International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Improved Email Spam Detection Using Machine Learning Techniques

[1]Ram Shankar, [2]Santosh Nagar AP, [3]Anurag Shrivastav AP

[1]MTech Research Scholar, NIRT, ram903536@gmail.com
[2]CSE, NIRT, santoshnagar9@gmail.com
[3]CSE, NIRT, anurag.shri08@gmail.com

ABSTRACT:

E-mail is one of the quickest and most professional ways to send messages from one location to another around the world; however, increased use of e-mail has increased to received messages in the mailbox, where the recipient receives a large number of messages, some of which cause significant and varied problems, such as the theft of the recipient's identity, the loss of vital information, and network damage. These communications are so harmful that the user has no way of avoiding them, especially when they come in a variety of forms, such as adverts and other types of messages. Spam is the term for these emails Filtering is used to delete these spam communications and prevent them from being viewed. This research intends to improve e-mail spam filtering by proposing a single objective evaluation algorithm issue that uses Deep Learning, based classifiers to build the optimal model for accurately categorizing e-mail messages. Text cleaning and feature selection are used as the initial stage in the modeling process to minimize the dimension of sparse text features obtained from spam and ham communications. The feature selection is used to choose the best features, and the third stage is to identify spam using a, Support Vector Machine and Long-Short Term Memory classifier.

Keywords: Cyber security, Spam, Natural Language Processing, Machine learning

1. Introduction:

Spam is an unwanted and unsolicited electronic communication delivered by a sender who has no established relationship with the recipient. Electronic spam is divided into numerous categories: e-mail, SMS, social media, and online commerce platforms. Spam takes users' time because they have to identify and delete unwanted communications; it also gobbles up inbox capacity and buries valuable personal e-mails. On the other hand, SMS spam is often sent through a mobile network. Due to a large number of spammers and the possible detrimental consequences of social network spam on the convenience and comprehension of all users, social network spam has recently gained greater attention from both researchers and practitioners. As a result, spam filtering is given a lot of thought in the above communication routes. Spam communications can be manually or automatically screened. Manual spam filtering, which involves recognizing spam messages and eliminating them, is time-consuming. Furthermore, spam communications may contain a security risk, such as links to phishing websites or malware-hosting servers. As a result, researchers and practitioners have worked for decades to improve automatic spam filtering systems. Machine learning algorithms are known for being very good at detecting spam emails. The basic idea behind machine learning algorithms is to create a word list and then give a weight to each word. Spammers, on the other hand, frequently incorporate common valid statements in spam messages to reduce the likelihood of being identified. Neural networks (NNs), support vector machines (SVMs), Naive Bayes (NB), and Random Forest (RF) are some of the known machine learning techniques used in spam filtering. Ensemble learning approaches such as bagging and Random Forest outperform traditional single classifiers, according to various research works. In comparison to single algorithms, ensemble approaches integrate the predictions of numerous underlying machine learning algorithms to enhance accuracy and precision. In previous studies, traditional classifiers like decision trees were used to effectively filter spam messages using ensemble approaches. Surprisingly little research has been done on Neural Networks (NNs) in ensemble learning. Recent data suggest that NNs using regularization approaches may detect spam in e-mail and SMS messages with excellent accuracy. This is due to improved optimization convergence and over fitting resistance. They also combined regularized Neural Networks with ensemble learning approaches for automatic spam filtering to make use of these properties. Rectified linear units (ReLu) and dropout regularization are utilized in deep feed forward NNs (DFFNNs) to improve the performance of the proposed technique by addressing the optimization convergence to a bad local minimum problem that is typical in classic shallow NN models. In general, spam filtering is a binary classification issue, in which each message must be classified as spam or ham. In addition to high accuracy, spam filtering algorithms should have a low false-positive rate (when a genuine communication is mistakenly classified as spam) to avoid instances where legitimate messages are not delivered to the intended recipient. Furthermore, a standard classification performance metric based on accuracy ignores the various costs associated with type I and type II mistakes. To determine the outcome of Type I and Type II mistakes, the null hypothesis is used. Type I errors reject the null hypothesis even if it is true, whereas type II errors do not reject the null hypothesis even if the alternative hypothesis is right. These are both called as False Negatives. Because the minority class (typically the class of spam messages) has less influence on accuracy compared to the majority class of valid communications, using accuracy for sometimes 3 extremely unbalanced spam datasets might lead to incorrect results. Therefore, multiple performance measures must be considered when evaluating the spam filtering algorithms. As previously said, the primary principle behind content-based machine learning models is to create a word (list) and give weight to each word or phrase (bag-of-words) or word category (part-of-speech tagging) in the list. Such characteristics, on the other hand, are sparse, making it challenging to capture semantic representations of communications. This method used word embedding derived from the CBOW (continuous bag-of-words) model to map words to vectors depending on context. As a result, global semantic information may be gathered, and the problem of sparse data can be addressed to some extent. This method was said to be more successful than traditional tagging using a bag of words or a chunk of speech [14]. This study was based on these recent results and used word embedding to extract the semantic representation of e-mails, SMS, social network communications, and online reviews. This research aims to explore machine learning models based on traditional algorithms and ensembles using a high-dimensional feature representation for spam filtering using two major open-source Spam Datasets which are Enron and Spam Assassin.

2. E-mail Spam Filtering

E-mail Spam Filtering Spammers (those who send spam messages) collect e-mail addresses from a variety of places, including websites and chatrooms, and send unwanted messages in mass. This hurts the receiver, resulting in a waste of time and money. E-mail spam, in particular, has a severe impact on the memory of the email server, CPU performance, and user time. Furthermore, spammers' deceptive techniques may cause victims to suffer significant financial losses. Even though global spam volume (as a percentage of total e-mail traffic) has decreased to roughly 55% over the last decade [20], the number of e-mail messages carrying hazardous attachments (viruses, ransomware, and other malware) has continually climbed [20]. China is the country with the greatest e-mail spam, accounting for roughly 20% of all e-mail spam. Spam senders are compelled to send spam that avoids spam filters to maximize income. As a result, spam filtering is a difficult operation since spammers employ various strategies to reduce spam detection rate. To get through spam filters, you can use a variety of techniques, such as utilizing irrelevant, odd, or misspelled phrases.

3. Literature Survey

Authors [1] survey is unique in the sense that it relates works to their openly available tools and resources. The analysis of the presented works revealed that not much work had been performed on phishing email detection using NLP techniques. Therefore, many open issues are associated with this phishing email detection. An evolving research area is illustrated by the phishing email detection The outcomes shown that further work is required to employ modernized DL techniques in phishing email detection studies, for instance, Recurrent Neural Networks (RNNs), Convolutional neural networks (CNN), and Deep Reinforcement Learning models. The tools and resources are not sufficient in this research area. Hence, the researchers are in dire need to perform more research efforts to assess DL techniques in the phishing email detection domain.

The Authors [2] proposed a machine learning model which will detect spam mail and nonspam emails, and also this system will optimize the data by removing the unwanted mails which contain the advertisement mails and also some useless emails and also some fraud mails. This proposed system will detect the spam mails and ham emails with the dataset consisting of spam mails and after identifying spam mails this system will remove that spam emails and this proposed system will calculate the amount of storage before and after the removal of spam mails.

The Authors [3] proposed a model to solve the issue of classifying messages as spam or ham by experimenting and analyzing the relative strengths of several machine learning algorithms such as K-Nearest Neighbors (KNN), Decision Tree Classifier, Random Forest Classifier, Logistic Regression, SGD Classifier, Multinomial Naive Bayes(NB), Support Vector Machine(SVM) to have a logical comparison of the performance measures of the methods we utilized in this research. The algorithm we proposed achieved an average accuracy of 98.49% with SVM model on ‘SMS Spam Collection’ dataset.

The aim of this work [10] is to reduce the amount of spam using a classifier to detect it. The most accurate spam classification can be achieved using machine learning methods. A natural language processing approach was chosen to analyze the text of an email in order to detect spam. For comparison, the following machine learning algorithms were selected: Naive Bayes, K-Nearest Neighbors, SVM, Logistic regression, Decision tree, Random forest. Training took place on a ready-made dataset. Logistic regression and NB give the highest level of accuracy – up to 99%. The results can be used to create a more intelligent spam detection classifier by combining algorithms or filtering methods.

4. Proposed Methods

In Proposed Model has been carried out in following different. After cleaning the text data, features are created using feature engineering techniques. intext mining, common feature selection methods involve bag of words or bow and term frequency-inverse document frequency. A feature selection, also known as a variable selection, is a method of selecting a subset of features from data. This method is often used in machine learning to address issues with high dimensionality. To simplify and summarize data representation, it selects a subset of significant characteristics and rejects redundant, irrelevant, and all noisy features. Filter models, Wrappers models, and embedding techniques are examples of feature selection approaches. For feature selection, we employed evolutionary algorithms and greedy search approaches. Enron dataset has 21 features for each employee in the datasets and Spam Assassin does not have as many features as Enron dataset. While we are comparing both the datasets, we have applied feature selection. Feature selection improves the machine learning process and increases the predictive power of machine learning algorithms by selecting the most relevant features and eliminating redundant and irrelevant features. Also it reduces over fitting, eliminates noise and improves accuracy. Classification Techniques Features from the above approaches are used to train classifiers i.e., Naive Bayes, Bayesian Network, Support Vector Machine (SVM), Genetic classifier, Random Forest, XGBoost, and Long Short-Term Memory (LSTM). Proposed work model shown in figure 4.1

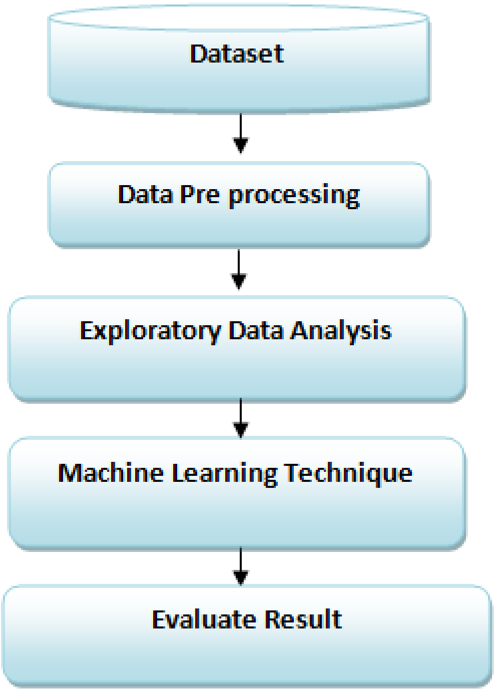

Figure 4.1 Proposed Model for Spam Detection

Support Vector Machine (SVM)

A supervised learning algorithm that offers an alternative view of logistic regression, the simplest classification algorithm, is the support vector machines or SVMs. Support vector machines try to find a model that precisely divides the classes with the same amount of margin on either side, where the support vectors are called samples on the margin. Support Vector Machine (SVM) also known as Support Vector Network in machine learning is a supervised learning technique used for classification **and** regression.

Long Short-Term Memory (LSTM)

Long Short-Term Memory, or LSTM, is a kind of recurrent neural network (RNN) that has been recognized for its ability to process short-term time series data. Furthermore, when compared to other deep neural networks (DNN), LSTM has demonstrated greater temporal series learning capacity. Long Short-Term Memory is a system that uses a cell state and a carry state to maintain information in the form of a gradient that is guaranteed not to be lost when the input sequence is stored deeper in the architecture. The current state, the carry state, and the cell state are all taken into account at each time step in the LSTM.

Dataset

Spam Assassin Dataset The Spam Assassin dataset was obtained from the public corpus website of Spam Assassin. It is made up of two data sets: training and testing. Each dataset comprises a series of emails in plain text format that has been classified as ham or spam at random. The total number of emails in the sample is 3435. The data is ham 78% of the time, and spam 22% of the time. There are 2680 spam emails and 755 ham emails in the database. The emails are in plain text format in the data. As a result, we need them to turn plain text into characteristics that might represent emails. We may then perform a machine learning algorithm on the emails using these attributes. First, a variety of pre-processing processes are 36 carried out such lower-casing all the character, removal the numbers, removal of special characters, punctuations etc., stemming, and lemmatizing.

5. Result Analysis

The Performance of SVM machine learning models and LSTM for Spam Assassin dataset with feature selection are shown in table.

Model	Accuracy	Precision
SVM	0.96	0.97
LSTM	0.97	0.98

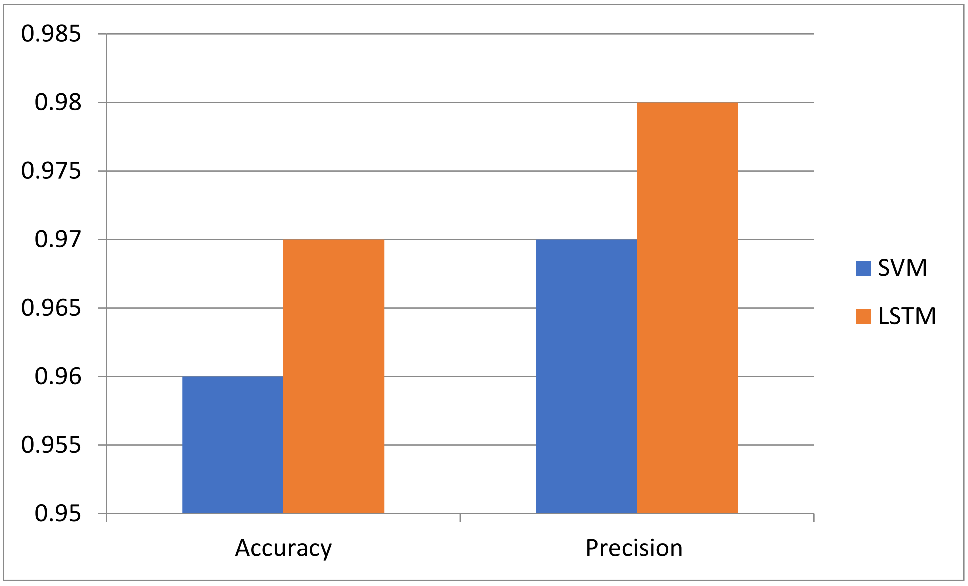

Figure 5.1 Comparison graph

6. Conclusion:

The integration of Support Vector Machines (SVM) and Long Short-Term Memory (LSTM) in email spam detection represents a significant advancement in enhancing the accuracy and efficiency of filtering unwanted messages. The combination of SVM's ability to create robust decision boundaries and LSTM's proficiency in capturing sequential patterns proves to be a formidable solution in mitigating the evolving sophistication of spam tactics. This hybrid approach capitalizes on the strengths of both algorithms, fostering a synergistic effect that outperforms traditional methods. The application of SVM ensures a reliable initial filtering process by efficiently categorizing emails based on their distinctive features, while LSTM contributes by recognizing subtle temporal dependencies within the email content. As a result, the amalgamation of these techniques not only elevates detection rates but also minimizes false positives, ultimately enhancing the overall user experience and security in the realm of electronic communication. This innovative approach stands as a promising stride towards a more resilient and adaptive defense against the persistent challenge of email spam.

Reference:

1. Isra'a AbdulNabi , Qussai Yaseen "Spam Email Detection Using Deep Learning Techniques" 1877-0509@2021, https://www.sciencedirect.com/
2. Saraswathi Morthala #1, Ms R Madhuri Devi #2 "Email Spam Classification via Machine Learning and Natural Language Processing" Vol 13 Issue 09,2022, ISSN:0377-9254 www.jespublication.com
3. BollamPragna, M.RamaBai "Spam Detection using NLP Techniques" International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8, Issue-2S11, September 2019
4. Alanazi Rayan1 , Ahmed I. Taloba1a "Detection of Email Spam using Natural Language Processing Based Random Forest Approach" : https://doi.org/10.21203/rs.3.rs-921426/v1
5. Thirumagal Dhivya S 1 , Nithya S1 , Sangavi Priya G1 , Pugazhendi E2 "Email Spam Detection and Data Optimization using NLP Techniques" International Journal of Engineering Research & Technology (IJERT) http://www.ijert.org ISSN: 2278-0181 IJERTV10IS080049, Vol. 10 Issue 08, August-2021
6. Vinodhini. M, Prithvi. D, Balaji. S "Spam Detection Framework using ML Algorithm" International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878 (Online), Volume-8 Issue-6, March 2020
7. Abishek Sharma, Arjun N "Spam Detection Using Machine Learning Techniques" International Journal of Research Publication and Reviews, Vol 4, no 7, pp 2478-2488 July 2023, www.ijrpr.com ISSN 2582-7421
8. Yaseen Khather Yaseen, Alaa Khudhair Abbas, Ahmed M. Sana "IMAGE SPAM DETECTION USING MACHINE LEARNING AND NATURAL LANGUAGE PROCESSING" ISSN: 0258-2724 DOI：10.35741/issn.0258-2724.55.2.41
9. Er. Farhana Siddiqui1 , Khan Suhail Nisar2 , Talha Atique Ansari3 , Kazi Zaki Haseeb4 "SPAM EMAIL CLASSIFICATION USING NLP" Volume 4- Issue 2, Paper 11, July 2021

10. Yuliya Kontsewayaa *, Evgeniy Antonova,b, Alexey Artamonovb "Evaluating the Effectiveness of Machine Learning Methods for Spam Detection" Procedia Computer Science 190 (2021) 479–486, t www.sciencedirect.com

11. CORMACK, G. V. Email spam filtering: a systematic review. Foundations and Trends® in Information Retrieval, 2006, vol. 1, no. 4, pp. 335–455. https://doi.org/10.1561/150 0000006

12. ZHANG, L., ZHU, J., YAO, T. An evaluation of statistical spam filtering techniques. ACM Transactions on Asian Language Information Processing, 2004, vol. 3, no. 4, pp. 243–269. DOI:10.1.1.109.7685

13. DELANY, S. J., BUCKLEY, M., GREENE, D. SMS spam filtering: methods and data. Expert Systems with Applications, 2012, vol. 39, no. 10, pp. 9899–9908. DOI: 10.1016/j.eswa.2012.02.053

14. ZHOU, B., YAO, Y., LUO, J. A three-way decision approach to email spam filtering. In:

15. Canadian Conference on Artificial Intelligence, Lecture Notes in Computer Science, vol. 6085 Springer, 2010, pp. 28–39. doi: 10.1007/978-3-642-13059-5_6