Enhancing Email Security: Abnormal Login Detection Through Machine Learning Algorithm

Ariyawansa M.M.T.R.
(Reg. No.: MS22040480)

A THESIS
SUBMITTED TO
SRI LANKA INSTITUTE OF INFORMATION TECHNOLOGY
IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE IN INFORMATION TECHNOLOGY
(CYBER SECURITY)

December 2024

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

.....

Prof Anuradha Jayakody

Approved for MSc. Research Project:

_____

MSc. Programme Co-ordinator, SLIIT

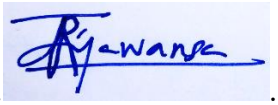
Approved for MSc:

_____

Head of Graduate Studies, FoC, SLIIT

DECLARATION

This is to certify that the work is entirely my own and not of any other person, unless explicitly acknowledged (including citation of published and unpublished sources). The work has not previously been submitted in any form to the Sri Lanka Institute of Information Technology or to any other institution for assessment for any other purpose.

Sign: ……  .....

Ariyawansa M.M.T.R.

Date: …… 12/11/2024 ……..

ABSTRACT

Enhancing Email Security: Abnormal Login Detection Through Machine Learning Algorithm

Ariyawansa M.M.T.R.
MSc. in Cyber Security
Supervisor: Prof. Anuradha Jayakody
December 2024

The research focuses on the application of random forest machine learning algorithm for the identification of non-standard authentication activities in email systems. The idea of the software is to strengthen email defenses by the means of the dynamical determination of the unusual login patterns and then responsively to the continuously changing threats in cyberspace. These projects use the most up-to-date machine learning approaches, meticulous hyperparameter tuning and comprehensive feature engineering, so that a strong barrier against unauthorized entry would be created.

The purpose is to design a machine learning model capable of differentiating between the most and least likely behaviors, based on the analysis of users' activity data. This includes steps of label encoding and timestamp processing targeted to clean the input data before model training for optimal efficiency. In the process of training, the Scikit-learn library is employed to implement the machine learning algorithms. Furthermore, hyperparameter optimization is performed using GridSearchCV to refine the model’s accuracy and efficiency.

The study puts its emphasis on user-friendly implementation with the development of an intuitive interface offering the users an understandable classification report illustrating the model of breach detection performance. The developed model that associated the random forest machine learning algorithm showed a accuracy of 83%, making it ideal for real world use. Instead of just enhancing user engagement, it also enables prompt reaction and mitigation measures. As a result, this thesis offers a practical and effective way of guarding email accounts from rapidly evolving threats.

ACKNOWLEDGEMENT

I am very grateful to Prof. Anuradha Jayakody, my supervisor at SLIIT, for all of the helpful advice and encouragement he gave me when I chose a research topic. I would also like to thank my coworkers and friends who were instrumental in getting this proposal together on time and providing invaluable feedback and guidance.

TABLE OF CONTENTS

List of Figures