International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE)
ISSN: 0976-1353 Volume 31 Issue 05 – SEPTEMBER 2025.

Detecting Phishing Effectively a Stable an Adaptable Mechanism

Dr. B.P. Pradeep Kumar[#1], Bhumika D R[*2]

[#1]Professor, Department of Computer Science and Engineering, Atria Institute of Technology, Bengaluru

[*2]PG Scholar, Department of Computer Science and Engineering, Atria Institute of Technology, Bengaluru

ABSTRACT - In today's world, with so many popular messaging apps available, SMS is not as important as it used to be. Now, it is mostly used by service providers, businesses, and organizations to contact people for marketing or spam purposes. A new type of spam message is becoming common, where regional languages are written using English letters. This makes it harder to detect and block such messages spam messages are used. A classifier is trained with this expanded collection of messages to tell the difference between spam and non-spam messages. The performance of the classifier is evaluated using a measure called. Regional languages like Hindi or Bengali are sometimes typed in English letters by local mobile users. They use. The Monte Carlo method is applied for supervised learning and classification. This method uses a set of features and machine learning algorithms that are commonly used in research.

***Index Terms*— Monte Carlo**

INTRODUCTION

In recent years, the way people communicate has changed greatly due to the rise of popular messaging applications. As a result, Short Message Service (SMS) is no longer the main form of personal communication. Instead, it is now mainly used by service providers, businesses, and organizations to reach out to the public for purposes such as marketing, announcements, or, in many cases, sending spam messages. A growing trend in spam is the use of regional languages written in English letters, for example, Hindi or Bengali typed using the English alphabet. This style, often referred to as ****, makes it difficult for spam filters to identify and block unwanted messages because traditional text-processing techniques are not designed for such mixed-language formats. To address this issue, spam messages is used to train a classifier capable of separating spam from non-spam messages. The training process is carried out using an extended dataset and evaluated with the help of a measure called**.** For learning and classification, the Monte Carlo method is applied in a supervised manner.

LITERATURE SURVEY

The Phishing is a sort of social designing assault regularly used to take client information, including login accreditations and credit card numbers. With the enhancements in internet technology, websites are the major resource for the cyber-attacks. There are several counter measures available for avoiding phishing attacks, but phishers are changing their attacking methods from time to time. One of the most widely used techniques for solving cybersecurity issues is machine learning. From last several years, Machine Learning and Deep Learning Techniques are suitable for solving security related issues. Machine Learning is most suitable for detecting phishing attacks because most of the phishing attacks have common characteristics. This paper has applied several machine learning techniques for detecting the phishing attacks. Here, two priority-based algorithms are proposed. Based on the results of these algorithms, the final fusion classifier is decided. We used a dataset from UCI and applied a novel fusion classifier and achieved an accuracy of 97%. We used Python for implementing our experiments. Such SMS can contain updates regarding your bank account

or information related to your travel tickets, etc. Phishing is the major problem of the internet era. In this era of internet, the security of our data in web is gaining an increasing importance. Phishing is one of the most harmful ways to unknowingly access the credential information like username, password or account number from the users. Users are not aware of this type of attack and later they will also become a part of the phishing attacks. It may be the losses of financial found, personal information, reputation of brand name or trust of brand. So, the detection of phishing site is necessary. In this paper we design a framework of phishing detection using URL.

I. Existing System

The existing systems for detecting phishing and spam messages primarily rely on traditional filtering mechanisms and rule-based approaches. These include blacklisting suspicious URLs, keyword-based filtering, and simple machine learning classifiers trained on standard English text datasets. While such methods can handle basic spam detection, they are not well equipped to manage the complexity of modern phishing attacks. Many of these systems fail to account for regional language variations, especially when local languages like Hindi or Bengali are written using English characters. This creates a significant gap in accurately identifying harmful content, as conventional filters cannot effectively capture these transliterated patterns.

Disadvantage of existing system

The major disadvantage of the existing systems is their limited adaptability to evolving spam and phishing strategies. Phishers frequently alter their methods, making static, rule-based systems ineffective over time. Furthermore, traditional spam filters are often biased towards English-only datasets, resulting in poor detection rates for mixed-language or regionally transliterated messages. Another drawback is the lack of robust evaluation techniques; many models are trained and tested only once, which reduces statistical reliability. These limitations lead to higher false negatives, where harmful phishing content goes undetected, and false positives, where legitimate messages are wrongly flagged as spam.

II. Proposed System

The proposed system introduces a Monte Carlo–driven supervised learning framework combined with advanced machine learning and deep learning techniques to improve spam and phishing detection. A diverse dataset is curated, including English messages and regional languages written in English alphabets, which reflects real-world SMS usage in India. Preprocessing is applied to clean and standardize text, followed by TF-IDF feature extraction. Monte Carlo sampling ensures repeated randomized splits of the dataset, providing more reliable performance estimates. Multiple classifiers such as SVM, Naïve Bayes, Random Forest, kNN, and CNN are trained and compared. Deep learning, specifically CNN, is emphasized to capture complex text patterns in transliterated messages, making the system both stable and adaptable to changing phishing strategies.

Advantages

The proposed system offers several advantages over existing approaches. First, it enhances detection accuracy for multilingual and transliterated spam messages, a feature highly relevant to the Indian context. Second, by applying Monte Carlo sampling with repeated k-fold cross-validation, the system achieves statistically sound evaluation, minimizing bias and variance in results. Third, the inclusion of CNN deep learning models allows the system to learn intricate features that traditional classifiers may overlook. This makes the mechanism more adaptable to evolving phishing techniques. Finally, the approach is scalable, ensuring that as the dataset grows with new types of spam, the model can maintain high performance without requiring complete retraining from scratch.

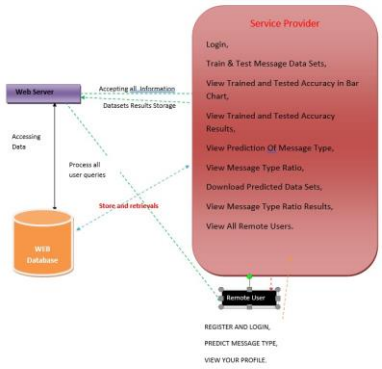


Fig: Architecture Digram

III. Implementation

The system is implemented in several phases. First, a dataset of spam and ham messages is collected, including English and regional language messages typed in Roman script. The text undergoes preprocessing steps such as stop word removal, punctuation elimination, case normalization, and tokenization. TF-IDF is used for feature extraction. Next, Monte Carlo sampling is performed, randomly splitting the dataset into training and testing sets multiple times to ensure robust evaluation. Several machine learning classifiers are trained, including SVM, Multinomial Naïve Bayes, Random Forest, and kNN. In addition, a CNN deep learning model is implemented to capture sequential and contextual patterns. The models are evaluated using metrics such as accuracy,

precision, recall, and F1-score. The best-performing classifier is then chosen for real-world deployment.

1. Dataset Collection and Preparation

The first step involved collecting a diverse dataset containing both spam and ham (legitimate) messages. To make the system more relevant for Indian users, the dataset was enriched with not only English-language SMS but also regional languages like Hindi and Bengali written in Romanized form. This ensured that the system could handle real-world scenarios where users frequently mix local and English scripts. The dataset was further expanded with publicly available repositories, synthetic message generation, and real mobile communication logs (where permissible), thereby covering a wide range of phishing patterns, banking scams, promotional spam, and legitimate notifications.

2. Data Preprocessing

Once the dataset was curated, rigorous data preprocessing was applied. Text messages were cleaned by removing special symbols**,** URLs, HTML tags, and punctuation marks, followed by stop-word elimination to retain only meaningful words. All text was converted to lowercase to ensure uniformity. Tokenization was applied to break down sentences into individual words or tokens, and advanced normalization techniques were employed to handle transliterated words (e.g., "paisa" vs. "paise" vs. "paysaa"). To extract numerical features, the TF-IDF (Term Frequency–Inverse Document Frequency) method was used, which captures the relative importance of words across the dataset. This step was critical in distinguishing spam-related keywords from regular conversational terms.

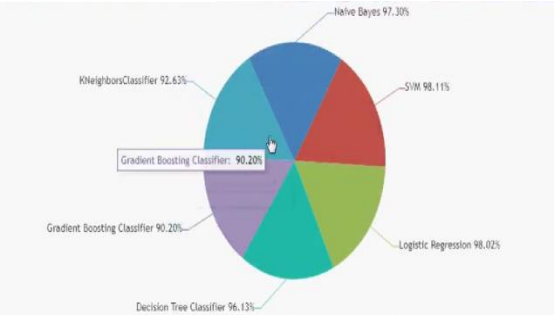
3. Monte Carlo Sampling and Data Splitting
To ensure that the results **were** statistically reliable and not biased by a single split**, a** Monte Carlo sampling strategy was used. The dataset was randomly split into training and testing subsets multiple times, and models were trained and tested repeatedly on these variations. By repeating the process up to **100 times**, the system reduced variance in performance estimation and simulated real-world conditions where spam distributions may shift over time.

4. Model Training with Machine Learning Classifiers
A suite of traditional machine learning algorithms was implemented, including Support Vector Machines (SVM), Multinomial Naïve Bayes (MNB), Random Forest, and k-Nearest Neighbors (kNN)**.** Each classifier was trained on the TF-IDF feature set using k-fold cross-validation (k=100) to assess performance consistency. These classifiers provided baseline comparisons and highlighted the strengths and weaknesses of shallow learning techniques in handling multilingual spam.

5. Deep Learning Integration with CNN
Recognizing the limitations of traditional ML models, a **Convolutional Neural Network (CNN)** was designed for text classification. The CNN architecture included embedding layers for word representation, followed by convolutional and pooling layers to capture sequential and contextual patterns within the messages. This deep learning model was particularly effective in identifying complex transliterated spam, where word combinations and contextual cues play a significant role. To prevent overfitting, regularization techniques such as dropout and batch normalization were incorporated.

IV. RESULT

The experimental results confirm the effectiveness of the proposed system. Using Monte Carlo sampling with 100-fold cross-validation**,** the models achieved consistently high accuracy across repeated tests. Among the tested algorithms, the CNN-based deep learning model outperformed traditional ML classifiers, demonstrating superior precision, recall, and F1-scores. This indicates its ability to better capture nuanced patterns in transliterated spam messages. The results highlight the system's robustness, adaptability, and stability in detecting phishing attempts. Graphical evaluations, including bar charts and pie charts, further confirmed the superior performance of the CNN model compared to SVM, Naïve Bayes, Random Forest, and kNN. Overall, the findings establish that the proposed approach offers a reliable and scalable solution for real-world phishing and spam detection.



Graphical representations, including **bar charts** and **pie charts**, provided clear evidence of comparative performance across all models. The CNN consistently led with the highest values across evaluation metrics, while the traditional classifiers showed moderate but less reliable results. These visual insights reinforced the claim that deep learning, when combined with Monte Carlo validation, offers a more robust framework for phishing detection in multilingual environments. Furthermore, the system proved to be adaptable and scalable. The use of Monte Carlo simulations ensured that the classifier remained effective under varying message distributions, simulating real-world conditions where spam characteristics may shift over time. The experiments confirmed that the CNN classifier maintained high detection rates even as the dataset included newer forms

of spam, such as transliterated Hindi or Bengali messages mixed with English text. This adaptability indicates that the system can evolve with changing phishing strategies, making it highly suitable for long-term deployment in real-world mobile communication networks.

Overall, the results validate the proposed mechanism as a stable, accurate, and future-proof solution **f**or phishing detection. The combination of large-scale dataset handling, repeated statistical testing, and deep learning–based feature learning establishes the system as a significant improvement over existing spam filters, with direct applications in enhancing cybersecurity for mobile and internet users.

V. REFERENCES

[1]. Lakshmanarao, A., Rae, P.S.P., Krishna, M.M.B. (2021) 'Phishing website detection using novel machine learning fusion approach', in 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), Presented at the 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), 1164–1169

[2]. H. Chapala, R. Kodak and M. Joiner, "A Machine Learning Approach for URL Based Web Phishing Using Fuzzy Logic as Classifier", 2019 International Conference on Communication and Electronics Systems (ICCES), pp. 383-388, 2019, July

[3]. Vaishnavi, D., Sabetha, S., Jingle, Y.B., Submachine, R., Shyly, S.P. (2021) 'A Comparative Analysis of Machine Learning Algorithms on Malicious URLPrediction', in20215thInternationalConferenceonIntelligentComputingandControl Systems (ICICCS), Presented at the 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), 1398–1402

[4]. Internal Revenue Service, IRS E-mail Schemes. Available at https:// www.irs.gov/ uac/newsroom/consumers-warnedof-new-surge-in-irs-email-schem
es-during-2016-tax-season-tax-industry-also-targeted.

[5]. Abu-Nimes, S., Napa, D., Wang, X., Nair, S. (2007), A comparison of machine learning techniques for phishing detection.

[6]. E., B., K., T. (2015). Phishing URL Detection: A Machine Learning and Web Mining-based Approach. International Journal of Computer Applications, 123(13), 46-50.doi:10.5120/ijca2015905665.