DEVELOPMENT OF A WEB BROWSER EXTENSION FOR PHISHING WEBSITE DETECTION USING MACHINE LEARNING

DUROJAIYE, PEACE DUROJAIYE
(15CH03729)
B.Sc. Management Information system, Covenant University, Ogun state

AUGUST, 2023

DEVELOPMENT OF A WEB BROWSER EXTENSION FOR PHISHING WEBSITE DETECTION USING MACHINE LEARNING

BY

DUROJAIYE, PEACE DUROJAIYE
(15CH03729)
B.Sc. Management Information system, Covenant University, Ogun state

A DISSERTATION SUBMITTED TO THE SCHOOL OF POSTGRADUATE STUDIES IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF MASTER OF SCIENCE (M.Sc) DEGREE IN MANAGEMENT INFORMATION SYSTEMS IN THE DEPARTMENT OF COMPUTER AND INFORMATION SCIENCES, COLLEGE OF SCIENCE AND TECHNOLOGY, COVENANT UNIVERSITY, OTA, OGUN STATE, NIGERIA

AUGUST, 2023

ACCEPTANCE

This is to attest that this dissertation is accepted in partial fulfilment of the requirements for the award of the degree of Master of Science in Management Information System in the Department of Computer and Information Sciences, College of Science and Technology, Covenant University, Ota, Nigeria.

Miss Adefunke F. Oyinloye
(Secretary, School of Postgraduate Studies) **Signature and Date**

Prof. Akan B. Williams
(Dean, School of Postgraduate Studies) **Signature and Date**

DECLARATION

I**, DUROJAIYE, PEACE OLUWASEYI (15CH03729),** declare that this research was carried out by me under the supervision of Dr. Aderonke A. Oni of the Department of Computer and Information Sciences, College of Science and Technology, Covenant University, Ota, Ogun State, Nigeria. I attest that the dissertation has not been presented either wholly or partially for the award of any degree elsewhere. All sources of data and scholarly information used in this dissertation are duly acknowledged.

DUROJAIYE, PEACE OLUWASEYI

Signature and Date

CERTIFICATION

We certify that this dissertation titled “**DEVELOPMENT OF A WEB BROWSER EXTENSION FOR PHISHING WEBSITE DETECTION USING MACHINE LEARNING**” is an original research work carried out by **DUROJAIYE, PEACE OLUWASEYI (15CH03729)** in the Department of Computer and Information Sciences, College of Science and Technology, Covenant University, Ota, Ogun State, Nigeria under the supervision of Dr. Aderonke A. Oni. We have examined and found this work acceptable as part of the requirements for the award of Master of Science (M.Sc.) in Management Information System.

Dr. Aderonke A. Oni
(Supervisor) **Signature and Date**

Prof. Olufunke O. Oladipupo
(Head of Department) **Signature and Date**

Prof. Olufunke O. Vincent
(External Examiner) **Signature and Date**

Prof. Akan B. Williams
(Dean, School of Postgraduate Studies) **Signature and Date**

DEDICATION

This dissertation is dedicated to God, who is my source of strength, wisdom, inspiration and knowledge. My heartfelt gratitude goes to my parents, Pastor and Pastor (Mrs.) B.A Durojaiye and siblings; Victor, Joy, my twin brother (John) and Victoria for their unending support and encouragement in diverse ways possible. To my colleagues (Favour, Dami, Ope, Emma, Jumoke, Faith and Paul) and loved ones who have been instrumental in their way towards fulfilling this quest, I appreciate you all and may God bless you all abundantly.

ACKNOWLEDGEMENTS

First, I want to acknowledge God Almighty for his mercy and strength that enabled me to carry out this research study efficiently and effectively. I want to specially thank my parents and siblings for all their love and support. Secondly, I want to specially thank my supervisor Dr. Aderonke A. Oni for her effort in making sure this research study was a success and also her guidance and patience and the understanding she impacted unto me, thank you so much ma, May God Almighty continue to bless and protect you. Thirdly, I want to thank and give my regards to the Management and Chancellor of Covenant University for the calm and adventurous environment that has enabled me to pursue my dreams and passion in academics. I also want to appreciate my postgraduate colleagues and friends: Favour Folorunso, Damilola Adeniji, Opeyemi Odetola, Emmanuel Adedire, Faith Adegoke, Jumoke Adeyemi, Paul Owolabi and many others for their support and encouragement in times when things were tough, I celebrate you all.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATION

ABBREIVATIONS	**MEANING**
APWG	Anti-Phishing Working Group
ANOVA	Analysis of Variance
CFS	Correlation based feature selection
CSS	Cascading Style Sheet
DNS	Domain Name System
DT	Decision Tree
ET	Extra Tree
GA	Genetic Algorithm
GBC	Gradient Boosting Classifiers
HTML	Hypertext Markup Language
IC3	Internet Crime Complaint Center
IDE	Integrated Development Environment
IG	Information gain
IP	Internet Protocol
KNN	K-Nearest Neighbor
LR	Logistic Regression
MAE	Mean Absolute Error
NB	Naïve Bayes
NN	Neural Network
PCA	Principal component analysis
RF	Random Forest
RFE	Recursive Feature Elimination
RFSSA	Recursive Features Subset Selection Algorithm
RMSE	Root Mean Square Error
SAAS	Software as a Service
SVM	Support Vector Machine
TLD	Top-level Domain
URL	Uniform Resource Locator

ABSTRACT

Online platforms play a critical role in daily life; however, they expose users to cybersecurity threats, including phishing attacks. This study focuses on developing a web browser extension that utilizes machine learning techniques to identify phishing websites with enhanced accuracy. Five machine learning algorithms - Decision Tree, Random Forest, Support Vector Machine (SVM), Logistic Regression, and Gaussian Naive Bayes - were evaluated for phishing detection using a dataset of 11,430 URLs consisting of 87 features such as URL length, domain age, and web traffic. The study also engaged Exploratory Data Analysis to extract key insights from the dataset. The evaluation reveals the effectiveness of different machine learning models. Metrics like accuracy, precision, recall, and F1 score are provided for each model, highlighting their strengths and limitations. Through cross-validation and careful hyperparameter tuning, the Random model emerges as the most accurate. Rule extraction is then applied to this model, yielding understandable rules that illuminate its decision-making process. Additionally, the study practically applies the developed model through a phishing detection Web Browser Extension. This extension offers real-time website validation and alerts users about potential phishing risks. By seamlessly integrating machine learning into a user-friendly interface, the browser extension empowers users to assess website legitimacy, thereby enhancing online security. This study offers valuable insights into cybersecurity by presenting an efficient machine learning method for the identification and classification of phishing websites. The findings underscore the potential of this model to safeguard sensitive information and counter the rising threat of phishing attacks.

Keywords: Phishing detection, machine learning, Web-based Platform, real-time detection, cybersecurity, browser extension.