Comparative Analysis of Naive Bias and Support Vector Machine for SMS Spam Classification

Abubakarsiddik Desai
Department Electronics and Communication
KLE Technological University Dr. M. S. Sheshgiri Campus
Belagavi, India
abubakarsiddikdesai37@gmail.com

Bhaskar Kamble
Department Electronics and Communication
KLE Technological University Dr. M. S. Sheshgiri Campus
Belagavi, India
bhaskarvk8806@gmail.com

Mayuri Hegade
Department Electronics and Communication
KLE Technological University Dr. M. S. Sheshgiri Campus
Belagavi, India
mayurihegade2003@gmail.com

Rakshita Mathapati
Department Electronics and Communication
KLE Technological University Dr. M. S. Sheshgiri Campus
Belagavi, India
rakshitamathapati68@gmail.com

Gujanatti Rudrappa
Department Electronics and Communication
KLE Technological University Dr. M. S. Sheshgiri Campus
Belagavi, India
rudraguj@gmail.com
ORCID: https://orcid.org/0000-0003-2968-0698

***Abstract*—The number of mobile phone users is growing steadily each day. Short Message Service (SMS) is a text messaging application accessible on smartphones and entry-level mobile devices. Consequently, the usage of SMS has increased in recent years. Spam messages have increased in tandem with this trend. Spammers send spam messages to achieve financial or business advantages, such as promoting market growth, spreading details of lottery tickets. Thus, spam classification has gained particular attention in recent times. In the presented work, Machine Learning (ML) methods viz., Support Vector Machine (SVM) and Naive Bayes (NB) are are applied for SMS spam classification. The comparative study showed that and accuracy of 94% and 88.8% was achieved for SVM and NB respectively. The SVM outperformed NB in terms accuracy, precision, recall and F1 score with values of 0.94, 0.95, 0.93 and 0.94 respectively.**

***Index Terms*—Spam, Ham, Classification, Support Vector Machine (SVM), Naive Bayes (NB).**

I. Introduction

In today's world, everyone is connected to each other because of internet and SMS. It is a method that can be used for many purposes like including communication purpose, customer service, marketing.Individuals transfer funds and engage in numerous actions that simplify day-to-day living. However, people also encounter a multitude of unwanted messages on a daily basis, commonly referred to as spam. Spammer means a person or a company who is responsible for unwanted messages. For their organization's benefits or personal benefits. Spammers send a huge number of spam texts sent to users [1]. SMS spam filtering methods [2], this issue with sophisticated methods. Spam communications can irritate users of mobile devices. Spam communications can There are two kinds of financial or commercial benefits, such market promotion: Spam by email or SMS [3]. Email spam and SMS spam serve the same objective. The majority of the time, spammers use these communications to advertise their utilities or businesses [4]. These spam communications might occasionally cause consumers to lose money as well. The objective here is to utilize various machine learning algorithms for tackling the SMS spam classification issue, comparing their performances to derive insights, and delving deeper into the problem. Furthermore, the aim is to develop an application rooted in one of these algorithms, which is capable of filtering spam messages in SMS with a high degree of accuracy. In classification scenarios the performances of classifiers depends on the amount of accurately labeled datas [5]. There is a need in the current circumstance of SMS classifier because so many people are undergoing loses.To reduce the financial loss the study is coming with solution like SMS classifier [6]. The models discussed in this paper have been designed to categorize data from text messages as either spam or non-spam. This classification is achieved after the models undergo training utilizing datasets that were collected beforehand, employing machine learning techniques. Specifically, a some of the commonly known machine learning algorithms [7] have been used, which are NB, SVM, Logistic Regression (LR) and Decision Tree Method (DT), Random Forest (RF), K-Nearest Neighbor (KNN), Recurrent Neural Network (RNN), Convolutional Neural Networks (CNN).

An overview of the present techniques [18], difficulties,

979-8-3315-0845-6/25/$31.00 ©2025 IEEE

and possibilities for further research on mobile SMS spam detection and removal methods is given in this study. .Firstly, it gives classification methods used to detect and filter mobile SMS spam [8] secondly, it provides an in depth an examination of these methods concerning the metrics used for evaluating performance [10] Thirdly, it conducts an examination of the research datasets currently available that are pertinent to both current and prospective research efforts. This model can be used in ECU To ECU Communication [12] and bank transactions, telecom industry [15] and IOT and automation. In this paper conducts the a comparative analysis of support vector machine and Naive Bayes and focusing on their performance, accuracy and efficiency in detecting spam message. The analysis aims to provide the a comprehensive understanding of the strength and weakness of these algorithms and evaluate their applicability in SMS spam detection. It is contributes in comprehensive comparison of machine learning algorithms. In depth analysis of the of each model balancing performance and efficiency. Its include the model Accuracy, Precision,Recall and F1-Score which aligns with ongoing effort to improve the spam detection in communication systems and machine learning field.

The paper contains different sections. Section I is the introduction of the SMS spam and machine learing algorithm; Section II presents literature survey, Section III briefs about the methodology employed for classifiction of SMS, Section IV gives the Results and Disussion about comparative analysis, and Section V is the Conclusion and Future scope of the analysis.

II. Literature Survey

The usage of mobile device and SMS increasing day by day. Additionally, the number of spam messages increased. Spammers attempt to send unsolicited messages for their own gain. Therefore, particular attention is paid to spam classification. Different forms of deep learning and machine learning Methods for detecting SMS spam were used. V Dharani et.al [1] used KNN, NB ,RF,LR, are the algorithms. This proposes a machine learning model using the NB algorithm and TF-IDF vectorization to detect and classify spam SMS, achieving 95% accuracy which is less than NB algorithm. Anikait Kapoor et.al [2] used techniques like NB for feature extraction and KNN algorithm. It has accuracy of 93%. NB is suitable because both the training sample error and the test sample error fall within a reasonable range and are fairly close to one another. It has the accuracy of 97.50%. The hardware requirements specified in the paper as a limitation. E Shankar Chavali et.al [3] used the technique called NB algorithm. The used algorithm NB has the accuracy of 98.03 %. Sridevi Gadde et.al [4] are compared all the algorithms like LR, NB , DT, KNN and RF. Here LSTM (long short-term memory) technique is used , which is a RNN method with accuracy 98.5%. Tarandeep Singh et.al [5] used several machine learning classification models to identify spam SMS in a dataset, using a combination of TF-IDF and Count Vectorization characteristics, and then compared the accuracy of the different models to determine the most accurate one for spam detection. Nithisha Sharma et.al [6] evaluates and compares the accuracy of different machine learning models (LR, SVM, and Multinomial NB) for SMS spam classification, using various pre-processing methods such as stemming and TF-IDF extraction.. Aaryan Sharma et.al [7] studied various techniques like NB, RF, KNN, SVM for SMS spam classification and found that NB algorithm performs the best interms of accuracy, precision and recall. But there are some challenges such as limited message size, use of local terminology. Shikha Mundra et.al [9] compared detection of SMS spam using NB, LR , LSTM, and CNN models. It used the data cleaning and feature extraction methods like TF-IDF and evaluated, its models with accuracy, precision, recall, and F1 score. However, it did not clearly reveal which model was the best. Tarun Jain et.al [10] used various ML techniques, including SVM, to classify SMS messages as spam or not spam with highest accuracy. Highest accuracy (98.79%) is achieved with SVM. It was also successful in reaching the highest F1 Score, Precision, and Recall. Spam messages are best detected and classified by SVM.William Sigian et.al [11] studied machine learning algorithms to improve SMS spam detection. Bidirectional Encoder Representations from Transformers, is a pre-trained natural language processing (NLP) model is used and achieved 99.0166% accuracy and 99.0179% precision. P. Roy et.al [12] presents deep learning (DL) approach such as CNN to classification of SMS as spam or not-spam, achieving a highest accuracy of 99.44% on a benchmark dataset.Neelam Choudary et.al [13] studied machine learning based techniques to detect and filter spam messages. Used Random Forest algorithm and achieved 96.5% true positive rate and 1.02% false positive rate. But this paper is unclear about size of the dataset. H. Jain [16]done analysis of SMS Spam Detection using Machine Learning Model. E. Wijaya [17] studied on Spam Detection in Short Message Service (SMS) Using Naive Bayes,SVM, LSTM, and CNN, which achieved 96.59% accuracy. It points out the applicability of machine learning towards the effective filtering of SMS spam. R. B. K [19] studied the comparision of supervised machine learning to classify the given message is spam or ham.

III. Methodology

This is the section of methodology which includes the methods used for SMS spam classifier. Fig.1 is referring to describe SMS Spam Classification workflow.

A. Input

The dataset [13] utilized in this study is sourced from the UCI ML Repository. It consists of two attributes: v1, which serves as the label, and v2, which contains the sample text messages. Both attributes are of the text data type. The dataset includes a total of 5572 message samples that will be used for classification purposes. The dataset utilized in this study can presently be found on Kaggle, titled ”SMS spam collection database.” Kaggle serves as a transparent platform for public datasets, and this particular dataset was contributed

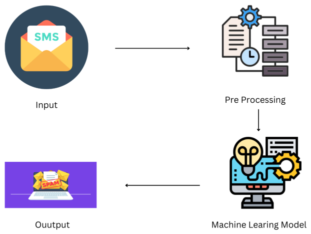

Fig. 1. SMS Spam Classification Workflow

by the UCL ML repository along with other publicly available datasets. The above dataset is imbalanced dataset. The used dataset is secondary dataset which is extracted from the UCI ML Repository. The dataset has the 5572 samples. Out of which some 4852 are Ham (not spam) messages and 749 are spam messages.

TABLE 1 depicts the dataset information.The dataset has 5572 samples:

TABLE I
DATASET DETAILS

Sl.No	Dataset attributes	No of messages
1	Spam	749
2	Ham	4852

B. Data Pre-Processing

In order to prepare the data for classifying it as spam or not spam, the following pre-processing methods are applied.

1. Text Cleaning: The first step is to clean the data [14] by removing unnecessary characters such as symbols, extra spaces, punctuations, and unwanted URLs. To standardize the data, all letters are need to convert to lowercase. Words like “the”, “is”, “or” are removed because these carry little meaning. Also simplifying the words to their root words for example cleaning to clean. This process is called stemming.

2.Tokenization After cleaning, the messages are split into smaller pieces are called tokens. This means breaking down each SMS into individual words or phrases. NLTK, SpaCy or even simple Python methods can be used for Tokenization.

3.Text Vectorization Machine learning models cannot directly work on text, the next step is to convert to tokens into numerical data. This can be done by method TF-IDF (Term Frequency – Inverse Document Frequency), which assigns the importance to words based on how unique they are in the dataset.

4.Label Encoding: Finally, the labels (”spam” and ”Not-spam”) are converted into numerical values. This is needed because ML algorithms work with only numbers, not text data. For example, ”spam” will be encoded as 1 and ”Not-spam” as 0, which gives the clear information that the model can understand during training.

C. Machine learning model:

For catching the spam and not spam messages dataset need to be split into two parts as Training and Testing sets. The Training set comprises 75% of the data, while the Testing set contains the remaining 25% of the data. SVM and NB algorithms are applied on training and testing sets of dataset and both algorithms are compared.

D. Output:

The study will helps to get the best method to classify the sms spam by comparing the algorithms like SVM and NB.

IV. RESULTS AND DISCUSSION

Technology stack used for this paper; python based Flask Platform. python module Dependencies NumPy scikit- learn scipy, sklearn and pandas Some terminologies need to be defined in order to analyze the performance of the model. For comparison python language kaggle notebook is used with library numpy, pandas and matplotlib. For simplicity, here it has been assumed that 1 represents true and 0 is false. For easy understanding, the following terminologies are used: True Positive (TP): The model correctly identifies a spam message as spam. True Negative (TN): Correctly marking a non-spam message as non-spam. False Positive (FP): The model incorrectly marks a non-spam message as spam. False Negative (FN): The model incorrectly labeled the spam message as being a non-spam message. These terms help understand and evaluate the performance of a model in identifying spam and understand how well it filters the spam and non-spam, focusing on reducing false positives and decreasing the false negatives number.

The metrics [10], taking into account all of these terms, are defined as:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \tag{1}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

$$\text{F1-score} = 2 \times \left(\frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \right) \tag{4}$$

Equations (1), (2), (3), and (4) provide the formulas for accuracy, precision, recall, and F1-score, respectively. These are the performance metrics used to assess the model’s effectiveness. SVM and Naive Bayes were employed to categorize the data , and the resulting confusion matrix is plotted Fig. 2 and Fig. 3.

The SVM model has a training accuracy of 99.49%, and a testing accuracy of 97.68%. The classification report indicates precision of 0.97, recall of 1.00, and F1-score of 0.99 for class 0 (non-spam). There are 885 samples. For class 1 (spam), the precision is 0.99, recall is 0.85, and F1-score is 0.91 with 149 samples. The model has an total accuracy of 94% and a macro mean of precision, recall, and F1-score as 0.98, 0.92, and 0.95,

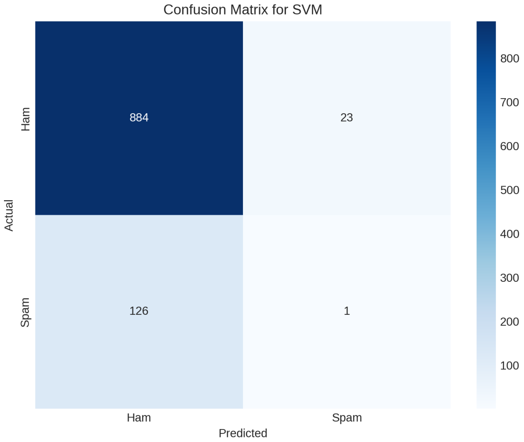

Fig. 2. Confusion Matrix Of SVM

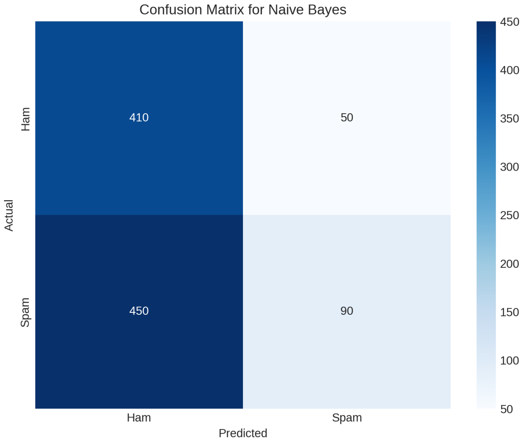

Fig. 3. Confusion Matrix Of Naive Bayes

respectively. The weighted average F1-score was 0.98 and the SVM ROC Accuracy was 0.9223.

The TABLE II which represents the calculations of accuracy, precision, recall, f1-score for NB, SVM, linear regression(LR), non-linear regression(NLR):

TABLE II
PERFORMANCE COMPARATIVE ANALYSIS

S.No	Model	Precision%	Accuracy%	Recall%	F1-score%
1	LR	52%	56%	59.8%	61%
2	NLR	64%	67%	71.8%	69%
3	NB	82%	88.8%	90%	89%
4	SVM	95%	94%	93%	94%

Here SVM and NB algorithms are compared and analyazed. This study gives the that SVM is better suitable to SMS Spam Classification than NB. This comparision is shown in Fig. 6

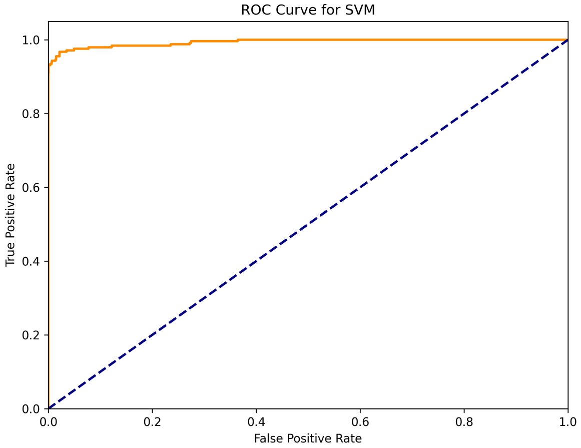

Fig. 4. ROC of SVM

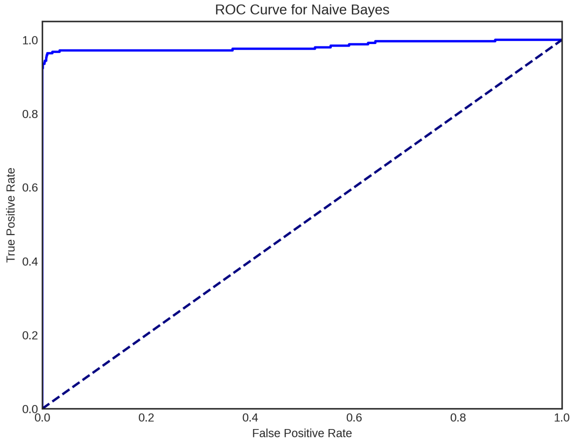

Fig. 5. ROC of NB

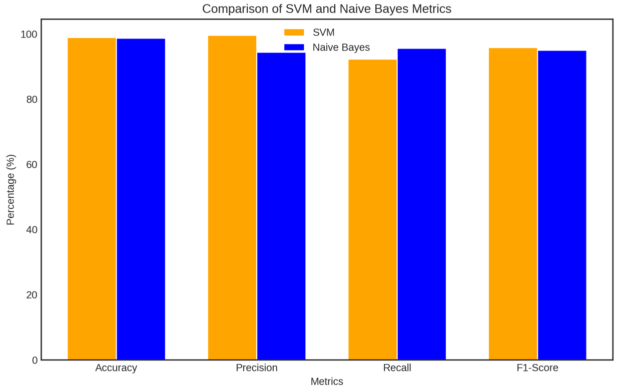

Fig. 6. Performance Comparison of NB and SVM

The Receiver Operating Characteristic (ROC) curve is a visual tool utilized to evaluate the effectiveness of a binary classification model. Fig. 4 and Fig.5 shows the ROC curves for SVM and NB respectively. It illustrates the trade-off between two key metrics: the True Positive Rate (TPR) and also the False Positive Rate (FPR), as shown in Fig. 4 and Fig. 5. The higher area under the curve means the better the model is doing, indicating that the SVM model has done a great job distinguishing spam from non-spam messages. ROC curve helps visualize how well the model balances the detection of spam with the least possible false alarms. From the Fig. 6 it It is noted that the SVM model achieves the highest accuracy compared to the other classification models. Additionally, it effectively attained the highest Precision, Recall, and F1-Score, followed by NB respectively.

The above analysis highlights the several key differences between the SVM and Naive Bayes. Where the SVM demonstrated with high performance evaluation . This analysis is better in terms of evaluation of performance metrics like accuracy, precision, recall and F1-score this metrics are critical in assessing the efficiency of spam detection. Additionally the ROC curves to visualise the trade between True positive rate and false positive rate further enhance the analysis by providing an intuitive way to compare model performance. This analysis done with collected dataset and this study may not be representative of all SMS spam classification tasks. The dataset does not fully capture the complexities of SMS spam patterns across different regions and languages, this could the affect the results. This analysis is implicated in choosing the algorithms for SMS Spam detection, contextual application and future engineering.

In conclusion, SVM effectively detect and classify spam messages, making them the most suitable method for application to the dataset compared to other approaches. The explanation for the relatively reduced accuracy could be due to the precision of NB is less compared to SVM.

V. CONCLUSION AND FUTURE SCOPE

In the NB and SVM model applied in this experiment to classify SMS spam, the later came out as the most efficient model. Not only was the accuracy in SVM higher than NB, but also more responsive to feature interaction and therefore more accurate in the classification. The graph in Fig.6 shows the comparison of the models based on the calculation data.The SVM performed much better than NB, especially with regard to their ability to classify new data. Further, the comparison justified why SVM was able to make good performance while Naive Bayes was slightly limited when handling more complex patterns. Therefore, analyzing the results obtained, it can be stated that SVM is the most accurate algorithm for SMS Spam classification providing better accuracy, flexibility and model performance.

The above discussion and experimentation have came to conclude that the ML algorithms can play an important role in classifying the SMS as spam or not. To gather all the necessary information, the research involved a detailed examination of various filtering algorithms and current anti-spam tools. The accuracy obtained in this work is 94%. Still, there remain areas that could be improve: for instance, incorporating additional filtering methods or adjusting elements of the current ones. Modifications like increasing or decreasing the count of significant terms in the message could lead to improved accuracy. The improvements in SMS spam classification utilizing SVM and NB methods sorround enhancing the accuracy of models through the addition of more comprehensive and varied datasets. This would facilitate the detection of spam across multiple languages. Furthermore, incorporating adaptive learning strategies could effectively address the issue of changing spam trends. Moreover, implementing real-time applications on messaging services could significantly improve the efficiency of spam filtering. .

REFERENCES

[1] V. Dharani, Divyashree Hegade and Mohana, ”Spam SMS (or) Email Detection and classification using ML”,2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT),https://doi.org/10.1109/ICSSIT55814.2023.10060908.

[2] Anikait Kapoor, D. Saikia, and Ishaan Dhawan, ”SMS Spam Detection Using Machine Learning Approach.”, International journal of research in science and technology(2024), https://doi.org/10.37648/ijrst.v14i01.002.

[3] E. Shankar Chavali, ”SMS SPAM DETECTION USING ML”, INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT(2023), https://doi.org/10.55041/ijsrem18832.

[4] Sridevi Gadde and A. Lakshmanarao, ”SMS Spam Detection Using ML and DL techniques” 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS).10.1109/ICACCS51430.2021.9441783.

[5] Tarandeep Singh, Tushar Anupam Kumar, P. G. Shambharkar, ”Enhancing Spam Detection on SMS performance using several ML classification models”, 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI), https://doi.org/10.1109/ICOEI53556.2022.9777157.

[6] Nitisha Sharma, ”A Methodological Study of SMS Spam Classification Using ML Algorithms” ,2022 2nd International Conference on Intelligent Technologies (CONIT),https://doi.org/10.1109/CONIT55038.2022.9848171.

[7] Aaryan Sharma, Harshit KUMAR SIMBAL, Smriti Kumari, Gautam Kumar, ”Spam SMS Classifier using ML algorithms”, International Journal For Multidisciplinary Research(2024),https://doi.org/10.36948/ijfmr.2024.v06i02.19483.

[8] Shikha Mundra, Ankit Mundra, Anshul Saigal, Punit Gupta, Josh Agarwal, M. Goyal ”ML Approaches for the Classification of Spammed Text in Messages”, https://doi.org/10.1007/978-981-16-2877-156(2021)

[9] Tarun Jain, Payal Garg, Namita Chalil, Aditya Sinha, Vivek Kumar Verma and Rishi Gupta, ”SMS Spam Classification Using Machine Learning Techniques”, Confluence(2022).10.1109/Confluence52989.2022.9734128.

[10] P. K. Roy and S. Banerjee, “Deep learning to filter SMS spam,” Future generation computer systems, vol. 102,p. 524–533, 2020

[11] William Siagian, Melisa Rachel Setiadi, Simeon Yuda Prasetyo, Improving SMS spam Detection through ML Techniques: An Investigation of Feature Extraction Techniques, https://doi.org/10.1109/ICIMTech59029.2023.10277999

[12] @articlegorikhan2022ecu, title=ECU To ECU Communication in Automotive Grade Processors and Security of Messages Exchanged, author=Gorikhan, Moin I and Patil, Bhagyashri S and Panhalkar, Mrunali N and Patil, Kapila J and Jadhav, Sushant and Rudrappa, Gujanatti and Mutkekar, Harshal and Anand, Shreyas and Vijapur, Nataraj, journal=Journal of Pharmaceutical Negative Results, pages=166–172, year=2022

[13] ”https://www.kaggle.com/datasets/uciml/sms-spam-collection-dataset/data”, accessed on 2nd January 2024.

[14] S. Kusumanjali, T. Anurag, K. P. Kumar, Mr. K. Vignesh, U. Vamsi Spam Detection in Text Using Machine Learning https://doi.org/10.1109/ICIMTech59029.2023.10277999 url=https://api.semanticscholar.org/CorpusID:69914711

[15] Oyeyemi, Dare Azeez, and Adebola K. Ojo. ”SMS Spam Detection and Classification to Combat abuse in Telephone Networks Using Natural Language Processing.” arXiv preprint arXiv:2406.06578 (2024).

[16] H. Jain and M. Mahadev, ”An Analysis of SMS Spam Detection using Machine Learning Model,” 2022 Fifth International Conference on Computational Intelligence and Communication Technologies (CCICT), Sonepat, India, 2022, pp. 151-156, doi: 10.1109/CCiCT56684.2022.00038.

[17] E. Wijaya, G. Noveliora, K. D. Utami, Rojali and G. Z. Nabiilah, ”Spam Detection in Short Message Service (SMS) Using Naïve Bayes, SVM, LSTM, and CNN,” 2023 10th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE), Semarang, Indonesia, 2023, pp. 431-436, doi: 10.1109/ICITACEE58587.2023.10277368.

[18] N. E. Majd and M. S. Hanchate, ”Spam SMS Classification Using Machine Learning,” 2023 32nd International Conference on Computer Communications and Networks (ICCCN), Honolulu, HI, USA, 2023, pp. 1-7, doi: 10.1109/ICCCN58024.2023.10230203.10.1109/ICITACEE58587.2023.10277368.

[19] R. B. K and D. N, ”Accurate SMS Spam Detection Using Support Vector Machine In Comparison With Linear Regression,” 2023 Fifth International Conference on Electrical, Computer and Communication Technologies (ICECCT), Erode, India, 2023, pp. 1-4, doi: 10.1109/ICECCT56650.2023.10179827.ICIC