PHISHING WEBSITE SCANNER USING MACHINE LEARNING

Mini Project Report

Submitted to the APJ Abdul Kalam Technological University in partial fulfillment of requirements for the award of degree

Bachelor of Technology

in

Computer Science and Engineering

by

AGNES MARY ALLEN (LBT21CS005)

CHAITHANYA S (LBT21CS031)

DEVU S THIRUMANGAL (LBT21CS033)

GANGA S (LBT21CS036)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

LBS INSTITUTE OF TECHNOLOGY FOR WOMEN

TRIVANDRUM, KERALA

MAY 2024

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

LBS INSTITUTE OF TECHNOLOGY FOR WOMEN

TRIVANDRUM, KERALA

CERTIFICATE

This is to certify that the report entitled '**PHISHING WEBSITE SCANNER USING MACHINE LEARNING**' submitted by **AGNES MARY ALLEN (LBT21CS005), CHAITHANYA S (LBT21CS031), DEVU S THIRUMANGAL (LBT21CS033), GANGA S (LBT21CS036)** to the APJ Abdul Kalam Technological University in partial fulfillment of the requirements for the award of degree of Bachelor of Technology in Computer Science and Engineering is a bonafide record of the project work carried out by them under our guidance and supervision.

Prof. Rehna R. S	**Dr. Lekshmy P. L**	**Dr. Anitha Kumari S**
Assistant Professor	Assistant Professor	Head of Department
Department of CSE	Department of CSE	Department of CSE
LBSITW, TVM	LBSITW, TVM	LBSITW, TVM
(Mini Project Guide)	(Mini Project Coordinator)	

DECLARATION

We, the undersigned, hereby declare that the project report titled '**PHISHING WEBSITE SCANNER USING MACHINE LEARNING**' submitted in partial fulfillment of the requirements for the Bachelor of Technology degree at APJ Abdul Kalam Technological University, Kerala, represents our original work conducted under the supervision of **Prof. Rehna R S**, Assistant Professor, Department of Computer Science and Engineering, L B S Institute of Technology for Women, Poojappura. We affirm that this submission reflects our own ideas and that any contributions from external sources have been accurately cited and referenced. We attest to our adherence to the principles of academic honesty and integrity, ensuring that all data, ideas, facts, and sources have been presented truthfully and ethically. We acknowledge that any breach of academic integrity or misrepresentation of data may result in disciplinary action by the institute and/or the University. Furthermore, we confirm that this report has not been previously used to obtain any degree, diploma, or similar title from any other academic institution.

Place: Thiruvananthapuram

Date: May 2024

AGNES MARY ALLEN

CHAITHANYA S

DEVU S THIRUMANGAL

GANGA S

ACKNOWLEDGEMENT

We extend our heartfelt gratitude to **Dr. Jayamohan J**, our Principal, for the provision of essential facilities and infrastructure vital for the completion of this project. Our sincere appreciation also goes to **Dr. Anitha Kumari S**, Head of the Department of Computer Science and Engineering, for facilitating the necessary resources throughout this endeavour. Special thanks are due to **Dr. Lekshmy P L**, our project coordinator and Assistant Professor in the Department of Computer Science and Engineering, for her invaluable guidance and assistance throughout the project's duration. We are equally grateful to **Prof. Rehna R S**, our project guide and Assistant Professor in the Department of Computer Science and Engineering, for offering us the opportunity and expert guidance to undertake our third-year project. We would like to acknowledge the contributions of **Prof. Diana Walts**, Assistant Professor, Department of Computer Science and Engineering, and **Prof. Shameema Latheef R**, Assistant Professor, Department of Computer Science and Engineering. We would also like to recognize the collective efforts of all staff members in the department for their continuous support and encouragement. Finally, we extend our gratitude to our friends and well-wishers for their cooperation, support, and insightful suggestions, which have contributed significantly to the advancement of this project.

AGNES MARY ALLEN

CHAITHANYA S

DEVU S THIRUMANGAL

GANGA S

ABSTRACT

In today's digital landscape, cybersecurity stands as a critical imperative in light of escalating threats such as ransomware and phishing. Within this context, Machine Learning (ML), a subset of Artificial Intelligence (AI), emerges as a potent tool for fortifying security measures by proactively predicting and countering malicious activities. Phishing attacks, designed to deceive users through counterfeit websites, represent a particularly menacing hazard to both individuals and organizations alike. DataProt's alarming statistic reveals the creation of a new phishing site every 20 seconds on average, with an overwhelming 90% of corporate breaches attributed to such insidious attacks according to Digital Guardian. In response to this pressing need, Machine Learning algorithms have assumed a pivotal role in analyzing vast datasets to detect and preempt cyber threats, including phishing attacks, in real-time. By harnessing the power of ML, organizations can effectively stay ahead of evolving risks and safeguard their digital assets. This paper proposes a novel approach to phishing website detection through the utilization of a hybrid LSD model, which amalgamates Logistic Regression, Support Vector Machine (SVM), and Decision Tree algorithms for URL classification. By synergistically leveraging the distinctive strengths of each algorithm, this model promises to deliver superior accuracy and generalization capabilities, thereby bolstering cybersecurity defenses. The efficacy of the proposed model is evaluated rigorously through the application of performance metrics such as accuracy, precision, recall, and F1-score. Furthermore, to provide a comprehensive assessment, a comparative analysis is conducted with the Random Forest classifier, identified as the second-best model as per the base paper. Through this comparative evaluation, valuable insights are gleaned into the relative strengths and weaknesses of different machine learning approaches in effectively distinguishing phishing websites from legitimate ones. In conclusion, this research underscores the indispensable role of Machine Learning in fortifying cybersecurity defenses against phishing attacks and other malicious activities. By advancing innovative methodologies and conducting thorough evaluations, we aim to empower organizations with the tools and insights necessary to mitigate cyber risks and safeguard their digital assets in an increasingly hostile digital landscape.

TABLE OF CONTENTS