectors. The phishing attacks were saved with screenshots. Since all phishing tweets were

² Mbox (Mail Box): A file format for holding email.

³ <https://www.twitter.com/>

⁴ <https://tweetdeck.twitter.com/>

with text messages in jpeg format, the screenshots of these attacks were recorded and after that written manually in CSV format.

As common in phishing attacks, the attackers can send the same text many times. Although all of the attack URLs are not the same, tweets have the same text. Therefore, these attacks were included to dataset only once. The texts were assumed to be different in case they involved at least one different word (except for the Named Entities). Most frequently observed phishing tweets have been shown in Table 5.

Table 5 Most Phishing Tweets

	<i>Phishing Tweet in Turkish</i>	<i>Translation in English</i>	<i>Frequency</i>
1	Kamuoyuna Önemli Duyuru Bugünden itibaren yeni internet şubemiz hizmete girmiştir. Yukarıdaki linkten ilk kez giriş sağlayan tüm müşterilerimize banka kriterlerinize göre 7,000TL ye kadar bonus anında hediye.	Important Announcement to the Public Since today our new internet application has been in service. All of our customers who have access from the above link for the first time will receive a bonus of up to 7,000TL according to your bank criteria.	7
2	Önemli Duyuru Tüm Müşterilerimize haftasonu hediyesi! Yukarıdaki linkten Şubemize giriş yapan tüm müşterilerimize 200 TL ile 2,000 TL arasında haftasonu bonusu anında hediye.	Important Announcement A weekend gift for all our customers! From the link above, all our customers entering our application will receive an instant gift of a weekend bonus between 200 TL and 2,000 TL.	5
3	Sayınbank müşterimiz, 421 TL kart aidatınız hesabınıza aktarılacaktır. Onay için	Dear ...bank customer, your 421 TL card fee will be transferred to your account. For approval	3
4	Kamuoyuna Duyuru Kredi kartı kullanan tüm vatandaşlarımızın aidat iadesi 4395 sayılı kanun gereğince 10,000TL ye kadar anında iadesi yapılmaktadır. Kredi kartlarınızın aidat iadesini almak için Yukarıdaki linke tıklayıp işleminizi gerçekleştirebilirsiniz.	Announcement to the Public According to the law no. 4,395, all citizens who use credit cards are refunded up to 10,000 TL. To get the dues refund of your credit cards, you can click on the link above.	5
5	bank Kamuoyuna Önemli Duyuru Bugünden itibaren yeni internet şubemiz hizmete girmiştir. Yukarıdaki Linkten ilk kez giriş sağlayan tüm müşterilerimize banka kriterlerinize göre 7,000 TL'ye kadar bonus anında hediye.	Bank Announcement to the Public Since today, our new internet application has been launched. All of our customers who log in for the first time from the above link will receive a bonus up to 7,000 TL according to your bank criteria.	8

6	Kamuoyuna Duyuru Kredi kartı kullanan tüm vatandaşlarımızın aidat iadesi 4395 sayılı kanun gereğince 10,000TL ye kadar anında iadesi yapılmaktadır. Kredi kartlarınızın aidat iadesini almak için	Announcement to the Public According to the law no. 4395, all citizens who use credit cards are refunded up to 10,000 TL. To get a refund of your credit cards	2
---	---	--	---

In total, the dataset included 119 different phishing examples. Since we had not many phishing samples, the dataset we created was an *imbalanced dataset*. The counts of non-phishing and phishing samples can be seen in Table 6. The ratio of total Phishing samples to Non-Phishing samples is 0.033. After dividing the dataset into training and test set with the ratio 0.33. The training dataset holds 2,363 non-phishing samples and 79 phishing samples with a ratio of 0.033. On the other hand, the test dataset holds 1,163 non-phishing samples and 40 phishing samples with a ratio of 0,034.

Table 6 Dataset

	Non-Phishing Samples	Phishing Samples	Ratio
Total	3,526	119	0,033
Training Data	2,363	79	0,033
Test Data	1,163	40	0,034

As a result of this study, the dataset containing phishing attacks will be public on the site www.turkceoltalama.org, which is created for supporting academic research in that area.

3.2. Preprocessing

This section of methodology contains four subsections which are respectively, clearing syntax, tokenization, removing stop words and lemmatization. All subsections have been implemented to the dataset respectively. The implementation of preprocessing can be seen in Figure 9.

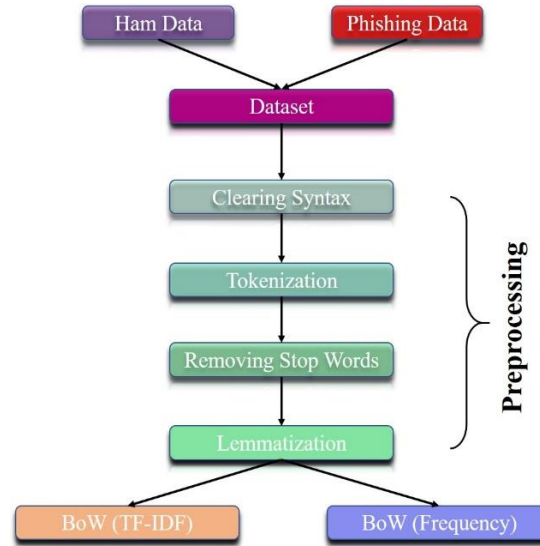


Figure 9 Implementation of Preprocessing

3.2.1. Clearing Syntax

This was the first step in the preprocessing section. In order to find the importance of words, this step was implied in the dataset. Since some of the emails contained HTML tags, it was important to find clear text inside emails. For clearing HTML tags, BeautifulSoup⁵ Library was used. BeautifulSoup, is a python library which is used for clearing text from unwanted syntax, parsing XML and HTML files and obtaining clean text from them (“BeautifulSoup,” 2019). After clearing HTML tags in dataset, other syntax like {‘,’, ‘\’, ‘”’, ‘|’, ‘/’, ‘\t’, ‘?’, ‘!’, ‘ ’, ‘%’, ‘#’, ‘@’} were cleared. Since the study includes only text features, the URL and other attachments to texts were cleared also in this section.

3.2.2. Tokenization

In order to implement the methodology, sentences must be converted to a vector. Therefore, tokenization came after clearing syntax. The function which was applied to texts takes the sentence and with the help of whitespaces by using as delimiters divide sentences into words (He et al., 2011a; Zareapoor & K. R, 2015; Zhai & Massung, 2016).

So, if in our dataset we have a simple sentence like ‘Bu bir oltalama saldırısıdır.’ (It is a phishing attack.), it will be shown in [‘Bu’, ‘bir’, ‘oltalama’, ‘saldırısıdır’, ‘.’] (‘It’, ‘is’, ‘a’, ‘phishing’, ‘attack’, ‘.’).

⁵ <https://pypi.org/project/beautifulsoup4/>

3.2.3. Removing Stop Words

Stop words are words that most frequently seen in a language. Stop words should be removed in pre-processing phase, because they have no effects in building classifiers (Basnet & Sung, 2010; He et al., 2011b; Meng, Lin, & Yu, 2011; Uğuz, 2011). We used NLTK⁶ library in Python, for removing these words. The list of stop words provided in Figure 10.

['acaba', 'ama', 'aslında', 'az', 'bazı', 'belki', 'biri', 'birkaç', 'birşey', 'biz', 'bu', 'çok', 'çünkü', 'da', 'daha', 'de', 'defa', 'diye', 'eğer', 'en', 'gibi', 'hem', 'hep', 'hepsi', 'her', 'hiç', 'için', 'ile', 'ise', 'kez', 'ki', 'kim', 'mı', 'mu', 'mü', 'nasıl', 'ne', 'neden', 'nerde', 'nerede', 'nereye', 'niçin', 'niye', 'o', 'sanki', 'şey', 'siz', 'şu', 'tüm', 've', 'veya', 'ya', 'yani']

('wonder', 'but', 'actually', 'few', 'some', 'maybe', 'one', 'a few', 'something', 'we', 'this', 'very', 'because', 'also', 'more', 'also', 'times', 'for that', 'if', '-est suffix', 'like', 'both', 'always', 'all', 'every', 'no', 'for', 'with', 'if', 'time', 'in', 'who', 'question suffix', 'question suffix', 'how', 'what', 'for what', 'where', 'where', 'to where', 'for what reason', 'why', 'it', 'as if', 'thing', 'you', 'there', 'all', 'and', 'or', 'or', 'that means')

Figure 10 Stop Words in Turkish

3.2.4. Lemmatization

Lemmatization means finding the root of a given word by eliminating its suffixes and derivative form, in order to represent the word as single item (Martin & Jurafsky, 2000). Lemmatization process decreases the feature count in dataset, thus reducing the possibility of overfitting. Turkish is an agglutinative language which means that words can be created with root and suffixes (Goksel & Kerslake, 2005).

In this study, the Zemberek⁷ tool was applied to the dataset in order to implement lemmatization. Zemberek is a tool created for Turkish NLP (Akın & Akın, 2007). In lemmatization, Zemberek finds the root of the word by cleaning suffixes of it. So, the only root of word stays in text. By doing that the repetitions of the same root with other suffixes can be stopped. For example, 'kitabımız' (our book) and 'kitaplara' (to the books) have the same root. It is 'kitap' (book). The root of the word can be found like that 'kita(b)p-ımız' and 'kitap-lar-a'. An example of a real phishing attack and the result of lemmatization have been shown in Figure 11.

⁶ <https://www.nltk.org/>

⁷ <https://github.com/ahmetaa/zemberek-nlp>

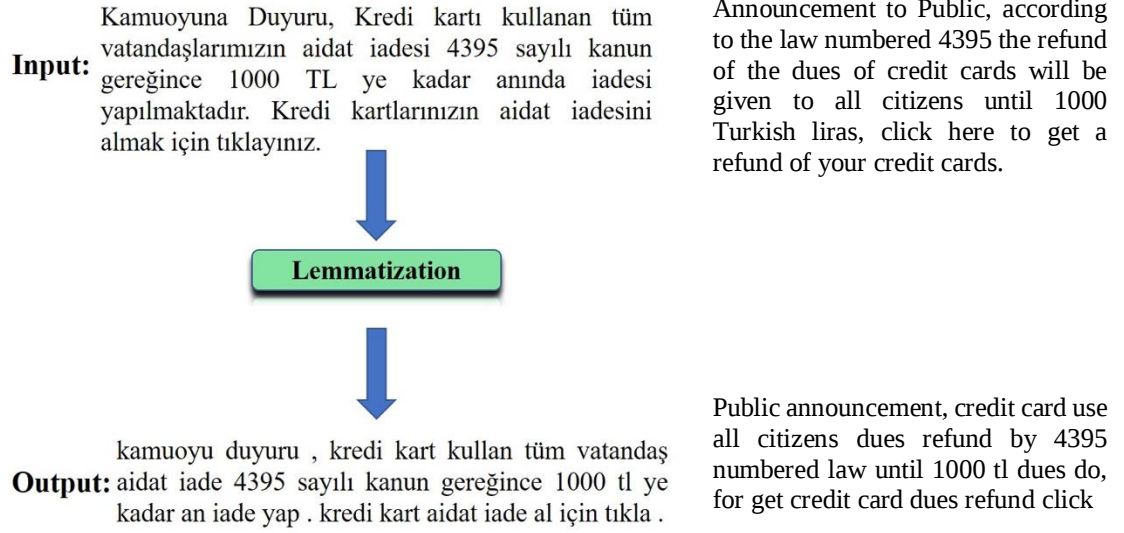


Figure 11 Lemmatization Example

As seen in the example the word ‘kartlarınız’ (your cards) was changed to ‘kart’(card) which is the root of the word. The suffixes of the words were eliminated.

3.3.Sampling Methods

The binary classification has two groups: one of them is tagged as a positive group and the other group is tagged as a negative group. In real-world problems, researches mostly counter with imbalanced data in which one group contains much more data than the other group (Johnson & Khoshgoftaar, 2019). As it is shown in Table 6 Dataset, our dataset is a typical imbalanced dataset with the ratio 0.034, where the number of phishing samples (119) is much smaller than the number of non-phishing samples (3,526). When dealing with imbalanced dataset, the available methods are Data-Level methods, Algorithm-Level methods and Hybrid methods. In this study, the Data-Level methods were employed. These were respectively *Under Sampling*, *Over Sampling*, and *Synthetic Minority Over-Sampling Method (SMOTE)*. In order to see the effects of data-level methods, we included non-sampling into our methodology, as explained below.

3.3.1 Non-Sampling

The term *non-sampling* is used to refer to the use of dataset without sampling. After taking preprocessed text vectors in training data, the sampling methods were not implemented, and they were directly given to machine learning algorithms. Therefore, the training size in Table 6 did not change in this case.

3.3.2 Under Sampling

Under Sampling means randomly eliminating data from majority class so that minority class size and majority class size become equal. While this method decreases the computing load, which may be an advantage in intense datasets, it also causes losing information from the majority class. The size of the dataset after applying under sampling is given in Table 7. Implementation of under sampling was done via the Scikit⁸ learn library.

Table 7 Under Sampling Data Size

	<i>Non-Phishing</i>	<i>Phishing</i>
<i>Under Sampling Training Dataset</i>	79	79

3.3.3 Over Sampling

Over Sampling is another type of re-sampling method. In this method, instances of majority class stay the same, but the minority class instances are increased randomly. Here no new instances are created, only the instances of the minority class were multiplied randomly. While this method avoids losing information on the majority class, repeated instances of minority class may result in overfitting (Chawla, 2009; Chawla, Japkowicz, & Kotcz, 2004). The size of the dataset after the application of over sampling is presented in Table 8. Implementation of over sampling was done via the Scikit learn library.

Table 8 Over Sampling Data Size

	<i>Non-Phishing</i>	<i>Phishing</i>
<i>Over Sampling Training Dataset</i>	2,363	2,363

3.3.4 Synthetic Minority Over Sampling Technique (SMOTE)

The application of random over sampling method may cause overfitting. One alternative to avoid the issues caused by the application of random over sampling is to create new training data during the training of the model without adding the same instances of the minority class. SMOTE sampling method adds new instance to minority class for

⁸ <https://scikit-learn.org/stable/>

compensating with majority class. It takes k neighbor minority class instances and creates line segments between these instances. Lastly, it creates new minority class instances on these line segments (Chawla, Bowyer, Hall, & Kegelmeyer, 2002). The implementation of SMOTE was done by the Scikit learn library. The parameter of k was left in default values, which is 5. The size of the dataset after the application of SMOTE is given in Table 9.

Table 9 SMOTE Dataset Size

	<i>Non-Phishing</i>	<i>Phishing</i>
<i>SMOTE Sampling Training Dataset</i>	2,363	2,363

3.4. Information Retrieval and Feature Selection

After the preprocessing of data, the information retrieval method was applied to the data. For the implementation, two methods were employed: BoW (TF-IDF) and BoW (Frequency). The Scikit learn library was used (Gupta, Tewari, Jain, & Agrawal, 2017; Hong & Cranor, 2007; Ramos, 2003; Zhang et al., 2010). Before choosing the parameters of information retrieval methods, the texts were processed in steps. First, the frequency of all words was calculated. After calculating the frequency of all words in the dataset, 10 most used words were removed from the dataset. These words can be seen in Table 10.

Table 10 Top Ten words of Dataset

	Word	Translation	Frequency
1.	bildirim	notification	9567
2.	odtu	M.E.T.U	5861
3.	destek	support	5713
4.	aşağıdaki	the following	3950
5.	tarihi	historical	3866
6.	lütfen	please	3860
7.	alınmıştır	has been taken	3828
8.	bildirim	the notification	3828
9.	bildiriminiz	your report	3825
10.	aksiyonlar	actions	3822

Also, the words which had less frequency (the bottom 10) were removed from the dataset.

As stated above, our dataset was an imbalanced dataset. Therefore, using an accuracy performance metric may lead to a misinterpretation of the results (Jeni, Cohn, & De La Torre, 2013; Johnson & Khoshgoftaar, 2019). For example, if the classifier tags all phishing samples as non-phishing, then our $TN = 1163$ and $FN = 40$. So, when we use equation 2.2, we find $Accuracy = \frac{0+1163}{0+1163+40+0} = \%96.7$. In other words, we end up with a very high accuracy score even without tagging one single phishing email. Because of this reason, we preferred to use F1-Score performance metric instead of default accuracy metrics.

BoW (TF-IDF) and BoW (Frequency) functions have `max_feature` parameter in Scikit library. This parameter value transforms most seen words in datasets to vector. For example, if we give ten to this parameter, BoW (TF-IDF) and BoW (Frequency) functions in Scikit learn will find mostly seen ten words in dataset and create vectors with these words.

For feature selection and finding optimum `max_feature` parameter of BoW (TF-IDF) and BoW (Frequency), binary search was implemented manually on the dataset. Binary search is an algorithm which is used for searching. It is applied by splitting the set into two equal subsets and comparing it with other for finding appropriate value (Rosen, 2007). For all samples, the process which is shown in Figure 12 was used. Firstly, a feature count was taken and given as a parameter into BoW (TF-IDF) and BoW (Frequency). Secondly, with the feature counts of text given in Table 11, the optimum parameters of all sampling methods and classifiers were found with the help of the Grid Search method in Scikit learn. Later, optimum parameters of ML classifiers were given as parameters and again all models with new parameters were created. Lastly, the means of F1 scores in the test dataset were taken and examined. Table 12 shows the optimum number of feature count given into `max_feature` parameter to BoW (TF-IDF) and BoW (Frequency).

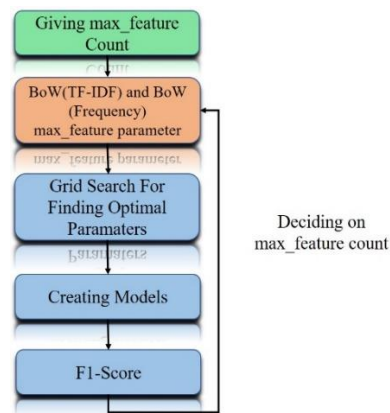


Figure 12 Finding Feature Count

Table 11 Binary Search of Feature Count and F1 scores.

<i>Number of Features</i>	<i>Algorithms</i>	<i>Non Sampling</i>	<i>Under Sampling</i>	<i>Random Over Sampling</i>	<i>SMOTE Sampling</i>
950	RF BoW (TF-IDF)	0.769	0.713	0.904	0.877
	RF BoW (Frequency)	0.824	0.587	0.904	0.620
	LR BoW (TF-IDF)	0.892	0.692	0.937	0.923
	LR BoW (Frequency)	0.868	0.718	0.872	0.604
	AdaBoost BoW (TF-IDF)	0.822	0.562	0.895	0.900
	AdaBoost BoW (Frequency)	0.846	0.486	0.872	0.661
	SVM BoW (TF-IDF)	0.909	0.747	0.909	0.909
	SVM BoW (Frequency)	0.857	0.685	0.842	0.508
932	RF BoW (TF-IDF)	0.841	0.578	0.919	0.829
	RF BoW (Frequency)	0.719	0.597	0.857	0.614
	LR BoW (TF-IDF)	0.892	0.698	0.937	0.923
	LR BoW (Frequency)	0.883	0.726	0.872	0.593
	AdaBoost BoW (TF-IDF)	0.880	0.585	0.909	0.872
	AdaBoost BoW (Frequency)	0.880	0.469	0.923	0.642
	SVM BoW (TF-IDF)	0.895	0.755	0.895	0.895
	SVM BoW (Frequency)	0.842	0.632	0.857	0.512
913	RF BoW (TF-IDF)	0.857	0.655	0.889	0.892
	RF BoW (Frequency)	0.877	0.649	0.873	0.642
	LR BoW (TF-IDF)	0.880	0.698	0.937	0.923
	LR BoW (Frequency)	0.883	0.725	0.871	0.594
	AdaBoost BoW (TF-IDF)	0.865	0.574	0.865	0.878
	AdaBoost BoW (Frequency)	0.868	0.479	0.887	0.654
	SVM BoW (TF-IDF)	0.895	0.763	0.895	0.895
	SVM BoW (Frequency)	0.842	0.626	0.857	0.517
875	RF BoW (TF-IDF)	0.806	0.747	0.873	0.845
	RF BoW (Frequency)	0.806	0.718	0.857	0.617
	LR BoW (TF-IDF)	0.892	0.692	0.937	0.923
	LR BoW (Frequency)	0.895	0.733	0.857	0.609
	AdaBoost BoW (TF-IDF)	0.829	0.571	0.895	0.838
	AdaBoost BoW (Frequency)	0.911	0.479	0.897	0.569
	SVM BoW (TF-IDF)	0.895	0.763	0.895	0.895
	SVM BoW (Frequency)	0.857	0.632	0.857	0.517

RF: Random Forest

LR: Logistic Regression

BoW: Bag Of Words

Table 12 Results of F1-Scores and Means

<i>Number of Features</i>	<i>Non-Sampling</i>	<i>Under Sampling</i>	<i>Random Over Sampling</i>	<i>SMOTE Sampling</i>	<i>Mean</i>
<i>All Features</i>	0.829	0.651	0.884	0.731	0.774
<i>2000</i>	0.835	0.629	0.880	0.738	0.770
<i>1500</i>	0.851	0.620	0.808	0.726	0.751
<i>1000</i>	0.857	0.652	0.892	0.730	0.783
<i>950</i>	0.848	0.649	0.892	0.750	0.785
<i>932</i>	0.854	0.630	0.896	0.735	0.779
<i>913</i>	0.871	0.646	0.884	0.748	0.787
<i>875</i>	0.861	0.667	0.884	0.727	0.786
<i>750</i>	0.857	0.638	0.883	0.741	0.780
<i>500</i>	0.874	0.625	0.874	0.734	0.777
<i>200</i>	0.751	0.397	0.546	0.548	0.560

The dataset was analyzed with feature selection, as a result, the optimum feature count was given into BoW (TF-IDF) and BoW (Frequency) as `max_feature` parameter. With unigram features, BoW (TF-IDF) and BoW (Frequency) vectors were created. An example of a BoW (TF-IDF) vector is presented in Table and BoW (Frequency) in Table 14.

Table 13 View of BoW (TF-IDF) Vector

	<i>alışveriş</i>	<i>an</i>	<i>analiz</i>	<i>ancak</i>	<i>ankara</i>	<i>anket</i>	<i>anla</i>	<i>antet</i>	<i>ara</i>	<i>arabirim</i>
0	0.0	0.0	0.000	0.0	0.0	0.073	0.0	0.0	0.0	0.0
1	0.0	0.0	0.140	0.0	0.0	0.102	0.0	0.0	0.0	0.0
2	0.0	0.0	0.155	0.0	0.0	0.114	0.0	0.0	0.0	0.0
3	0.0	0.0	0.139	0.0	0.0	0.102	0.0	0.0	0.0	0.0
4	0.0	0.0	0.135	0.0	0.0	0.010	0.0	0.0	0.0	0.0
5	0.0	0.0	0.092	0.0	0.0	0.068	0.0	0.0	0.0	0.0
6	0.0	0.0	0.133	0.0	0.0	0.098	0.0	0.0	0.0	0.0

Table 14 View of BoW (Frequency) Vector

	<i>alışveriş</i>	<i>an</i>	<i>analiz</i>	<i>ancak</i>	<i>ankara</i>	<i>anket</i>	<i>anla</i>	<i>antet</i>	<i>ara</i>	<i>arabirim</i>
0	0	0	0	0	0	1	0	0	0	0
1	0	0	1	0	0	1	0	0	0	0
2	0	0	1	0	0	1	0	0	0	0
3	0	0	1	0	0	1	0	0	0	0
4	0	0	1	0	0	1	0	0	0	0
5	0	0	1	0	0	1	0	0	0	0
6	0	0	1	0	0	1	0	0	0	0

It can be seen BoW (TF-IDF) vector ‘analiz’ (analyses) have different values that change due to sentences. However, the BoW (Frequency) vector holds the frequency of ‘analiz’ (analyses). Accordingly, BoW (TF-IDF) vector holds much more information than a BoW (Frequency) vector.

3.5. Model Creation

During our study, we have used four machine learning classifiers. In this section we have presented how we have found their optimal parameters during feature selection and model creation phase.

3.5.1 Random Forest

Random forest is an ensemble learning method that can be used for classification. During training, it creates decision trees and via using them it classifies the data. Random forest was applied by using Scikit learn library (Marchal, 2016). Random forest has several parameters, one is `n_estimator` and the other one is the `max_feature` parameter. During finding the optimal parameters for feature selection with grid search, `n_estimator` values was [5, 10, 20, 25, 30, 35, 50, 100, 150, 200], scoring method was `f1_score`, cross-validation parameter 5, `n_jobs` was -1 for using all cores of processor.

After finding optimal number of features, which is 913 in the recent study, grid search was applied again for finding optimal values. This time `n_estimator` was increased and set to these values [5, 10, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 100, 150, 175, 200, 250, 300, 350, 400, 450, 500]. The other parameters were same.

3.5.2 Logistic Regression

A popular method of binary classification is Logistic Regression. Logistic regression in our methodology was implied by Scikit learn library. For both BoW (TF-IDF) and BoW (Frequency) information retrieval methods it was used. During finding optimum feature count, regularization parameter in grid search was set to [0.0001, 0.001, 0.01, 1, 2, 5, 10, 20, 30, 50, 100, 200, 500], cross-validation parameter 5 and scoring method `f1_score`.

After finding optimal feature in grid search, regularization parameter was set to [0.0001, 0.001, 0.01, 1, 2, 5, 10, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 100, 150, 175, 200, 250, 300, 350, 400, 450, 500] cross validation parameter 5, scoring method `f1_score` and `n_jobs` was -1 for using all cores of processor.

3.5.3 AdaBoost

AdaBoost is a boosting method that fits weak classifiers and training them in order to increase success with random guessing. It was implied by Scikit learn python library. In grid search, Base estimator was set to None which means Decision Tree Algorithm, `n_estimator` parameters [1, 2, 5, 10, 20, 30, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 400, 500] and scoring parameter to `f1_score`.

After finding optimal feature count Base estimator stayed same, `n_estimator` parameters [5, 10, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 100, 150, 175, 200, 250, 300, 350, 400, 450, 500] and scoring parameter to `f1_score`.

3.5.4 Support Vector Machine

Support Vector Machines are based on finding the lowest true error. SVM works well on text classification due to their dealing mechanism with overfitting and many features (Joachims, 1998). During finding features, Grid search parameters were selected radial basic function (RBF) and linear as kernel and normalization parameters as [1, 10, 100, 1000] with `cross-validation` as 5.

After finding optimum features, for training the model with optimal parameters in Grid Search kernel was selected rbf and linear, regularization parameters were selected [0.0001, 0.001, 0.01, 1, 5, 10, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 100, 150, 175, 200, 250, 300, 350, 400, 450, 500] with `cross validation` as 5.

The results are presented in the next chapter.

CHAPTER 4

RESULTS

4.1.Experiment Setup

For evaluation of the classifiers, first, the dataset was separated into a training set and a test set were separated respectively 66% and 33%. After separating the dataset, the methodology in Section 3 was implemented and the optimal number of features found. Later, hypermeter tuning was done with parameters given in Section 3.5 for all sampling and classifiers. The evaluation was done on a computer with Intel Core i7-8550U CPU 1.80 GHz, 20 GB RAM and Linux Ubuntu 16.04 operating system.

4.2.Results

The dataset is an imbalanced dataset. As it is stated in Section 3.4 evaluating results solely on accuracy scores may be misleading. Therefore, performance metrics of classifiers have been evaluated on F1 score. The general view of all classifiers' F1 scores and AUC scores is presented in Table 15. The results show that the under sampling methods did not perform well on the selected classifiers. On the other hand, non-Sampling, over sampling, and SMOTE sampling training data were learned by the classifiers. In addition, the BoW (Frequency) as the information retrieval model did not show high performance, especially on the SMOTE sampling method.

In order to compare results of classifiers, we created dummy classifiers by using Scikit Library in python. All classifiers gave better results than dummy classifiers. Therefore, we can think that all classifiers learned to detect Turkish phishing attacks.

Table 15 F1 and AUC Scores of Classifiers

	Non-Sampling Test Data		Under Sampling Test Data		Over Sampling Test Data		SMOTE Test Data	
	F1	AUC	F1	AUC	F1	AUC	F1	AUC
RF BoW (TF-IDF)	0.806	0.837	0.698	0.889	0.907	0.900	0.892	0.862
RF BoW (Frequency)	0.823	0.850	0.673	0.889	0.873	0.900	0.596	0.922
LR BoW (TF-IDF)	0.892	0.912	0.755	0.903	0.889	0.947	0.923	0.949
LR BoW (Frequency)	0.883	0.923	0.758	0.891	0.872	0.923	0.584	0.895
AdaBoost BoW (TF-IDF)	0.880	0.911	0.550	0.875	0.883	0.911	0.886	0.922
AdaBoost BoW (Frequency)	0.883	0.923	0.522	0.862	0.886	0.935	0.496	0.897
SVM BoW (TF-IDF)	0.895	0.924	0.847	0.896	0.909	0.936	0.909	0.936
SVM BoW (Frequency)	0.842	0.910	0.643	0.884	0.857	0.910	0.525	0.878

Dummy BoW (TF-IDF)	0.067	0.461	0.049	0.547	0.062	0.520	0.060	0.564
Dummy BoW (Frequency)	0.050	0.462	0.046	0.470	0.071	0.504	0.064	0.531

RF: Random Forest

LR: Logistic Regression

BoW: Bag of Word

The results show that the Logistic Regression BoW (TF-IDF) with SMOTE Sampling Method is the one that showed the highest performance, returning the highest F1 score (0.923). It is followed by SVM BoW (TF-IDF) SMOTE Sampling (F1 score of 0.909), SVM BoW (TF-IDF) Over Sampling (0.909), and finally Random BoW (TF-IDF) Over Sampling (0.907). On the other hand, the BoW (Frequency) method gave the highest F1 score for AdaBoost BoW (Frequency) Over Sampling (0,886) and the AdaBoost BoW (Frequency) Non-Sampling (0,883), and then Logistic Regression BoW (Frequency) Non-Sampling (0.883).

Table 16 shows precision and recall values of all classifiers. Logistic Regression BoW (TF-IDF) with SMOTE Sampling Method precision is 0.947 and recall value is 0.90.

Table 16 Precision and Recall Values of Classifiers

	Non-Sampling Test Data		Under Sampling Test Data		Over Sampling Test Data		SMOTE Test Data		Total (Mean)	
	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
RF BoW (TF-IDF)	1.00	0.675	0.560	0.925	0.971	0.850	0.970	0.825	0.875	0.818
RF BoW (Frequency)	1.00	0.700	0.537	0.900	1.00	0.775	0.459	0.850	0.749	0.806
LR BoW (TF-IDF)	0.970	0.825	0.637	0.925	0.878	0.900	0.947	0.900	0.858	0.887
LR BoW (Frequency)	0.918	0.850	0.654	0.900	0.894	0.850	0.452	0.825	0.729	0.856
AdaBoost BoW (TF-IDF)	0.942	0.825	0.395	0.900	0.918	0.850	0.897	0.875	0.788	0.862
AdaBoost BoW (Frequency)	0.918	0.850	0.372	0.875	0.897	0.875	0.350	0.850	0.634	0.862
SVM BoW (TF-IDF)	0.944	0.850	0.800	0.900	0.945	0.875	0.945	0.875	0.908	0.875
SVM BoW (Frequency)	0.888	0.800	0.500	0.900	0.891	0.825	0.390	0.800	0.667	0.831

Dummy BoW (TF-IDF)	0.035	0.55	0.026	0.4	0.033	0.525	0.032	0.475	0.031	0.487
Dummy BoW (Frequency)	0.026	0.4	0.024	0.375	0.038	0.6	0.034	0.5	0.030	0.468
Mean	0.947	0.796	0.556	0.903	0.924	0.85	0.676	0.85		

The test data contained 1,163 Non-Phishing instances and 40 Phishing instances. The results are presented in Table 17. The results reveal that all the classifiers on all the sampling methods (except the Under Sampling) performed well in predicting Non-Phishing instances.

Table 17 Confusion Matrix of Classifiers

	Non-Sampling Test Data				Under Sampling Test Data				Over Sampling Test Data				SMOTE Test Data			
	N-Ph		Ph		N-Ph		Ph		N-Ph		Ph		N-Ph		Ph	
	Pr	N-Pr	Pr	N-Pr	Pr	N-Pr	Pr	N-Pr	Pr	N-Pr	Pr	N-Pr	Pr	N-Pr	Pr	N-Pr
RF BoW (TF-IDF)	1,163	0	27	13	1,134	29	37	3	1,162	1	34	6	1,162	1	33	7
RF BoW (Frequency)	1,163	0	28	12	1,132	31	36	4	1,163	0	31	9	1,123	40	34	6
LR BoW (TF-IDF)	1,162	1	33	7	1,142	21	37	3	1,158	5	36	4	1,161	2	36	4
LR BoW (Frequency)	1,160	3	34	6	1,144	19	36	4	1,159	4	34	6	1,123	40	33	6
Ada BoW (TF-IDF)	1,161	2	33	7	1,108	55	36	4	1,160	3	34	6	1,159	4	35	5
Ada BoW (Frequency)	1,160	3	34	6	1,104	59	35	5	1,159	4	35	5	1,100	63	34	6
SVM BoW (TF-IDF)	1,161	2	34	6	1,154	9	36	4	1,161	2	35	5	1,161	2	35	5
SVM BoW (Frequency)	1,159	4	32	8	1,127	36	36	4	1,159	4	33	7	1,113	50	32	8
Dummy BoW (TF-IDF)	569	594	18	22	566	597	24	16	556	607	19	21	597	566	21	19

Dummy BoW (Frequency)	584	579	24	16	573	590	25	15	558	605	16	24	601	562	20	20
TOTAL	1,163				40				1,163				40			

N-Ph: Non-Phishing

Ph: Phishing

N-Pr: Not Predicted

Pr: Predicted

Most of the classifiers gave false alarms to non-phishing instances while detecting most of the phishing attacks. On the other hand, among 40 phishing instances in the test dataset, 30 of 32 classifier models detected over 30 phishing instances (i.e., %75 of all the phishing instances were successfully detected).

As seen in Table 17, the Logistic Regression BoW (TF-IDF) with SMOTE Sampling Method from classified 1,161 as Non-Phishing instance out of 1,163 Phishing instances (99.8%). Also, out of 40 Phishing instances, the model classified 36 of them as Phishing instances.

CHAPTER 5

DISCUSSION AND CONCLUSION

5.1. Discussion

In our thesis, we studied four machine learning classifiers which are respectively Random Forest, Logistic Regression, AdaBoost and Support Vector Machine with two different information retrieval methods BoW (TF-IDF) and BoW (Frequency). Due to the imbalanced dataset which we collected during our work, we employed sampling methods. These were under sampling, over sampling, SMOTE sampling and non-sampling for comparing results. Therefore, totally we created 32 machine learning models.

In the literature, there are three methods to process imbalanced datasets: data-level methods, algorithm level methods and hybrid methods. In our study we used data-level methods. Therefore, we used sampling methods on the training dataset. The algorithm-level methods do not add new samples to a dataset. They effect classifiers during learning process. The algorithm-level methods are mostly employed through cost-sensitive approaches. This is employed by setting values in the cost matrix based on experiences (Johnson & Khoshgoftaar, 2019; Krawczyk, 2016). Since it is a hard process which based on experiences, we have preferred to use data-level methods, thus leaving the application of algorithm level methods to a further study.

As seen on Table 16, the average value of all the over sampling method precisions is 0.924, whereas the average value for the non-sampling method precisions is 0.947. On the other hand, the average recall values of under sampling (0.903), over sampling (0.85) and SMOTE sampling (0.85) are higher than the average recall of the non-sampling (0.796) methods. The results also show that BoW (TF-IDF) performs better than the BoW (Frequency) methods. In particular, the SVM BoW (TF-IDF) performs better than the other classifiers with an average precision value of (0.908). On the other hand, the Random Forest BoW (Frequency) has an average precision value of 0.749, and it performs better than the other classifiers that employ the BoW (Frequency) method. It can be inferred that BoW (Frequency) methods work well with Random Forest.

Cyber attackers mostly use the latest agenda items and the cultural values of the country to influence the people in phishing attacks. In order to affect their victims, they choose their phishing text words more carefully. However, when we analyzed our models results in Table 18. We can understand that some cyber criminals by translating English or other language phishing emails in to Turkish have attacked. First phishing attack in Table 18 is written in Turkish but it has some semantical mistakes and did not detected by our models. Also, second phishing email also have semantical mistakes but just Support Vector

Machine with BoW (TF-IDF) and AdaBoost BoW (TF-IDF) detected it. Therefore, in future works English phishing emails can be translated into Turkish and can be used with addition to our dataset.

Our dataset contains tweets and SMS Turkish phishing attacks. In Table 18, third phishing attack only detected by Logistic Regression BoW (Frequency) and fourth phishing attack only missed by AdaBoost BoW (TF-IDF). These samples are both semantically correct and look like real bank tweets.

Phishing attacks mostly contains malicious URLs inside them and tries to steal our personal information, credit card numbers, credentials, etc. On the other hand, Spams are mostly used in advertisement purposes without containing any malicious URLs. But both can be spread via same media types like emails, SMS etc. In literature, there is also some Turkish spam detection researches(Ergin & Isik, 2014; Özgür, Güngör, & Gürgen, 2004). As a future work, Turkish spam data can be used as training data and models can be tested with phishing dataset. Also, same features can be mapped between spam emails and phishing attacks.

Table 18 Phishing Attack Detections

	Phishing Attack	Translation	Lemmatization	Media Type	Results
1	Merhaba Sadece o güzel şeyler bir göz atın bu hiç havalı şeyler! Sadece şaşıracaksınız, şuna bak Web sitesi Take care,	Hello there Just take a look at those beautiful things that are the coolest things ever! Just be surprised, look at this website Take care,	merhaba sadece güzel şey bir göz at hava şey sadece şaşır şu bak site	Email	All classifiers have not detected that phishing attack.
2	Merhaba! Uzun bir süre için bu işler için aradığınızı biliyorum, ı kurmak o halde, burada bir göz atın açık bağlantı Sent from a prehistoric stone tablet, imoelvisudo	Hello! I know you've been looking for these jobs for a long time, I found it open connection, though, take a look here	merhaba uzun bir süre iş ara bil ı kur hal bura bir göz at açık bağlantı	Email	Only SVC BoW (TF-IDF) AdaBoost BoW (TF-IDF) have detected.
3	Haziran Ayında Yaptığınız Kargo alışverişi sebebiyle akıllı cep telefonlarında indirim kazandınız. Sipariş için	You gained discounts on smart mobile phones due to the Cargo Shopping in June. To order	haziran ay yap kargo alışveriş sebep akıl telefon indirim kazan sipariş	Tweet	Only Logistic Regression BoW (Frequency) have detected.

4	3 sorudan oluşan anketimize katılarak 10 GB internet kazanın	Earn 10 GB internet by participating in our 3-question survey	soru oluş anket katıl gb internet kazan	Tweet	Only AdaBoost BoW (TF-IDF) have missed.
---	--	---	---	-------	---

Table 19 shows results of other studies. (Basnet & Sung, 2010; Zareapoor & Seeja, 2015) investigated email content and they used a balanced dataset (see Jose Nazario, 2019 for a recent version of the dataset). (Zareapoor & Seeja, 2015) studies feature selection methods, respectively Chi-Square, Gain Ratio, Information Gain and without feature selection for comparing results. They have calculated the F1 scores for the SVM as 0.994, for AdaBoost 0.995 and for Random Forest 0.995. As seen in Table 15, our models have almost reached their value with Random Forest BoW (TF-IDF) Over Sampling (0.906), AdaBoost BoW (TF-IDF) SMOTE Sampling (0.8861) and SVM BoW (TF-IDF) SMOTE (0.909).

Table 19 Results of Other Studies Authors

	Features	Results		Dataset	
		Classifiers	F1 Scores	Phishing	Non-Phishing
(Zareapoor & Seeja, 2015)	Email Content	SVM	0.994	500	1,400
		Ada Boost	0.995		
		RF	0.995		
(Miyamoto, Hazeyama, & Kadobayashi, 2009)	URLs	AdaBoost	0.858	1,500	1,500
		LR	0.855		
		SVM	0.856		
		RF	0.855		
(Basnet & Sung, 2010)	Email Content	CWLC	0.998	1,409	5,152
		SVM	0.996		

(Basnet & Sung, 2010) compared two machine learning algorithms on the balanced dataset. They are respectively, Confidence-Weighted Linear Classifiers and SVM with a linear kernel. As a result, they reached respectively 0.998 and 0.997. On the other hand, (Miyamoto et al., 2009) created a dataset that contains only URLs and their experiment results lower than our results. So, we can infer that taking URLs as the only feature cannot work well.

5.2.Conclusion

This thesis presented a Turkish Phishing attack detection by comparing 32 machine learning models.

During our study, we collected 119 Turkish Phishing Attacks including emails, tweets, and SMS. In order to create more realistic dataset, we used Turkish ham emails and ham bank tweets. As a result, we had a real imbalanced dataset with a ratio of 0.033. To eliminate this data problem, we used different sampling methods respectively, under sampling, random over sampling, SMOTE over sampling and non-sampling for comparing results.

In the preprocessing section, firstly we cleaned syntax and removed Turkish stop words like *ve* (and), *evet* (yes), *tamam* (okey), etc. Turkish is agglutinative language, so having the same root but with different suffixes could give us more features. So, this could be resulted in overfitting problems and dimensionality. In order to eliminate this problem, we used the Zemberek Turkish NLP tool and found the roots of words with lemmatization.

For transforming texts into vectors, we employed BoW (TF-IDF) and BoW (Frequency) information retrieval methods. In order to find `max_feature` parameters of these methods, binary search was implemented. So, we found 913 optimal features. Finally, after finding optimal parameters for classifiers with grid search using cross-validation, 32 machine learning classifiers were trained.

As a result, after comparing 32 machine learning models, we found Logistic Regression with SMOTE sampling using BoW (TF-IDF) vectors as a better model on our dataset. In our dataset, BoW (TF-IDF) information retrieval methods gave better results than BoW (Frequency) method. Also, found that SMOTE and over sampling methods gave better results than under sampling method.

5.3.Future Work

As future work, dataset phishing size can be increased, and this dataset can be used in other studies. After increasing data samples, better model can be developed. URLs of phishing attacks, time, etc, can be added as new features to dataset, and this can increase accuracy results of classifiers. During creating vectors, other information retrieval methods can be deployed. In addition to data-level methods on imbalanced dataset, algorithm level methods or hybrid methods can be implemented.

In our research we have employed information retrieval methods with unigram features. Bigram features also can be used during vectorization of text. By employing word embeddings, deep learning algorithms can be used.

(Jose Nazario, 2019) English phishing dataset can be translated into Turkish and this dataset can be used as training dataset on machine learning classifiers. After testing with our dataset, the cultural differences and latest agenda items affects in emails can be found between English and Turkish phishing attacks.

Finding Turkish spam email sample is easier than finding Turkish phishing attack sample. Although spams and phishing attacks different form each other, they could have some same features. By employing transfer learning methods, features of Turkish spams could help our models to learn Turkish phishing attacks. Also, same features of these datasets can be mapped.

On the other hand, different feature selection methods like Wrapper Feature Selection, Chi-Square etc. can be implemented. These feature selection methods can be compared, and the best feature selection method can be found.

5.4. Limitations of Thesis

In this thesis, collected data from the internet have been used and this dataset is not a public dataset. While the size of the dataset with ratio 0.033 reflects the real world, the dimensionality of phishing attacks not reflects the real world. All data have been collected, checked manually and saved in CSV files. However, emails and tweets traffic have more topics than this dataset. Since most of the tweets contain pictures of texts, OCR implementation to models must be done.

REFERENCES

- Abu-Nimeh, S., Nappa, D., Wang, X., & Nair, S. (2009). Distributed phishing detection by applying variable selection using Bayesian additive regression trees. *IEEE International Conference on Communications*.
<https://doi.org/10.1109/ICC.2009.5198931>
- Akın, A. A., & Akın, M. D. (2007). Zemberek, An Open Source Nlp Framework for Turkic Languages. *Structure*, 10, 1–5. <https://doi.org/10.1.1.556.69>
- Aleroud, A., & Zhou, L. (2017). Phishing environments, techniques, and countermeasures: A survey. *Computers and Security*, Vol. 68, pp. 160–196.
<https://doi.org/10.1016/j.cose.2017.04.006>
- Alsharnouby, M., Alaca, F., & Chiasson, S. (2015). Why phishing still works: User strategies for combating phishing attacks. *International Journal of Human Computer Studies*. <https://doi.org/10.1016/j.ijhcs.2015.05.005>
- Basnet, R. B., & Sung, A. H. (2010). Classifying Phishing Emails Using Confidence-Weighted Linear Classifiers. *2010 International Conference on Information Security and Artificial Intelligence*, (Isai), 108–112.
- BeautifulSoup. (2019). Retrieved October 13, 2019, from
<https://pypi.org/project/beautifulsoup4/>
- Birk, D., Gajek, S., Gröbert, F., & Sadeghi, A. R. (2007). Phishing phishers - Observing and tracing organized cybercrime. *Second International Conference on Internet Monitoring and Protection, ICIMP 2007*. <https://doi.org/10.1109/ICIMP.2007.33>
- Chandrasekaran, M., Chinchani, R., & Upadhyaya, S. (2006). PHONEY: Mimicking user response to detect phishing attacks. *Proceedings - WoWMoM 2006: 2006 International Symposium on a World of Wireless, Mobile and Multimedia Networks*. <https://doi.org/10.1109/WOWMOM.2006.87>
- Chawla, N. V. (2009). Data Mining for Imbalanced Datasets: An Overview. In *Data Mining and Knowledge Discovery Handbook* (pp. 875–886).

https://doi.org/10.1007/978-0-387-09823-4_45

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321–357.

Chawla, N. V, Japkowicz, N., & Kotcz, A. (2004). Editorial. *ACM SIGKDD Explorations Newsletter*, 6(1), 19. <https://doi.org/10.1145/1007730.1007733>

Dhamija, R., & Tygar, J. D. (2005). The battle against phishing: Dynamic security skins. *ACM International Conference Proceeding Series*, 93, 77–88. <https://doi.org/10.1145/1073001.1073009>

Ergin, S., & Isik, S. (2014). The assessment of feature selection methods on agglutinative language for spam email detection: A special case for Turkish. *INISTA 2014 - IEEE International Symposium on Innovations in Intelligent Systems and Applications, Proceedings*, 122–125. <https://doi.org/10.1109/INISTA.2014.6873607>

European Union. Agency for Network and Information Security. (2019). *ENISA threat landscape report 2018 : 15 top cyberthreats and trends*. 138.

Ferolin, R. J. (2012). *A Proactive Anti-Phishing Tool Using Fuzzy Logic and RIPPER Data Mining Classification Algorithm*. 292–304.

Fette, I., Sadeh, N., & Tomasic, A. (2007, June 6). *Learning to detect phishing emails*. 649. <https://doi.org/10.1145/1242572.1242660>

Garera, S., Provos, N., Chew, M., & Rubin, A. D. (2007). A framework for detection and measurement of phishing attacks. *Proceedings of the 2007 ACM Workshop on Recurring Malcode - WORM '07*. <https://doi.org/10.1145/1314389.1314391>

Goksel, A., & Kerslake, C. (2005). *Turkish: A Comprehensive Grammar*.

Goodrich, M. T. (Michael T. ., & Tamassia, R. (2014). *Introduction to Computer Security: Pearson New International Edition*. Pearson Education Limited.

- Gupta, B. B., Tewari, A., Jain, A. K., & Agrawal, D. P. (2017). Fighting against phishing attacks: state of the art and future challenges. *Neural Computing and Applications*. <https://doi.org/10.1007/s00521-016-2275-y>
- He, M., Horng, S. J., Fan, P., Khan, M. K., Run, R. S., Lai, J. L., ... Sutanto, A. (2011a). An efficient phishing webpage detector. *Expert Systems with Applications*, 38(10), 12018–12027. <https://doi.org/10.1016/j.eswa.2011.01.046>
- He, M., Horng, S. J., Fan, P., Khan, M. K., Run, R. S., Lai, J. L., ... Sutanto, A. (2011b). An efficient phishing webpage detector. *Expert Systems with Applications*, 38(10), 12018–12027. <https://doi.org/10.1016/j.eswa.2011.01.046>
- Hong, J., & Cranor, L. (2007). CANTINA : A Content-Based Approach to Detecting Phishing Web Sites. *Www 2007*, 639–648.
- Jain, A. K., & Gupta, B. B. (2016). A novel approach to protect against phishing attacks at client side using auto-updated white-list. *Eurasip Journal on Information Security*. <https://doi.org/10.1186/s13635-016-0034-3>
- Jakobsson, M., & Myers, S. (2006). Phishing and Countermeasures: Understanding the Increasing Problem of Electronic Identity Theft. In *Phishing and Countermeasures: Understanding the Increasing Problem of Electronic Identity Theft*. <https://doi.org/10.1002/9780470086100>
- Jeni, L. A., Cohn, J. F., & De La Torre, F. (2013). Facing imbalanced data - Recommendations for the use of performance metrics. *Proceedings - 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, ACII 2013*, 245–251. <https://doi.org/10.1109/ACII.2013.47>
- Joachims, T. (1998). Text categorization with support vector machines: Learning with many relevant features. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 1398, 137–142. <https://doi.org/10.1007/s13928716>
- Johnson, J. M., & Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. *Journal of Big Data*, 6(1). <https://doi.org/10.1186/s40537-019-0192-5>

- Jose Nazario. (2019). Index of /~jose/phishing. Retrieved July 3, 2019, from <https://monkey.org/~jose/phishing/>
- Khan, A. A. (2013). Preventing Phishing Attacks using One Time Password and User Machine Identification. In *International Journal of Computer Applications* (Vol. 68). Retrieved from <https://www.mysite.com/getpassword.aspx>
- Kishan Gajera, M. J. (2018). A Novel Approach to Detect Phishing Attack Using Artificial Neural Networks Combined with Pharming Detection. *Proceedings of the 2nd International Conference on Electronics, Communication and Aerospace Technology, ICECA 2018*.
- Knickerbocker, P., Yu, D., & Li, J. (2009). Humboldt: A distributed phishing disruption system. *2009 ECrime Researchers Summit, ECRIME '09*. <https://doi.org/10.1109/ECRIME.2009.5342620>
- Krawczyk, B. (2016). Learning from imbalanced data: open challenges and future directions. *Progress in Artificial Intelligence*, 5(4), 221–232. <https://doi.org/10.1007/s13748-016-0094-0>
- Li, S., & Schmitz, R. (2009). A novel anti-phishing framework based on honeypots. *2009 ECrime Researchers Summit, ECRIME '09*. <https://doi.org/10.1109/ECRIME.2009.5342609>
- Marchal, S. (2016). *DNS and Semantic Analysis for Phishing Detection Network traffic profiling View project Semantic DNS analysis View project*. Retrieved from <https://www.researchgate.net/publication/303458208>
- Martin, J. H., & Jurafsky, D. (2000). Speech and language processing. In *Speech and language processing*.
- McRae, C. M., & Vaughn, R. B. (2007). Phighting the phisher: Using Web bugs and honeytokens to investigate the source of phishing attacks. *Proceedings of the Annual Hawaii International Conference on System Sciences*. <https://doi.org/10.1109/HICSS.2007.435>
- Medvet, E., & Kruegel, C. (2008). *Visual-Similarity-Based Phishing Detection*.

Retrieved from www.attacker.

Meng, J., Lin, H., & Yu, Y. (2011). A two-stage feature selection method for text categorization. *Computers and Mathematics with Applications*, 62(7), 2793–2800. <https://doi.org/10.1016/j.camwa.2011.07.045>

Miyamoto, D., Hazeyama, H., & Kadobayashi, Y. (2009). An evaluation of machine learning-based methods for detection of phishing sites. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5506 LNCS(PART 1), 539–546. https://doi.org/10.1007/978-3-642-02490-0_66

Özgür, L., Güngör, T., & Gürgen, F. (2004). Adaptive anti-spam filtering for agglutinative languages: A special case for Turkish. *Pattern Recognition Letters*, 25(16), 1819–1831. <https://doi.org/10.1016/j.patrec.2004.07.004>

Prakash, P., Kumar, M., Rao Kompella, R., & Gupta, M. (2010). PhishNet: Predictive blacklisting to detect phishing attacks. *Proceedings - IEEE INFOCOM*. <https://doi.org/10.1109/INFCOM.2010.5462216>

Ramos, J. (2003). *Using TF-IDF to Determine Word Relevance in Document Queries*.

Ramzan, Z., & Wuest, C. (2007). Phishing Attacks : Analyzing Trends in 2006. *Conference on Email and Anti-Spam*, 8. Retrieved from <https://pdfs.semanticscholar.org/ad58/0f4be151ef7b57cdbe29b838e23b44335674.pdf>

Rosen, K. H. (2007). *Discrete mathematics and its applications*. (903).

Salahdine, F., & Kaabouch, N. (2019). Social Engineering Attacks: A Survey. *Future Internet*, 11(4), 89. <https://doi.org/10.3390/fi11040089>

Sponchioni, R. (2015). The phishing economy: How phishing kits make scams easier to operate. *Phishing.Org*. Retrieved from <https://www.symantec.com/connect/blogs/phishing-economy-how-phishing-kits-make-scams-easier-operate>

- Thorsten Joachims. (1998). Text Categorization with Support Vector Machines: Learning with Many Relevant Features. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 1398, 119–124. <https://doi.org/10.1007/s13928716>
- Toolan, F., & Carthy, J. (2010). Feature selection for Spam and Phishing detection. *General Members Meeting and ECrime Researchers Summit, ECrime 2010*, 1–12. <https://doi.org/10.1109/ecrime.2010.5706696>
- Uğuz, H. (2011). A two-stage feature selection method for text categorization by using information gain, principal component analysis and genetic algorithm. *Knowledge-Based Systems*, 24(7), 1024–1032. <https://doi.org/10.1016/j.knosys.2011.04.014>
- Varol, A., Karabatak, M., Varol, C., Fırat Üniversitesi, Institute of Electrical and Electronics Engineers. Turkey Section, & Institute of Electrical and Electronics Engineers. (2018). *6th International Symposium on Digital Forensic and Security : proceeding book : 22-25 March 2018, Antalya, Turkey*.
- Verma, R., Shashidhar, N., & Hossain, N. (2012). Detecting phishing emails the natural language way. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7459 LNCS, 824–841. https://doi.org/10.1007/978-3-642-33167-1_47
- Wu, M., Miller, R. C., & Little, G. (2006). *Proceedings of the second symposium on ..., 2006 -cite141.pdf*.
- Zareapoor, M., & K. R. S. (2015). Feature Extraction or Feature Selection for Text Classification: A Case Study on Phishing Email Detection. *International Journal of Information Engineering and Electronic Business*, 7(2), 60–65. <https://doi.org/10.5815/ijieeb.2015.02.08>
- Zareapoor, M., & Seeja, K. R. (2015). Text Mining for Phishing E-mail Detection. *Advances in Intelligent Systems and Computing*. https://doi.org/10.1007/978-81-322-2012-1_8
- Zhai, C., & Massung, S. (2016). *Text Data Management and Analysis*.

Zhang, Y., Jin, R., & Zhou, Z. H. (2010). Understanding bag-of-words model: A statistical framework. *International Journal of Machine Learning and Cybernetics*, 1(1–4), 43–52. <https://doi.org/10.1007/s13042-010-0001-0>

APPENDICES

APPENDIX A

Codes of Classifiers and Sampling Methods

#Required Library

```
from bs4 import BeautifulSoup
from sklearn.metrics import classification_report
import pandas as pd
import email
import re, string, unicodedata
import nltk
import contractions
from bs4 import BeautifulSoup
from nltk import word_tokenize, sent_tokenize
from nltk.corpus import stopwords
from sklearn.metrics import roc_auc_score, accuracy_score, roc_curve
from sklearn.metrics import accuracy_score, f1_score, confusion_matrix, recall_score
from sklearn.model_selection import cross_val_score
import matplotlib.pyplot as plt
import numpy as np
import jpy as jp
from sklearn.ensemble import AdaBoostClassifier
import seaborn as sns
import matplotlib.pyplot as graph
from sklearn.model_selection import train_test_split
from sklearn.dummy import DummyClassifier
from sklearn.neighbors import KNeighborsClassifier
from sklearn.preprocessing import StandardScaler, OneHotEncoder
from sklearn.model_selection import GridSearchCV
from IPython.display import display
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import AdaBoostClassifier

ZEMBEREK_PATH = 'zemberek-full.jar'
jp.startJVM(jp.getDefaultJVMPath(), '-ea', '-Djava.class.path=%s' % (ZEMBEREK_PATH))
TurkishMorphology = jp.JClass('zemberek.morphology.TurkishMorphology')
Paths = jp.JClass('java.nio.file.Paths')
morphology = TurkishMorphology.createWithDefaults()
```

Reading Dataset and Cleaning Syntax

```
emails_df_ODTU_Oltalama = pd.read_csv('ODTU_Turkce_Oltalama.csv')
emails_df_ODTU_Oltalama['Body'].replace("\n", '', regex=True, inplace=True)
emails_df_ODTU_Oltalama['Body'].replace('https?:\\/(www\\.)?[-a-zA-Z0-9@:%_\\+~#={2,256}\\.[a-z]{2,6}\\b([-a-zA-Z0-9@:%_\\+~#?&/=]*)', '', regex=True, inplace=True)
emails_df_ODTU_Oltalama['Body'].replace('(https?:\\/(?:www\\.|(?!www\\.)[a-zA-Z0-9][a-zA-Z0-9-]+[a-zA-Z0-9]\\.[^\\s]{2,}|www\\.([a-zA-Z0-9][a-zA-Z0-9-]+[a-zA-Z0-9]\\.[^\\s]{2,}|https?:\\/(?:www\\.|(?!www\\.)[a-zA-Z0-9][a-zA-Z0-9-]+[a-zA-Z0-9]\\.[^\\s]{2,}|www\\.([a-zA-Z0-9][a-zA-Z0-9-]+[a-zA-Z0-9]\\.[^\\s]{2,}'))', regex=True, inplace=True)
emails_df_ODTU_Oltalama['Body'].replace('=', '', regex=True, inplace=True)
emails_df_ODTU_Oltalama['Body'].replace('\t', '', regex=True, inplace=True)
emails_df_ODTU_Oltalama = emails_df_ODTU_Oltalama.drop_duplicates(subset='Body', keep='first', inplace=False)
emails_df_ODTU_Oltalama['Body'].replace('\t', '', regex=True, inplace=True)
emails_df_ODTU_Oltalama['Body'].replace('|', '', regex=True, inplace=True)
emails_df_ODTU_Oltalama['Body'].replace('&nbsp', '', regex=True, inplace=True)
emails_df_ODTU_Oltalama['Body'].replace('<[^>]*>', '', regex=True, inplace=True)
emails_df_ODTU_Oltalama['Body'].replace('[-|[]|<|>|_ ', '', regex=True, inplace=True)
#emails_df_ODTU_Oltalama['Body'].replace('[^A-Za-z]+', '', regex=True, inplace=True)
emails_df_ODTU_Oltalama = emails_df_ODTU_Oltalama.rename(index=str, columns={"Body": "Text", "Class": "Class"})
```

```
emails_df_ODTU_Oltalama['Class'] = 1
# Taking emails which word counts are more than 5
emails_df_ODTU_Oltalama['word_count'] = emails_df_ODTU_Oltalama['Text'].apply(lambda x:
len(str(x).split(" ")))
#cond = emails_df_ODTU['word_count'] > 5
#emails_df_ODTU = emails_df_ODTU[cond]
emails_df_ODTU_Oltalama = emails_df_ODTU_Oltalama.reset_index(drop=True)
```

```
data2 = pd.read_csv('ODTU_Turkce_Ham.csv')
```

```
emails=[]
for item in data2['email_body']:
    emails_dict = {}
    full_email = email.message_from_string(item)
    body = full_email.get_payload()
    try:
        emails_dict["email_body"] = re.split(r"Kime:",body)[1]#"email body here"
    except IndexError:
        emails_dict["email_body"] = re.split(r"Kime:",body)[0]#"email body here"
    emails.append(emails_dict)
```

```

data2 = pd.DataFrame(emails)
data2['email_body'] = [BeautifulSoup(text).get_text() for text in data2['email_body']]
#data2 = data2.sample(frac=1).reset_index(drop = True)
data2['email_body'].replace('\n', ' ', regex=True, inplace=True)
data2['email_body'].replace('\w\S*@\.*\w', '', regex=True, inplace=True)
data2['email_body'].replace('https?:\V/(www\.)?[-a-zA-Z0-9@:%_\+~#&\/=]*', '', regex=True, inplace=True)
data2['email_body'].replace('(https?:\V/(?:www\.|(?:!www))[-a-zA-Z0-9]([-a-zA-Z0-9]+[-a-zA-Z0-9])\.\.[^s]{2,}|www\.[a-zA-Z0-9]([-a-zA-Z0-9]+[-a-zA-Z0-9])\.\.[^s]{2,})|https?:\V/(?:www\.|(?:!www))[-a-zA-Z0-9]([-a-zA-Z0-9]+[-a-zA-Z0-9])\.\.[^s]{2,}|www\.[a-zA-Z0-9]([-a-zA-Z0-9]+[-a-zA-Z0-9])\.\.[^s]{2,})', '', regex=True, inplace=True)
data2['email_body'].replace('=', '', regex=True, inplace=True)
data2['email_body'].replace('\t', '', regex=True, inplace=True)
data2 = data2.drop_duplicates(subset='email_body', keep='first', inplace=False)
data2['email_body'].replace('\t', '', regex=True, inplace=True)
data2['email_body'].replace('\'', '', regex=True, inplace=True)
data2['email_body'].replace('&nbsp', '', regex=True, inplace=True)
data2['email_body'].replace('<[>]*>', '', regex=True, inplace=True)
data2['email_body'].replace('-[|]|<|>|_', '', regex=True, inplace=True)
data2 = data2.rename(index=str, columns={"email_body": "Text", "Class": "Class"})

data2['Class'] = 0
data2['word_count'] = data2['Text'].apply(lambda x: len(str(x).split(" ")))
data2 = data2.reset_index(drop=True)

```

Merging Two Dataset

```

calisma = pd.concat([emails_df_ODTU_Oltalama ,data2])
calisma = calisma.reset_index(drop=True)
calisma['Text'].replace('[^A-Za-zıöüşçğİİÖÜŞÇĞ.]+' , ' ', regex=True, inplace=True)
stop = stopwords.words('turkish')
calisma['StopWords'] = calisma['Text'].apply(lambda x: " ".join(x for x in x.split() if x in stop))
calisma['StopWordsCounts'] = calisma['Text'].apply(lambda x: len([x for x in x.split() if x in stop]))
calisma['Text'] = calisma['Text'].apply(lambda x: " ".join(x for x in x.split() if x not in stop))
calisma['Text'] = calisma['Text'].apply(lambda x: " ".join(x.lower() for x in x.split()))

```

Lemmatization

```

def lemma(text):
    toplam = ""
    cumle = nltk.sent_tokenize(text)
    for sentence in cumle:
        results = morphology.analyzeAndDisambiguate(sentence).bestAnalysis()
        for word in results:
            #if str(word.getLemmas()[0]) == 'UNK':
            #    cumle += word + ' '
            #else:
            toplam += word.getLemmas()[0] + ' '

```

```
return toplam
```

```
newStopWords = ['UNK', '.', '...']
```

```
calisma['Text'] = calisma['Text'].apply(lemma)
```

```
calisma['Text'] = calisma['Text'].apply(lambda x: " ".join(x for x in x.split() if x not in newStopWords))
```

Splitting Dataset

```
from sklearn.model_selection import train_test_split
```

```
X_train, X_test, y_train, y_test = train_test_split(calisma['Text'], calisma['Class'], random_state=1,  
test_size=0.33)
```

Tf-IDF

```
from sklearn.feature_extraction.text import TfidfVectorizer
```

```
tfidf = TfidfVectorizer(max_features=913, lowercase=True, analyzer='word', ngram_range=(1,1))
```

```
train_vect_tfidf = tfidf.fit_transform(X_train)
```

```
train_data_features_tfidf = train_vect_tfidf.toarray()
```

```
test_data_features_tfidf = tfidf.transform(X_test)
```

```
test_data_features_tfidf = test_data_features_tfidf.toarray()
```

Bag of Words

```
from sklearn.feature_extraction.text import CountVectorizer
```

```
cv = CountVectorizer(max_features=913, analyzer="word",  
                    tokenizer=None,  
                    preprocessor=None,  
                    stop_words=None)
```

```
train_vect_cv = cv.fit_transform(X_train)
```

```
train_data_features_cv = train_vect_cv.toarray()
```

```
test_data_features_cv = cv.transform(X_test)
```

```
test_data_features_cv = test_data_features_cv.toarray()
```

Random Under Sampling

```
from imblearn.under_sampling import RandomUnderSampler
```

```
rus = RandomUnderSampler(random_state=56)
```

```
X_rusampled_tfidf, y_rusampled_tfidf = rus.fit_resample(train_data_features_tfidf, y_train)
```

```
X_rusampled_cv, y_rusampled_cv = rus.fit_resample(train_data_features_cv, y_train)
```

Random Over Sampling

```
from imblearn.over_sampling import RandomOverSampler
ros = RandomOverSampler(random_state=42)
X_rosampled_tfidf, y_rosampled_tfidf = ros.fit_resample(train_data_features_tfidf, y_train)
X_rosampled_cv, y_rosampled_cv = ros.fit_resample(train_data_features_cv, y_train)
```

SMOTE

```
from imblearn.over_sampling import SMOTE
sm = SMOTE(random_state=8)
X_smsampled_tfidf, y_smsampled_tfidf = sm.fit_resample(train_data_features_tfidf, y_train)
X_smsampled_cv, y_smsampled_cv = sm.fit_resample(train_data_features_cv, y_train)
```

Random Forest

```
forest_tfidf = RandomForestClassifier(n_estimators=75)
_ = forest_tfidf.fit(train_data_features_tfidf, y_train)

forest_cv = RandomForestClassifier(n_estimators=55)
_ = forest_cv.fit(train_data_features_cv, y_train)
```

Random Forest Under Sampling

```
forest_tfidf_rus = RandomForestClassifier(n_estimators=250)
_ = forest_tfidf_rus.fit(X_rusampled_tfidf, y_rusampled_tfidf)

forest_cv_rus = RandomForestClassifier(n_estimators=80)
_ = forest_cv_rus.fit(X_rusampled_cv, y_rusampled_cv)
```

Random Forest Over Sampling

```
forest_tfidf_ros = RandomForestClassifier(n_estimators=55)
_ = forest_tfidf_ros.fit(X_rosampled_tfidf, y_rosampled_tfidf)

forest_cv_ros = RandomForestClassifier(n_estimators=20)
_ = forest_cv_ros.fit(X_rosampled_cv, y_rosampled_cv)
```

Random Forest SMOTE

```
forest_tfidf_sm = RandomForestClassifier(n_estimators=45)
_ = forest_tfidf_sm.fit(X_smsampled_tfidf, y_smsampled_tfidf)
forest_cv_sm = RandomForestClassifier(n_estimators=10)
```

```
_ = forest_cv_sm.fit(X_smsampled_cv, y_smsampled_cv)
```

Random Forest Results

```
forest_tfidf_base_pred = forest_tfidf.predict(test_data_features_tfidf)
forest_tfidf_rus_pred = forest_tfidf_rus.predict(test_data_features_tfidf)
forest_tfidf_ros_pred = forest_tfidf_ros.predict(test_data_features_tfidf)
forest_tfidf_sm_pred = forest_tfidf_sm.predict(test_data_features_tfidf)
```

```
forest_cv_base_pred = forest_cv.predict(test_data_features_cv)
forest_cv_rus_pred = forest_cv_rus.predict(test_data_features_cv)
forest_cv_ros_pred = forest_cv_ros.predict(test_data_features_cv)
forest_cv_sm_pred = forest_cv_sm.predict(test_data_features_cv)
```

Support Vector Machine

```
svc_tfidf = SVC(C=5, cache_size=200, class_weight=None, coef0=0.0,
    decision_function_shape='ovr', degree=3, gamma='auto_deprecated',
    kernel='linear', max_iter=-1, probability=False, random_state=None,
    shrinking=True, tol=0.001, verbose=False)
```

```
svc_cv = SVC(C=350, cache_size=200, class_weight=None, coef0=0.0,
    decision_function_shape='ovr', degree=3, gamma=0.001, kernel='rbf',
    max_iter=-1, probability=False, random_state=None, shrinking=True,
    tol=0.001, verbose=False)
```

```
_ = svc_cv.fit(train_data_features_cv, y_train)
```

Support Vector Machine Under Sampling

```
svc_tfidf_rus = SVC(C=20, cache_size=200, class_weight=None, coef0=0.0,
    decision_function_shape='ovr', degree=3, gamma='auto_deprecated',
    kernel='linear', max_iter=-1, probability=False, random_state=None,
    shrinking=True, tol=0.001, verbose=False)
```

```
svc_cv_rus = SVC(C=400, cache_size=200, class_weight=None, coef0=0.0,
    decision_function_shape='ovr', degree=3, gamma=0.001, kernel='rbf',
    max_iter=-1, probability=False, random_state=None, shrinking=True,
    tol=0.001, verbose=False)
```

```
_ = svc_cv_rus.fit(X_rusampled_cv, y_rusampled_cv)
```

Support Vector Machine Over Sampling

```
svc_tfidf_ros = SVC(C=250, cache_size=250, class_weight=None, coef0=0.0,
    decision_function_shape='ovr', degree=3, gamma=0.001, kernel='rbf',
    max_iter=-1, probability=False, random_state=None, shrinking=True,
    tol=0.001, verbose=False)
```

```

svc_cv_ros = SVC(C=75, cache_size=200, class_weight=None, coef0=0.0,
    decision_function_shape='ovr', degree=3, gamma=0.001, kernel='rbf',
    max_iter=-1, probability=False, random_state=None, shrinking=True,
    tol=0.001, verbose=False)
_ = svc_cv_ros.fit(X_rosampled_cv, y_rosampled_cv)

```

Support Vector Machine SMOTE

```

svc_tfidf_sm = SVC(C=250, cache_size=200, class_weight=None, coef0=0.0,
    decision_function_shape='ovr', degree=3, gamma=0.001, kernel='rbf',
    max_iter=-1, probability=False, random_state=None, shrinking=True,
    tol=0.001, verbose=False)
_ = svc_tfidf_sm.fit(X_smsampled_tfidf, y_smsampled_tfidf)

```

```

svc_cv_sm = SVC(C=20, cache_size=200, class_weight=None, coef0=0.0,
    decision_function_shape='ovr', degree=3, gamma=0.001, kernel='rbf',
    max_iter=-1, probability=False, random_state=None, shrinking=True,
    tol=0.001, verbose=False)

```

```

_ = svc_cv_sm.fit(X_smsampled_cv, y_smsampled_cv)

```

Support Vector Machine Results

```

svc_tfidf_base_pred = svc_tfidf.predict(test_data_features_tfidf)
svc_tfidf_rus_pred = svc_tfidf_rus.predict(test_data_features_tfidf)
svc_tfidf_ros_pred = svc_tfidf_ros.predict(test_data_features_tfidf)
svc_tfidf_sm_pred = svc_tfidf_sm.predict(test_data_features_tfidf)

```

```

svc_cv_base_pred = svc_cv.predict(test_data_features_cv)
svc_cv_rus_pred = svc_cv_rus.predict(test_data_features_cv)
svc_cv_ros_pred = svc_cv_ros.predict(test_data_features_cv)
svc_cv_sm_pred = svc_cv_sm.predict(test_data_features_cv)

```

AdaBoost

```

ada_tfidf = AdaBoostClassifier(n_estimators=350)
_ = ada_tfidf.fit(train_data_features_tfidf, y_train)
ada_cv = AdaBoostClassifier(n_estimators=500)
_ = ada_cv.fit(train_data_features_cv, y_train)

```

AdaBoost Under Sampling

```

ada_tfidf_rus = AdaBoostClassifier(n_estimators=75)
_ = ada_tfidf_rus.fit(X_rusampled_tfidf, y_rusampled_tfidf)
ada_cv_rus = AdaBoostClassifier(n_estimators=350)
_ = ada_cv_rus.fit(X_rusampled_cv, y_rusampled_cv)

```

AdaBoost Over Sampling

```
ada_tfidf_ros = AdaBoostClassifier(n_estimators=250)
_ = ada_tfidf_ros.fit(X_rosampled_tfidf, y_rosampled_tfidf)

ada_cv_ros = AdaBoostClassifier(n_estimators=300)
_ = ada_cv_ros.fit(X_rosampled_cv, y_rosampled_cv)
```

AdaBoost SMOTE

```
ada_tfidf_sm = AdaBoostClassifier(n_estimators=500)
_ = ada_tfidf_sm.fit(X_smsampled_tfidf, y_smsampled_tfidf)

ada_cv_sm = AdaBoostClassifier(n_estimators=50)
_ = ada_cv_sm.fit(X_smsampled_cv, y_smsampled_cv)
```

AdaBoost Results

```
ada_tfidf_base_pred = ada_tfidf.predict(test_data_features_tfidf)
ada_tfidf_rus_pred = ada_tfidf_rus.predict(test_data_features_tfidf)
ada_tfidf_ros_pred = ada_tfidf_ros.predict(test_data_features_tfidf)
ada_tfidf_sm_pred = ada_tfidf_sm.predict(test_data_features_tfidf)

ada_cv_base_pred = ada_cv.predict(test_data_features_cv)
ada_cv_rus_pred = ada_cv_rus.predict(test_data_features_cv)
ada_cv_ros_pred = ada_cv_ros.predict(test_data_features_cv)
ada_cv_sm_pred = ada_cv_sm.predict(test_data_features_cv)
```

Logistic Regression

```
from sklearn.linear_model import LogisticRegression

log_tfidf = LogisticRegression(C=150)
_ = log_tfidf.fit(train_data_features_tfidf, y_train)
log_cv = LogisticRegression(C=300)
_ = log_cv.fit(train_data_features_cv, y_train)
```

Logistic Regression Under Sampling

```
log_tfidf_rus = LogisticRegression(C=1)
_ = log_tfidf_rus.fit(X_rusampled_tfidf, y_rusampled_tfidf)
log_cv_rus = LogisticRegression(C=10)
_ = log_cv_rus.fit(X_rusampled_cv, y_rusampled_cv)
```

Logistic Regression Over Sampling

```
log_tfidf_ros = LogisticRegression(C=1)
_ = log_tfidf_ros.fit(X_rosampled_tfidf, y_rosampled_tfidf)

log_cv_ros = LogisticRegression(C=2)
_ = log_cv_ros.fit(X_rosampled_cv, y_rosampled_cv)
```

Logistic Regression SMOTE

```
log_tfidf_sm = LogisticRegression(C=10)
_ = log_tfidf_sm.fit(X_smsampled_tfidf, y_smsampled_tfidf)

log_cv_sm = LogisticRegression(C=0.1)
_ = log_cv_sm.fit(X_smsampled_cv, y_smsampled_cv)
```

Logistic Regression Results

```
log_tfidf_base_pred = log_tfidf.predict(test_data_features_tfidf)
log_tfidf_rus_pred = log_tfidf_rus.predict(test_data_features_tfidf)
log_tfidf_ros_pred = log_tfidf_ros.predict(test_data_features_tfidf)
log_tfidf_sm_pred = log_tfidf_sm.predict(test_data_features_tfidf)

log_cv_base_pred = log_cv.predict(test_data_features_cv)
log_cv_rus_pred = log_cv_rus.predict(test_data_features_cv)
log_cv_ros_pred = log_cv_ros.predict(test_data_features_cv)
log_cv_sm_pred = log_cv_sm.predict(test_data_features_cv)
```


APPENDIX B

Turkish Phishing Attacks

Body
Sevgili Webmail Kullanıcı, Bizim veri tabanı ve tüm hesapların yeniden onayında hesabınızda görüntülenen bazı şüpheli olaylar gerçekleştirilebilir yükseltmek Anında tüm (Edu Webmail) kullanıcılardan istendi, biz sizin bilgilerinizi aldı ama hata devam etti. Aşağıdaki bağlantıyı takip ve bize düzgün doğrulamak ve bu sadece kısa bir süre içinde yapılabilir olarak hesabınızı yükseltmek yardımcı olmak için gerekli bilgileri doldurmak için tavsiye vardır. Hızla cevap verin. WEBMAIL ARAYÜZÜ V 1.4.22 http://mailserverupgrade.tripod.com Takım Yönetimi Sistemi, Copyright © 2015 Webmaster ubi.pt tarafından desteklenmektedir
-- Sevgili (ODTÜ) kullanıcı Webmail, Kimliğinizi doğrulamak. Biz bazı hesap bilgileriniz gibi görünüyor eksik veya yanlış olarak 0,99 GB üzerinde nakit postane güncelleştirene kadar zaman yeni e-posta gönderip mümkün olmayabilir çalışmakta olan yönetici tarafından belirlenen 100 GB depolama sınırını aştı, posta kutunuzun kullanmaya devam etmek için ve hesap bilgilerinizin doğrulanması gerekiyor fark ettik , Lütfen aşağıdaki linke tıklayarak ve hesap bilgilerinizin doğrulanması Bilgilerinizi güncellemek için buraya tıklayın: http://techupgradde.ucoz.net/metu-edu.html.html 24 saat içinde doğrulanmadı tüm hesapları veritabanından kaldırılır. İlginiz için teşekkür ederiz. Teknik destek, Yardım Masası sistemi. @ Doğru 2015 (ODTÜ) telif hakkı. Bütün hakları, Web e-posta yönetim ekibi.
Posta kota seti kotası / sınırını aştı ve şu anda nedeniyle posta kutusuna gizli dosya ve klasör Düşük GB Açık çalışıyor. Alıyorsunuz ya da web posta klasörlerde yer sağlamak için yeniden doğrulamak kadar yeni posta göndermek için mümkün olmayabilir. Bu aynı zamanda, daha önce tavsiye olarak webmail doğrulayarak değil neden olabilir.

tıklayarak işe yaramazsa, kopyalama ve doğrulamak için bir web tarayıcısında aşağıdaki URL'yi yapıştırın.

Lütfen hesabınızı yeniden doğrulamak ve bizim web posta veritabanından DE-aktive ölemek için tıklayın veya kopyalayıp tarayıcınıza bağlantıyı :::
<http://webupgrade2016microsoft-zim365.jimdo.com/yapıştırın>.

E-postanızı doğrulamak için başarısızlık, webmail kota / sınır webmail olarak önemli bilgilerin Of kaybına neden olabilir ve bu kendi web posta hesabınıza sınırlı erişim neden olacaktır ..

Genel uyarı,
Bu bizim sunucu üzerinde bir değişiklik ve güncelleme olduğunu size bildirmek içindir, tüm müşterilerin hesaplarının sonlandırma veya askıya alınmasını önlemek için kendi bilgi ve webmail hesabınızı güncellemek için gereklidir.
Bilgilerinizi güncellemek için aşağıdaki gerekli güncelleme bilgilerini doldurun
Ad Soyad:
E-posta:
Kullanıcı adı:
Parola:
Şifreyi Onayla:
Cep numarası:
Yukarıdaki bilgileri doldurun doğru webmail hesabınız otomatik olarak güncellenecek.
Selamlar
Yönetici.

Merhaba!

Uzun bir süre için bu işler için aradığımı biliyorum, ı kurmak o halde, burada bir göz atın açık bağlantı

Sent from a prehistoric stone tablet, imoelvisudo

MS-PHD mailing list
MS-PHD@metu.edu.tr
<https://mailman.metu.edu.tr/mailman/listinfo/ms-phd>

Merhaba!

Sana kaç ı takdir etmek senin yardım etmek söylemek istiyorum ve bu nedenle çok iyi şeyler paylaşmak, istiyorum sadece burada bir göz atın İleti

Best Wishes,

Merhaba

Sadece o güzel şeyler bir göz atın bu hiç havalı şeyler! Sadece şaşıracaksınız, şuna bak Web sitesi

Take care,

Veri depolama ve güvenliğini artırmak için şu an posta sunucumuzu sürdürüyoruz. Veri tabanını korumak, güvenliğini arttırmak ve tüm hesaplarını bilgisayar korsanlarından korumak için kullanılır. Bu nedenle, kesintiyi, veri kaybını veya askıya alınmayı önlemek için posta hesabınızı manuel olarak onaylamanız gerekiyor.

Ad Soyad:

Kullanıcı Adı:

Şifre:

Doğum tarihi:

Yukarıdaki bilgileri gönderdikten sonra hesabınız kesintiye uğramaz ve bakım periyodu sırasında ve sonrasında normal şekilde devam eder. Yukarıdaki bilgilerin gönderilmemesi, bilgilerin kaybolmasına veya posta hesabınızın askıya alınmasına neden olabilir ve sorumlu tutulamayacağız.

METU Hesap Yönetimi terim

Dava numarası: 345287

Mülk: Hesap Bakımı ve Güvenliği

Veri depolama ve güvenliğini artırmak için şu an posta sunucumuzu sürdürüyoruz. Veri tabanını korumak, güvenliğini arttırmak ve tüm hesaplarını bilgisayar korsanlarından korumak için kullanılır. Bu nedenle, kesintiyi, veri kaybını veya askıya alınmayı önlemek için posta hesabınızı manuel olarak onaylamanız gerekiyor.

Ad Soyad:

Kullanıcı Adı:

Şifre:

Doğum tarihi:

Yukarıdaki bilgileri gönderdikten sonra hesabınız kesintiye uğramaz ve bakım periyodu sırasında ve sonrasında normal şekilde devam eder. Yukarıdaki bilgilerin gönderilmemesi, bilgilerin kaybolmasına veya posta hesabınızın askıya alınmasına neden olabilir ve sorumlu tutulamayacağız.

METU Hesap Yönetimi terim

Dava numarası: 345287

Mülk: Hesap Bakımı ve Güvenliği

Konu:

Veri depolama ve güvenliğini artırmak için şu an posta sunucumuzu sürdürüyoruz. Veri tabanını korumak, güvenliğini arttırmak ve tüm hesaplarını bilgisayar korsanlarından korumak için kullanılır. Bu nedenle, kesintiye, veri kaybını veya askıya alınmayı önlemek için posta hesabınızı manuel olarak onaylamanız gerekiyor.

Ad Soyad:

Kullanıcı Adı:

Şifre:

Doğum tarihi:

Yukarıdaki bilgileri gönderdikten sonra hesabınız kesintiye uğramaz ve bakım periyodu sırasında ve sonrasında normal şekilde devam eder. Yukarıdaki bilgilerin gönderilmemesi, bilgilerin kaybolmasına veya posta hesabınızın askıya alınmasına neden olabilir ve sorumlu tutulamayacağız.

METU Hesap Yönetimi terim

Dava numarası: 345287

Mülk: Hesap Bakımı ve Güvenliği

E-postanız depolama alanı sınırı 2.GB'yi aştı

Yönetici tarafından kurulan şu an 2.30 GB, yapamıyor

Hesabınızı yeniden doğrulayana kadar yeni mesaj gönderebilir veya alabilirsiniz.

E-postanızı doğrulamak için aşağıdaki bağlantıyı tıklayın

<http://e-posta-hizmeti.my-free.website/>

teşekkür ederim

Sistem yöneticisi

E-postanız depolama alanı sınırı 2.GB'yi aştı
Yönetici tarafından kurulan şu an 2.30 GB, yapamıyor
Hesabınızı yeniden doğrulayana kadar yeni mesaj gönderebilir veya alabilirsiniz.
E-postanızı doğrulamak için aşağıdaki bağlantıyı tıklayın
<http://e-posta-hizmeti.my-free.website/>
teşekkür ederim
Sistem yöneticisi

131 MB

senin

METU posta kutusunun kontenjanı doludur ve şu anda bu bağlantıya göz atarak güncellenmelidir >> >>> <http://e-postaportal.weebly.com/>
Not: Bağlantıyı tıklayamazsanız, güncelleme için Posta Kontenjanı miktarınızı artırmak için yeni bir tarayıcıda kopyalayıp yapıştırmanız önerilir. Güncelleme için gerekli doğru doğrulama bilgilerini göndermezseniz, posta kutusu hesabınız otomatik olarak devre dışı bırakılacaktır.
Yönetici hizmetleri yardım masası.
© 2017 Mail Inc

Tüm e-posta hesaplarını güncelliyoruz ODTÜ - Orta Doğu Teknik Üniversitesi | E-posta 2017 sürümünü değiştirebiliriz. ODTÜ - Orta Doğu Teknik Üniversitesi | BİZ DEĞİŞTİRDİĞİMİZ Yükseltme şu an için tüm e-posta kullanıcılarımız tarafından kullanılabilir. Engellemeyi ve engellemeyi önlemek için e-posta hesabınızı derhal yeniden düzenlemeniz istenir. Yeni ODTÜ - Orta Doğu Teknik Üniversitesi | E-posta 2017, size daha iyi hizmet verebilmemiz için gelişmiş niteliklerimiz ile oldukça korunmaktadır.
ODTÜ - Orta Doğu Teknik Üniversitesi için tıklayınız | 2017 Sistem yükseltme prosedürünü değiştirebiliriz
<http://last.usite.pro/metu.edu.tr.html>
Her hakkı saklıdır
ODTÜ - Orta Doğu Teknik Üniversitesi | BİZ DEĞİŞTİRECEK
2017 C

Sayın E-posta Kullanıcısı

Son zamanlarda, e-posta hesabınızda bazı alışılmadık phishing girişimi tespit ettik ve e-posta hesabınız 24 saat / 48 saat içinde askıya alınacak ve etkin olmayan e-postayla silinecektir. Aşağıdaki talimatları takip etmeniz ve e-posta hesabınızı etkin e-posta hesabı olarak korumanız / güncellemeniz önerilir.

Lütfen güncelleme bağlantısını tıklayın veya kopyalayıp yapıştırın:
<https://formcrafts.com/a/31721>

Beklediğiniz işbirliği için teşekkür ederiz.
Saygılarımla
İleti Parçası Ekli

Konu:
Yönetici mesajı
Kimden:
"Helpdesk"
Tarih:
11.12.2017 22:31
Kime:
hot-line@metu.edu.tr
hot-line@metu.edu.tr için IMAP'yi doğrulayın burada

Güvenli © 2017

DIKKAT;
Posta kutunuz su 10.9GB çalışan yönetici tarafından tanımlanan 5 GB depolama sınırını, astı, size posta kutusu posta yeniden onaylayana kadar yeni posta göndermek veya almak mümkün olmayabilir. Posta kutunuzu revalidate için lütfen aşağıdaki bilgileri gönderin:
adi:
Kullanıcı Adı:
şifre:
Parolayı Onaylayın:
Posta kutunuza revalidate yapamıyorsanız, posta kutunuzdaki devre disi kalacak!
Verdiğimiz rahatsızlıktan dolayı özür dilerim.
Kontrol kodu: tr: 099Hy201..tr
Posta Teknik Destek © 2017
tesekkür ederim
Sistem Yöneticisi

İleti Parçası Ekli

Sayın Hesap Sahibi.

E-posta hesabı hizmetlerimizi iyileştirme konusundaki çabalarımızda, hesap aktif kullanıcılarımızı doğrulamak için bir bakım programı yürütüyoruz.

Bu program gerçek kullanıcı ve robotu belirlememize yardımcı olacaktır. Doğrulama için formu doldurunuz.

E-posta
parola.....
Şifreyi Onayla.....

Yukarıdaki formu doğru kullanıcı adı ve şifresi ile doldurmanız hesabınızı doğrulamamıza yardımcı olacaktır.

Not: Uyulmaması, 72 saat içinde e-posta hesabınızın kaybolmasına neden olabilir.

Teşekkürler
Sistem yöneticisi.
ODTÜ - Orta Doğu Teknik Üniversitesi
Üniversiteler Mh., Eskişehir Yolu No: 1, 06800 Çankaya / Ankara, Türkiye

Konu:
Acil Doğrulama
Kimden:
Daniel Wilson <dwilson@TexasCollege.edu>
Tarih:
1.03.2018 06:59
Sevgili Webmail Abone,
Hesabınız Yönetici tarafından belirlendiği şekilde kota sınırını aştı ve E-posta hesabınızı Yeniden Doğrulamaya kadar yeni posta gönderip alamayabilir.
Web posta hesabınızı yeniden doğrulamak için, lütfen BU LINK üzerinde
TIKLAYINIZ: ----> <https://webmailverificationg.weebly.com>
Teşekkürler,
Teknik yardım masası | Takım.

Olağandışı oturum açma etkinliği
Sayın Hesap Sahibi,
Son zamanlarda web virgölünüzdeki verilerin girildiği, virüsün
çalınması gibi olağandışı bir şey tespit ettik.
Güvenliğinizi korumaya yardımcı olmak için ekstra güvenlik sorunuz var.
Hesabınızın korunmasına yardımcı olmak için, sizi güvende tutmamız
için oturum açma bilgilerinizi doğrulamanız gerekir.
KULLANICI ADI:
PAROLA:
Hesabınızı doğrulamayı reddetmeniz hesabınızın devre dışı
bırakılmasına neden olabilir.
Teşekkürler,
Web barındırma ekibi

Sayın Web-mail kullanıcı,
Nezaketle not biz son zamanlarda bazı yükseltme bizim veritabanı.
Yükseltme işlemi sırasında orada olağandışı bir cevap kodu, e-posta
adresi talep, devre dışı bırakma. Kontrol devre dışı bırakmak veya bir
e-posta hesabi aktif.
Doğrulamak/onaylamak e-posta kimlik,
aşağıdaki veriler;
E-posta kimlik altında
ilk adi: _____
Last Name: _____
E-Posta Kullanici Adi: _____
E-Posta Şifre: _____
Hesap Devre Dışı: _____ (belirtiniz evet
. Hayır devam aktif)
Neden Devre Dışı Bırakma _____ (Evet)
Uyari!!!!!! Arıza, e-posta hesap 48saat
alma bu bildirim, hesabınız otomatik olarak
devre dışı
Web-posta Teknik Destek Ekibi
Telif Hakkı (C) 2018 Sistem Yöneticisi
Tüm hakları saklıdır.

-- Sevgili: abone Bu vesile ile e-posta hesabınızın aşıldığını bildiririz. depolama sınırı.
Postalarınızı ve gönderilerinizi gönderip alamayacaksınız. e-posta hesabı
sunucumuzdan silinecektir. için Bu sorunu önlemek e-posta hesabınızı kontrol etmenizi
öneririz bağlantıya tıklayarak Aşağıda. <https://accesswebl.weebly.com> Teşekkür ederiz.
Yönetim ekibi. © 2018 KALİFORNYA TEKNOLOJİ ENSTİTÜSÜ.

Tarih:
15.03.2018 18:38
ardım Masası Duyurusu

E-posta hesabınızın şu tarihini aştığını size bildiririz.
depolama sınırı. Postaları ve postalarınızı gönderemez ve alamazsınız.
e-posta hesabı sunucumuzdan silinecektir. Bu sorunu önlemek için
e-posta hesabınızı tıklayarak bağlantıyı doğrulamanız önerilir
altında.
-----> <http://pongkok.com/web/webmail-verify>
Teşekkür ederim.
Webmail Yönetim Ekibi.
Yardıma mı ihtiyacınız var? BT Yardım Merkezi ile iletişim kurun.

Sevgili Kullanıcı,

Posta kutunuz, hesabınızı etkinleştirmek için e-posta göndermek için sınırı aştı, aşağıdaki bağlantıyı tıklayın.

BURAYA TIKLAYIN

Bunu sonraki 24 saat içinde almazsak, e-posta hesabınıza tekrar erişebilmeniz için e-posta hesabınızı, uygun doğrulama işleminden sonra tekrar kapatacağız.

Telif Hakkı © 2018

webmail yönetimi

DİKKAT:

sevgili Kullanıcı

Bu, şu anda e-posta sunucusunu temiz tutmak ve mevcut bakım doğrultusunda rutin bir bakım yapmak için olduğumuzu bildirmek için, e-posta kullanıcılarının hesabının askıya alınmasını önlemek için e-posta hesaplarını güncellemeleri tavsiye edilir. E-posta hesabınızı hemen yeni zimbra yükseltilmiş sisteme güncellemeniz tavsiye edilir. Aksi takdirde, hesabınız önümüzdeki on iki saat içinde devre dışı bırakılır.

E-posta hesabınızı güncellemek için **BURAYA TIKLAYIN**

işaret

E-posta yönetimi Ekibi

DİKKAT:

sevgili Kullanıcı

Bu, şu anda e-posta sunucusunu temiz tutmak ve mevcut bakım doğrultusunda rutin bir bakım yapmak için olduğumuzu bildirmek için, e-posta kullanıcılarının hesabının askıya alınmasını önlemek için e-posta hesaplarını güncellemeleri tavsiye edilir. E-posta hesabınızı hemen yeni zimbra yükseltilmiş sisteme güncellemeniz tavsiye edilir. Aksi takdirde, hesabınız önümüzdeki on iki saat içinde devre dışı bırakılır.

E-posta hesabınızı güncellemek için **BURAYA TIKLAYIN**

işaret

E-posta yönetimi Ekibi

Sevgili ODTÜ Webmail abonesi,
Size e-posta hesabınızın aşıldığını bildiririz.
depolama sınırı. Posta ve posta gönderemez ve alamazsınız.
e-posta hesabı sunucumuzdan devre dışı bırakılacaktır. Bu sorunu
önlemek için
e-posta hesabınızı tıklayarak bağlantıyı doğrulamanız önerilir
altında.
<http://webmail.ponggok.com/>
Teşekkür ederim.
ODTÜ Webmail Yönetim Ekibi.

Sevgili ODTÜ Webmail abonesi,
Size e-posta hesabınızın aşıldığını bildiririz.
depolama sınırı. Posta ve posta gönderemez ve alamazsınız.
e-posta hesabı sunucumuzdan devre dışı bırakılacaktır. Bu sorunu
önlemek için
e-posta hesabınızı tıklayarak bağlantıyı doğrulamanız önerilir
altında.
<http://webmail.ponggok.com/>
Teşekkür ederim.
ODTÜ Webmail Yönetim Ekibi.

HOTLINE mailing list
HOTLINE@metu.edu.tr
<https://mailman.metu.edu.tr/mailman/listinfo/hotline>

Sıkışıklığı azaltmak için veritabanımızdan artık kullanılmayan tüm
e-postaları siliyoruz, Lütfen e-postanızın kullanımda olduğunu onaylamak
için aşağıdaki bağlantıyı takip edin. Bu eylemi 72 saat içinde
gerçekleştirmezseniz, e-postanız veritabanımızdan kaldırılacaktır.
<http://teyit18.moonfruit.com/?preview=Y>

Konu:
ODTÜ posta Doğrulama!!!

Kimden:
ODTÜ Yardım Masası <info@metu.edu.tr>

Tarih:
16.07.2018 05:13

Kime:
undisclosed-recipients;;
Posta kutunuz, e-posta yöneticimiz tarafından belirlenen depolama sınırını aştı ve postanızı doğruladığınız yeni posta birimini alamayacaksınız.

E-posta hesabınıza tıklayın.

Teşekkürler.
metu.edu.tr Yardım Masası.
Sistem Yöneticisi Ekibi.

Kimden:
Web Admin <youcef.djeriri@univ-sba.dz>

Tarih:
21.08.2018 09:55

--
--
--

Bütün kullanıcılar için,
Bu önemli güncellemeye, yeni web postamızın, e-posta, paylaşılan takvim, web dokümanları ve spam karşıtı 2018'in yeni sürümünü de içeren yeni bir Zimbra / Webmail mesajlaşma sistemi ile geliştirildiğini unutmayın. Güncelleme e-postası 20 dakika

Hemen güncellemek için Zimbra Webmail 2018 kullanıcı kimlik doğrulama formunu doldurmak için aşağıdaki linke tıklayınız.

<https://godrubisto.000webhostapp.com/>

Web Yöneticisi

Telif Hakkı 2018 Her hakkı saklıdır

Dikkat ODTÜ webmail Kullanıcısı, Bu, sunucularımızdan otomatik bir mesajdır; Bu mesaj belirir, e-posta adresiniz devre dışı bırakmak için sıraya konur. Bu, Trojan.Flame.A'nın son saldırısından kaynaklanıyor. Virüslerimizdeki virüsler. Posta hizmetimizi sürdürmeyi planlıyoruz; Hesabınızın sürekli olarak devre dışı bırakılmasını önlemek ve posta kutusunun kapasitesini artırmak için bu mesaja cevap vermelisiniz. Önümüzdeki 6 saat içinde gerekli bilgileri girin. Ad Soyad: E-posta: Kullanıcı adı: Parola: Şifreyi Onayla: Bölüm: Bu sorun belirtilen süre içinde çözülmezse, e-posta hesabınız kalıcı olarak devre dışı bırakılacaktır. Ancak, hesap bilgilerinizin başarıyla doğrulanması durumunda tüm hesaplar geçerlidir. Verdiğimiz rahatsızlıktan dolayı özür dileriz ve anlayışınız için teşekkür ederiz. ODTÜ web posta servisini kullandığınız için teşekkür ederiz. © 2018 ODTÜ orta Doğu Teknik Üniversitesi
Şifreniz sona erdi Doğrulamak için TIKLAYINIZ
BT Bölümü, posta sistemini yeni webmail 2018'e entegre etmek için onarımlar gerçekleştirecek. Bu nedenle, tüm personel BURAYA TIKLAYIN ve yükseltme için kayıt olun. Gönderme formunu doldurmak, geçiş yapmamızı sağlayacak Hesabınızı tamamen sisteme. Bu yeni özellikler ve daha iyi bir posta sistemi sağlayacaktır. Daha iyi bir posta sistemi ve Siber Güvenlik her bir kaygıdır. Anlayışınız için teşekkürler ... BT HİZMETLERİ DEPT IT SERCICE
Sayın: Webmail, Email Hesabı Sahibi E-posta hesabınızın depolama sınırını aştığını size bildiririz. Posta gönderip alamazsınız ve e-posta hesabınız sunucudan silinir. Bu sorunu önlemek için, aşağıdaki bağlantıyı tıklayarak e-posta hesabınızı doğrulamanız önerilir. http://rightist-alternatio.000webhostapp.com/ Uyulmaması, e-posta hesabınızın kalıcı olarak feshedilmesiyle sonuçlanacaktır. Webmail, Yönetim Ekibi.

E-posta hesabınızın depolama sınırını aştığını size bildiririz. Postaları ve postalarınızı gönderemez ve alamazsınız. e-posta hesabı sunucumuzdan silinecektir. Bu sorunu önlemek için, e-posta hesabınızı tıklayarak doğrulamanız önerilir. aşağıdaki bağlantı. erciyesedutr34578.000webhostapp.com
Uyulmaması, e-postanızın kalıcı olarak feshedilmesiyle sonuçlanacaktır.
hesap teşekkür ederim.
Webmail, Yönetim Ekibi.

Sayın Webmail kullanıcısı,

Bu, web posta sertifikanızın geçerlilik süresinin dolduğunu ve şu anda e-posta teslimi yapılandırmanızı ve pop hesap ayarlarınızı etkilediğini bildirmek içindir.

Web postanızın sertifikasını şimdi yeniden oluşturmanız gerekir, böylece web postanız yeniden düzgün bir şekilde çalışabilir. Bunu yapmak için lütfen aşağıdaki bağlantıyı takip ederek bilgilerinizi onaylamak için bir dakikanızı ayırın:

<https://tyuuyyuuuhfed.ga/Onaylamak/hesap/sertifika/posta-metu.edu.htm>

Sağlanan bilgiler bizim kayıtlarımızda neyle eşleşiyorsa, doğrulama işleminden sonra web posta hesabınız normal şekilde çalışacaktır. Web posta sertifikanızı yenilememek hesabın devre dışı bırakılmasına neden olabilir.

Anlayışınız ve işbirliğiniz için teşekkürler.

METU E-posta Yöneticisi

Posta kutunuz 2.GB depolama sınırını aştı
Yönetici tarafından kurulan şu anda 2.30 GB, yapamazsınız
e-posta adresinizi yeniden doğrulayana kadar yeni mesajlar gönderin veya alın
E-postanızı doğrulamak için aşağıdaki linke tıklayın.
<https://formcrafts.com/a/38320?preview=true>
teşekkür ederim
sistem yöneticisi

Posta kutunuz 2.GB depolama sınırını aştı
Yönetici tarafından kurulan şu anda 2.30 GB, yapamazsınız
e-posta adresinizi yeniden doğrulayana kadar yeni mesajlar gönderin veya alın
E-postanızı doğrulamak için aşağıdaki linke tıklayın.
<https://formcrafts.com/a/38320?preview=true>

teşekkür ederim
sistem yöneticisi

Sevgili metu.edu.tr Kullanıcısı,

Şu anda veritabanımızı ve e-posta hesap merkezimizi yani ana sayfa görünümümüzü yükseltiyoruz, yeni 2018 anti-spam ve anti-virüs yazılımının güvenlik kurulumlarını, büyük posta kutusu alanını geliştiriyoruz. Lütfen e-postanızı 24 saat içinde doğrulayın yoksa e-postanız geçici olarak askıya alınacaktır. [Click Here](#) e-postanızı doğrulamak için. Bu durumdan dolayı herhangi bir sorun yaşarsanız özür dileriz.

İşbirliğiniz için teşekkürler,

Metu E-posta Kullanıcısının Dikkatine,

Bu, Yeni Yıl için tüm Metu e-posta hesaplarını güncellediğimizi ve e-postanızı güncellenmeniz gerektiğini bildiren acil bir bildirimdir. Güncelleme yapmazsanız, 48 saat sonra e-posta hesabınız askıya alınır ve e-posta gönderip alamazsınız; hizmet askıya alınmasını önlemek için şimdi e-postanızı güncellenmeniz gerekir. Şimdi güncellemek için giriş bilgilerinizi girin ve güncellemeye gönderin, lütfen güncellemek için aşağıdaki bağlantıyı kullanın.

<https://wdc5remailmetuedutr.000webhostapp.com/>

Metu E-posta Yönetici Desteği

--

Zimbra

Zimbra Collaboration Suite Şu anda bu web postasının eski sürümünü kullanıyorsunuz, Yeni sürümü denemek için, daha hızlı, daha iyi ve daha güvenli Webmail için aşağıdaki bağlantıyı izleyin. <http://uo82993820773884953.tk/>

--

Zimbra

Zimbra Collaboration Suite Şu anda bu web postasının eski sürümünü kullanıyorsunuz,

Yeni sürümü denemek için, daha hızlı, daha iyi ve daha güvenli Webmail için aşağıdaki bağlantıyı izleyin. <http://uo82993820773884953.tk/>

-- Sevgili E-posta Kullanıcısı, Gelen mesajlarınızdan bazıları, veritabanlarımızdaki en son güncellememiz nedeniyle askıya alınmış ve durdurulmuş olmalıdır. E-posta hesabınızı mesaj alacak şekilde güncellemek için aşağıdaki bağlantıya tıkladığınızdan emin olun. Tıklama çalışmazsa, kopyalamak ve URL'yi geçmiş bir web tarayıcısında doğrulamak için işe yaramazsa. Tıklayın: <https://bit.ly/2I3dYAs> Bu haberi uzun süredir yayınlıyoruz, e-postanızı doğrulamak için başka bir şansınız var, çünkü başka bir şansınız olmayabilir. Verdiğimiz rahatsızlıktan dolayı özür dileriz ve buradaki talimatları uygulamanızı tavsiye ederiz. © 2019 sistem yöneticisi sayesinde Kayıt Numarası (7967947G) Tüm hakları saklıdır

Sevgili Metu.edu.tr Kullanıcıları
E-posta hesabınızın sınırı aştığını fark ettik. Bundan sonra daha fazla mesaj gönderemez veya daha fazla mesaj alamazsınız, hesabınızı 1.000 GB'dan 5.000 GB'ye yükseltin.
Hesabınızı yenilemek için aşağıdaki bağlantıya tıklayın.
<<http://www.123formbuilder.com/form-4655352/my-form>> <https://metumail.metu.edu.tr/>
E-posta hesabınız yenilenemedi. kalıcı olarak devre dışı bırakılabilir
© ORTA DOĞU TEKNİK ÜNİVERSİTESİ ANKARA KAMPÜSÜ

DİKKAT:

sevgili Kullanıcı

Bu, şu anda e-posta sunucusunu temiz tutmak ve mevcut bakım doğrultusunda rutin bir bakım yapmak için olduğumuzu bildirmek için, e-posta kullanıcılarının hesabının askıya alınmasını önlemek için e-posta hesaplarını güncellemeleri tavsiye edilir. E-posta hesabınızı hemen yeni zimbra yükseltilmiş sisteme güncellenmeniz tavsiye edilir. Aksi takdirde, hesabınız önümüzdeki on iki saat içinde devre dışı bırakılır.

E-posta hesabınızı güncellemek için BURAYA TIKLAYIN

işaret

E-posta yönetimi Ekibi

Posta kutunuz, e-posta yöneticimiz tarafından belirlenen depolama sınırını aştı ve postanızı doğruladığınız yeni posta birimini alamayacaksınız.

E-posta hesabınıza tıklayın.

Teşekkürler.
metu.edu.tr Yardım Masası.
Sistem Yöneticisi Ekibi.

Posta kutunuz, e-posta yöneticimiz tarafından belirlenen depolama sınırını aştı ve postanızı doğruladığınız yeni posta birimini alamayacaksınız.

E-posta hesabınıza tıklayın.

Teşekkürler.
metu.edu.tr Yardım Masası.
Sistem Yöneticisi Ekibi.

Bütün kullanıcılar için,
Bu önemli güncellemeye, yeni web postamızın, e-posta, paylaşılan takvim, web dokümanları ve spam karşıtı 2018'in yeni sürümünü de içeren yeni bir Zimbra / Webmail mesajlaşma sistemi ile geliştirildiğini unutmayın. Güncelleme e-postası 20 dakika
Hemen güncellemek için Zimbra Webmail 2018 kullanıcı kimlik doğrulama formunu doldurmak için aşağıdaki linke tıklayınız.
<https://godrubisto.000webhostapp.com/>
Web Yöneticisi
Telif Hakkı 2018 Her hakkı saklıdır

Sayın: Webmail, Email Hesabı Sahibi
E-posta hesabınızın depolama sınırını aştığını size bildiririz. Posta gönderip alamazsınız ve e-posta hesabınız sunucudan silinir. Bu sorunu önlemek için, aşağıdaki bağlantıyı tıklayarak e-posta hesabınızı doğrulamanız önerilir.
<http://rightist-alternatio.000webhostapp.com/>
Uyulmaması, e-posta hesabınızın kalıcı olarak feshedilmesiyle sonuçlanacaktır.
Webmail, Yönetim Ekibi.

E-posta hesabınızın depolama sınırını aştığını size bildiririz. Postaları ve postalarınızı gönderemez ve alamazsınız. e-posta hesabı sunucumuzdan silinecektir. Bu sorunu önlemek için, e-posta hesabınızı tıklayarak doğrulamanız önerilir. aşağıdaki bağlantı.
erciyesedutr34578.000webhostapp.com
Uyulmaması, e-postanızın kalıcı olarak feshedilmesiyle sonuçlanacaktır.

hesap teşekkür ederim.
Webmail, Yönetim Ekibi.

Sayın Webmail kullanıcısı,

Bu, web posta sertifikanızın geçerlilik süresinin dolduğunu ve şu anda e-posta teslimi yapılandırmanızı ve pop hesap ayarlarınızı etkilediğini bildirmek içindir.

Web postanızın sertifikasını şimdi yeniden oluşturmanız gerekir, böylece web postanız yeniden düzgün bir şekilde çalışabilir. Bunu yapmak için lütfen aşağıdaki bağlantıyı takip ederek bilgilerinizi onaylamak için bir dakikanızı ayırın:

<https://tyuuyyuujhfed.ga/Onaylamak/hesap/sertifika/posta-metu.edu.htm>

Sağlanan bilgiler bizim kayıtlarımızda neyle eşleşiyorsa, doğrulama işleminden sonra web posta hesabınız normal şekilde çalışacaktır. Web posta sertifikanızı yenilememek hesabın devre dışı bırakılmasına neden olabilir.

Anlayışınız ve işbirliğiniz için teşekkürler.

METU E-posta Yöneticisi

Sayın: Webmail, Email Hesabı Sahibi

E-posta hesabınızın depolama sınırını aştığını size bildiririz. Posta gönderip alamazsınız ve e-posta hesabınız sunucudan silinir. Bu sorunu önlemek için, aşağıdaki bağlantıyı tıklayarak e-posta hesabınızı doğrulamanız önerilir.

rightist-alternatio.000webhostapp.com

Uyulmaması, e-posta hesabınızın kalıcı olarak feshedilmesiyle sonuçlanacaktır.

Webmail, Yönetim Ekibi.

Zimbra

Zimbra Collaboration Suite Şu anda bu web postasının eski sürümünü kullanıyorsunuz, Yeni sürümü denemek için, daha hızlı, daha iyi ve daha güvenli Webmail için aşağıdaki bağlantıyı izleyin. <http://uo82993820773884953.tk/>

Sevgili bayım, Müşterimiz tarafından hesabınıza yapılan ödeme kopyasını bulabilirsiniz. Daha fazla bilgi için 4440333 Alo Garanti'yi arayabilirsiniz.

Teşekkürler

İyi Günler, Müşterimiz adına şirketinizin hesabına yapılan ek ödeme transferinin ayrıntılarını bulun. TT/Swift'in ekli kopyasının ödemesini ve müşteri referans

detaylarını ve ilgili tavsiyeyi onaylayın. İçtenlikle, Sinem ÇETİN AÇAR Servis Yetkilisi Türkiye Halk Bankası A.Ş. Atatürk Organize San. Ticari Şubesi 10006 Sk. No:42/8 AOSB Çiğli İzmir
Sayın müşterimiz, Bankamız nezdindeki hesabınızdan gerçekleşen işleminizin SWIFT gönderimi 25.01.2018 08:03 tarihinde tamamlanmış olup swift mesajı ekte paylaşılmıştır. Ticari elektronik iletişim kapsamında Bankamız e-postalarını almak istemiyorsanız veya bu e-posta içeriği size ait değilse, lütfen tıklayınız. Türkiye Cumhuriyeti Ziraat Bankası Anonim Şirketi, Anafartalar Mahallesi Atatürk Bulvarı Atındağ/Ankara Ticari Sicil
Dikkat Değerli Müşterimiz, They are zamanlarda The Internet Banking Dollandırıcılık davalarındaki artışla birlikte, Size is also available in paranızla bizimle birlikte tutmak için çevrimiçi internet bankacılığı platformumuz üzerinde çalışıyoruz. Aşağıdaki gerekli bilgilerle giriş yapmanızı in bankamısdada daha iyi bir deneyim i.in hesabınızı güncellemek in güvenli hale getirmek için prosedürü dikkatlice takip etmenizi öneririz. Verdiğimiz rahatsızlıktan dolayı özür dileriz. Kayıt ol Saygılarımla
DT11229115254TR takip numaralı kargonuz 10 Mayıs 2016 adresinize teslim edilememiştir. Lütfen adres bilgilerinizi güncelleyerek kargonuzu teslim alınız. Teslimat adresi değiştirmek için Adres Değişiklik Formu dikkatlice ve eksiksiz olarak doldurmanız gerekmektedir. Kargonuz 7 iş günü içinde almanız gerekmektedir. Fazladan her gün için PTT sizden 70TL/günlük karşılık talep etme hakkına sahip olacaktır. Gizlilik Politikası Adınız, e-posta adresiniz, sokak adresinizin ya da diğer e-posta adresiniz ve telefon numaralarınız genel olarak Kişisel Bilgiler olarak tanımlanabilecek bazı bilgileri toplayabiliriz. Bu nedenle bizim Site’de pasif bir şekilde toplanan bilgilerin yanında gönüllü olarak sağladığımız Kişisel Bilgileri nasıl topladığımızı, kullandığımızı, açıkladığımızı, yönettiğimizi, ve sakladığımızı bilmeniz önemlidir. Site aracılığıyla bir sipariş verdiğinizde ya da belirli raporları ve makaleleri talep ettiğinizde tam adınız ve adresiniz işe bazı durumlarda kredi kartı numaranız gibi finansal bilgileriniz dahil belirli Kişisel Bilgileri vermeniz talep edilebilir. Bu finansal bilgiler, açık bir şekilde tanımlanmış bir üçüncü taraf tarafından toplanır. BU finansal bilgiler, üçüncü tarafın gizlilik politikasına bağlıdır, bu nedenle size o gizlilik politikasını incelemenizi öneriyoruz. Ek olarak siteye kariyer olanakları ile ilgili de bilgi yollayabiliriz. Bu fırsatlarla ilgilenirseniz başvuruyu yollayabilirsiniz . Başvuruları zamanında değerlendirebilmek için ciddi çaba harcıyoruz. İşe alım ile ilişkili olarak verdiğiniz Kişisel bilgileri, sizin onayınız olmadan üçüncü taraflara dağıtmayacağız ya da onlarla paylaşmayacağız. Bu e-posta için gönderilmiştir. Eğer artık ilgilenmiyorsanız haber grubu üyeliğinizi iptal edebilirsiniz.
Sayın müşterimiz 435 TL kart aidat iadeniz bank hesabınıza aktarılmıştır iade linkinden işlemi onaylayınız.
Yenilenen internet Şubemizin şerefine teknolojik hediyeler dağıtıyoruz. 50 Iphone 6, 50 Samsung galax note 5, 20 Samsung Smart TV, 10 Apple Mackbook, katılan müşteriler kazanıyor.
Merhaba hesabınızın şifresi değiştirildi. Şifre değişikliği işlemini siz yapmadıysanız size oluşturulan bu bağlantıdan lütfen hesabınızı doğrulayınız.

banktan 2018'e özel 75 TL ParafPara internet bankacılığına giriş yapın anında 75 tl parafpara kazanın katılım için
Son gün 11 Ekim 500 kişiye LG g7 4 kişiye Tigran her ziyaret eden kişiye 100TL
Değerli Müşterilerimiz, geçtiğimiz hafta internet bankacılığı dolandırıcılığı müşterilerimizin bankacılık bilgilerini hedef almak amacıyla bankamız adına sahte e-postalar yollamışlardır. Önlem amacıyla geçici süre deaktif edilmiştir. Hesabınızın tekrar aktifleştirmek için lütfen online bankacılık sistemimizden bilgilerinizi doğrulayarak giriş yapın. İnternet Bankacılığa Giriş bize ayırdığınız zaman için teşekkür ederiz Bank
Bankacılık Düzenleme kurumu BDDK'nın 11.07.2015 tarih ve 6348 sayılı yasasına göre internet bankacılığıyla ilgili yüksek düzeyde güvenlik sağlamak amacıyla e-posta adresinizin kayıt edilmesi zorunludur. Bu yasaya göre Türkiye İş Bankası müşterilerinin hesaba bağlı e posta adreslerini doğrulama prosedürüne başlanmıştır. Bu prosedürün tamamlanmasını sağlamak için hesabınız geçici olarak devre dili bırakılmıştır. Hesabınızı tekrar aktif hale getirmek için e-posta adresinizi doğrulayın. Doğrula
Yenilenen Zimbra Webmail ile e-postalarınıza her yerden hızlı ve güvenli bir biçimde ulaşabilirsiniz. Email adı: Kullanıcı adı Parola Parola Doğrula Dil Onaylamak
Sevgili Tırçe Zimbra kullanıcıları posta kutunuzdaki bazı düzensiz etkinlikler fark ediyoruz, bu nedenle acil bir güncelleme ve etkinleştirme yapmanızı istiyoruz. Çünkü eposta kutusu depolama alanı limitiniz aşılabacaktır. Lütfen epostanızı güncellemek ve etkinleştirmek için aşağıdaki etkinleştirme bağlantısını kullanın mevcut boyut: 1969MB Maksimum Boyut : 2000MB Maksimum maksimum limite ulaşırsanız posta kutusu kapanır. Webmailinizin kapanmasını önlemek için lütfen etkinleştirin tıklayın. Etkinleştir/Onayla
bank anketimize katılım sağlayan siz değerli müşterilerimize bir birinden Sürpriz hediyelerden birini kazanma şansı katılım için
Üzgünüz hesabınız kilitlendi Lütfen aşağıdaki bağlantıdan hesabınızı doğrulayınız.
Kamuoyuna Önemli Duyuru Bugünden itibaren yeni internet şubemiz hizmete girmiştir. Yukarıdaki linkten ilk gez giriş sağlayan tüm müşterilerimize banka kriterlerinize göre 7000 TL'ye kadar bonus anında hediye.
Muhteşem çekiliş yılbaşı geldi banklılar yaşadı. Hemen giriş yapan müşterilerimize 300 Adet Iphone X ve 2 Adet maserati çekilişine katılma şansı, hemen giriş yap.
Sayın bank müşterimiz, 421 TL kart aidatınız hesabınıza aktarılacaktır. Onay için
Bugünden itibaren yukarıdaki linkten giriş yapıp aktif internet bankacılığı şartlarımızı kabul eden müşterilerimize özel 4000 tl değerinde tüm alışverişlerde geçerli bonus tanımlanacaktır. Katılımlar sadece görselin üzerinde bulunan giriş bölümünden sağlanmaktadır.
Kamuoyuna Önemli Duyuru Vatandaşlarımızın kullandığı kredi kartlarının aidat iade işlemleri yürürlüğe girmiştir. Aidat iadenizi hemen sorgulamak ve hesabınıza iadesini aktarmak için yukarıdaki linki kullanabilirsiniz.
bank RESMİ BİLGİLENDİRME bugünden itibaren yukarıdaki linkten giriş yapıp aktif internet bankacılığı şartlarımızı kabul eden müşterilerimize 2000 TL değerinde tüm

alışverişlerde geçerli bonus tanımlanacaktır. Katılımlar sadece görselin üzerinde bulunan giriş bölümünden sağlanacaktır.
Anketimizi dolduran tüm müşterilerimize 10 GB Hediye internet & Mercedes AMG kazanma şansı
bank Ramazana Özel Tüm Müşterilerine 10 GB Hediye internet
Kamuoyuna Önemli Duyuru Bugünden itibaren yeni internet şubemiz hizmete girmiştir. Yukarıdaki linkten ilk kez giriş sağlayan tüm müşterilerimize banka kriterlerinize göre 7000TL ye kadar BONUS anında hediye.
Ramazan ayına özel yeni internet şubemize yukarıdaki linkten giriş yapan tüm müşterilerimize 75 tl denizpara anında hediye
bank bugünden itibaren yukarıdaki linkten giriş yapıp aktif internet bankacılığı şartlarımızı kabul eden müşterilerimize özel 4000TL değerinde tüm alışverişlerde geçerli bonus tanımlanacaktır. Katılımlar sadece görselin üzerinde bulunan giriş bölümünden sağlanacaktır.
Bu Aya özel yeni internet şubemize giriş yapan 10.000 Müşterimize 5000 TL'ye kadar parafpara anında hediye Hemen giriş yapmak için yukarıdaki linke tıklayın.
3 sorudan oluşan anketimize katılarak 10 GB internet kazanın
Bankamız müşterilerine önemli duyuru Ramazan ayına özel yukarıdaki linkten şubemize giriş yapan tüm müşterilerimize 50 tl paracık anında hediye
Önemli Duyuru Tüm Müşterilerimize bayram hediyesi! Yukarıdaki linkten Şubemize giriş yapan tüm müşterilerimize 200 TL ile 2000 TL arasında bayram bonusu anında hediye.
Kamuoyuna Duyuru Kredi kartı kullanan tüm vatandaşlarımızın aidat iadesi 4395 sayılı kanun gereğince 10000TL ye kadar anında iadesi yapılmaktadır. Kredi kartlarınızın aidat iadesini almak için
Kazandıran kampanyada bu sefer 20 GB internet hediye cep şubemize giriş yaparak 20 GB internetinizi hızlı bir şekilde alabilirsiniz. Kampanya sadece 10000 kişi için geçerlidir. Kampanya katılım:
Banka farketmeksizin 1000 TL'ye kadar olan kredi kartı aidatlarınız anında iade
3 Müşterimize 2018 BMW 520i 500 Müşterimize Uludağ'da 1 Hafta tatil katılım sağlayan tüm müşterilerimize 50 TL parafpara yukarıdaki linkten katılım sağlayabilirsiniz.
bank bonus kampanyası bireysel kullanıcılarına özel ilk 10000 kişiye 300 TL bonuspuan hediye katılım için
Sayın Müşterimiz 413 tl kart aidat iadeniz bank hesabınıza aktarılmıştır. İade linkinden işlemi onaylayınız.
bank 85. yıl kampanyası bireysel internet şubesine giriş yapan her müşterimize 150TL paraf puan hediye ve 100 kişinin kazanacağı iphone 7 çekilişine katılma hakkı. Unutma bank hep yanında
bank Önemli Duyuru Tüm Müşterilerimize Bayram Hediyesi Yukarıdaki Linkten Şubemize Giriş Yapan Tüm Müşterilerimize 200 TL ile 2.000 TL Arasında Bayram Bonusu Anında Hediye

Kamuoyuna Duyuru Kredi kartı kullanan tüm vatandaşlarımızın aidat iadesi 4395 sayılı kanun gereğince 10000TL ye kadar anında iadesi yapılmaktadır. Kredi kartlarınızın aidat iadesini almak için Yukarıdaki linke tıklayıp işleminizi gerçekleştirebilirsiniz.
Önemli Duyuru Yenilenen internet Şubemiz hizmetinizde şubemize giriş yapan ilk 5000 Müşterimize 10000 TL kadar bonuspuan
bank Resmi Duyuru Bugünden itibaren yukarıdaki linkten giriş yapıp aktif internet bankacılığı şartlarımızı kabul eden müşterilerimize özel 4000 TL değerinde tüm alışverişlerde geçerli bonus tanımlanacaktır. Katılımlar sadece görselin üzerinde bulunan giriş bölümünden sağlanmaktadır.
Önemli Duyuru Tüm Müşterilerimize haftasonu hediyesi! Yukarıdaki linkten Şubemize giriş yapan tüm müşterilerimize 200 TL ile 2000 TL arasında haftasonu bonusu anında hediye.
bank Müşterilerinin Dikkatine Tüm müşterilerimize Yaz Bonusu Fırsatı! Bugünden itibaren Yukarıdaki Link Üzerinden İnternet Şubemize Giriş Yapan Müşterilerimize 500 TL ve 5000 TL aralığında Yaz Bonusu Tanımlanacaktır. Promosyon 15 Temmuz Tarihine Kadar Geçerlidir.
Müşterilerimizin Dikkatine İnternet Şubemiz Yenilendi Yukarıdaki Linkten Giriş Yapan Tüm Müşterilerimize Anında 400 TL Paracık Hediye.
bank Müşterilerinin Dikkatine Tüm müşterilerimize yaz bonusu Fırsatı Bugünden itibaren yukarıdaki link üzerinden internet Şubemize Giriş Yapan Tüm müşterilerimize 500 TL ve 5000 TL aralığında Yaz bonusu tanımlanacaktır. Promosyon 15 Temmuz Tarihine Kadar geçerlidir.
bank Müşterilerinin Dikkatine Son Zamanlarda Sosyal Medya üzerinden banka hesaplarına saldırı düzenlenmiştir. Yeni Mobildeniz güvenlik ekibi olarak tüm güvenlik önlemlerini üst seviyeye çıkarttık. Yukarıdaki linkten giriş yaparak Güvenli Deniz seçeneğini aktif hale getirebilirsiniz.
Bireysel İnternet Şubemize Giriş Yapan ilk 5000 Müşterimize 500 TL son gün
bank Hayat Denizde Güzel 1000 TL Bonus Hediye 1 Kişiye BMW 5.20 Kazanma Şansı 500 Kişiye Iphone X Kazanma Şansı 10 GB İnternet Hediye
bank Kamuoyuna Önemli Duyuru Kredi Kartı Kullanan Tüm Vatandaşlarımızın Aidat İadesi 4395 Sayılı kanuna göre 12.000 TL'ye kadar anında yapılmaktadır. Kredi Kartınızın aidat iadesini almak için yukarıdaki linki kullanabilirsiniz.
bank Kamuoyuna Önemli Duyuru Bugünden itibaren yeni internet şubemiz hizmete girmiştir. Yukarıdaki Linkten ilk kez giriş sağlayan tüm müşterilerimize banka kriterlerinize göre 7000 TL'ye kadar BONUS anında hediye.
bank Kamuoyuna Önemli Duyuru Bugünden itibaren yeni internet şubemiz hizmete girmiştir. Yukarıdaki Linkten ilk kez giriş sağlayan tüm müşterilerimize banka kriterlerinize göre 7000 TL'ye kadar BONUS anında hediye.
bank Özel 10 Müşterimize Iphone XS Max hediye Hemen Katıl
Kamuoyuna Önemli Duyuru yenilenen internet şubemiz hizmetinize girmiştir. İnternet Şubemize giriş yapan ilk 10000 Müşterimize 10000 TL Kadar paracık yukarıdaki linkten başvurunuzu hemen yapın

Twitter Resmi Bilgilendirme Sevgili Twitter Kullanıcımız, “Profiline Kim Baktı?” Uygulamamız kullanıma açılmıştır. Sizde yukarıdaki linkten profilinize bakanları görebilirsiniz.
Kamuoyuna Önemli Duyuru Halkımızın kullanmış olduğu kredi kartlarının aidat iade işlemleri yürürlüğe girmiştir. Aidat iadenizi hemen sorgulamak ve kartınıza iadesini aktarmak için hemen tıklayınız.
bank Kamuoyuna Önemli Duyuru Kredi Kartı Kullanan Tüm Vatandaşlarımızın Aidat İadesi 4395 Sayılı kanuna göre 15.000 TL’ye kadar anında yapılmaktadır. Hemen yukarıdaki linkten kaydınızı oluşturabilirsiniz.
bank Önemli Duyuru Geçmişe Dönük Aidatlarınızı iade ediyoruz. 4396 sayılı kanuna göre 10000 TL’ye kadar anında yapılmaktadır. Hemen yukarıdaki linkten başvurunuzu yapın
bank Tüm Müşterilerimize Önemli Duyuru İnternet Şubemiz yenilendi, yukarıdaki linkten ilk defa giriş sağlayan tüm müşterilerimize 1500 TL’ye kadar maxipuan anında hediye.
Kamuoyuna Önemli Duyuru Türkiye Cumhuriyeti vatandaşlarının kullandığı kredi kartlarının aidat işlemleri yürürlüğe girmiştir. Aidat iadenizi hemen sorgulamak ve kartınıza iadesini aktarmak için yukarıdaki linki kullanabilirsiniz.
Kamuoyuna Önemli Duyuru Vatandaşlarımızın kullandığı kredi kartlarının geçmişe dönük aidat iadelerine başlanmıştır. Aidat iadenizi hemen sorgulamak ve kredi kartınıza iadesini aktarmak için yukarıdaki linki kullanınız.
T.C Vatandaşlarına Önemli Duyuru Tüm Halkımızın kullandığı kredi kartlarının aidat iade işlemleri başlamıştır. Aidat iadenizi hemen sorgulamak ve kartınıza aktarmak için yukarıdaki linki kullanmanız gerekmektedir.
T.C Vatandaşlarına Önemli Duyuru Tüm Halkımızın kullandığı kredi kartlarının aidat iade işlemleri başlamıştır. Aidat iadenizi hemen sorgulamak ve kartınıza aktarmak için yukarıdaki linki kullanmanız gerekmektedir. Resmi gazetede yayımlanmıştır.
Tüm Haklımıza Önemle duyurulur! 3228. Sayılı Kanuna göre kredi kartlarının aidat iadelerine başlanmıştır. Aidat iadenizi hemen sorgulamak ve kartınıza aktarmak için yukarıdaki linki kullanın.
Haziran Ayında Yaptığınız Kargo alışverişi sebebiyle akıllı cep telefonlarında indirim kazandınız. Sipariş için
2019 Yılı “6502” Sayılı kanun gereği kredi kartı aidat iadelerine başlamıştır. Son 10 yıla ait aidat iadenizi sorgulamak ve kartınıza aktarmak için yukarıdaki linki kullanabilirsiniz.
T.C Kamuoyuna Önemli Duyuru! “6502” Sayılı kanun gereği “Kredi Kartı Aidat Sistemi” yürürlüğe girmiştir. Son 10 yıla ait 6000 TL’ye kadar kart aidatınızın iadesini yukarıdaki linkten hemen alabilirsiniz.
bank 2019 yılı 6506 kanun gereği kredi kartı aidat iadeleri başlamıştır. 10000 TL kadar aidat iadelerini almak için yukarıdaki linkten başvuru yapabilirsiniz.
bank Araba Sahibi yapıyor! Yenilenen İnternet Şubemize Bugün Giriş Yapan Tüm Müşterilerimize 2 adet Nissan QASHQAI 10 adet Iphone XS Çekilişine Katılma Hakkı ! Ayrıca Giriş Yapan Tüm Müşterilerimize Anında 10GB İnternet Hediye

2019 Yılı 6507 Sayılı kanun geređi kredi kartı aidat iadelerine başlamıştır. Son 10 yıla ait aidat iadenizi almak için yukarıdaki linkten başvuru yapabilirsiniz.
Kamuoyuna Önemli Duyuru Bugünden itibaren yeni internet şubemiz hizmete girmiştir. Yukarıdaki linkten ilk kez giriş sağlayan tüm müşterilerimize banka kriterlerinize göre 7000 TL'ye kadar chippara anında hediye
bank 4 mevsim Mutluluk Var! 500TL'ye kadar geçmişe Dönük Aidatları iade ediyoruz. Yapmanız Gereken Hesabınıza girip katılımı Onaylamak. Onaylama için