IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

Spam Classification Using Machine Learning: A Survey

[1]Wasim Yasin, [2]N Govind Prasad, [3]Jnanashree T R, [4]Vibha Datta
[1]Assistant Professor, [2]UG Student, [3]UG Student, [4]UG Student
[1]Department of Computer Science & Design
[1] K S Institute of Technology, Bengaluru, India

Abstract: In this Generation of emails, messages spam continues to pose several challenges to email ecosystem. Spam detection in emails have been a concern because the user security depends on the classification of emails as spam or ham. The existing methods for spam detection lack in precision and is a time consuming process. This paper provides a spam detecting model that accounts for the dynamic nature of spam mails and learning based clustering techniques for classifying spam and ham messages. The model contains various Machine Learning (ML) algorithms used for detection and classification of spam emails. The model is integrated with Artificial Intelligence (AI) for automatic detection of spam or ham messages, which is most advanced form of detecting spam compared to other methods. The model present a novel approach to detect spam using Random forest (RM) classifier which is further enhanced by the designed methodology. The model claims the effective methodology with robust and interpretable features for detecting the spam messages.

Index Terms - **Deep Learning . Email Spam Detection . Machine Learning**

1.INTRODUCTION

Email is the fastest and most cost-effective way to share information . It is widely used tool globally .It is Spam means unwanted or inappropriate emails sent to users , which can contain malware like viruses. Spam also has the ability to impact computers system security and integrity. The spam detection was first introduced in early 1990s , since then spam detection remains as one of the challenge. A study says where in 2020 , out of 300 billion emails daily,170 billion emails were identified as spam emails. But this has gradually decreased in recent years.,170 billion emails were identified as spam emails. But this has gradually decreased in recent years. Spam detection methods are done through text classification, word frequency -based techniques. Successful deployment to production environments becomes challenging.

Pattern recognition classification models generally take the word frequency of email content as input. Highly robust word frequency-based classification methods, such as Naïve Bayes, Support Vector Machines, and later Neural Networks , have been proposed in several studies in the literature .To improve the detection accuracy researchers have shifted to machine learning methods like SVMs , decision trees, Naïve Bytes and neural networks .These algorithms are used to classify the given message as spam or not spam(ham). Spam detection can be done in various methods even though none of them will give 100% accurate results. Detecting spam message in email systems is challenging Natural Language Processing (NLP) [1] task because spam messages can be often short , misleading text, making it difficult to classy them. NLP is a combination of computational statistics, machine learning, deep learning , and even used in detecting emotions in text. NLP is also used in speech recognition and real time data processing .

In today’s networked age, email has turned into a critical communication medium in every sphere of life, i.e., personal, professional, corporate, and government. Ironically, the usage of email has also led to a rise in spam emails.

To stop spam messages reaching the users spam detection and spam filtration techniques are necessary. Even though there are methods to do this even more improvements are needed.

Text categorization and word frequency-based methods are the foundation of spam detection, however implementing these models in practical settings is difficult. Word frequency is analyzed by conventional classification techniques like Naive Bayes, Support Vector Machines (SVMs), and Neural Networks to detect spam.

One key issue in spam filtering is the nature of spam communications, which are generally brief, deceptive, and designed to elude detection. Because of this, spam detection is a challenging Natural Language Processing (NLP) task. NLP analyzes human language, including speech recognition and sentiment detection, by combining computational statistics, machine learning, and deep learning. NLP is also essential for processing data in real time, which improves spam detection systems even more. Text categorization and word frequency-based methods are the foundation of spam detection, however implementing these models in practical settings is difficult. The deceptive and succinct style of spam communications, which are purposefully written to evade detection systems, presents a significant difficulty in spam filtering. Because of this, spam identification in Natural Language Processing (NLP) is a challenging task. By using computational statistics, machine learning, and deep learning to examine human language, including speech recognition and sentiment analysis, natural language processing (NLP) improves spam identification. Furthermore, real-time data processing capabilities make spam filtering systems much more robust.

Word frequency-based methods and text classification are the mainstays of spam detection. However, because spammers are often changing their tactics to avoid detection, it is difficult to successfully implement these models in practical applications.

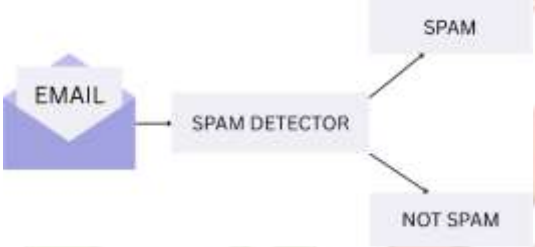

Fig. 1: Spam classification by a spam detector

The figure above shows a flowchart of how the spam emails are detected by a spam detector.

This model is all about creating facial animations through the control of one image by [5] another. Rather than requiring extensive video training, FOMM can animate a single static image by employing motion data from another video. This technique is commonly used to animate portraits and generate realistic talking faces with minimal input data.

Related Study:

The classification of email spam has drawn a lot of attention in the literature. Only a small subset of methods, nevertheless, have been widely used. Important email Providers like Yahoo and Gmail can swiftly identify and notify users of spam emails thanks to their extensive data centre infrastructure, which powers their spam detection techniques . NLP [1]is also used in speech recognition and real time data processing.Within the same email provider, thousands of users can be swiftly targeted by a spam campaign. Finding similarity patterns in email content is essential to spotting these kinds of campaigns. Large email providers, however, are unable to detect similarities by comparing every email to every other email. As a result, spam detection is essential for big email gateways.

It is simpler to recognize other emails after they have been marked as spam. Some methods, called knowledge engineering techniques, are based on keyword filters (which can utilize regular expressions), rules, and blacklists of known sources of spam.[2] TabNet, a deep learning model, can be used for spam detection by leveraging its ability to learn from tabular data and its attention mechanism to identify important features This technique has been used in the literature. But it is easy for spammers to bypass such methods, and such lists are harder and harder to maintain over time as the number of rules to be created and updated increases due to spam proliferation. Other methods, like grey listing and SPF try to classify spam based on the behaviour of the spammer. A recent paper illustrates how this happens.

Grey listing, for instance, operates by rejecting the first delivery attempt of an email message to the end user. It expects that a valid mail transfer agent (MTA) will retry the delivery.[3]Optimized approach on the second delivery attempt, the message is accepted. This behaviour normally 2 indicates legitimate senders and not spammers . Still grey listing might not be effective if the spammers resent messages quickly enough within the designated time limit for the second try. Besides, if a valid e-mail provider is misconfigured or fails to have a resend mechanism , it can generate false positives. SPF demands that all IP addresses that are in charge of sending e-mail on behalf of a specific domain be included in the DNS records[4]A hybrid machine learning algorithm for spam detection combines the strengths of multiple algorithms, like Naive Bayes and Support Vector Machines (SVM), to achieve higher accuracy and robustness in identifying spam messages When an e mail is received, the MTA at the destination will verify the mail exchange record in the DNS registries to determine whether the IP of the sender corresponds with the domain connected with the e-mail address. SPF may work in identifying phishing e-mails with spoofed legitimate domains. [7]Existing methods for spam identification suffer from time-consuming processes and lack precision. To tackle these limitations, this study introduces the Octave Convolutional Multi-Head Capsule Nutcracker Network (OCMCNN).

Yet, it is likely to fail when phishers trick users by employing similar domain names (e.g., playpal.com). Furthermore, not all authentic email servers publish SPF records in the DNS. Advanced attackers might have a registered MX record that evades SPF checks .Even with its limitations in more complex scenarios, SPF checking remains prevalent due to its success in simple cases.

Besides SPF, there are other commonly employed methods based on email authentication. One of them is DomainKeys Identified Mail (DKIM) . This employs public key infrastructure to implement an email authentication mechanism. In this process, the source of the email uses cryptographic signatures on its messages, and the destination checks the message's authenticity with validation keys that are saved in the DNS.

Sender Policy Framework (SPF) has shortcomings when it comes to complex phishing attempts, even though it works well in simple situations. Using domain names that closely resemble authentic ones (such as "playpal.com" instead of "paypal.com") allows attackers to trick victims. Furthermore, not all trustworthy email servers post SPF records in the DNS, and in order to get around SPF verification, skilled attackers could register their own Mail Exchange (MX) entries.

Other techniques, such as DomainKeys Identified Mail (DKIM), are also employed to improve email authentication. DKIM uses public key infrastructure to confirm the legitimacy of emails. [6]Emails are cryptographically signed by the sender and validated by the recipient's server using public keys that are kept in the DNS. This helps stop email spoofing and guarantees the message's integrity.

[5]Another method that has been developed more recently is Domain-Based Message Authentication (DMARC). This process employs SPF and DKIM to authenticate email . The sending domain is able to set its authentication policy and the way the destination domain will treat messages that disavow it. The destination domain is also able to report the results back to the original domain . Email authentication methods are a potential solution with a good prospect to fight spam. Still, today there is no accepted standard for authentication. Moreover, the SMTP protocol enables email to be transmitted without authentication, which leads to spam propagation. [8]Consequently, on premise commercial tools mostly include one or more of the above authentication techniques to improve spam detection and prevention efforts. In recent research activities focused on enhancing the effectiveness of spam filters used for email categorization, various machine-learning methods have come into the limelight. Importantly, researchers like Temidayo et al., employed the Enron1 dataset for setting up ground-level models using the extreme gradient boost ensemble and random forest algorithms, providing sophisticated abilities in recognizing and classifying spam emails.

[11]In order to improve email security, the recipient domain can also provide feedback to the sender regarding authentication results. Although email authentication methods offer promising solutions to combat spam, there is currently no widely accepted standard. Additionally, the SMTP protocol permits emails to be sent without authentication, which makes it easier for spammers to exploit the system. [10]As a result, commercial on-premise email security tools frequently incorporate one or more authentication techniques to improve spam detection and prevention. Recent research has focused on improving spam filters through advanced machine learning techniques. Studies have used datasets like Enron1 to develop robust spam classification models, and techniques like random forest algorithms and extreme gradient boosting have proven to be highly effective in correctly identifying and classifying spam emails.

Outcomes of their empirical study highlighted the effectiveness of hyperparameter tuning, training both the RF and XG Boost models to a remarkable 97.78% accuracy. Simultaneously ,Naeem Ahmed et al., performed a far reaching survey including machine learning techniques employed in spam filtering on email and IoT platforms. This thorough review included a careful comparison of various performance evaluation metrics. Venturing into the field of text- and voice-enabled emails, Safaa S. I. Ismail, presented a [4]new hybrid system for data processing, a Genetic decision tree with natural language processing GDTPNLP. Not only did this new method improve text extraction speed but also displayed increased performance, effectiveness, and accuracy in spam filtering. cost Additionally, Qinglin Qi , presented the Markov based phishing ensemble detection FMPED technique, supported by its ensemble detection method. These new under-sampling algorithms leveraged ensemble learning techniques, carefully separating benign emails from overlapping areas, and then discretionary under-sampling of the remaining benign emails. [9]The resulting combination of benign and phishing emails created a new training dataset, showing a promising approach in phish detection methods. In recent studies focused on enhancing the effectiveness of spam filters for email classification, several machine-learning strategies have gained importance. Additionally, a number of parameter tuning methods for optimal feature selection and extraction exist. These hybrid methods provide important results for the classification model.

Methodology:

This section describes the methodical process used to create a very successful spam detection system based on deep learning. To increase precision and effectiveness, the system combines feature extraction, selection, and classification algorithms.

1. Information Gathering

Three well-known public datasets are utilized by the system: Dataset of Ling Spam Dataset for Spam Assassin Dataset of Enron Spam The combination of spam and valid emails in these datasets ensures linguistic diversity and improves the model's capacity for generalization.

2. Preprocessing Data

The following preprocessing procedures are used to improve classification accuracy and enhance the text data: Tokenization is the process of breaking up text into discrete words or tokens. Eliminating common terms that don't aid in classification is known as "stop-word removal." Filtering: Eliminating extraneous symbols, punctuation, and special characters. Lemmatization and stemming are the processes of reducing words to their most basic form for uniformity. challenges fall into three general categories: technical challenges, ethical challenges, and legal challenges.

3. Feature Extraction

Feature extraction is performed using the ADF-CTF (Advanced Deep Feature-Contextual Term Frequency) method, which ensures that only the most meaningful features are extracted from the text for classification.

4. Selection of Features

A feature selection strategy is used to minimize dimensionality and maximize efficiency. The retrieved features are improved by the suggested OKOA (Oppositional-Based Kepler Optimization Algorithm) by utilizing: Exploration of search spaces is improved by oppositional-based learning (OBL). The Kepler Optimization Algorithm (KOA) increases the precision of feature selection. Only the most important traits are kept in this stage, guaranteeing effective classification.

5. Classification with OCMCNN

The selected features are fed into the OCMCNN (Optimized Convolutional Multi-Class Neural Network) for classification. To further improve performance, the model is optimized using the NCO (Newly Created Optimization) minimizing classification errors

6. Training and Assessing the Model approach

Training and testing sets are separated from the dataset to ensure a fair assessment. Using deep learning techniques, the classification model is trained, and important metrics are used to evaluate its performance, such Accurate Keep in mind as: Correctness

7. Decision Insights & Model Explainability

SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) are utilized to guarantee transparency in spam classification. These methods aid in the analysis of word frequency patterns and the evaluation of the impact of particular attributes on categorization results. A high-performance spam detection system with improved robustness, accuracy, and interpretability is the outcome of this methodology.

Conclusion:

Though with high prospects, machine-learning methods for spam detection are seldom used in production environments. This paper fully explains and assesses the challenges hindering such methods from deploying in real-world environments. We suggest and assess a new, robust model for spam detection to meet these challenges. Our model seeks to enhance detection reliability without heavily affecting detection throughput using shallow and deep classifiers. Our analyses indicate that the model suggested can enhance detection accuracy over time, identify model obsolescence, and have a high detection throughput. Our analysis shows that this approach improves detection accuracy over time, effectively identifies model obsolescence, and sustains high detection efficiency.

This model aims to bridge the gap between research advancements and practical implementation in spam filtering systems by addressing the limitations of existing methods. Despite their strong potential, machine-learning methods for spam detection are rarely implemented in real-world production environments. This paper explores the key challenges preventing their deployment and proposes a robust spam detection model to address these issues. Looking forward, the future research motivated by our findings is ambitious and required.

We hope to investigate the flexibility of our model to a wider set of datasets, including those that reflect the diversity of worldwide email traffic and the continuing sophistication of spamming methods. The quest for further XAI methods and the investigation of ethical approaches to AI in spam detection reflect our commitment to developing the field in a way that is both effective and ethical. In addition, responding to the methodological shortcomings noted in our research, future work will explore other data balancing approaches and the potential of ensemble models to improve both performance and understandability.

References:

1. Reddy, K.R. and Joshi, G., 2024, December. Innovative Development of Sophisticated Text Mining Architectures for Precision Spam Detection Leveraging NLP Techniques, Neural Networks, and Ensemble Classifiers. In 2024 International Conference on Emerging Research in Computational Science 6). IEEE. (ICERCS) (pp. 1 -6)

2. Naseer, M., Ullah, F., Saeed, S., Algarni, F. and Zhao, Y., 2025. Explainable TabNet ensemble model for identification of obfuscated URLs with features selection to ensure secure web browsing. Scientific Reports, 15(1), p.9496.

3. Fatima, R., Fareed, M.M.S., Ullah, S., Ahmad, G. and Mahmood, S., 2024. An Optimized Approach for Detection and Classification of Spam Email's Using Ensemble Methods. Wireless Personal Communications, pp.1-27.

4. Oluchukwu, U.W., Sylvanus, O.A., Asogwa, D., Chinedu, E., Chibuogu, A. and Sylvanus, A.K., 2024. Hybrid machine learning algorithms for email and malware spam filtering: A review. Eur. J. Theor. Appl. Sci, 2, pp.76-86.

5.Shawly, T., Alsheikhy, A.A., Said, Y., Shaaban, S.M., Lahza, H., AbuEid, A.I. and Alzahrani, A., 2025. DaC-GANSAEBF: Divide and Conquer Generative Adversarial Network—Squeeze and Excitation-Based Framework for Spam Email Identification. Computer Modeling in Engineering & Sciences (CMES), 142(3).

6.Truong, C.K., Hao Do, P. and Duc Le, T., 2023. A comparative analysis of email phishing detection methods: a deep learning perspective.

7. Ratmele, A., Dhanare, R. and Parte, S., 2025. Octave convolutional multi-head capsule nutcracker network with oppositional Kepler algorithm based spam email detection. Wireless Networks, 31(2), pp.1625-1644.

8. Rashed, A., Abdulazeem, Y., Farrag, T.A., Bamaqa, A., Almaliki, M., Badawy, M. and Elhosseini, M.A., 2025. Toward Inclusive Smart Cities: Sound-Based Vehicle Diagnostics, Emergency Signal Recognition, and Beyond. Machines, 13(4), p.258.

9. Crawford, M., Khoshgoftaar, T.M., Prusa, J.D., Richter, A.N. and Al Najada, H., 2015. Survey of review spam detection using machine learning techniques. Journal of Big Data, 2, pp.1-24.

10. Kumar, S., Kar, A.K. and Ilavarasan, P.V., 2021. Applications of text mining in services management: A systematic literature review. International Journal of Information Management Data Insights, 1(1), p.100008.

11. Thudumu, S., Branch, P., Jin, J. and Singh, J., 2020. A comprehensive survey of anomaly detection techniques for high dimensional big data. Journal of big data, 7, pp.1-30.