MACHINE LEARNING FOR WEB VULNERABILITY DETECTION

A MINI PROJECT REPORT

Submitted by

APPIDI GIRIDHAR REDDY

(20841A0546)

PATHAPATI BHANUTEJA

(20841A0532)

S SAI SREEKAR

(20841A0560)

GUIDED BY

Mrs. V. APARNA VARALAKSHMI

in partial fulfillment for the award of the degree

of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING

AURORA'S TECHNOLOGICAL AND RESEARCH INSTITUTE

(Affiliated to JNTU, Hyderabad and Accredited by NAAC with 'A' grade)

Parvathapur, Uppal, Hyderabad - 500039

2023 - 2024

AURORA’S TECHNOLOGICAL AND RESEARCH INSTITUTE

(Affiliated to JNTU, Hyderabad and Accredited by NAAC with ‘A’ grade)

Parvathapur, Uppal, Hyderabad - 500 039

DECLARATION

We hereby declare that the work described in this project, entitled **“MACHINE LEARNING FOR WEB VULNERABILITY DETECTION”** which is being submitted by us in partial fulfilment for the award of Bachelor of Technology in Computer Science and Engineering to **AURORA’S TECHNOLOGICAL AND RESEARCH INSTITUTE** is the result of investigation carried by us under the guidance of **Mrs. V. Aparna Varalakshmi, Associate Professor, CSE.**

The work is original and has not been submitted for any degree of this or any other university.

Place: Hyderabad

Date:

Appidi Giridhar Reddy (20841A0546)

Pathapati Bhanuteja (20841A0532)

S Sai Sreekar (20841A0560)

AURORA’S TECHNOLOGICAL AND RESEARCH INSTITUTE

(Affiliated to JNTU, Hyderabad and Accredited by NAAC with ‘A’ grade)

Parvathapur, Uppal, Hyderabad - 500 039

CERTIFICATE

Certified that this project report **“MACHINE LEARNING FOR WEB VULNERABILITY DETECTION”** is the bonafide work of **“APPIDI GIRIDHAR REDDY - 20841A0546, PATHAPATI BHANUTEJA-20841A0532, S SAI SREEKAR - 20841A0560”** who carried out the project work under my supervision.

GUIDE	**MINI PROJECT COORDINATOR**
Mrs. V. Aparna Varalakshmi	**Dr. B. T. R. Naresh Reddy**
Associate Professor	Associate Professor
	Department of CSE / IT
HEAD OF THE DEPARTMENT	**PRINCIPAL**
Mrs. A. Durga Pavani	**Dr. A. Mahesh Babu**
Department of CSE	

EXTERNAL EXAMINAR

ACKNOWLEDGMENT

This work has been done during the project period and it was a very good opportunity to put theoretical knowledge into planned exercise with an aim to solve a real time problem and also to develop confidence to face various practical situations.

We convey thanks to our project guide **Mrs. V. Aparna Varalakshmi,** Department of Computer Science and Engineering, for providing encouragement, constant support and guidance which was of great help to complete this project successfully.

We express our sincere thanks to our Project Coordinator **Dr. B. T. R. Naresh Reddy** for helping us to complete our project work by giving valuable suggestions.

We express our sincere thanks to Head of the Department **Mrs. A. Durga Pavani** for giving us the support and her kind attention and valuable guidance to us throughout this course.

We would also like to express our gratitude to **Dr. A. Mahesh Babu, Principal,** Aurora's Technological and Research Institute for providing us with a congenial atmosphere and encouragement.

Finally, we would also like to thank the people who have directly or indirectly helped us, our parents, and our friends for their cooperation in completing the Mini Project work.

Appidi Giridhar Reddy (20841A0546)
Pathapati Bhanuteja (20841A0532)
S Sai Sreekar (20841A0560)

ABSTRACT

In this project, we propose a methodology to leverage Machine Learning (ML) for the detection of web application vulnerabilities. Web applications are particularly challenging to analyses, due to their diversity and the widespread adoption of custom programming practices. ML is thus very helpful for web application security: it can take advantage of manually labeled data to bring the human understanding of the web application semantics into automated analysis tools. We use our methodology in the design of Mitch, the first ML solution for the black-box detection of Cross-Site Request Forgery (CSRF) vulnerabilities. Mitch allowed us to identify 35 new CSRFs on 20 major websites and 3 new CSRFs on production software. Web applications have become an integral part of modern digital interactions, yet they remain susceptible to a wide range of security vulnerabilities that can compromise user data and system integrity. This project presents an innovative approach to enhancing web application security through the application of machine learning techniques for web vulnerability detection. By leveraging a diverse dataset of simulated attack scenarios and legitimate interactions, a robust and adaptive model is developed to identify and classify vulnerabilities such as SQL injection, cross-site scripting (XSS), and cross-site request forgery (CSRF). Feature engineering, selection, and extraction methods are employed to effectively capture intricate patterns indicative of vulnerabilities.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVATIONS

ML	-	Machine Learning
CSRF	-	Cross Site Request Forgery
MySQL	-	My Structured Query Language
HTML	-	Hypertext Markup Language
CSS	-	Cascading Style Sheets
GUI	-	Graphical User Interface
XSS	-	Cross Site Scripting
HTTP	-	Hypertext Transfer Protocol
WAF	-	Web Application Firewall