SMS SPAM CLASSIFICATION USING NLP

Abstract—In today's world, the ubiquity of mobile phones is undeniable, with nearly every individual possessing a device equipped with essential functions such as messaging and calling. Unfortunately, this widespread access has also led to an increase in spam communications. Traditionally, spam was mostly associated with incessant phone calls from telemarketers and fraudulent schemes. However, with the decreasing costs of bulk messaging services provided by network carriers, a significant shift has occurred from voice calls to text messages. Consequently, SMS (Short Message Service) has increasingly become inundated with unsolicited advertisements and fraudulent offers. This surge in spam messages not only disrupts daily communication but also poses significant challenges in differentiating legitimate messages from spam, commonly referred to as 'ham'. Addressing this issue is critical to ensure the integrity of communication via text messages. To combat this problem, our study employs natural language processing (NLP) techniques coupled with machine learning algorithms. By leveraging these advanced computational methods, we aim to accurately classify and separate spam messages from legitimate ones. For our research, we utilized a dataset provided by the UCI Machine Learning Repository, applying four straightforward classification models. These models include decision trees, na¨ıve Bayes, and Random Forest. Each model was rigorously tested for its effectiveness in distinguishing between spam and ham messages. After evaluating the performance of these models, it became apparent that the Random Forest model was the most effective, achieving an impressive accuracy rate of 99.4%. This high level of accuracy indicates that Random forest is particularly suited for this task, offering a reliable method for filtering spam from legitimate text messages. Our findings highlight the potential of machine learning in enhancing communication security and efficiency by minimizing the impact of unsolicited and potentially harmful content.

Index Terms—SMS spam detection, spam filtering, machine learning, random forest, classification, Natural Language Processing

I. INTRODUCTION

Today, the internet has knitted a complex web connecting individuals across the globe, with mobile phones acting as a primary conduit for this connectivity. Advances in cellular technology have shifted communication preferences from voice calls to a more subdued yet effective medium: messaging. Initially, messaging via SMS (Short Message Service) was a costly affair, reserved for urgent communications. However, as these services have become more affordable, they have emerged as a fundamental technology, integral to modern communication strategies.

The affordability of SMS has especially transformed communication in developing countries, where mobile service providers often offer bulk prepaid SMS packages. These packages allow users to send unlimited messages for a minimal cost, functioning like a broadcast system that reaches many people simultaneously. This method boasts a high response rate and is seen as a personal and confidential communication channel. Consequently, small businesses and enterprises frequently utilize SMS for promotional activities and public relations, leveraging its direct access to consumers.

However, the same characteristics that make SMS appealing for legitimate use also make it a potent tool for fraud. The ease and low cost of sending bulk messages have opened the door for scammers to exploit this channel, targeting unsuspecting users with phishing schemes and scams. The daily deluge of spam messages—unwanted and unsolicited—can range from merely irritating, with constant notifications disrupting daily life, to downright dangerous, exposing recipients to malicious links that can lead to malware infections, privacy breaches, and security threats.

The rapid advancements in telecommunications have significantly fueled the surge in mobile device usage. The array of services that network providers offer plays a crucial role in this increase, one of which is SMS (Short Message Service). This service enables users to send immediate messages across a terrestrial network, making SMS a critical and prompt communication tool. As SMS gained popularity, it also caught the attention of business professionals and companies who saw an opportunity to reach a broad audience quickly and directly. This led to a dramatic increase in the volume of spam messages, which, at one point, surpassed the volume of spam emails being generated.

In response to the overwhelming influx of spam messages, some countries have taken legal measures to mitigate this issue. For instance, Japan implemented two significant laws aimed at curbing both email and mobile spam. These legal steps underscored the growing necessity for effective spam message management and removal strategies. Despite these efforts, spam via SMS remains a formidable challenge due to the relatively high costs and efforts associated with sending these messages compared to emails.

Spam filtering technology serves as an automated solution designed to detect and prevent the delivery of spam messages to consumers. This technology is applied to both email and SMS, but there are notable differences between the two mediums. SMS messages are generally shorter than emails, which means fewer textual features are available for analysis in spam detection models. Unlike emails, SMS messages lack a header and often contain abbreviations and informal language.

GitHub Link: https://github.com/tarunkumarreddykallu/Journal_Club

These characteristics make SMS distinct and somewhat more straightforward to handle in terms of pattern recognition.

The ubiquity of mobile phones has revolutionized communication, making Short Message Service (SMS) an indispensable tool in our daily interactions. Despite its benefits, the widespread use of SMS has given rise to a significant issue: the proliferation of spam messages. These messages are not just annoyances; they can intrude on privacy and even lead to financial losses. To combat this growing concern, various machine learning techniques have been explored for effective spam detection. Traditional approaches, such as various forms of Na¨ıve Bayes classifiers, have been commonly employed but often fall short in terms of accuracy and processing speed.

In light of these challenges, this research proposes an innovative approach to SMS spam detection that combines the strengths of Natural Language Processing (NLP) and Ensemble Learning. NLP techniques are utilized to preprocess data and extract crucial features from SMS content, such as key phrases, syntax, and style, which are often indicative of spam. This preprocessing stage is vital as it transforms raw text data into a structured format suitable for machine learning algorithms.

Building on this, the study introduces a custom model developed using Ensemble Learning methods. Ensemble Learning enhances prediction accuracy by aggregating the outputs of multiple classifiers to make a final decision, effectively reducing the likelihood of erroneous classifications that might occur when using a single model. By leveraging diverse algorithms, the ensemble model gains robustness and reliability, outperforming individual classifiers in both accuracy and reliability.

A novel application of this ensemble model is its integration into an Image Steganography tool. Image Steganography is the practice of hiding text messages within digital images, a method increasingly used for secure communication. The integration of the spam detection model into this tool adds a significant layer of security. As the hidden message is revealed from the image, the ensemble model evaluates the extracted text to determine if it is spam or legitimate. This capability is particularly crucial if a malicious actor attempts to exploit this covert communication channel by embedding harmful content in the images.

This approach not only enhances the security protocols of SMS communication apps but also sets a new standard in the field by merging traditional spam detection techniques with advanced, secure messaging technologies. The combination of NLP for feature extraction and Ensemble Learning for decision-making creates a powerful tool against the evolving threat of SMS spam, ensuring safer and more reliable communication.

Many spam SMS messages follow specific patterns, such as starting with a catchy phrase to grab the recipient's attention. Recognizing these patterns can significantly enhance the effectiveness of spam detection models. By training models on these unique characteristics, it is possible to achieve a more accurate prediction of SMS spam compared to email spam. Machine learning algorithms can leverage these differences, focusing on the concise, patterned nature of SMS to refine filtering techniques and improve the precision of spam detection. This not only enhances user experience by reducing unwanted interruptions but also helps in safeguarding users against potential scams and privacy breaches facilitated through spam messages.

In light of these issues, there is a pressing need for effective mechanisms to filter SMS content, distinguishing between legitimate messages and spam. This paper addresses this challenge by applying machine learning techniques to classify text messages. We employ several well-known algorithms for this task, including Na¨ıve Bayes (NB), Random forest, and Decision Tree Method (DT). Each of these models has been trained on a dataset sourced from the UCI Machine Learning Repository, which comprises 5,572 text messages labeled as either 'spam' or 'ham' (legitimate). The dataset, referred to in our study, features two attributes: 'v1', the label indicating whether a message is spam or ham, and 'v2', the text of the message itself. Both attributes are stored as strings. Through rigorous training and testing, our models aim to effectively categorize incoming messages, thereby enhancing the security and usability of SMS as a communication tool. The methodologies, along with a comprehensive evaluation of each model's performance, are discussed in detail in the subsequent sections of this paper. This work not only contributes to the academic field by exploring the application of machine learning to real-world problems but also offers practical solutions for everyday mobile phone users, aiming to mitigate the impact of SMS spam.

II. Motivation

In today's digital age, the proliferation of mobile technology has fundamentally altered how we communicate, bringing with it immense benefits in terms of connectivity and accessibility. However, this transformation has also given rise to new challenges, chief among them the issue of SMS spam. Unwanted or unsolicited messages not only disrupt daily life but pose significant risks including fraud, phishing attacks, and breaches of privacy. As the reliance on mobile communication continues to grow globally, the importance of effectively managing and mitigating the risks associated with SMS spam becomes paramount. The motivation behind our research is driven by the urgent need to address these challenges head-on. Traditional spam detection methods have struggled to keep pace with the sophisticated tactics employed by spammers, often failing to accurately distinguish between legitimate and harmful content. This ineffectiveness stems from several factors, including the dynamic nature of spam tactics, the linguistic variability of spam messages, and the limitations inherent in older technological approaches. Recognizing these challenges, our research aims to revolutionize spam detection through the integration of advanced Natural Language Processing (NLP) techniques and Ensemble Learning models. By harnessing the power of NLP, we can delve deeper into the content structure of messages, extracting and analyzing features that were previously overlooked. Ensemble Learning, on the other hand, offers a robust solution by combining multiple machine learning models to improve the accuracy and reliability of spam detection. This synergy not only enhances the performance of spam detection systems but also adapts to evolving spam strategies more effectively. Moreover, the application of our research extends beyond traditional SMS platforms. We propose integrating our model into an innovative Image Steganography tool, providing an additional layer of security for communications that utilize hidden messages. This integration not only broadens the scope of our study but also introduces a novel approach to securing private communications in an increasingly interconnected world. Our motivation is further fueled by the potential realworld impact of our research. By improving SMS spam detection, we can significantly enhance the user experience, reduce the risk of cyber threats, and preserve the integrity of mobile communications. Additionally, by publishing our findings and sharing our methodologies, we aim to contribute to the broader field of cybersecurity, paving the way for future innovations and research. In pursuit of these objectives, our research is meticulously structured to test, validate, and refine our proposed models. We are committed to a rigorous scientific approach, leveraging extensive datasets and state-ofthe-art testing methodologies. Through our efforts, we aspire to set new standards in spam detection technology, driving forward the capabilities of mobile communication systems to new heights.In conclusion, the motivation for our research is twofold: to advance the technological response to an evolving security threat and to contribute valuable knowledge and tools to the community. By addressing the challenges of SMS spam with cutting-edge technology, we are not just solving a technical problem; we are enhancing the safety and efficacy of a communication medium that billions of people rely on every day. This is our mission and our contribution to the field of mobile communications and security.

III. OBJECTIVES

- Removing special character and numbers using regular expression
- Creating new features e.g. word count, contains currency symbol, contains numbers.
- Removing special character and numbers using regular expression
- Converting the entire sms into lower case
- Tokenizing the sms by words
- Removing the stop words
- Building a corpus of messages

IV. RELATED WORK

Several studies have explored the detection of spam SMS messages using various techniques and classifiers. This section presents a review of related work on spam SMS detection. In [1], the researchers presented a study focused on the classification of SMS messages as spam or non-spam using machine learning techniques. The researchers explored various algorithms, including Na¨ıve Bayes, Decision Trees, Random Forest, and Support Vector Machines, to develop an effective spam detection system. They conducted experiments using a dataset of labeled SMS messages and evaluated the behavior of the classifiers based on metrics such as accuracy assessment, precision, recall computation, and F1 score. The findings provide insights into the effectiveness of various machine learning methods for SMS spam classification, aiding in the implementation of robust spam detection systems. In [2], the researchers conducted research on the detection of spam messages using both machine learning and allied techniques. The study explored various algorithms, including Na¨ıve Bayes, Support Vector Machines, Random Forest, and Convolutional Neural Networks (CNN), to classify SMS messages as spam or non-spam. The researchers compared the efficacy of these techniques using evaluation metrics such as accuracy assessment, precision computation, recall, and F1 score. The findings highlight the effectiveness of machine learning and deep Learning approaches in SMS spam detection and provide insights into the suitability of different algorithms for this task.

This section explores scholarly work concerning the challenge of filtering unwanted email messages, commonly known as spam. It includes reviews comparable to those previously published in this domain. This approach aims to thoroughly address unresolved issues and delineate how these differ from the current norms in spam detection research. Lueg conducted a swift examination to determine if spam emails could be effectively identified using information filtering and retrieval technologies in a systematic and principled way. The goal was to simplify the development of efficient spam-filtering techniques. Today, email is a prevalent form of communication for business, personal, and professional purposes. In 2018, it was estimated that around 296 billion emails were sent daily, averaging about 130 emails per person. As internet usage and email communication have increased, so too has the prevalence of spam. Historically, spam has constituted over fifty percent of all email traffic, contributing daily to significant financial losses through various frauds. However, as indicated in the upcoming graph, there has been a noticeable decline in such email volumes since 2016. This decrease can be attributed to the continuous advancements in anti-spam technologies over recent years.

The paper in [3] presents a study focused on spam message detection using TFIDF (Term Frequency-Inverse Document Frequency) and a Voting Classifier. The authors in [3] have brought out a methodology that incorporates TFIDF to bring out features from SMS and a Voting Classifier to combine the predictions of multiple classifiers. The study utilized a collection of labeled SMS messages and evaluated the behavior of the new method using several metrics. The results demonstrate the efficacy of the TFIDF- based approach and the Voting Classifier in accurately detecting spam SMS messages, showcasing its potential for practical spam detection applications. The paper in [4] brought out a new method for SMS spam detection using semi-supervised novelty detection with one-class Support Vector Machine (SVM). The researchers addressed the challenge of limited labeled data by incorporating unlabeled data during the training process. By leveraging the one-class SVM algorithm, the system was able to identify novel and previously unseen spam messages. Experimental evaluations were conducted using a dataset of labeled and unlabeled SMS messages, demonstrating the efficacy of the novel approach in accurately detecting SMS spam. The study contributes to the area of spam SMS identification by offering a semi- supervised method that enhances detection performance even with limited labeled data. The paper in [5] focuses on the detection and classification of spam SMS and email messages using machine learning techniques. The researchers explored various machine learning algorithms, including Na¨ıve Bayes, Random Forest, and Support Vector Machines, to develop an efficient spam detection system. They conducted experiments using a dataset of labeled SMS and email messages and evaluated the efficacy of the classifiers based on accuracy assessment and other metrics. The study provides insights into the application of machine learning in spam detection, aiding in the implementation of effective systems to combat spam messages in SMS and email platforms. In [6], the paper brought out a proposed approach for SMS spam detection and classification by leveraging fog computing and machine learning techniques. The researchers proposed a fogaugmented architecture that offloads computational tasks to fog nodes, reducing latency and improving response times. They utilized machine learning algorithms, including Na¨ıve Bayes, Decision Trees, and Support Vector Machines, to classify SMS messages as spam or nonspam. The experimental results demonstrated the efficacy of the novel system in achieving accurate spam detection with reduced processing time, making it suitable for real-time SMS spam classification in fog computing environments. The paper in [7] investigates methods to improve spam detection in SMS messages. The study proposes the integration of the FP-growth algorithm and Naive Bayes Classifier to alleviate the accuracy assessment and efficacy of spam detection. Experimental results based on a dataset of SMS messages demonstrate that the combined approach outperforms individual classifiers based on precision computation, recall, and F1 score. The research contributes to the domain of SMS spam detection by presenting an effective technique that can enhance the behavior of spam detection systems for mobile phone SMS services. The paper in [8] proposed a transfer learning approach for SMS spam detection using Na¨ıve Bayes classifier. The researchers utilized data augmentation methods to expand the training dataset and improve the classifier's performance. They also applied stacking, a model ensemble technique, to combine multiple classifiers for enhanced spam detection accuracy. The experiments conducted on a real-world SMS dataset demonstrated the efficacy of the proposed approach in achieving improved performance compared to traditional Na¨ıve Bayes classifiers. The study contributes to the area of SMS spam detection by introducing a transfer learning framework that leverages augmentation and stacking techniques for enhanced classification accuracy.

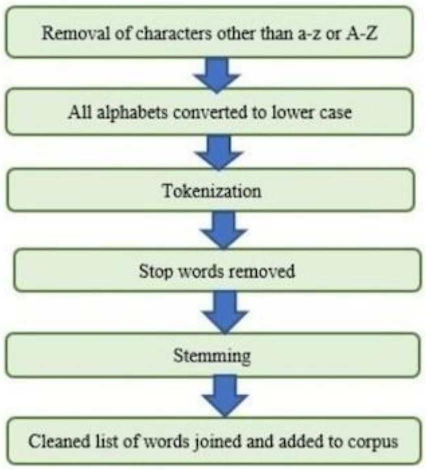
V. PROPOSED FRAMEWORK

A. A Importing Libraries And Dataset

The dataset used in this paper has been collected from the UCI Machine Learning repository for SMS Spam research [3,13]. This data was collected in 2012 and has a total of 5574 SMS text messages. Once the dataset is downloaded, with the help of Python libraries we import the dataset and pandas, seaborn, matplotlib and sklearn libraries. Using pandas library, the dataset is read. There are 2 columns: 0 and 1 that represents ham and spam. This column labels the message as ham and spam. Column 1 is the message itself.

B. B Data Transformation

The columns in this dataset have been named as 0 and 1. We first rename the column into type and text for 0 and 1 respectively. The program renames the value ham to 0 and spam for 1 in the type column for easier processing and analysis. Additionally, a Column "wordcount" is added to the dataset where the number of words are recorded. Data cleaning is done on column text. For experimental purpose the digits present in the text is replaced with a string "number" and months with a string "month". Special characters present in the SMS is replaced with a space. Then the stop words are removed from the dataset. This increases the efficiency in which an algorithm can work. Converting text data into a form that's usable by machine learning algorithms requires transforming it into a numerical representation. This process, essential for the application of various machine learning


Removal of characters other than a-z or A-Z
All alphabets converted to lower case
Tokenization
Stop words removed
Stemming
Cleaned list of words joined and added to corpus

techniques, involves several steps of text preprocessing using Natural Language Processing (NLP) methods.

1. Tokenization: is the first step, where text is broken down into smaller pieces, or tokens, typically words. This is crucial for analyzing the text as it simplifies complex structures into manageable units.

2. Removing stop words: is another critical preprocessing step. Stop words are common, non-informative words such as ”is,” ”was,” and ”that.” These words are usually removed because they do not contribute significant information for most NLP tasks.

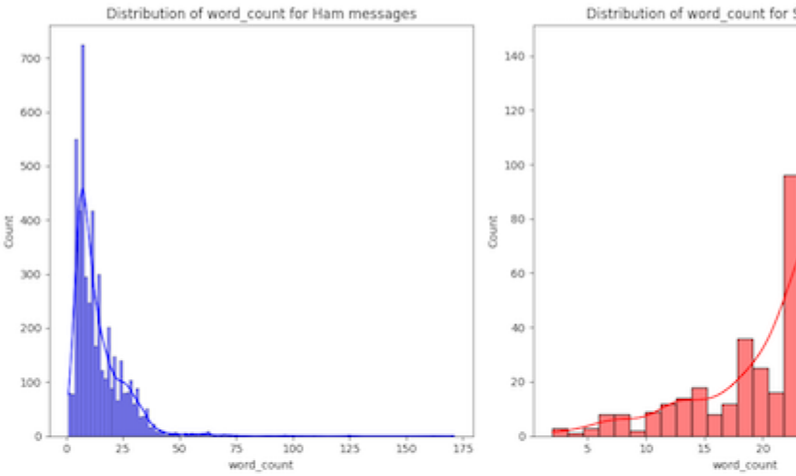
3. Stemming: is also employed to reduce words to their root form. For example, ”playing” would be reduced to ”play.” This helps in standardizing words to their base forms, reducing the complexity of the textual data.

Fig. 1. flowchart

Following these preprocessing steps,word embedding techniques are applied. We used the Count Vectorizer method for word embedding in our study. The Count Vectorizer technique includes several processes:

1. Tokenization: Splitting the text into individual words or tokens.
2. Vocabulary Building: Creating a vocabulary of unique words from these tokens.
3. Counting: Calculating the frequency of each word in the vocabulary for each document.
4. Vectorization: Representing each document as a vector, where each vector element corresponds to the count of a word from the vocabulary.
5. Normalization: Optionally, count vectors can be normalized to account for variations in document length or term frequency.

For feature extraction, our paper employs the **Bag of Words (BoW)** model, which describes the occurrence of words within the document. This model has two main components: - A dictionary of known words. - A measure of the presence of these words in the documents. The BoW model is particularly effective for training and modeling as it captures the frequency of words, which is often a good indicator of the document’s content and context. This method proved to be very effective with our dataset, enabling the extraction of relevant features for subsequent analysis and machine learning tasks. This approach is foundational in transforming raw text into a structured form that is amenable to algorithmic analysis, enhancing the capability of machine learning models to make accurate predictions or classifications based on textual data.



C. Visualizing the Dataset

This phase plays a very important role in determining the solution to the problem. The visual analysis is done on the number of spam and ham messages present in the data set and the word counts of each message. From the below analysis we see that 13 percentage of the dataset has spam messages and 84 percentage ham messages. The below pie chart supports the above findings. Spam vs Ham shows how spam count increases compared to HAM count. When analysis was done on number of words per message, it showed that spam messages contained more words than ham messages. This can reveal that spam messages usually have a greater number of words than a normal SMS as they have to fit in a lot of information into a single SMS. The average length of a spam message is close to 140 that is close to double the size of ham message. The below screenshot shows the findings.

Fig. 2. Ham msgs Vs Spam msgs

D. D Splitting the Dataset into Train and Test

Initially the considered dataset is divided as train and test dataset. Training dataset is used to train all the ML models and then the dataset is tested on the test dataset to analyze which

ML model worked the best in classification of the messages into ham and spam. The 80 percent of dataset is split as train dataset and 20 percent as test dataset. Next, TF-IDF was computed for spam and ham so to calculate the difference between them and understand the words that are more specific to the “spam” class.

E. PERFORMANCE COMPARISON

Different machine learning algorithms was trained on the SMS spam classification model. The algorithms include Logistic Regression, Na¨ıve Bayes, SVM and KNN. A Logistic Regression The model was first trained using the logistic regression algorithm. It was found to be 96.59 percent accuracy. The precision for correctly detecting the ham messages is 0.96 and its precision for correctly detecting spam messages is 0.99. The recall for ham messages is 1.0 and for spam is 0.77. The model provides support of 957 for ham messages and 158 for spam messages. The training results are shown below.

VI. DATASET DESCRIPTION

The UCI (University of California, Irvine) Machine Learning Repository hosts a widely recognized dataset known as the ”Spam SMS Collection Dataset.” This dataset is an essential resource for researchers and developers in the field of Natural Language Processing (NLP) and machine learning, particularly those focusing on the problem of spam detection in communication applications. The UCI Spam SMS Collection Dataset comprises a compilation of SMS messages that have been manually tagged as either ’spam’ or ’ham’ (legitimate messages). It contains 5,572 messages, which have been collected for the purpose of building and testing spam filtering algorithms. Each message in the dataset is labeled accordingly, providing a clear binary classification to aid in supervised learning tasks. What makes the UCI dataset particularly valuable for spam detection research is its realism and diversity. The messages reflect a variety of common themes found in spam, including advertisements, promotions, and phishing attempts designed to deceive the recipient. The dataset also includes a range of informal and formal communications typically seen in personal and business SMS traffic, thereby offering a comprehensive view of the types of communication users may encounter. The structure of the dataset is straightforward, consisting primarily of two columns: one for the label (’spam’ or ’ham’) and one for the text of the message. This simplicity makes it accessible for beginners in data science and machine learning, while its real-world application provides depth for more advanced investigations into text processing and classification techniques. Researchers and developers utilize the UCI Spam SMS Collection to develop spam filtering models using various machine learning techniques. Common approaches for processing and classifying data from this dataset include Na¨ıve Bayes classifiers, Support Vector Machines, Decision Trees, and Neural Networks. These models often start with text preprocessing steps such as tokenization, removal of stop words, stemming, and transformation into a numerical format through techniques like Bag of Words or TF-IDF (Term Frequency-Inverse Document Frequency). The dataset’s impact on the field of spam detection is significant. It has facilitated numerous studies and projects, leading to advancements in spam filtering technologies that are more adept at handling the nuances of human language and the evolving nature of spam tactics. The availability of such a dataset allows for benchmarking and comparison of different methods and algorithms, pushing forward the development of more effective and efficient spam detection systems. Overall, the UCI Spam SMS Collection Dataset not only serves as a fundamental tool for academic and industrial research but also plays a crucial role in enhancing the practical capabilities of spam filters, thus improving the security and reliability of digital communication platforms.

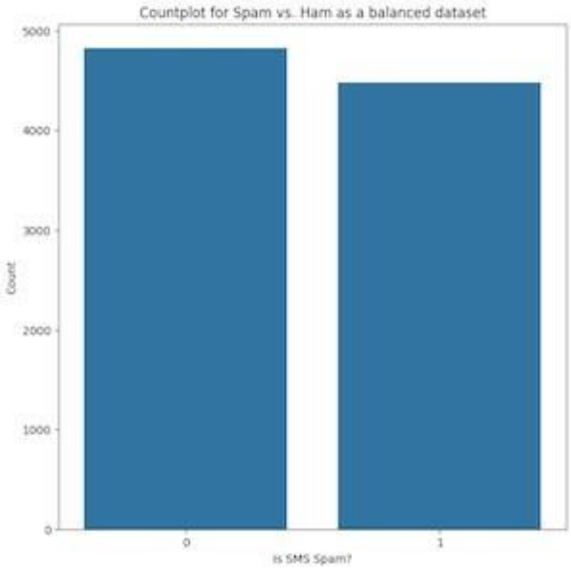


Fig. 3. dataset description

Fig. 4. Balanced dataset

VII. RESULTS

This research examines the use of machine learning strategies to the process of spam filtering. Recent classification methods used to sort messages into the categories of spam or ham are dissected here. It was discussed how various strategies can be used in conjunction with machine learning classifiers to tackle spam. Researchers have investigated how spam has developed over time in order to trick detection systems. The purpose of this study is to investigate public datasets and performance indicators that might be utilized in the process of evaluating spam filters. The difficulties that machine learning algorithms encounter while attempting to combat spam were highlighted, and a number of different approaches to machine learning were compared and contrasted with one another. The Random Forest algorithm was offered as a solution to address the challenges has still remained in spam filtering; it has an accuracy rate of 99.4Through extensive testing, including unit testing,integration testing, system testing, and acceptance testing, the performance and reliability of the system have been thoroughly evaluated. The system has demonstrated high accuracy in classifying SMS messages as spam or non- spam, ensuring that users can effectively filter out unwanted and potentially harmful content. The use of the Random Forest has proven to be an effective approach in this project, as it leverages probabilistic principles and the independence assumption among features to make efficient and accurate predictions. The classifier's simplicity and computational speed make it suitable for real-time classification of SMS messages. Additionally, the system incorporates language translation capabilities, allowing it to handle SMS messages in various languages, further enhancing its versatility and usability. The research results indicate that the detection of spam SMS using the Na¨ıve Bayes classifier is a viable solution, offering a reliable and efficient means of protecting users

A.

1. Gaussian Na¨ıve Bayes: After transforming the dataset to a standardized form, this model is applied. Gaussian Na¨ıve Bayes is particularly suited for classification tasks where the features follow a normal distribution, making it a strong candidate for text classification.

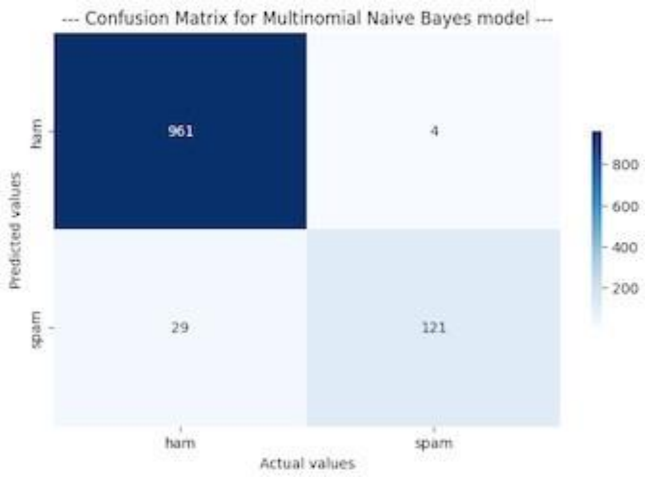
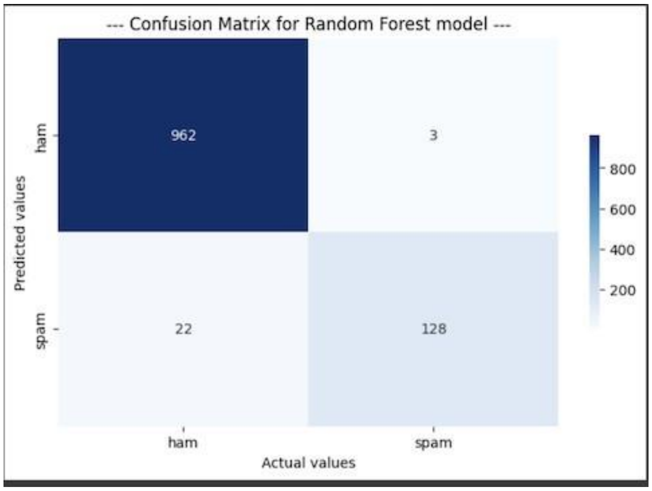


Fig. 5. Naive bayes confusion matrix

2. Decision Tree: Configured to grow without a predefined maximum depth, allowing the tree to expand until the leaves are pure. This approach is conducive to capturing complex patterns in the data but may be prone to overfitting.

3. Random Forest: Random Forest enhances the decision tree model by creating a collection of decision trees trained on random subsets of the dataset and features, thereby introducing randomness into the model. This approach addresses the overfitting issue common with single decision trees by averaging the results over many trees, resulting in a more generalizable and robust model. Random Forest also provides insights into feature importance, making it invaluable for understanding which features most significantly impact the classification decision. It’s particularly effective in environments with complex interactions between variables and can handle high-dimensional datasets with ease. Each model undergoes training on the dataset, followed by testing to assess its efficacy in accurately classifying SMS messages as either spam or ham. This comparative analysis of machine learning models, coupled with thorough data preprocessing and insightful visualizations, forms the cornerstone of this research, offering a comprehensive overview of spam detection techniques and their applicability to SMS messages.



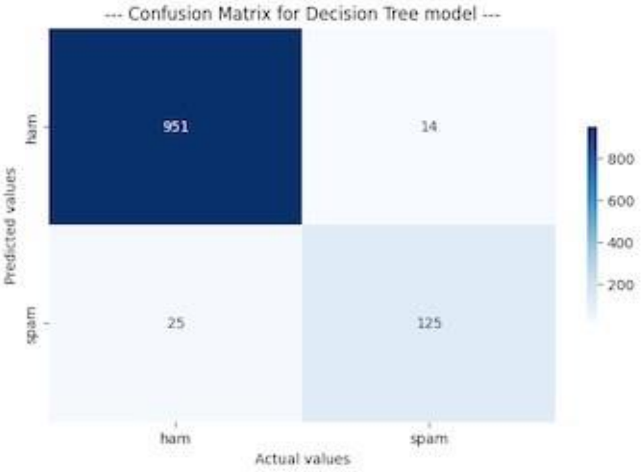
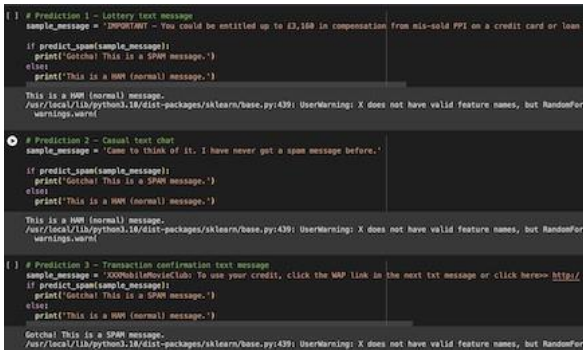


Fig. 6. Decision tree confusion matrix

Fig. 7. Random Forest confusion matrix



```
[ ] # Prediction 1 - Lottery text message
    sample_message = 'IMPORTANT - You could be entitled up to £3,160 in compensation from mis-sold PPI on a credit card or loan

    if predict_spam(sample_message):
      print('Gotcha! This is a SPAM message.')
    else:
      print('This is a HAM (normal) message.')

    This is a HAM (normal) message.
    /usr/local/lib/python3.10/dist-packages/sklearn/base.py:439: UserWarning: X does not have valid feature names, but RandomFor
      warnings.warn(

    # Prediction 2 - Casual text chat
    sample_message = 'Came to think of it. I have never got a spam message before.'

    if predict_spam(sample_message):
      print('Gotcha! This is a SPAM message.')
    else:
      print('This is a HAM (normal) message.')

    This is a HAM (normal) message.
    /usr/local/lib/python3.10/dist-packages/sklearn/base.py:439: UserWarning: X does not have valid feature names, but RandomFor
      warnings.warn(

[ ] # Prediction 3 - Transaction confirmation text message
    sample_message = 'XXXMobileMovieClub: To use your credit, click the WAP link in the next txt message or click here>> http:/
    if predict_spam(sample_message):
      print('Gotcha! This is a SPAM message.')
    else:
      print('This is a HAM (normal) message.')

    Gotcha! This is a SPAM message.
    /usr/local/lib/python3.10/dist-packages/sklearn/base.py:439: UserWarning: X does not have valid feature names, but RandomFor
```

Fig. 8. Classification Output

REFERENCES

[1] T. Jain, P. Garg, N. Chalil, A. Sinha, V. K. Verma and R. Gupta, ”SMS Spam Classification Using Machine Learning Techniques,” 2022 12th International Conference on Cloud Computing, Data Science Engineering (Confluence), Noida, India, 2022, pp. 273-279, doi: 10.1109/Confluence52989.2022.9734128. keywords: Support vector machines;Machine learning algorithms;Costs;Machine learning;Probability;Message service;Natural language processing;Spam detection;SMS spam;machine learning,

[2] S. V P, V. V, K. R and T. T. T, ”Performance Comparison of Machine Learning Algorithms in Short Message Service Spam Classification,” 2023 2nd International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA), Coimbatore, India, 2023, pp. 1-4, doi: 10.1109/ICAECA56562.2023.10199265. keywords: Support vector machines;Training;Logistic regression;Machine learning algorithms;Forestry;Filtering algorithms;Message services;SMS spam detection;spam filtering;machine learning;random forest;classification,

[3] A. Kumar and C. Fancy, ”Enhancing Security in SMS by Combining NLP Models Using Ensemble Learning for Spam Detection with Image Steganography Integration,” 2023 2nd International Conference on Edge Computing and Applications (ICECAA), Namakkal, India, 2023, pp. 583-586, doi: 10.1109/ICECAA58104.2023.10212103. keywords: Support vector machines;Steganography;Machine learning algorithms;Computational modeling;Receivers;Feature extraction;Natural language processing;Natural Language Processing;Ensemble Learning;Spam Detection;Image Steganography,

[4] P. Joseph and S. Y. Yerima, ”A comparative study of word embedding techniques for SMS spam detection,” 2022 14th International Conference on Computational Intelligence and Communication Networks (CICN), Al-Khobar, Saudi Arabia, 2022, pp. 149-155, doi: 10.1109/CICN56167.2022.10008245. keywords: Support vector machines;Unsolicited e-mail;Digital communication;Communication networks;Organizational aspects;Random forests;Computational intelligence;Spam detection;machine learning;word embedding;bag-of-words;term frequency-inverse document frequency;n-grams;word2vec;doc2vec,

[5] K. Debnath and N. Kar, ”Email Spam Detection using Deep Learning Approach,” 2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON), Faridabad, India, 2022, pp. 37-41, doi: 10.1109/COM-ITCON54601.2022.9850588. keywords: Deep learning;Support vector machines;Radio frequency;Unsolicited e-mail;Computational modeling;Bit error rate;Data preprocessing;Email Spam detection;Deep Learning;Machine Learning;LSTM;BERT,

[6] A. K and S. Halder, ”Detection of Multilingual Spam SMS Using Na¨ıveBayes Classifier,” 2023 IEEE 5th International Conference on Cybernetics, Cognition and Machine Learning Applications (ICCCMLA), Hamburg, Germany, 2023, pp. 89-94, doi: 10.1109/ICCCMLA58983.2023.10346960. keywords: Maximum likelihood estimation;System performance;User interfaces;Probability;Mobile communication;Tokenization;Real-time systems;Spam SMS;Multilingual Detection;Naive Bayes Classifier;Text Preprocessing;Feature Extraction;Language Translation,

[7] S. Gadde, A. Lakshmanarao and S. Satyanarayana, ”SMS Spam Detection using Machine Learning and Deep Learning Techniques,” 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2021, pp. 358-362, doi: 10.1109/ICACCS51430.2021.9441783. keywords: Deep learning;Communication systems;Computational modeling;Credit cards;Message service;Smart phones;Business;Short Message Service;Spam;Machine Learning;Deep Learning;LSTM;UCI,

[8] B. Sultana, Z. Afrin, F. R. Kabir and D. M. Farid, ”Bilingual Spam SMS detection using Machine Learning,” 2023 26th International Conference on Computer and Information Technology (ICCIT), Cox’s Bazar, Bangladesh, 2023, pp. 16, doi: 10.1109/ICCIT60459.2023.10441338. keywords: Support vector machines;Machine learning algorithms;Filtering;Machine learning;Forestry;Message services;Information technology;Spam SMS;Bengali Text;TF-IDF;SVM;Random Forest;Decision Tree, [9] D. Komarasamy, O. Duraisamy, M. S. S, S. Krishnamoorthy, S. Rajendran and D. M. K, ”Spam Email Filtering using Machine Learning Algorithm,” 2023 7th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2023, pp. 1-5, doi: 10.1109/ICCMC56507.2023.10083607. keywords: Support vector machines;Industries;Machine learning algorithms;Filtering;Unsolicited e-mail;Neural networks;Knowledge based systems;Spam;Ham;Probability;Filtering Techniques;Classification Algorithms,

[10] N. Sharma, ”A Methodological Study of SMS Spam Classification Using Machine Learning Algorithms,” 2022 2nd International Conference on Intelligent Technologies (CONIT), Hubli, India, 2022, pp. 1-5, doi: 10.1109/CONIT55038.2022.9848171. keywords: Support vector machines;Recurrent neural networks;Machine learning algorithms;Boosting;Data models;Mobile handsets;Random forests;Lemmatization;SMS Spam detection;Stemming;TF-IDF,

[11] A. Theodorus, T. K. Prasetyo, R. Hartono and D. Suhartono, ”Short Message Service (SMS) Spam Filtering using Machine Learning in Bahasa Indonesia,” 2021 3rd East Indonesia Conference on Computer and Information Technology (EIConCIT), Surabaya, Indonesia, 2021, pp. 199-203, doi: 10.1109/EIConCIT50028.2021.9431859. keywords: Support vector machines;Machine learning algorithms;Computational modeling;Training data;Tools;Data models;Message service;short message;spam;comparative study;machine learning;natural language processing,

[12] I. S. Mambina, J. D. Ndibwile, D. Uwimpuhwe and K. F. Michael, ”Uncovering SMS Spam in Swahili Text Using Deep Learning Approaches,” in IEEE Access, vol. 12, pp. 25164-25175, 2024, doi: 10.1109/ACCESS.2024.3365193. keywords: Phishing;Message services;Deep learning;Filtering;Training;Mobile handsets;Natural language processing;Unsolicited email;Message services;Deep learning;natural language processing;Swahili;SMS;spam detection,

[13] S. M. Gowri, G. Sharang Ramana, M. Sree Ranjani and T. Tharani, ”Detection of Telephony Spam and Scams using Recurrent Neural Network (RNN) Algorithm,” 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2021, pp. 1284-1288, doi: 10.1109/ICACCS51430.2021.9441982. keywords: Support vector machines;Deep learning;Recurrent neural networks;Machine learning algorithms;Buildings;Telephony;Prediction algorithms;Malicious;Prediction;RNN;Latency;Accuracy;Analysis,

[14] S. M. Abdulhamid et al., ”A Review on Mobile SMS Spam Filtering Techniques,” in IEEE Access, vol. 5, pp. 1565015666, 2017, doi: 10.1109/ACCESS.2017.2666785. keywords: Mobile communication;Filtering;Measurement;Unsolicited electronic mail;Databases;Benchmark testing;Review;spam;mobile SMS;access layer;service provider layer,

[15] A. R. Yeruva, D. Kamboj, P. Shankar, U. S. Aswal, A. K. Rao and C. S. Somu, ”E-mail Spam Detection Using Machine Learning – KNN,” 2022 5th International Conference on Contemporary Computing and Informatics (IC3I), Uttar Pradesh, India, 2022, pp. 1024-1028, doi: 10.1109/IC3I56241.2022.10072628. keywords: Machine learning algorithms;Filtering;Unsolicited email;Phishing;Machine learning;Filtering algorithms;Software;Email;Spam Classification;Machine Learning - KNN,

[16] A. K. Singh, S. Bhushan and S. Vij, ”Filtering spam messages and mails using fuzzy C means algorithm,” 2019 4th International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU), Ghaziabad, India, 2019, pp. 1-5, doi: 10.1109/IoTSIU.2019.8777483. keywords: Postal services;Unsolicited email;Filtering;Machine learning;Feature extraction;Classification algorithms;Spam;E-mail;Detecting;Filtering;Classification,

[17] F. Ji-Hui, L. Xu-Yao and T. Shao-Hua, ”Research on spam message recognition algorithm based on improved naive Bayes,” 2022 International Conference on Intelligent Transportation, Big Data Smart City (ICITBS), Hengyang, China, 2022, pp. 241-244, doi: 10.1109/ICITBS55627.2022.00059. keywords: Training;Text recognition;Filtering;Smart cities;Text categorization;Programming;Classification algorithms;Gaussian Bayesian classification;Spam SMS;Python;Accuracy,

[18] H. Jain and R. K. Maurya, ”A Review of SMS Spam Detection Using Features Selection,” 2022 Fifth International Conference on Computational Intelligence and Communication Technologies (CCICT), Sonepat, India, 2022, pp. 101-106, doi: 10.1109/CCiCT56684.2022.00030. keywords: Support vector machines;Social networking (online);Computational modeling;Bibliographies;Feature extraction;Communications technology;Electronic mail;Mobile;Machine learning;SMS;Classification;Feature Selection,

[19] E. Wijaya, G. Noveliora, K. D. Utami, Rojali and G. Z. Nabiilah, ”Spam Detection in Short Message Service (SMS) Using Na¨ıve Bayes, SVM, LSTM, and CNN,” 2023 10th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE), Semarang, Indonesia, 2023, pp. 431436, doi: 10.1109/ICITACEE58587.2023.10277368. keywords: Support vector machines;Machine learning algorithms;Text categorization;Neural networks;Support vector machine classification;Message services;Convolutional neural networks;Short Message Service;SMS;Spam;Machine Learning;Deep Learning;Na¨ıve Bayes;Support Vector Machine;Long Short Term Memory;Convolutional Neural Networks,

[20] N. Ramya, M. K. Devi, N. K, H. V and T. A. W. R, ”Detection of Malicious Messages from Mobile Computing Devices Using NLP and Slack Integration,” 2023 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), Chennai, India, 2023, pp. 1-6, doi: 10.1109/ICSES60034.2023.10465341. keywords: Social networking (online);Phishing;Organizations;User interfaces;Message services;Natural language processing;Mobile applications;Spam;Ham;Natural Language Processing (NLP);Internet;Slack,