Grenze International Journal of Engineering and Technology, January Issue

# Enhanced Spam Detection in Short Message Service using Hybrid Techniques

E. Babu Raj[1], R. Barona[2], N.Ansgar Mary[3] and Shiju George[4]
[1, 4]Christ University, Bengaluru /CSE, Bengaluru, India
Email: baburaj.e@christuniversity.in
[2-3]St. Xavier's Catholic College of Engineering /CSE, IT, Chunkankadai, India
Email: {barona,ansgar}@sxcce.edu.in

***Abstract*— Receiving unwanted text messages, or SMS spam, costs consumers time and money and poses a security concern. To address this issue, we can deploy a system that recognizes and automatically filters out undesirable messages. This method, a testament to the advancement in technology, employs machine learning algorithms that gain knowledge from a pool of communications classified as spam or not. Managing various message contents and languages is one of the system's unique challenges. Notwithstanding these challenges, the approach may be effective in reducing unsolicited communications, improving the security of people's mobile devices and saving them time and money. To address this issue, a variety of machine learning approaches have been employed, ranging from more modern deep learning methods like Convolutional Neural Networks (CNNs) to more traditional ones like Naive Bayes. It is common practice to assess the effectiveness of SMS spam classifiers using measures like as F1-score, precision, and recall. All things considered, SMS spam classification is crucial for protecting the security and privacy of mobile phone users and has useful applications in everyday situations.**

***Index Terms*— SMS Spam, Machine Learning, Naïve Bayes, Convolutional Neural Network (CNN).**

## I. INTRODUCTION

People communicate with one other using mobile messaging, which are exchanged by billions of users. However, because there are insufficient message-filtering techniques, such communication is insecure. Spam is seen as one of the major issues with instant messaging and email systems. At the moment, spam accounts for 85% of emails and messages that mobile users get. The cost of sending and receiving emails and messages is low for senders but high for recipients. Human time lost and vital messages lost are two ways to quantify the cost of spam messages. Because each user has a limited amount of time, memory, and Internet access, these spam emails and messages impact valuable users. The primary objective of SMS (Short Message Service) spam classification is to automatically identify and categorize unsolicited and unwanted messages sent to mobile phone users via SMS or text. The classification process aims to distinguish between spam messages and legitimate messages accurately. SMS spam classification aims to protect mobile phone users from unwanted messages that can be time-consuming, annoying, and potentially harmful to their personal information. By identifying and filtering out SMS spam messages, users can have a better mobile phone experience and avoid potential security risks such as phishing scams or fraudulent activities.

*Grenze ID: 01.GIJET.11.1.332*

The goal of classification is to give a given input a label or category based on a collection of attributes. The process of classification involves several steps. First, a training dataset is used to train the algorithm to recognize the patterns and features associated with each class. The algorithm then applies this knowledge to new, unlabelled data to predict the class of each sample. There are several types of classification algorithms: Decision Trees, Naive Bayes, Logistic Regression, and Support Vector Machines. Supervised Learning is a type of Machine Learning where an algorithm is trained on labeled data to learn the mapping between inputs and outputs. In supervised learning, a training dataset with the proper output labeled for each sample is given to the algorithm. The algorithm learns to generalize from the labeled examples to make accurate predictions on new, unlabelled data. Naive Bayes is a probabilistic algorithm used for classification tasks in Machine Learning. It is based on Bayes' theorem and the assumption that the features are conditionally independent of each other given the class, hence the name "naive. "Naive Bayes is a simple and efficient algorithm that works well on large datasets with high dimensionality. It is commonly used in text classification, spam filtering, sentiment analysis, and recommendation systems. Given the input features, the Naive Bayes method determines the likelihood of each class and chooses the class with the highest probability to be the predicted class. Given the input features, the Naive Bayes method determines the likelihood of each class and chooses the class with the highest probability to be the predicted class. This is because the algorithm can learn from a relatively small amount of labeled data and make predictions quickly. Convolutional neural networks, or CNNs, are frequently employed in applications involving the processing of images and videos. From input data, such photographs, it is intended to automatically and adaptively learn spatial hierarchies of features. The main feature of a CNN is its use of convolutional layers that apply a set of filters to the input data to extract relevant features. Each filter is applied to a small local region of the input data and produces a single output value, which is then passed on to the next layer. This process results in a feature map set that captures different aspects of the input data. The final layers of a CNN are typically fully connected layers that take the flattened output of the pooling layers and produce the final output of the network. The Naive Bayes Classification Algorithm is used to classify spam into real messages.The process normally begins with the collection of data from the Dataset; in this Dataset, the data will be labeled as either spam or ham, and it will be represented as the population of messages that the classification system will encounter. Then, the data from the Dataset will be Preprocessed by cleaning the messages, removing stop words (common words that do not carry much meaning), and converting the text into a numerical format that the classification algorithm can use. Then, the features will be extracted from the preprocessed messages. TF-IDF (Term Frequency-Inverse Document Frequency) can be used for this. The model will then be trained and tested using the Naïve-Bayes Classifier Algorithm, which is used because it is easy to use, quick, scalable, and effective even with a small amount of training data. It can be applied to a variety of tasks, including document classification, sentiment analysis, and spam filtering. Convolutional neural networks are used as the artificial neural networks, and the ReLU function is used as the active function. The ReLU function returns the input value if it is positive and returns zero if it is negative. In the training process, we will teach the algorithm by way of the chosen quality values. In the testing process, they will experiment with the recent inwards data by teaching the model .75% of the data is used for training, and 25% is used for testing. This involves splitting the Dataset into training and validation sets and using techniques such as cross-validation to tune the model's hyperparameters. Then, the trained model will be evaluated on a separate test dataset. Then, the performance will be measured, and this will calculate the accuracy and precision score. Then, the active function will be implied, classifying whether the given message in the textbox is spam or a Ham message.

## II. LITERATURE REVIEW

Karim et al. [1] described a comprehensive survey of the state-of-the-art intelligent spam email detection techniques. The survey covers various aspects of spam email detection, including feature selection, classification algorithms, and evaluation metrics. They review and compare different machine learning algorithms used for spam email detection, such as Decision Trees, Random Forest, and Naïve Bayes. They also discuss the different feature selection techniques, such as statistical methods and natural language processing techniques, that have been used to extract relevant features from emails for classification. Ghatasheh et al. [2] suggested a modified genetic algorithm for feature selection and hyperparameter optimization using the XGBoost algorithm for spam prediction. They contend that choosing pertinent features and fine-tuning hyper parameters are essential to getting spam detection systems to operate well. A mutation operator that can produce new candidate feature subsets not present in the existing population was presented in the work. The XGBoost algorithm's performance on the chosen feature subset serves as the basis for the fitness function. A method for spam identification utilizing an LSTM neural network with an improved semantic representation was presented by Jain et al. [3]. By putting forth a novel strategy that makes use of semantic representations to increase spam detection accuracy, the work makes a significant contribution to the area. Liu et al. [4] suggested an approach for SMS spam detection using a transformer-based neural network architecture called the Spam Transformer Model (STM). They contend that the incapacity of conventional machine learning models to identify SMS spam stems from their failure to recognize the contextual and semantic connections between words and phrases in the messages. On the other hand, these linkages have been well represented by transformer-based models in tasks involving natural language processing. By putting out a novel strategy that makes use of transformer-based models to capture the semantic and contextual links between words and phrases in the messages, the work makes a significant contribution to the field of SMS spam identification. Mohammad et al. [5] explained how to use the Grasshopper Optimization Algorithm (GOA), a multiobjective optimization technique, for feature selection in spam detection systems. The experimental results demonstrate the approach's effectiveness and its potential for application in real-world spam detection systems. Jun et al. [6] suggested a machine learning-based approach for detecting spam messages in secure mobile message communication. Using machine learning methods, the work offers a viable method for identifying spam communications in secure mobile communication. To test the strategy on bigger and more varied datasets and determine its scalability, more research is required. Pan et al. [7] described an approach for spam email classification using a semantic graph neural network. The work presents a promising approach to spam email classification using semantic graph neural networks, which can effectively capture the complex relationships between email attributes and improve classification accuracy. CEKIK et al.[8] described a study on detecting and classifying unwanted SMS messages, commonly known as spam, using text mining techniques. Using machine learning methods, the work offers a viable method for identifying spam communications in secure mobile communication. To test the strategy on bigger and more varied datasets and determine its scalability, more research is required. Fayaz et al. [9] described a study on detecting and classifying spam product reviews using an ensemble machine learning model. The approach that they have used involves preprocessing using the Term Frequency-Inverse Document Frequency (TF-IDF) method. The work presents a promising approach to detect and classify spam product reviews using an ensemble machine learning model, which can effectively improve the credibility of online product reviews and benefit both consumers and businesses. Siddique et al. [10] suggested a study on detecting and classifying spam emails using machine learning algorithms. To determine whether the emails were spam or not, they used a number of machine learning methods, including Naïve Bayes, Decision Trees, and Support Vector Machines (SVMs). This study offers a viable method for identifying and categorizing spam emails using machine learning algorithms, which can successfully lessen the harm that spam emails cause to both people and businesses.

## III. METHODOLOGY

The work, which consists of modules for data collecting, data preprocessing, feature extraction, training, testing, evaluation, and classification, is centered on categorizing and identifying the messages in the text box.

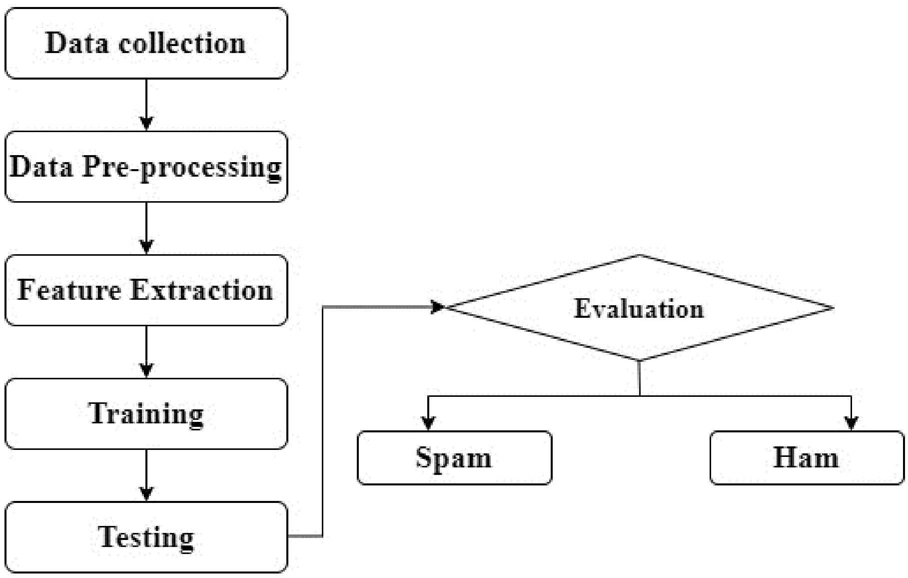


Figure 1. Work Flow Diagram

*A. Data Collection*

Gathering a collection of SMS messages that have been classified as either legitimate or spam is the first step. The SMS spam collection dataset is the one we used for this study. 6,284 English-language SMS messages classified as spam or ham (legal) are included in the dataset. It has 4843 Ham (or Not Spam) Messages and 1441 Spam Messages. The texts were gathered from a number of sources, including as websites that report SMS spam and internet forums.

| | |
|---|---|
| ham | Go until jurong point, crazy.. Available only in bugis n great world la e buffet... Cine there got amore wat... |
| ham | Ok lar... Joking wif u oni... |
| spam | Free entry in 2 a wkly comp to win FA Cup final tkts 21st May 2005. Text FA to 87121 to receive entry question(std txt rate)T&C's apply 08452810075over18's |
| ham | U dun say so early hor... U c already then say... |
| ham | Nah I don't think he goes to usf, he lives around here though |
| spam | FreeMsg Hey there darling it's been 3 week's now and no word back! I'd like some fun you up for it still? Tb ok! XxX std chgs to send, �1.50 to rcv |
| ham | Even my brother is not like to speak with me. They treat me like aids patent. |
| ham | As per your request 'Melle Melle (Oru Minnaminunginte Nurungu Vettam)' has been set as your callertune for all Callers. Press *9 to copy your friends Callertune |
| spam | WINNER!! As a valued network customer you have been selected to receivea �900 prize reward! To claim call 09061701461. Claim code KL341. Valid 12 hours only. |
| spam | Had your mobile 11 months or more? U R entitled to Update to the latest colour mobiles with camera for Free! Call The Mobile Update Co FREE on 08002986030 |
| ham | I'm gonna be home soon and i don't want to talk about this stuff anymore tonight, k? I've cried enough today. |
| spam | SIX chances to win CASH! From 100 to 20,000 pounds txt> CSH11 and send to 87575. Cost 150p/day, 6days, 16+ TsandCs apply Reply HL 4 info |
| spam | URGENT! You have won a 1 week FREE membership in our �100,000 Prize Jackpot! Txt the word: CLAIM to No: 81010 T&C www.dbuk.net LCCLTD POBOX 4403LDNW1A7RW18 |
| ham | I've been searching for the right words to thank you for this breather. I promise i wont take your help for granted and will fulfil my promise. You have been wonderful and a blessing at all times. |
| ham | I HAVE A DATE ON SUNDAY WITH WILL!! |
| spam | XXXMobileMovieClub: To use your credit, click the WAP link in the next txt message or click here>> http://wap. xxxmobilemovieclub.com?n=QJKGIGHJJGCBL |
| ham | Oh k...i'm watching here:) |
| ham | Eh u remember how 2 spell his name... Yes i did. He v naughty make until i v wet. |
| ham | Fine if that��s the way u feel. That��s the way its gota b |
| spam | England v Macedonia - dont miss the goals/team news. Txt ur national team to 87077 eg ENGLAND to 87077 Try:WALES, SCOTLAND 4txt/1.20 POBOXox36504W45WQ 16+ |
| ham | Is that seriously how you spell his name? |
| ham | I���m going to try for 2 months ha ha only joking |
| ham | So �_ pay first lar... Then when is da stock comin... |
| ham | Aft i finish my lunch then i go str down lor. Ard 3 smth lor. U finish ur lunch already? |
| ham | Ffffffffff. Alright no way I can meet up with you sooner? |
| ham | Just forced myself to eat a slice. I'm really not hungry tho. This sucks. Mark is getting worried. He knows I'm sick when I turn down pizza. Lol |

Figure 2. Part of the SMS Spam Collection Dataset

*B. Data Preprocessing*

The second step is to preprocess the data by cleaning it and transforming it into a format that can be used to train the machine learning algorithms.

This involves Text Cleaning, i.e., Removing any unnecessary or irrelevant information from the text data such as Punctuations, Special Characters, and Stop Words; tokenization, i.e., breaking down text into individual words and individual sentences; Stemming/Lemmatization, i.e., normalize different forms of a word so that they can be treated as the same word, and Vectorization.

Preprocessing can be done using CNN Neural Network. It improves the data quality and makes it more suitable for machine learning algorithms.

*C. Feature Extraction*

The third stage is to extract pertinent features, like word frequencies, character-based features, and linguistic features, from the preprocessed data that can be used to identify spam communications.

This entails choosing characteristics that can differentiate between spam and non-spam communications. The Naive Bayes Classification Algorithm, which is based on Baye's theorem, is used in this step to divide the dataset into the training dataset and the test dataset.

*D. Training*

The preprocessed and feature-extracted dataset is used to train the machine learning algorithms in the fourth stage. To train the data, the preprocessed data will be used to generate the feature matrix. The count values of every word in the text messages will be included in this feature matrix. Next, initialize the Naive Bayes model using the Naive Bayes module in Scikit-Learn.

A Multinomial Naive Bayes model, which is appropriate for discrete count-based features like word counts and will have high precision and accuracy, uses the Multinomial NB class. We can forecast fresh SMS messages after the model has been trained.

*E. Testing*

The fifth phase involves evaluating the SMS spam classification system's performance on a different dataset of SMS messages. This is primarily done to assess the system's performance on a different collection of data known as the testing data. Throughout testing, the model is given the test data and is asked to estimate if a given message is spam or not. A variety of evaluation measures, including accuracy and precision, are then computed by comparing the projected labels with the actual labels of the test data.

### *F. Evaluation*

Assessing the SMS spam categorization system's performance using a variety of measures, including accuracy and precision, is the sixth and last phase. The evaluation data can be used to further enhance and refine the system. The primary goal of evaluation is to ascertain how well the model can distinguish between spam and non-spam communications. The model's performance on the testing set is then used to compute the evaluation measures.

## IV. Results and Discussions

In this section, we discuss the results obtained through our work. We have implemented KNN(K-Nearest Neighbor), SVM(Support et al.), RF(Random Forest), BgC(Bagging Classifier), ETC(ExtraTreesClassifier), LR(LogisticRegression), XGB(Extreme et al.), AdaBoost, GBDT(GradientBoostingClassifier), NB(Naïve Baye’s) and DT(Decision Tree Classifier). Among these algorithms, Nive Baye’s algorithm performs better than other algorithms.The accuracy value of the NB algorithm is 0.998414, which is higher than other algorithms.

The pair plot representation of the Dataset is shown in Fig 4. Several characters, words, and sentences in the Dataset obtain the pair plot representation. Target value 0 indicates spam messages, and 1 indicates ham messages.

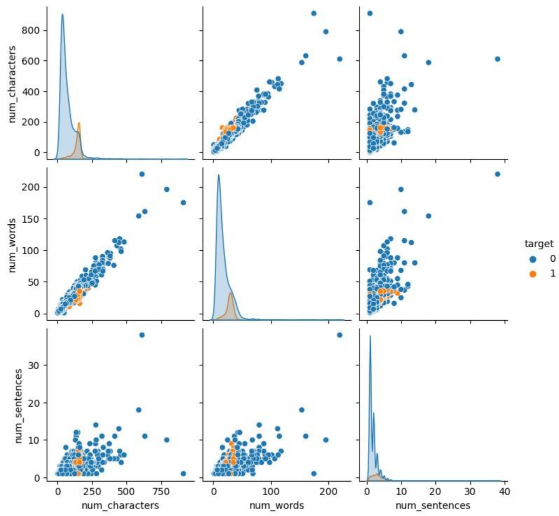


Figure 3. Ham and Spam data distributed in two classes

The heat map representation of the data is shown in Fig 5. It shows the correlation of each feature with the target.The performance metrics values for each algorithm are given in Table 1. The Naïve Baye’s algorithm has higher Accuracy,Precision, F1-score, and Recall values. Thus, our work shows that the NB algorithm is best for classifying messages as spam or ham. Using performance criteria including accuracy, precision, F1-score, and recall values, Fig. 6 compares different algorithms.


Figure 5. Correlation of each feature with the target

TABLE 1 PERFORMANCE METRICS VALUES

| Algorithm | Accuracy | Precision | F1-score | Recall |
|---|---|---|---|---|
| KN | 0.905222 | 0.98 | 0.449438 | 0.289855 |
| SVM | 0.970986 | 0.97 | 0.878049 | 0.782609 |
| RF | 0.975822 | 0.982906 | 0.901961 | 0.833333 |
| BgC | 0.975822 | 0.97479 | 0.902724 | 0.84058 |
| ETC | 0.974855 | 0.974576 | 0.898438 | 0.833333 |
| LR | 0.958414 | 0.970297 | 0.820084 | 0.710145 |
| xgb | 0.967118 | 0.933333 | 0.868217 | 0.811594 |
| AdaBoost | 0.960348 | 0.929204 | 0.836653 | 0.76087 |
| GBDT | 0.946809 | 0.919192 | 0.767932 | 0.65942 |
| **NB** | **0.998414** | **0.988217** | **0.988951** | **0.981594** |
| DT | 0.930368 | 0.836735 | 0.694915 | 0.594203 |

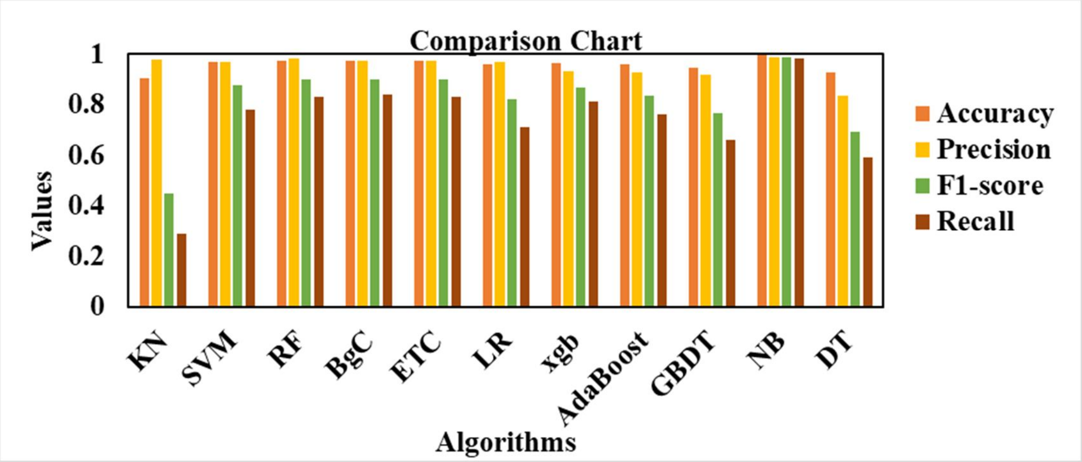


Figure 6. Comparison chart of various algorithms

## V. CONCLUSION

For users of mobile devices, SMS spam has become a recurring problem that interferes with communication and frequently causes annoyance. The demand for automated methods to efficiently filter out unsolicited messages has increased as a result of this difficulty. By identifying trends and accurately differentiating between spam and valid messages, machine learning (ML) algorithms provide a potent solution to this issue. By guaranteeing that only pertinent, meaningful, or authentic communications make it to the user's inbox, the implementation of an ML-based SMS spam filtering system has the potential to greatly enhance the messaging experience. Despite its promise, ML-based SMS spam filtering systems still face difficulties. Large and varied training data is necessary to build a model that correctly detects spam across languages and dialects, but obtaining this data can be resource-intensive. Furthermore, to prevent the difficulty of classifying valid messages as spam, it is crucial to strike a balance between high accuracy and low false-positive rates. More novel architectures are the need of the hour which can adjust to the ever-changing terrain of spam policies by utilizing sophisticated natural language processing techniques and constantly changing models will be our future work.

## REFERENCES

[1] A. Karim, S. Azam, B. Shanmugam, K. Kannoorpatti and M. Alazab, "A Comprehensive Survey for Intelligent Spam Email Detection," in IEEE Access, vol. 7, pp. 168261-168295, 2019.
[2] N. Ghatasheh, I. Altaharwa, and K. Aldebei, "Modified Genetic Algorithm for Feature Selection and Hyper Parameter Optimization: Case of XGBoost in Spam Prediction," in IEEE Access, vol. 10, pp. 84365-84383, 2022.
[3] Gauri Jain, Manisha Sharma & Basant Agarwal “Optimizing semantic LSTM for spam detection”. In the International Journal of Information Technology-2019.

[4] X. Liu, H. Lu and A. Nayak, "A Spam Transformer Model for SMS Spam Detection," in IEEE Access, vol. 9, pp. 80253-80263, 2021.
[5] S. A. A. Ghaleb et al., "Feature Selection by Multiobjective Optimization: Application to Spam Detection System by Neural Networks and Grasshopper Optimization Algorithm," in IEEE Access, vol. 10, pp. 98475-98489, 2022.
[6] Luo Gung Jun, Shah Nazir and Amin Ul Haq, “Spam detection approach for secure mobile message communication using machine learning algorithms”, Security and Communication Networks, Hindawi, Vol 2020, No 1 PP 887639.
[7] WeisenPan, Jian Li, Lisa Gao, Yan Yang “Semantic Graph Neural Network, A conversion from Spam email classification to Graph classification”, Scientific Programming, Hindawi, Vol 2022, No 11 PP 1-8.
[8] Rasim ÇEKİK1 “Classification of Unwanted SMS Data (Spam) with Text Mining Techniques”. Journal of Soft Computing and Artificial Intelligence 03(02): 41-50, 2022.
[9] Muhammad Fayaz, Atif Khan, Javid Ur Rahman, Badar Alouffi ” Ensemble Machine Learning Model for Classification of Spam product Reviews” , Complexity Journal, 2020-12.
[10] Zeeshan Bin Siddique, Mudassar Ali Khan, Ikram Ud Din, Ahmad Almogren, Irfan Mohiuddin, Shah Nazir “Machine Learning-Based Detection of Spam Emails”, Scientific Programming, Wiley, 2021.