



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

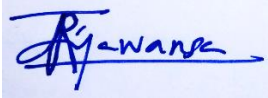
Approved for MSc:



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is entirely my own and not of any other person, unless explicitly acknowledged (including citation of published and unpublished sources).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

DECLARATION	ii
ABSTRACT.....	iii
ACKNOWLEDGEMENT	iv
TABLE OF CONTENTS.....	v
List of Figures	viii
Chapter 1 Introduction	1
1.1 Demand for more advanced security systems	2
1.2 Restrictions on traditional security methods	3
1.3 Innovative methods to enhance Email cryptography	4
1.4 Email Security Critical for Businesses.....	4
1.4.1 Preservation of Brand Trust	4
1.4.2 Regulatory and Compliance Imperatives	5
1.4.3 Protection Against Financial Fraud.....	6
1.4.4 Data Protection and Intellectual Property	7
1.4.5 Business Continuity	8
1.4.6 Protection from Insider Threats	9
1.5 Benefits of Email Security	9
1.6 Real-World Examples of Email Security Breaches	11
1.7 Email Security Policies	14
1.7.1 Incident Response Guidelines	15
1.7.2 Data Retention and Archiving.....	15
1.7.3 Password Management Requirements	16
1.7.4 User Access Reviews	16
1.7.5 Incident Reporting Framework	17
1.8 Key Features Considered in Email Security Solution.....	17
1.8.1 Advanced Threat Detection	17
1.8.2 Integration with Broader Security Stack.....	18
1.8.3 Email Quarantine and Filtering.....	18
1.8.1 Threat Intelligence Integration.....	19
1.8.2 User Behavior Analytics	19
1.8.3 Reporting and Compliance Tools	20
1.9 Research Gap	21
1.10 Research Questions	22
1.11 Research Objectives	22
1.11.1 Main Objective.....	23
1.11.2 Sub Objectives	23

Chapter 2 Literature review	24
Chapter 3 Methodology	30
3.1 Structure of the research	30
3.1.1 Data Collection and Preprocessing:	30
3.1.2 Feature Selection and Engineering.....	30
3.1.3 Model Development and Optimization:.....	31
3.1.4 Pattern Identification and Analysis	31
3.1.5 Real-time Response Mechanisms	31
3.1.6 Performance Evaluation and Impact Analysis:	32
3.1.7 Actionable Recommendations and Implementation:	32
3.2 Technologies used in research	33
3.3 Random Forest Classifier.....	34
Chapter 4 Testing and Evaluation.....	38
4.1 Model testing procedure	38
4.2 Evaluation metrics	38
4.2.1 Accuracy:	39
4.2.2 Precision.....	40
4.2.3 Recall (Sensitivity).....	41
4.2.4 F1-score.....	42
4.2.5 Area Under the ROC Curve (AUC)	43
4.3 Comparing model with different datasets procedure	43
4.4 Visual Evaluation Tools.....	46
4.4.1 Confusion Matrix	46
4.4.2 Precision-Recall Curve	46
4.4.3 Receiver Operating Characteristic (ROC) Curve and AUC.....	47
4.5 Results Interpretation	48
4.6 Used Parameter Grid.....	49
4.7 Security Impact Analysis	52
Chapter 5 Conclusion.....	54
References.....	56
Appendix.....	59
Appendix 1: Code used for Model Creation	59
Appendix 2: Code used for Prediction.....	61
Appendix 3: Turnitin report : 11%.....	62

List of Figures

Figure 3.1 The flow of the methodology	33
Figure 4.1 Python command used in generating the predictions	38
Figure 4.2 The confusion matrix creation and plotting using Python.....	46
Figure 4.3 The Precision-Recall curve creation and plotting using Python.....	47
Figure 4.4 The ROC curve and AUC creation and plotting using Python.....	47
Figure 4.5 The output : Result interpretation.....	48