SMS SPAM CLASSIFICATION USING NLP

Venkata Suraj, Gamini
University of Central Missouri
700744962

Rishitha, Putta
University of central Missouri
700744139

Shivani, Pothula
University of central Missouri
700745449

Abstract—In today's world, the ubiquity of mobile phones is undeniable, with nearly every individual possessing a device equipped with essen<al func<ons such as messaging and calling. Unfortunately, this widespread access has also led to an increase in spam communica<ons. Tradi<onally, 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 shiC has occurred from voice calls to text messages. Consequently, SMS (Short Message Service) has increasingly become inundated with unsolicited adver<sements and fraudulent offers. This surge in spam messages not only disrupts daily communica<on but also poses significant challenges in differen<a<ng legi<mate messages from spam, commonly referred to as 'ham'. Addressing this issue is cri<cal to ensure the integrity of communica<on via text messages. To combat this problem, our study employs natural language processing (NLP) techniques coupled with machine learning algorithms. By leveraging these advanced computa<onal methods, we aim to accurately classify and separate spam messages from legi<mate ones. For our research, we u<lized a dataset provided by the UCI Machine Learning Repository, applying four straighSorward classifica<on models. These models include decision trees, na¨ıve Bayes, and Random Forest. Each model was rigorously tested for its effec<veness in dis<nguishing between spam and ham messages. ACer evalua<ng the performance of these models, it became apparent that the Random Forest model was the most effec<ve, achieving an impressive accuracy rate of 99.4%. This high level of accuracy indicates that Random forest is par<cularly suited for this task, offering a reliable method for filtering spam from legi<mate text messages. Our findings highlight the poten<al of machine learning in enhancing communica<on security and efficiency by minimizing the impact of unsolicited and poten<ally harmful content.

Index Terms—SMS spam detec<on, spam filtering, machine learning, random forest, classifica<on, Natural Language Processing

I. Introduction

Today, the internet has kni,ed a complex web connecHng individuals across the globe, with mobile phones acHng as a primary conduit for this connecHvity. Advances in cellular technology have shiLed communicaHon preferences from voice calls to a more subdued yet effecHve medium: messaging. IniHally, messaging via SMS (Short Message Service) was a costly affair, reserved for urgent communicaHons. However, as these services have become more affordable, they have emerged as a fundamental technology, integral to modern communicaHon strategies.

The affordability of SMS has especially transformed communicaHon in developing countries, where mobile service providers oLen offer bulk prepaid SMS packages. These packages allow users to send unlimited messages for a minimal

cost, funcHoning like a broadcast system that reaches many

people simultaneously. This method boasts a high response rate and is seen as a personal and confidenHal communicaHon channel. Consequently, small businesses and enterprises frequently uHlize SMS for promoHonal acHviHes and public relaHons, leveraging its direct access to consumers.

However, the same characterisHcs that make SMS appealing for legiHmate 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, targeHng unsuspecHng users with phishing schemes and scams. The daily deluge of spam messages—unwanted and unsolicited—can range from merely irritaHng, with constant noHficaHons disrupHng daily life, to downright dangerous, exposing recipients to malicious links that can lead to malware infecHons, privacy breaches, and security threats.

The rapid advancements in telecommunicaHons 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 criHcal and prompt communicaHon tool. As SMS gained popularity, it also caught the a,enHon of business professionals and companies who saw an opportunity to reach a broad audience quickly and directly. This led to a dramaHc 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 miHgate 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 effecHve spam message management and removal strategies. Despite these efforts, spam via SMS remains a formidable challenge due to the relaHvely high costs and efforts associated with sending these messages compared to emails.

Spam filtering technology serves as an automated soluHon 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 detecHon models. Unlike emails, SMS messages lack a header and oLen contain abbreviaHons and informal language.

Github Link: https://github.com/shivani1789/NNDL_Project.git

These characterisHcs make SMS disHnct and somewhat more straigh[orward to handle in terms of pa,ern recogniHon.

The ubiquity of mobile phones has revoluHonized communicaHon, making Short Message Service (SMS) an indispensable tool in our daily interacHons. Despite its benefits, the widespread use of SMS has given rise to a significant issue: the proliferaHon 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 effecHve spam detecHon. TradiHonal approaches, such as various forms of Na¨ıve Bayes classifiers, have been commonly employed but oLen fall short in terms of accuracy and processing speed.

In light of these challenges, this research proposes an innovaHve approach to SMS spam detecHon that combines the strengths of Natural Language Processing (NLP) and Ensemble Learning. NLP techniques are uHlized to preprocess data and extract crucial features from SMS content, such as key phrases, syntax, and style, which are oLen indicaHve 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 predicHon accuracy by aggregaHng the outputs of mulHple classifiers to make a final decision, effecHvely reducing the likelihood of erroneous classificaHons 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 applicaHon of this ensemble model is its integraHon into an Image Steganography tool. Image Steganography is the pracHce of hiding text messages within digital images, a method increasingly used for secure communicaHon. The integraHon of the spam detecHon 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 legiHmate. This capability is parHcularly crucial if a malicious actor a,empts to exploit this covert communicaHon channel by embedding harmful content in the images.

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

Many spam SMS messages follow specific pa,erns, such as starHng with a catchy phrase to grab the recipient's a,enHon. Recognizing these pa,erns can significantly enhance the effecHveness of spam detecHon models. By training models on these unique characterisHcs, it is possible to achieve a more accurate predicHon of SMS spam compared to email spam. Machine learning algorithms can leverage these differences, focusing on the concise, pa,erned nature of SMS to refine filtering techniques and improve the precision of spam detecHon. This not only enhances user experience by reducing unwanted interrupHons but also helps in safeguarding users against potenHal scams and privacy breaches facilitated through spam messages.

In light of these issues, there is a pressing need for effecHve mechanisms to filter SMS content, disHnguishing between legiHmate 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' (legiHmate). The dataset, referred to in our study, features two a,ributes: 'v1', the label indicaHng whether a message is spam or ham, and 'v2', the text of the message itself. Both a,ributes are stored as strings. Through rigorous training and tesHng, our models aim to effecHvely categorize incoming messages, thereby enhancing the security and usability of SMS as a communicaHon tool. The methodologies, along with a comprehensive evaluaHon of each model's performance, are discussed in detail in the subsequent secHons of this paper. This work not only contributes to the academic field by exploring the applicaHon of machine learning to real-world problems but also offers pracHcal soluHons for everyday mobile phone users, aiming to miHgate the impact of SMS spam.

II. MOTIVATION

In today's digital age, the proliferaHon of mobile technology has fundamentally altered how we communicate, bringing with it immense benefits in terms of connecHvity and accessibility. However, this transformaHon 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 a,acks, and breaches of privacy. As the reliance on mobile communicaHon conHnues to grow globally, the importance of effecHvely managing and miHgaHng the risks associated with SMS spam becomes paramount. The moHvaHon behind our research is driven by the urgent need to address these challenges head-on. TradiHonal spam detecHon methods have struggled to keep pace with the sophisHcated tacHcs employed by spammers, oLen failing to accurately disHnguish between legiHmate and harmful content. This ineffecHveness stems from several factors, including the dynamic nature of spam tacHcs, the linguisHc variability of spam messages, and the limitaHons inherent in older technological approaches. Recognizing these challenges, our research aims to revoluHonize spam detecHon through the integraHon 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, extracHng and analyzing features that were previously overlooked. Ensemble Learning, on the other hand, offers a robust soluHon by combining mulHple machine learning models to improve the accuracy and reliability of spam detecHon. This synergy not only enhances the performance of spam detecHon systems but also adapts to evolving spam strategies more effecHvely. Moreover, the applicaHon of our research extends beyond tradiHonal SMS pla[orms. We propose integraHng our model into an innovaHve Image Steganography tool, providing an addiHonal layer of security for communicaHons that uHlize hidden messages. This integraHon not only broadens the scope of our study but also introduces a novel approach to securing private communicaHons in an increasingly interconnected world. Our moHvaHon is further fueled by the potenHal realworld impact of our research. By improving SMS spam detecHon, we can significantly enhance the user experience, reduce the risk of cyber threats, and preserve the integrity of mobile communicaHons. AddiHonally, by publishing our findings and sharing our methodologies, we aim to contribute to the broader field of cybersecurity, paving the way for future innovaHons and research. In pursuit of these objecHves, our research is meHculously structured to test, validate, and refine our proposed models. We are commi,ed to a rigorous scienHfic approach, leveraging extensive datasets and state-oLhe-art tesHng methodologies. Through our efforts, we aspire to set new standards in spam detecHon technology, driving forward the capabiliHes of mobile communicaHon systems to new heights.In conclusion, the moHvaHon 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 cuhng-edge technology, we are not just solving a technical problem; we are enhancing the safety and efficacy of a communicaHon medium that billions of people rely on every day. This is our mission and our contribuHon to the field of mobile communicaHons and security.

III. OBJECTIVES

- Removing special character and numbers using regular expression
- CreaHng new features e.g. word count, contains currency symbol, contains numbers.
- Removing special character and numbers using regular expression
- ConverHng the enHre 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 detecHon of spam SMS messages using various techniques and classifiers. This secHon presents a review of related work on spam SMS detecHon. In [1], the researchers presented a study focused on the classificaHon 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 effecHve spam detecHon 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 computaHon, and F1 score. The findings provide insights into the effecHveness of various machine learning methods for SMS spam classificaHon, aiding in the implementaHon of robust spam detecHon systems. In [2], the researchers conducted research on the detecHon 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 ConvoluHonal Neural Networks (CNN), to classify SMS messages as spam or non-spam. The researchers compared the efficacy of these techniques using evaluaHon metrics such as accuracy assessment, precision computaHon, recall, and F1 score. The findings highlight the effecHveness of machine learning and deep Learning approaches in SMS spam detecHon and provide insights into the suitability of different algorithms for this task.

This secHon 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 detecHon research. Lueg conducted a swiL examinaHon to determine if spam emails could be effecHvely idenHfied using informaHon filtering and retrieval technologies in a systemaHc and principled way. The goal was to simplify the development of efficient spam-filtering techniques. Today, email is a prevalent form of communicaHon for business, personal, and professional purposes. In 2018, it was esHmated that around 296 billion emails were sent daily, averaging about 130 emails per person. As internet usage and email communicaHon have increased, so too has the prevalence of spam. Historically, spam has consHtuted over fiLy percent of all email traffic, contribuHng daily to significant financial losses through various frauds. However, as indicated in the upcoming graph, there has been a noHceable decline in such email volumes since 2016. This decrease can be a,ributed to the conHnuous advancements in anH-spam technologies over recent years.

The paper in [3] presents a study focused on spam message detecHon using TFIDF (Term Frequency-Inverse Document Frequency) and a VoHng Classifier. The authors in [3] have brought out a methodology that incorporates TFIDF to bring out features from SMS and a VoHng Classifier to combine the predicHons of mulHple classifiers. The study uHlized a collecHon 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 VoHng Classifier in accurately detecHng spam SMS messages, showcasing its potenHal for pracHcal spam detecHon applicaHons. The paper in [4] brought out a new method for SMS spam detecHon using semi-supervised novelty detecHon with one-class Support Vector Machine (SVM). The researchers addressed the challenge of limited labeled data by incorporaHng unlabeled data during the training process. By leveraging the one-class SVM algorithm, the system was able to idenHfy novel and previously unseen spam messages. Experimental evaluaHons were conducted using a dataset of labeled and unlabeled SMS messages, demonstraHng the efficacy of the novel approach in accurately detecHng SMS spam. The study contributes to the area of spam SMS idenHficaHon by offering a semi- supervised method that enhances detecHon performance even with limited labeled data. The paper in [5] focuses on the detecHon and classificaHon 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 detecHon 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 applicaHon of machine learning in spam detecHon, aiding in the implementaHon of effecHve systems to combat spam messages in SMS and email pla[orms. In [6], the paper brought out a proposed approach for SMS spam detecHon and classificaHon by leveraging fog compuHng and machine learning techniques. The researchers proposed a fogaugmented architecture that offloads computaHonal tasks to fog nodes, reducing latency and improving response Hmes. They uHlized 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 detecHon with reduced processing Hme, making it suitable for real-Hme SMS spam classificaHon in fog compuHng environments. The paper in [7] invesHgates methods to improve spam detecHon in SMS messages. The study proposes the integraHon of the FP-growth algorithm and Naive Bayes Classifier to alleviate the accuracy assessment and efficacy of spam detecHon. Experimental results based on a dataset of SMS messages demonstrate that the combined approach outperforms individual classifiers based on precision computaHon, recall, and F1 score. The research contributes to the domain of SMS spam detecHon by presenHng an effecHve technique that can enhance the behavior of spam detecHon systems for mobile phone SMS services. The paper in [8] proposed a transfer learning approach for SMS spam detecHon using Na¨ıve Bayes classifier. The researchers uHlized data augmentaHon methods to expand the training dataset and improve the classifier's performance. They also applied stacking, a model ensemble technique, to combine mulHple classifiers for enhanced spam detecHon accuracy. The experiments conducted on a real-world SMS dataset demonstrated the efficacy of the proposed approach in achieving improved performance compared to tradiHonal Na¨ıve Bayes classifiers. The study contributes to the area of SMS spam detecHon by introducing a transfer learning framework that leverages augmentaHon and stacking techniques for enhanced classificaHon accuracy.

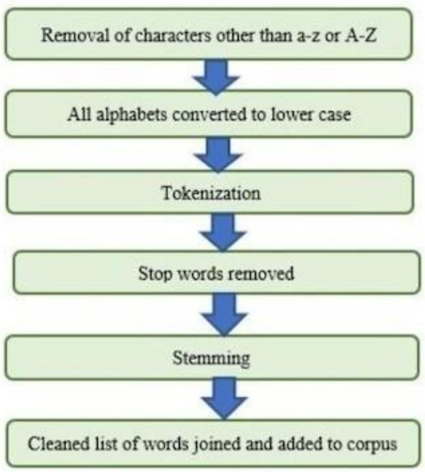
V. PROPOSED FRAMEWORK

A. A ImporAng 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 TransformaAon

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 respecHvely. The program renames the value ham to 0 and spam for 1 in the type column for easier processing and analysis. AddiHonally, 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. ConverHng text data into a form that's usable by machine learning algorithms requires transforming it into a numerical representaHon. This process, essenHal for the applicaHon of various machine learning



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

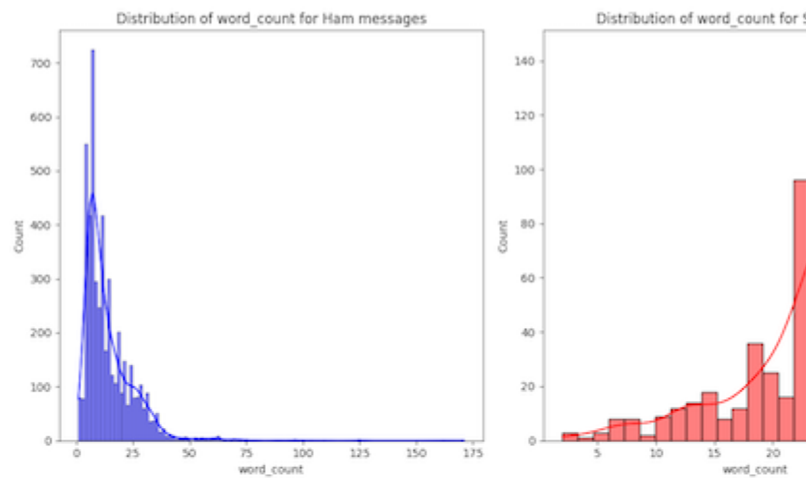
1. TokenizaHon: is the first step, where text is broken down into smaller pieces, or tokens, typically words. This is crucial for analyzing the text as it simplifies complex structures into manageable units.
2. Removing stop words: is another criHcal preprocessing step. Stop words are common, non-informaHve words such as ”is,” ”was,” and ”that.” These words are usually removed because they do not contribute significant informaHon 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. TokenizaHon: Splihng the text into individual words or tokens.
2. Vocabulary Building: CreaHng a vocabulary of unique words from these tokens.
3. CounHng: CalculaHng the frequency of each word in the vocabulary for each document.
4. VectorizaHon: RepresenHng each document as a vector, where each vector element corresponds to the count of a word from the vocabulary.
5. NormalizaHon: OpHonally, count vectors can be normalized to account for variaHons in document length or term frequency.

For feature extracHon, 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 dicHonary of known words. - A measure of the presence of these words in the documents. The BoW model is parHcularly effecHve for training and modeling as it captures the frequency of words, which is oLen a good indicator of the document’s content and context. This method proved to be very effecHve with our dataset, enabling the extracHon of relevant features for subsequent analysis and machine learning tasks. This approach is foundaHonal in transforming raw text into a structured form that is amenable to algorithmic analysis, enhancing the capability of machine learning models to make accurate predicHons or classificaHons based on textual data.



C. *Visualizing the Dataset*

This phase plays a very important role in determining the soluHon 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 informaHon 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 SpliMng the Dataset into Train and Test

IniHally 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 classificaHon 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 classificaHon model. The algorithms include LogisHc Regression, Na¨ıve Bayes, SVM and KNN. A LogisHc Regression The model was first trained using the logisHc regression algorithm. It was found to be 96.59 percent accuracy. The precision for correctly detecHng the ham messages is 0.96 and its precision for correctly detecHng 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 CollecHon Dataset.” This dataset is an essenHal resource for researchers and developers in the field of Natural Language Processing (NLP) and machine learning, parHcularly those focusing on the problem of spam detecHon in communicaHon applicaHons. The UCI Spam SMS CollecHon Dataset comprises a compilaHon of SMS messages that have been manually tagged as either ’spam’ or ’ham’ (legiHmate messages). It contains 5,572 messages, which have been collected for the purpose of building and tesHng spam filtering algorithms. Each message in the dataset is labeled accordingly, providing a clear binary classificaHon to aid in supervised learning tasks. What makes the UCI dataset parHcularly valuable for spam detecHon research is its realism and diversity. The messages reflect a variety of common themes found in spam, including adverHsements, promoHons, and phishing a,empts designed to deceive the recipient. The dataset also includes a range of informal and formal communicaHons typically seen in personal and business SMS traffic, thereby offering a comprehensive view of the types of communicaHon users may encounter. The structure of the dataset is straigh[orward, consisHng 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 applicaHon provides depth for more advanced invesHgaHons into text processing and classificaHon techniques. Researchers and developers uHlize the UCI Spam SMS CollecHon 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 oLen start with text preprocessing steps such as tokenizaHon, removal of stop words, stemming, and transformaHon 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 detecHon 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 tacHcs. The availability of such a dataset allows for benchmarking and comparison of different methods and algorithms, pushing forward the development of more effecHve and efficient spam detecHon systems. Overall, the UCI Spam SMS CollecHon Dataset not only serves as a fundamental tool for academic and industrial research but also plays a crucial role in enhancing the pracHcal capabiliHes of spam filters, thus improving the security and reliability of digital communicaHon pla[orms.

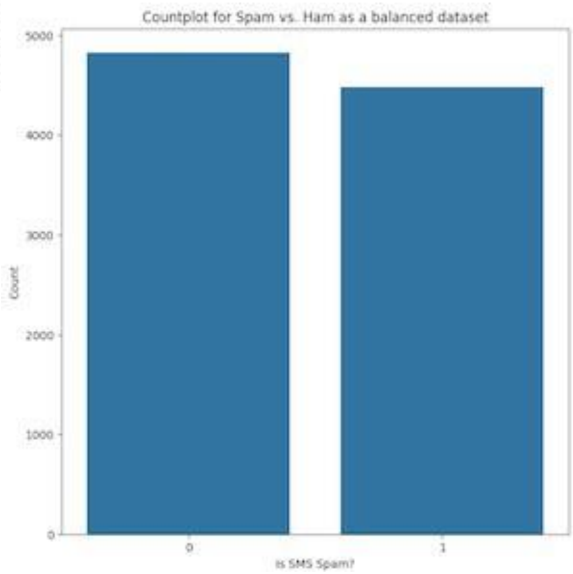


Fig. 3. dataset descripBon

Fig. 4. Balanced dataset

VII. RESULTS

This research examines the use of machine learning strategies to the process of spam filtering. Recent classificaHon 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 conjuncHon with machine learning classifiers to tackle spam. Researchers have invesHgated how spam has developed over Hme in order to trick detecHon systems. The purpose of this study is to invesHgate public datasets and performance indicators that might be uHlized in the process of evaluaHng spam filters. The difficulHes that machine learning algorithms encounter while a,empHng 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 soluHon to address the challenges has sHll remained in spam filtering; it has an accuracy rate of 99.4Through extensive tesHng, including unit tesHng,integraHon tesHng, system tesHng, and acceptance tesHng, 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 effecHvely filter out unwanted and potenHally harmful content. The use of the Random Forest has proven to be an effecHve approach in this project, as it leverages probabilisHc principles and the independence assumpHon among features to make efficient and accurate predicHons. The classifier’s simplicity and computaHonal speed make it suitable for real-Hme classificaHon of SMS messages. AddiHonally, the system incorporates language translaHon capabiliHes, allowing it to handle SMS messages in various languages, further enhancing its versaHlity and usability. The research results indicate that the detecHon of spam SMS using the Na¨ıve Bayes classifier is a viable soluHon, offering a reliable and efficient means of protecHng users

A.

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

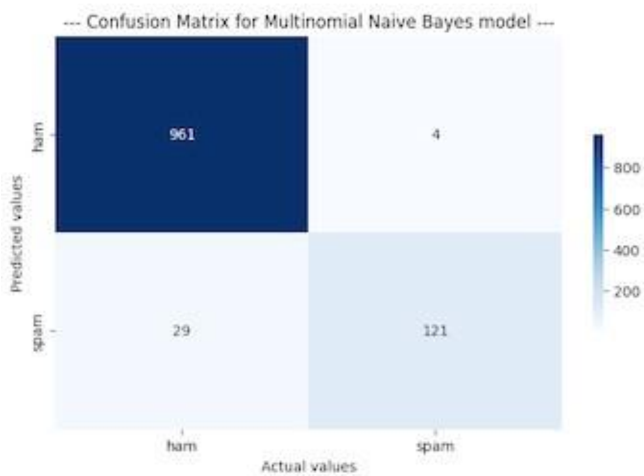


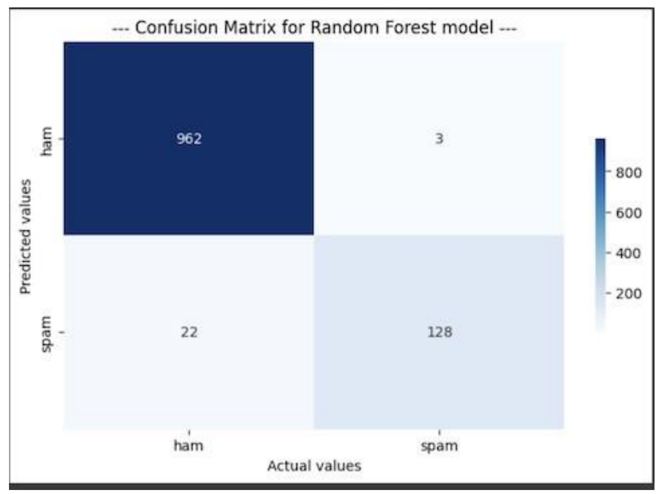
Fig. 5. Naive bayes confusion matrix

classificaHon.

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

3. Random Forest: Random Forest enhances the decision tree model by creaHng a collecHon of decision trees trained on random subsets of the dataset and features, thereby introducing randomness into the model. This approach addresses the overfihng issue common with single decision trees by averaging the results over many trees, resulHng in a more generalizable

its efficacy in accurately classifying SMS messages as either spam or ham. This comparative analysis of machine learning models, coupled with thorough data preprocess-ing and insightful visualizations, forms the cornerstone of this research, offering a comprehensive overview of spam detection techniques and their applicability to SMS messages.



and robust model. Random Forest also provides insights into feature importance, making it invaluable for understanding which features most significantly impact the classificaHon decision. It's parHcularly effecHve in environments with complex interacHons between variables and can handle high-dimensional datasets with ease. Each model undergoes training on the dataset, followed by tesHng to assess

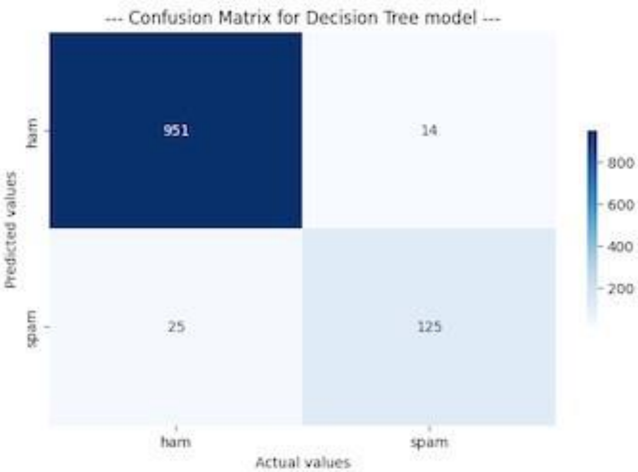
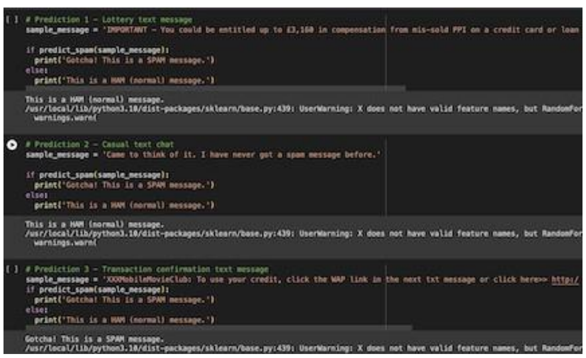


Fig. 6. Decision tree confusion matrix

Fig. 7. Random Forest confusion matrix



REFERENCES

[1] T. Jain, P. Garg, N. Chalil, A. Sinha, V. K. Verma and R. Gupta, ”SMS Spam ClassificaBon Using Machine Learning Techniques,” 2022 12th InternaBonal Conference on Cloud CompuBng, 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 detecBon;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 ClassificaBon,” 2023 2nd InternaBonal Conference on Advancements in Electrical, Electronics, CommunicaBon, CompuBng

Fig. 8. ClassificaBon Output

and AutomaBon (ICAECA), Coimbatore, India, 2023, pp. 1-4, doi: 10.1109/ICAECA56562.2023.10199265. keywords: Support vector machines;Training;LogisBc regression;Machine learning algorithms;Forestry;Filtering algorithms;Message services;SMS spam detecBon;spam filtering;machine learning;random forest;classificaBon,

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

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

[5] K. Debnath and N. Kar, ”Email Spam DetecBon using Deep Learning Approach,” 2022 InternaBonal Conference on Machine Learning, Big Data, Cloud and Parallel CompuBng (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;ComputaBonal modeling;Bit error rate;Data preprocessing;Email Spam detecBon;Deep Learning;Machine Learning;LSTM;BERT,

[6] A. K and S. Halder, ”DetecBon of MulBlingual Spam SMS Using Na¨ıveBayes Classifier,” 2023 IEEE 5th InternaBonal Conference on CyberneBcs, CogniBon and Machine Learning ApplicaBons (ICCCMLA), Hamburg, Germany, 2023, pp. 89-94, doi: 10.1109/ICCCMLA58983.2023.10346960. keywords: Maximum likelihood esBmaBon;System performance;User interfaces;Probability;Mobile communicaBon;TokenizaBon;Real-Bme systems;Spam SMS;MulBlingual DetecBon;Naive Bayes Classifier;Text Preprocessing;Feature ExtracBon;Language TranslaBon,

[7] S. Gadde, A. Lakshmanarao and S. Satyanarayana, ”SMS Spam DetecBon using Machine Learning and Deep Learning Techniques,” 2021 7th InternaBonal Conference on Advanced CompuBng and CommunicaBon Systems (ICACCS), Coimbatore, India, 2021, pp. 358-362, doi: 10.1109/ICACCS51430.2021.9441783. keywords: Deep learning;CommunicaBon systems;ComputaBonal 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 detecBon using Machine Learning,” 2023 26th InternaBonal Conference on Computer and InformaBon 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;InformaBon 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 InternaBonal Conference on CompuBng Methodologies and CommunicaBon (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;ClassificaBon Algorithms,

[10] N. Sharma, ”A Methodological Study of SMS Spam ClassificaBon Using Machine Learning Algorithms,” 2022 2nd InternaBonal 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;BoosBng;Data models;Mobile handsets;Random forests;LemmaBzaBon;SMS Spam detecBon;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 InformaBon Technology (EIConCIT), Surabaya, Indonesia, 2021, pp. 199-203, doi: 10.1109/EIConCIT50028.2021.9431859. keywords: Support vector machines;Machine learning algorithms;ComputaBonal modeling;Training data;Tools;Data models;Message service;short message;spam;comparaBve 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 detecBon,

[13] S. M. Gowri, G. Sharang Ramana, M. Sree Ranjani and T. Tharani, ”DetecBon of Telephony Spam and Scams using Recurrent Neural Network (RNN) Algorithm,” 2021 7th InternaBonal Conference on Advanced CompuBng and CommunicaBon 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;PredicBon algorithms;Malicious;PredicBon;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 communicaBon;Filtering;Measurement;Unsolicited electronic mail;Databases;Benchmark tesBng;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 DetecBon Using Machine Learning – KNN,” 2022 5th InternaBonal Conference on Contemporary CompuBng and InformaBcs (IC3I), Uoar Pradesh, India, 2022, pp. 1024-1028, doi: 10.1109/IC3I56241.2022.10072628. keywords: Machine learning algorithms;Filtering;Unsolicited email;Phishing;Machine learning;Filtering algorithms;Sopware;Email;Spam ClassificaBon;Machine Learning - KNN,

[16] A. K. Singh, S. Bhushan and S. Vij, ”Filtering spam messages and mails using fuzzy C means algorithm,” 2019 4th InternaBonal Conference on Internet of Things: Smart InnovaBon 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 extracBon;ClassificaBon algorithms;Spam;E-mail;DetecBng;Filtering;ClassificaBon,

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

[18] H. Jain and R. K. Maurya, ”A Review of SMS Spam DetecBon Using Features SelecBon,” 2022 Fiph InternaBonal Conference on ComputaBonal Intelligence and CommunicaBon Technologies (CCICT), Sonepat, India, 2022, pp. 101-106, doi: 10.1109/CCiCT56684.2022.00030. keywords: Support vector machines;Social networking (online);ComputaBonal modeling;Bibliographies;Feature extracBon;CommunicaBons technology;Electronic mail;Mobile;Machine learning;SMS;ClassificaBon;Feature SelecBon,

[19] E. Wijaya, G. Noveliora, K. D. Utami, Rojali and G. Z. Nabiilah, ”Spam DetecBon in Short Message Service (SMS) Using Na¨ıve Bayes, SVM, LSTM, and CNN,” 2023 10th InternaBonal Conference on InformaBon 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 categorizaBon;Neural networks;Support vector machine classificaBon;Message services;ConvoluBonal neural networks;Short Message Service;SMS;Spam;Machine Learning;Deep Learning;Na¨ıve Bayes;Support Vector Machine;Long Short Term Memory;ConvoluBonal Neural Networks,

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