Faculty of Computer Science and Information Technology

IMPLEMENTATION AND PERFORMANCE ANALYSIS OF MACHINE LEARNING MODELS FOR DETECTING PHISHING ATTACKS ON WEBSITES

LIONG KAH PONG

Bachelor of Computer Science with Honours

(Software Engineering)

2023

IMPLEMENTATION AND PERFORMANCE ANALYSIS OF MACHINE LEARNING MODELS FOR DETECTING PHISHING ATTACKS ON WEBSITES

LIONG KAH PONG

This project is submitted in partial fulfilment of the

requirements for the degree of

Bachelor of Computer Science with Honours

(Software Engineering)

Faculty of Computer Science and Information Technology

UNIVERSITI MALAYSIA SARAWAK 2023

IMPLEMENTASI DAN ANALISIS PRESTASI MODEL PEMBELAJARAN MESIN UNTUK MENGESAN SERANGAN PHISHING ATAS LAMAN WEB

LIONG KAH PONG

Projek ini merupakan salah satu keperluan untuk

Ijazah Sarjana Muda Sains Komputer dengan Kepujian

(Kejuruteraan Perisian)

Fakulti Sains Komputer dan Teknologi Maklumat

UNIVERSITI MALAYSIA SARAWAK 2023

UNIVERSITI MALAYSIA SARAWAK

THESIS STATUS ENDORSEMENT FORM

TITLE IMPLEMENTATION AND PERFORMANCE ANALYSIS OF MACHINE LEARNING MODELS FOR DETECTING PHISHING ATTACKS ON WEBSITES

ACADEMIC SESSION: 2022/2023

LIONG KAH PONG

(CAPITAL LETTERS)

hereby agree that this Thesis* shall be kept at the Centre for Academic Information Services, Universiti Malaysia Sarawak, subject to the following terms and conditions:

1. The Thesis is solely owned by Universiti Malaysia Sarawak
2. The Centre for Academic Information Services is given full rights to produce copies for educational purposes only
3. The Centre for Academic Information Services is given full rights to do digitization in order to develop local content database
4. The Centre for Academic Information Services is given full rights to produce copies of this Thesis as part of its exchange item program between Higher Learning Institutions [or for the purpose of interlibrary loan between HLI]
5. ** Please tick (√)

☐ CONFIDENTIAL (Contains classified information bounded by the OFFICIAL SECRETS ACT 1972)

☐ RESTRICTED (Contains restricted information as dictated by the body or organization where the research was conducted)

☑ UNRESTRICTED

Validated by

JH MOHAMAD IMRAN BIN HJ BANDAN
Senior Lecturer
Network Computing Programme
Faculty of Computer Science and Information Technology
Universiti Malaysia Sarawak

(AUTHOR'S SIGNATURE) (SUPERVISOR'S SIGNATURE)

Permanent Address

6642, TAMAN FIKRI
24000 KEMAMAN
TERENGGANU

Date: 24 July 2023 Date: 24 July 2023

Note * Thesis refers to PhD, Master, and Bachelor Degree
** For Confidential or Restricted materials, please attach relevant documents from relevant organizations / authorities

DECLARATION

I, Liong Kah Pong, hereby declare that the thesis with entitled “Implementation and Performance Analysis of Machine Learning Models for Detecting Phishing Attacks on Websites” is fully based on my original research work except for quotations and citations which have been appropriately acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at Universiti Malaysia Sarawak (UNIMAS).

Signature,

…………………………………………

(LIONG KAH PONG)

Faculty of Computer Science and Information Technology
Date: 17th January 2023
Universiti Malaysia Sarawak

ACKNOWLEDGEMENT

First, I would like to express my upmost gratitude towards my supervisor, Dr Mohamad Imran Hj Bandan, who have been providing me invaluable guidance and his professional knowledge to achieve the success of this project, especially in the scientific research related to the domain of artificial intelligence. I sincerely appreciate his effort in providing guidance, support, and constructive criticism that is given to me to ensure the highest quality of the project. This project would not be possible without the heartfelt guidance from my supervisor and his time and effort contributed to this project.

I would also like to express a warm appreciation to my family and friends who have been providing me with encouragement and reassurance. With their supports, it gave me strength in accomplishing this ever so important project. I am especially thankful to receive warm emotional supports from my family that keeps me going in my career journey.

As the symbol that represents one's qualifications and knowledge gained throughout the university program, this project bears a significant meaning. Therefore, I would like to thank the Faculty of Computer Science and Information Technology of Universiti Malaysia Sarawak for giving me this opportunity and platform to accomplish it with all sorts of amenities and assistance provided to me.

ABSTRACT

In the contemporary world, phishing attacks have become more apparent and caused tremendous financial loss to internet users. When attackers instrument these phishing attacks, an indispensable component frequently used together is a phishing website. Phishing websites are constructed to steal confidential information such as login credentials from victims. Usually, phishing websites are created resembling legitimate sources to deceive victims. To prevent users from falling victim to phishing websites, a machine-learning-based solution is proposed in this project. This project aims to detect phishing websites by implementing a tool that is built from a machine-learning model. This machine learning model is trained using known datasets on phishing websites and legitimate websites. So, features or attributes of phishing websites need to be discovered and this is achieved by looking at techniques phishing websites used to mimic legitimate sources. With the scholarly review of techniques employed by phishing websites, it is decided that they can be identified by their URLs and their SSL certificate information. Then, a machine learning tool is selected to build machine learning models that use three different machine learning algorithms, which are Support Vector Machine, Random Forest, and XGBoost. By having three different machine learning models, performance on how well these models classify phishing websites can be done. With the models successfully trained, they are deployed as a Chrome browser's Extension and subsequently tested. These models are then evaluated with accuracy, precision, and recall. Finally, the testing and evaluation is done, and XGBoost is proven to be the best performing model in terms of accuracy, precision, and recall.

ABSTRAK

Pada masa kini, serangan pancingan data atau phishing semakin meluas dan telah menyebabkan kerosakan dan kerugian terhadap pengguna internet. Apabila penjenayah siber menjalankan serangan phishing, komponen penting yang sering digunakan islah laman web phishing. Laman web phishing ini dibina untuk mencuri maklumat sulit daripada mangsa seperti kata kunci akaun bank. Biasanya, laman web phishing ini dibina untuk meniru sumber yang sah untuk mempedayakan mangsa. Untuk mengelakkan pengguna daripada menjadi mangsa, penyelesaian yang berasaskan pembelajaran mesin dicadangkan dalam projek ini. Projek ini bertujuan untuk mengesan laman web pancingan data dengan memperkenalkan alat yang dibina daripada model pembelajaran mesin. Model pembelajaran mesin ini dilatih menggunakan set data yang diketahui berkaitan dengan laman web pancingan data dan laman web sahih. Oleh itu, ciri-ciri laman web pancingan data adalah diperlukan dan ini dapat diperoleh dengan mengkaji teknik yang digunakan oleh laman web phishing untuk meniru sumber yang sah. Dengan melakukan kajian tentang teknik tersebut, laman web phishing boleh dikenal pasti melalui URLdan maklumat tentang sijil SSL laman web tersebut. Kemudian, alat pembelajaran mesin telah dipilih untuk membina model pembelajaran mesin yang menggunakan tiga algoritma pembelajaran mesin yang berbeza, iaitu Support Vector Machine, Random Forest, dan XGBoost. Dengan adanya tiga model pembelajaran mesin yang berbeza, prestasi mengenai model yang berbeza dalam mengklasifikasikan laman web pancingan data dapat dinilai dan dibanding. Dengan menggunakan model yang dilatih, ia diaplikasikan sebagai Extension Chrome dan seterusnya diuji. Model-model ini kemudiannya dinilai dengan Accuracy, Precision dan Recall. Akhirnya, kesimpulan bahawa model XGBoost adalah yang paling berprestasi telah ditetapkan.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

CHAPTER 1: INTRODUCTION

1.1 Background

In cybersecurity, phishing is defined as a social engineering technique that tricks Internet users into giving valid credentials or personal information (Merriam-Webster, n.d.). Nowadays, the number of phishing attacks has been increasing, parallel to the rise in internet users. This unprecedented connectivity across the globe allows cybercriminals to make use of this opportunity to reap confidential information from individuals and corporates. According to Warburton, phishing attacks skyrocketed to 222% in the year 2020 due to the outbreak of COVID-19 as people quarantined at home and did remote work. Therefore, it is observed that the spike in internet users caused an increase in cybercriminal activities, especially in phishing attacks, and it can get worse in the future because internet users will only increase over time.

The effect of phishing attacks is often detrimental to victims, be they individuals or corporates. By gaining confidential information such as credentials, cybercriminals usually aim for monetary gain using harvested illicit information. Besides, personal data can be collected from phishing techniques and sold to third parties. Sometimes, phishing can be also orchestrated by rival corporations or even countries to conduct espionage to gain invaluable intellectual property. With the severity of the damage phishing attacks can cause, a volatile and ingenious solution is essential to protect the interest of users.

With the advent of Artificial Intelligence in the early 2000s, the application of AI in various industries has seen limitless potential (Géron, 2017). With the ongoing threat of phishing attacks, it is proposed that AI can be utilised in assisting internet users from falling victim to such risks. For a user to prevent himself from falling victim to phishing websites, he must first be able to identify one. Therefore, an AI system has the potential to help the user in identifying phishing websites. By using machine-learning algorithms in supervised learning, a system is proposed to label phishing websites by utilising a carefully crafted feature engineering process. In this project, machine learning techniques are explored and evaluated to determine the best approach to labelling phishing websites and their potential in a scalable environment. A combination of two or more techniques is also expected to produce the results with desired performance.

This research is carried out to create a system that incorporates machine-learning techniques in identifying phishing websites to prevent users from falling victim to them. In this project, research is carried out to select machine-learning algorithms, and then design and implement the detection tool from the produced machine-learning models. Finally, the selected models' performance is evaluated to decide the best approach in labelling phishing websites. A thorough research methodology and product development are outlined in detail in upcoming chapters. The proposed system aims to run locally, relying on the performance of the computer.

1.2 Problem Statement

Phishing attacks involve the use of social engineering techniques to obtain information by posing as a trustworthy source (Google Search Central, n.d.). Nowadays, one of the most common techniques of phishing is by hosting phishing websites to lure users into giving up their confidential information. To achieve their malicious purpose, phishing websites are carefully crafted to imitate their legitimate counterparts. A phishing website has an identical appearance and an almost similar URL to its counterpart to deceive users. Nowadays, phishing websites employ more intricate techniques such as having legitimate SSL certificates, using

Unicode to imitate the URLs of their legitimate counterpart and so on (Mahajan & Siddavatam, 2018). Therefore, it has become increasingly hard for internet users to identify phishing websites on their own without prior knowledge of cybersecurity.

Besides, the number of victims who fell victim to the trickery of phishing websites has been increasing over the year. According to a report by SlashNext (2022), there are 255 million phishing attacks over six months in 2022, which is a whopping 61% increase as compared to the previous year. This could be caused by the intricate techniques used by modern phishing websites in luring victims. Besides, cybercriminals also exploit the fact that internet users have been increasing. Therefore, an ingenious solution is needed to prevent phishing attacks to safeguard the interest of individuals and corporates.

To achieve the purpose of the project, which is to identify phishing websites, the researcher must first identify the type of machine-learning problem of the project. This can be done by analysing the possible operation of the proposed solution and the characteristics of the gathered datasets. Since the purpose of the project is to predict a discrete outcome, which is to label websites into a binary result of being either a phishing website or not a phishing website, and considering that the datasets are labelled, the type of machine-learning problem of the project is a classification problem, particularly a binary classification problem.

1.3 Objectives

1. To implement three machine learning models for building an application to detect phishing websites.
2. To evaluate the performance of different machine learning models in detecting phishing websites in terms of different performance metrics.

1.4 Scope

In this research, the proposed outcome – the phishing detection tool takes the form of a browser extension, specifically a Chrome browser extension. The tool can scan through the website a person is visiting to check whether the website is a phishing website. The logic that runs the checking is powered by a machine-learning model.

To successfully carry out this research, the programming languages required are Python for machine learning purposes and JavaScript for developing the Chrome extension. Other than that, the required Python libraries for machine learning are Scikit-learn, Pandas, and Numpy. These libraries have built-in machine-learning algorithms that are ready to be applied to the project.

Since the project is based on machine learning models, a selection of machine learning algorithms and datasets is needed to build the models. The machine learning algorithms selected are the likes of the classification algorithm, which is appropriate for solving a binary classification problem. The classification algorithm is a supervised-learning algorithm that requires labelled datasets. So, the labelled datasets on phishing websites are gathered from PhishTank, Mendeley, and UNIMAS Repository.

The tool that is developed is a Chrome extension. Therefore, the tool requires a Chromium-based browser such as Google Chrome to run the training models. To ensure the relevancy of the tool to most computer users, it should be running on at least a 64-bit Windows 10. The phishing detection tool run its machine learning models natively on the user’s computer. Therefore, users need to have computers that use at least an Intel i7-7500U processor and 8GB of RAM (Singh, 2019).

1.5 Project Significance

Overall, this project is beneficial to internet users, especially to users who have little to no cybersecurity awareness. From the outcome of the research, a good machine learning model to detect phishing websites is fundamental for the future development of better phishing detection tools. As prevention is better than cure, the proposed phishing detection tool is the first line of defence against phishing attacks. The proposed tool identifies phishing websites and is capable of doing zero-hour detection without extra effort. Indirectly, it also helps educate internet users on how to safeguard their interests from phishing attacks.

1.6 Project Schedule

The project is scheduled to start on 7th October 2022, and it is expected to be completed on 9th June 2023. The details of the schedule and its associated Gantt Chart are shown in Appendix A.

1.7 Expected Outcome

For this project, a working phishing detection tool is developed. This tool can run on different machine-learning models in scanning and labelling websites as phishing or non-phishing websites. Then, a report containing the analysis of the performance of different machine learning models in detecting phishing websites is compiled. The objective of the project is achieved to provide a fruitful contribution to the field of AI and cybersecurity for the best interests of the users.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

According to Cisco (n.d.), Phishing Attacks involve sending fraudulent communication to imitate a reputable source to steal confidential information such as credit card and login credentials or to install malware on victims' machines. In essence, phishing attacks use spoofed emails and websites to bait victims into voluntarily handing over their confidential information (Phishing.org, n.d.-b). Spoofing is an integral part of facilitating phishing (Panigrahi, 2022). It is an identity theft technique where a criminal attempts to use a legitimate entity's identity in achieving malicious objectives (Lim et al., 2020). Thus, phishing websites that work as bait to lure victims can apply spoofing techniques to provide a better effect. For example, a phishing website used in a phishing attack can use a poofed URL where the URL highly resembles the legitimate source.

According to Phishing.org, the first ever recorded phishing attack dates to January 2, 1996. The website stated that the first malicious internet activity that first used the term phishing, happened in AOHell, a Usenet newsgroup. The term phishing is associated with the earliest hackers who were known as phreaks (Phishing.org, n.d.-a). In the earlier days of the internet, hackers target AOL communities such as the warez community that were popular with users trading pirated software with each other (Phishing.org, n.d.-a). They would use a program known as AOHell in generating randomised credit card numbers to create spam accounts on AOL. After this practice was put to an end, hackers resorted to impersonating AOL employees and deceiving users into giving out their credentials (Phishing.org, n.d.-a). Hence, this social engineering technique would eventually be known as phishing.

Nowadays, phishing attacks have become more sophisticated and are now propagated through different means. It can come in many forms, such as email phishing, spear phishing, whaling, smishing and vishing, and angler phishing (Irwin, 2022). To elaborate on the types of phishing, email phishing attacks are sent by email that used a fake email domain that looks like a legitimate source (Irwin, 2022). Spear phishing on the other hand also uses email, but the email sent is engineered to include a targeted victim’s information to gain the victim’s trust (Irwin, 2022). A more ambitious form of spear phishing is whaling, which targets important figures like senior executives (Irwin, 2022). Smishing and vishing are other forms of phishing attack that uses SMS and voice call (Irwin, 2022). Finally, angler phishing is a modern type of phishing that uses social media to propagate phishing attacks (Irwin, 2022).

2.2 Phishing Attacks

Most phishing attack techniques require a phishing website to steal information from victims (Gillis, n.d.). In the world wide web, technically, there is essentially no limitation on the website or webpage that can be hosted, given the required expertise in doing so. Therefore, people with malicious intentions can create websites to spoof a legitimate website from a reputable source. These websites are called phishing websites and are essential to different phishing attacks. To give a clear picture, in an email phishing attack, the crafted email would attach a link to a spoofed phishing website (Gillis, n.d.). As the victim let his guard down and trusted the email, the victim clicked on the link and was redirected to the phishing website (Gillis, n.d.). From there, any input by the user, especially login credentials or bank information keyed into the phishing website, would be sent to cybercriminals instead. Even worse, some