IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2347-6710

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 11, November 2023

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.423

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 11, November 2023 ||

| DOI:10.15680/IJIRSET.2023.1211036 |

Automated SMS Classification and Spam Analysis using topic modelling

Chaudhari Komal[1], Prof. N.S. Sapike[2]

P.G. Student, Dept. of Computer Engineering, Vishwabharati Academy's COE, Ahmednagar, Maharashtra India[1]

Assistant Prof., Dept. of Computer Engineering, Vishwabharati Academy's COE, Ahmednagar, Maharashtra, India[2]

ABSTRACT: Phishing is a type of extensive fraud that happens when a malicious website act like a real one keeping in mind that the end goal to obtain touchy data. In spite of the fact that there are a few contrary to phishing programming and methods for distinguishing potential phishing endeavours in messages and identifying phishing substance on messages. However, detecting phishing messages is a challenging task, as most of these techniques are not able to make an accurate decision dynamically as to whether the new message is phishing or legitimate. Propose system will compare message with spam keyword database ,if content in message matched with database contents, then message is detected as spam.

KEYWORDS: Phishing, Extreme Learning Machine, Feature Classification.

I. INTRODUCTION

Phishing is a technique of tricking people into giving sensitive information like usernames and passwords, credit card details, sensitive bank information, etc., by way of email spoofing, instant messaging, or using fake web sites whose look and feel gives the appearance of a legitimate website.System will decides whether a message is phishing thanks to the Bayesian classification algorithm and the scores added to the database. It is instantly perceived as a spam message by the words that are exciting, phrases that increase the desire for shopping, and which contain unwanted content.

Spammers have focused their attention on sending spam through short message service to mobile users. By way of instant messaging, whose look and feel gives the appearance of a legitimate SMS. The Designed System will decides whether a message is Spam thanks to the Bayesian classification algorithm and the scores added to the database. It is instantly perceived as a spam message by the words that are exciting, phrases that increase the desire for shopping, and which contain unwanted content. Model is designed with Automated framework for filtering SMS spam with the use of various classifiers in being developed.

Phishing is a type of extensive fraud that happens when a malicious message act like a real one , for example, passwords, account points of interest, or MasterCard numbers. In spite of the fact that there are a few contrary to phishing programming and methods for distinguishing potential phishing endeavors in messages and identifying phishing substance on sites, phishers think of new and half breed strategies to go around the accessible programming and systems.

Phishing is a trickery system that uses a blend of social designing what's more, innovation to assemble delicate and individual data, for example, passwords and charge card subtle elements by taking on the appearance of a dependable individual or business in an electronic correspondence. Phishing makes utilization of spoof messages that are made to look valid and implied to be originating from honest to goodness sources like money related foundations, e-commerce destinations and so forth, to draw clients to visit fake messages through joins gave in the phishing email.

Paper is organized as follows. Section gives the introduction topic with need for plant disease detection and overall graph for plants over world. Section II gives the related work till now in this field for plant disease detection. Section III gives the architecture for the system and inclination of more facilities included in the system. Finally, Section V presents conclusion.

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 11, November 2023 ||

| DOI:10.15680/IJIRSET.2023.1211036 |

II. RELATED WORK

In "**An Intelligent System for Spam Message Detection**"Authors present a system of spam message prediction based on appropriate lexical analysis like tokenization, stop-words removal, stemming, lemmatization, and feature extraction[1].
While in"Spam Message Detection Using Logistic Regression"with the help of Natural Language Processing and Machine Learning, we can quickly detect spam messages. One of the crucial aspects of research in the world of machine learning applications is "NLP"[2].In Intelligent Security Schema for SMS Spam Message Based on Machine Learning Algorithms authors try to resolve spam message issue, a new intelligent security system is proposed to reduce the number of spam messages. It can detect novel spam messages that have a direct and negative impact on networks. The proposed system is heavily based on machine learning to explore various types of messages[3].

In **Bayesian Spam Detection Framework on Mobile Device**,most of the spam detection schemes based on Bayesian are designed for communication providers. The scalability of filter policy is hard to control because the strict policy might filter the normal messages while some spam messages would not be detected due to the flexibility in the policy[4].In Robust Approach for Effective Spam Detection Using Supervised Learning Techniques.Recent trend is using regional language for spamming.to detect these spam messages author has used monto -carlo method.

Abdulhamit Subasi, Esraa Molah, Fatin Almkallawi, Touseef J. Chaudhery, "Intelligent Phishing Website Detection using Random Forest Classifier
Phishing is defined as mimicking a creditable company's website aiming to take private information of a user. In order to eliminate phishing, different solutions proposed. However, only one single magic bullet cannot eliminate this threat completely. Data mining is a promising technique used to detect phishing attacks. In this paper, an intelligent system to detect phishing attacks is presented. We used different data mining techniques to decide categories of websites: legitimate or phishing. Different classifiers were used to construct accurate intelligent system for phishing website detection. Classification accuracy, area under receiver operating characteristic (ROC) curves (AUC) and F-measure is used to evaluate the performance of the data mining techniques. Results showed that Random Forest has outperformed best among the classification methods by achieving the highest accuracy 97.36\%. Random forest runtimes are quite fast, and it can deal with different websites for phishing detection.

C. Emilin Shyni,Anesh D Sundar ,G.S.Edwin Ebby , "Phishing Detection in Websites using Parse Tree Validation
Phishing is a technique of tricking people into giving sensitive information like usernames and passwords, credit card details, sensitive bank information, etc., by way of email spoofing, instant messaging, or using fake web sites whose look and feel gives the appearance of a legitimate website. In this work, a technique named parse tree validation is proposed to determine whether a webpage is legitimate or phishing. It is a novel approach to detect the phishing web sites by intercepting all the hyperlinks of a current page through Google API, and constructing a parse tree with the intercepted hyperlinks. This technique is implemented and tested with 1000 phishing pages and 1000 legitimate pages. The false negative rate achieved was 7.3% and the false positive rate achieved was 5.2%.

Shraddha Parekh,Dhwanil Parikh ,Srushti Kotak, " A new method for Detection of Phishing Websites: URL Detection
Phishing is an unlawful activity wherein people are misled into the wrong sites by using various fraudulent methods.
The aim of these phishing websites is to confiscate personal information or other financial details for personal benefits or misuse. As technology advances, the phishing approaches used need to get progressed and there is a dire need for better security and better mechanisms to prevent as well as detect these phishing approaches. The primary focus of this paper is to put forth a model as a solution to detect phishing websites by using the URL detection method using random Forest algorithm. There are 3 major phases such as Parsing, Heuristic Classification of data, Performance Analysis in this model and each phase makes use of a different technique or algorithm for processing of data to give better results.

Chaoyue Niu,Zhenzhe Zheng,Fan Wu, "Achieving Data Truthfulness and Privacy Preservation in Data Markets
As a significant business paradigm, many online information platforms have emerged to satisfy society's needs for person-specific data, where a service provider collects raw data from data contributors, and then offers value-added data services to data consumers. However, in the data trading layer, the data consumers face a pressing problem, i.e., how to verify whether the service provider has truthfully collected and processed data? Furthermore, the data

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 11, November 2023 ||

| DOI:10.15680/IJIRSET.2023.1211036 |

contributors are usually unwilling to reveal their sensitive personal data and real identities to the data consumers. In this paper, we propose TPDM, which efficiently integrates Truthfulness and Privacy preservation in Data Markets.

III. METHODOLOGY

Phishing message detection is truly an unpredictable and element issue including numerous components and criteria that are not stable. However, detecting phishing messages is a challenging task, as most of these techniques are not able to make an accurate decision dynamically as to whether the new message is phishing or legitimate. Proposed model will for detect phishing message based on content.

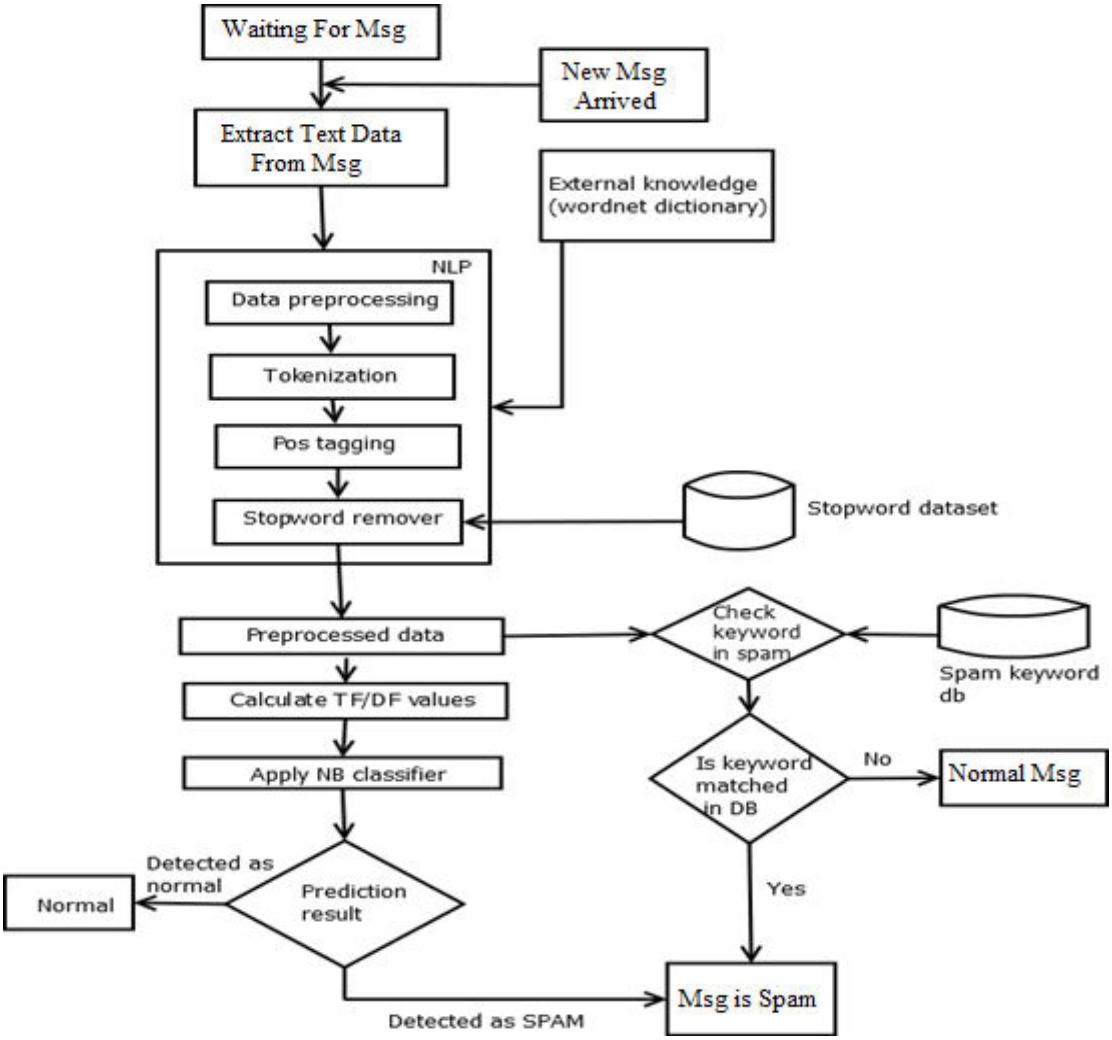

Fig1: System Architecture

Modules:

1. Feature Extraction

Data from message is pre processed , Tokenized and data feature are extracted from mail.

2. Classification

Naive Bayes classifier is applied to calculate TF/Df values

3. Prediction The message is predicted to be spam or normal..

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 11, November 2023 ||

| DOI:10.15680/IJIRSET.2023.1211036 |

IV. CONCLUSION

System will control the security of information and prevent infringements, to check whether spam is available from the current database, to enable the user to create his own spam list, and to check whether the incoming message has dangerous content. Analysis of deep learning algorithms can be done to check the best algorithm for spam email detection.

REFERENCES

[1] Sartaj S., Mollah A.F. (2021) An Intelligent System for Spam Message Detection. In: Sheth A., Sinhal A., Shrivastava A., Pandey A.K. (eds) Intelligent Systems. Algorithms for Intelligent Systems. Springer, Singapore. https://doi.org/10.1007/978-981-16-2248-937

[2] KUDUPUDI, N. and NAIR, S., Spam Message Detection Using Logistic Regression.2021

[3] Alshahrani, Ali. "Intelligent Security Schema for SMS Spam Message Based on Machine Learning Algorithms."iJIM15, no. 16 (2021): 53.

[4] Yang, Yitao, Guozi Sun, and Chengyan Qiu. "Bayesian Spam Detection Framework on Mobile Device."Recent Advances in Computer Science and Communications (Formerly: Recent Patents on Computer Science)14, no. 5 (2021): 1461-1469.

[5] Chakraborty, Amartya, Suvendu Chattaraj, Sangita Karmakar, and Shillpi Mishrra. "A Robust Approach for Effective Spam Detection Using Supervised Learning Techniques."Machine Learning Techniques and Analytics for Cloud Security(2021): 171.

[6] Abdulhamit Subasi, Esraa Molah, Fatin Almkallawi, Touseef J. Chaudhery "Intelligent Phishing Website Detection using Random Forest Classifier ".

[7] C. Emilin Shyni , Anesh D Sundar, G.S.Edwin Ebby "Phishing Detection in Websites using Parse Tree Validation ".

[8] Shraddha Parekh, Dhwanil Parikh , Srushti Kotak "A new method for Detection of Phishing Websites: URL Detection".

[9] N. Abdelhamid, A. Ayesh, F. Thabtah, "Phishing detection based associative classification data mining," Expert Systems with Applications, vol. 41(13), pp. 5948-5959, 2014.

[10] R. M. Mohammad, F. Thabtah, L. McCluskey, "Tutorial and critical analysis of phishing websites methods," Computer Science Review, vol. 17, pp. 1-24, 2015.

[11] R. M. Mohammad, F. Thabtah, L. McCluskey, "Predicting phishing websites based on self-structuring neural network," Neural Computing and Applications, vol. 25(2), pp. 443-458, 2014.

[12] M. A. U. H. Tahir, S. Asghar, A. Zafar, S. Gillani, "A Hybrid Model to Detect Phishing-Sites Using Supervised Learning Algorithms," International Conference on Computational Science and Computational Intelligence (CSCI), pp. 1126-1133, IEEE, 2016.

[13] R. M. Mohammad, F. Thabtah, L. McCluskey, "Intelligent Rule-based Phishing Websites Classification, " IET Information Security, vol. 8(3), pp. 153-160, 2014.

[14] A. Hodzic, J. Kevric, A. Karadag, "Comparison of Machine Learning Techniques in Phishing Website Classification," 2016.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.423

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

Scan to save the contact details