Machine Learning Techniques on Mobile SMS Spam Detection

Abstract

Unsolicited mass sms or fraudulent sms delivered to people or organisations are known as spam. To prevent data breaches and invasions of privacy, spam texts must be recognized and eliminated. Scholars are consistently investigating machine learning approaches and strategies to efficiently distinguish and categorise spam sms from authentic ones, often known as "ham" sms. Researchers have built systems that can accurately classify sms as spam or ham by analysing numerous textual elements. This study assesses the accuracy of several classification techniques in identifying spam from valid sms by analysing data gathered from multiple sources. sms are filtered and categorised using Natural Language Processing (NLP) algorithms according to their content.The Extreme Learning Machine (ELM) is one instance of a machine learning model used for this purpose. ELM is the state-of-the-art feedforward neural network technique with a single hidden layer. ELM avoids overfitting problems and has quick training times compared to standard neural networks. Because ELM only needs one iteration cycle, spam detection using it is both practical and efficient. This paper concludes by reviewing and contrasting a number of machine learning techniques for spam detection, emphasising the efficiency and adaptability of strategies like ELM in protecting against spam sms on a variety of domains.

Keywords: SMS detection, Spam Detection, Machine Learning Algorithms Analysis, Natural Language Processing.

I. INTRODUCTION

Technological advancement is intrinsically tied to modern progress. The use of SMS and the internet for communication and information sharing is continuously rising. But there's also a deluge of uninvited bulk messages, or spam, in addition to the important information. Oftentimes, these spam emails promote lotteries or incentives while simultaneously advertising products, questionable websites, or hoaxes. They cause security issues owing to potential malware infestations, impede internet speed, eat up precious memory, and deflect our focus from important communications. Spam detection requires a lot of time and effort to do manually. Because of this, large businesses rely on spam detection software, which often use techniques like Naive Bayesian analysis to identify spam phrases.

According to social websites experts, 40% of social website accounts are fraudulently utilised for spam. Spammers post articles with hidden links to review or fan pages, using popular technologies to target specific demographics. They create phoney accounts to market inappropriate goods and services and send pornographic SMS messages to groups. By using pattern analysis, we can make spam detection more accurate. Artificial intelligence helps distinguish between spam and non-spam groups in SMS messages by using features extracted from the subject, headers, and body of messages. Using this technique, we can effectively categorise communications based on their content.

Experts in social networking claim that 40% of social network accounts are used fraudulently for spam purposes. Spammers use common technology to target particular demographics, putting content on review or fan pages that contains hidden links. They send lewd SMS messages to groups and use fictitious accounts to promote improper goods and services. Through pattern analysis, we can improve spam detection. Using characteristics taken from the subject, headers, and body of messages, artificial intelligence assists in dividing SMS messages into spam and non-spam groups. With this method, we may efficiently group messages according to their content.

An example of a learning-based paradigm is Extreme. a learning machine (ELM). is a modern model for machine learning that only uses feedforward neural networks. It eliminates one hidden layer and the sluggish training pace. overfitting issues in comparison to typical neural There is only one iteration cycle required in ELM. Because of its enhanced resilience, capacity for generalisation, and Specifically, this method is now used in In this study, we investigate various machine learning techniques for spam identification.

1. The design of several machine learning-based spam filters is examined in this research, along with their benefits

and drawbacks. It also covers the basic elements of spam text messages.

2. After a thorough analysis of the suggested techniques and the makeup of spam, a number of interesting research gaps in the field of spam detection and filtering were found.
3. Using machine learning approaches, open research challenges and future prospects are investigated to improve SMS security and spam SMS filtration.
4. The paper also discusses the difficulties that spam filtering algorithms are now facing and how these difficulties affect their efficacy.
5. A thorough analysis of various machine learning ideas and techniques clarifies the function of machine learning in spam identification.

SPAM MESSAGES

Because different people have different attitudes about sms, the term "spam" is deceptive when referring to it. sms spam is the topic of everyone's attention right now.

Generally, sms spam is made up of specific, impulsive messages that people send in large quantities. You lack knowledge. The name "spam" originates from a Monty Python animation [23] that heavily emphasises the canned beef product from Hormel. Although the term "spam" was purportedly coined in 1978 to make an unwanted allusion As we approach the mid-1990s, sms usage increased dramatically. outside of academic and research groups is becoming more and more widespread [24]. One noteworthy model is the growth.

Techniques for screening spam on IoT systems and sms.

Spam SMS is becoming more and more prevalent in marketing, finance, politics, education, and chain messaging [24]. Different sectors create algorithms and filtering techniques to effectively detect spam and understand the filtering process. Read the next section, where we go over filtering strategies in various methodologies, to get a better understanding of this process.

Common Spam Filtering Technique

A rule-based classifier called standard spam filtering serves as an example of a typical spam detection strategy. The next stage is to implement content filters that recognise spam using AI techniques. Header information is extracted via an SMS header filter. After then, an SMS is filtered using a blacklist to identify spam. Following this phase, the sender is identified using rule-based filtering based on user-defined parameters and the subject line. The last step is to apply a job and permission filter.

filtered spam from the customer's perspective.

A filtration system that follows a set of rules and follows protocols to achieve A person who can send or receive SMS and has access to the Internet or an SMS network is received. Several guidelines and techniques are available for ensuring secure communication transmission between people and organisations through spam identification at the client point. A client needs to install several functional frameworks on their computer in order to communicate data. By connecting to client SMS agents and composing, receiving, and handling incoming SMS, these systems filter the client's inbox.

Commercial-Level Detection for Spam.

Installing several filtering frameworks on the server, corresponding with the SMS transfer agent, and classifying the collected SMS into a single spam or ham is system are all part of enterprise-level SMS spam detection. A criterion that is now employed by spam detection systems is utilised to rate the SMS.

This idea makes it possible to rate every post and create a ranking system. Every spam or unwanted SMS message receives a unique score. Since spammers use a range of tactics, the adoption of a list

-based method to automatically block the messages regularly modifies all jobs.

Table 1 : Spam Categories

Categories	Descriptions
Health	The widespread use of fake pharmaceuticals.
Products Promotion	The proliferation of phoney watches, purses, and apparel.
Adult content	The rise of adult content that features prostitution and pornography.
Marketing and accounts	The profusion of stock manipulation, tax scams, and loan offers.
Fraud	Fraudulent SMS messages intended to pilfer money.

II. LITERATURE SURVEY

Navaney, P., Dubey, G., & Rana, A. [1] "SMS Spam Filtering Using Supervised Machine Learning Algorithms," presented at the 2018 8th International Conference on Cloud Computing, Data Science & Engineering (Confluence) in Noida, India. This work stands as a notable contribution in the field of SMS spam detection, focusing on the implementation of supervised machine learning algorithms to address the pervasive issue of SMS spam. By leveraging these algorithms, the research aims to refine and optimize spam filtering techniques within the realm of SMS communication.

The study delves into the exploration and application of supervised machine learning methodologies tailored explicitly for SMS spam filtering. Through meticulous analysis and experimentation, Navaney, Dubey, and Rana provide a comprehensive understanding of the efficacy and adaptability of supervised machine learning algorithms in combatting SMS spam.

Their work serves as a valuable resource for researchers, practitioners, and industry experts involved in mobile communication security. The methodologies and findings outlined in this paper contribute significantly to the ongoing pursuit of more robust and effective SMS spam filters, thereby enhancing the overall user experience and security in mobile communications.

Ubale, G., & Gaikwad, S. [2] "SMS Spam Detection Using TFIDF and Voting Classifier," presented at the 2022 International Mobile and Embedded Technology Conference (MECON) in Noida, India. This research paper focuses on SMS spam detection and employs TFIDF (Term Frequency-Inverse Document Frequency) coupled with a Voting Classifier. The study meticulously explores the practical implementation of these techniques, shedding light on their potential for enhancing the efficacy of existing spam detection mechanisms.

By leveraging TFIDF, the research underscores the pivotal role of feature extraction in distinguishing spam content from legitimate messages. Additionally, the integration of a Voting Classifier highlights the effectiveness of ensemble learning in bolstering the precision of spam detection models. This integration accentuates the importance of amalgamating diverse approaches for more robust and accurate spam identification systems.

Subasi, A., Alzahrani, S., Aljuhani, A., & Aljedani, M. [3] "Comparison of Decision Tree Algorithms for Spam E-mail Filtering," presented at the 2018 1st International Conference on Computer Applications & Information Security (ICCAIS) in Riyadh, Saudi Arabia. This research undertook a meticulous comparative analysis, focusing on various decision tree algorithms employed in filtering spam emails. The primary objective was to discern the effectiveness and efficiency of these algorithms in combating the pervasive issue of email-based spam.

Through rigorous evaluation and comparison of diverse decision tree models, the study aimed to offer insights into their relative strengths and weaknesses in mitigating spam. By dissecting the performance metrics and nuances of these algorithms, the research provides valuable benchmarks and guidelines for researchers and practitioners in the realm of email security. The findings serve as a fundamental reference point, facilitating the development and optimization of robust spam filtering mechanisms.

III. METHODOLOGY USED

- **DATA PREPROCESSING**
- **EXPLORATORY DATA ANALYSIS**
- **FEATURE EXTRACTION**

- **EVALUATION OF MODEL PREDICTION**

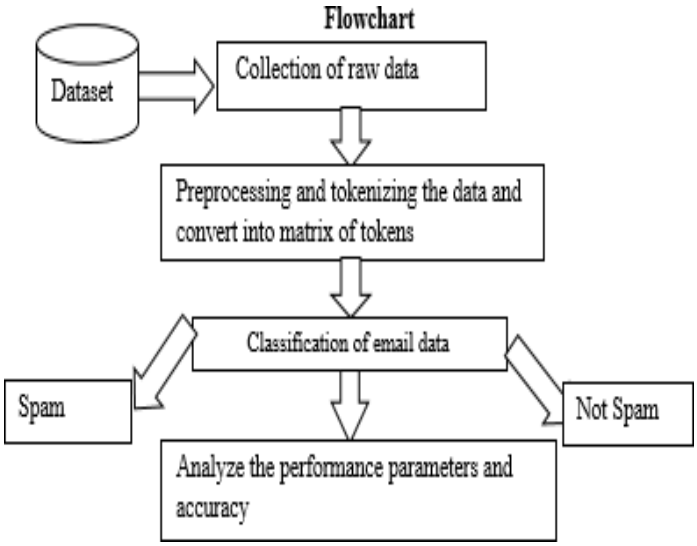


Figure 1: Flowchart for Identifying SMS Spam

DATASET

The study dataset can be found on Kaggle, a popular machine learning repository. There are 4, characteristics and 5,572 instances in the 'Spam' dataset. Of these, 672 SMS messages are classified as spam and 4,900 as ham. Table 2 contains the dataspecifics

.

DATA PREPROCESSING

In machine learning, data cleansing is essential; if done incorrectly, the results of the models are erroneous.

Among the advantages of data cleaning are:,

- enhanced ability to make decisions
- Time conservation
- Increased output
- streamlined methods of doing business
- Increased income

Operation for preprocessing your dataset

STEP 1: Remove any unnecessary data from the data set.

	v1	v2
760	spam	Romantic Paris. 2 nights, 2 flights from å£79 .....
196	ham	Romantic Paris. 2 nights, 2 flights from å£79 ...
4384	spam	Do you want a New Nokia 3510i Colour Phone Del...
5558	ham	Sorry, I'll call later
4958	ham	What i mean was i left too early to check, cos...

Table 2: After Preprocessing Data Table

STEP 2:

Upon renaming the data appropriately, we proceed with the task

	target	sms
3907	ham	Sounds like a plan! Cardiff is still here and ...
478	ham	K, can I pick up another 8th when you're done?
1484	ham	Sorry, I'll call later
3716	ham	I'm gonna rip out my uterus.
4147	spam	Please call Amanda with regard to renewing or ...

STEP 3:
In simplifying the dataset, we employ the LabelEncoder class to label the given data, reducing model complexity. Spam instances are encoded as 1, while ham instances are encoded as 0. By utilizing the LabelEncoder, we streamline the data representation, enhancing model interpretability and facilitating more effective classification.

	target	sms
0	0	Go until jurong point, crazy.. Available only ...
1	0	Ok lar... Joking wif u oni...
2	1	Free entry in 2 a wkly comp to win FA Cup fina...
3	0	U dun say so early hor... U c already then say...
4	0	Nah I don't think he goes to usf, he lives aro...

Table 3: After Labeling of target column

STEP 5: We eliminate duplicate data from the provided dataset throughout this procedure.

```
In [36]: df.duplicated().sum()
Out[36]: 403

In [37]: df = df.drop_duplicates(keep='first')

In [38]: df.duplicated().sum()
Out[38]: 0
```

EXPLORATORY DATA ANALYSIS

The process of looking over or comprehending the data and drawing conclusions or key features is known as exploratory data analysis. Graphical and non-graphical analysis are the two categories into which exploratory data analysis is divided.

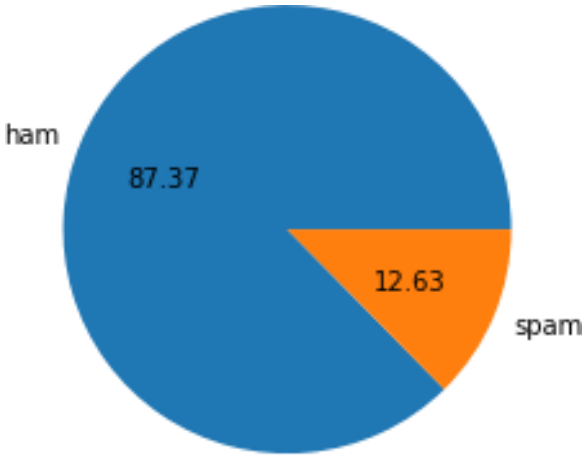


Figure 2: Pie plot of Ham and Spam Percentage

1. The proportion of spam and ham in our data set is displayed in the provided pie graphic.
2. To improve modelling, we augment our data set with other features. For instance, we may determine the overall character count, word count, and sentence count of a given SMS.

	Sender	Text	Unnamed: 4	num_characters	num_words	num_sentences
0	0	Go until jurong point, crazy.. Available only ...	NaN	111	24	2
1	0	Ok lar... Joking wif u oni...	NaN	29	8	2
2	1	Free entry in 2 a wkly comp to win FA Cup fina...	NaN	155	37	2
3	0	U dun say so early hor... U c already then say...	NaN	49	13	1
4	0	Nah I don't think he goes to usf, he lives aro...	NaN	61	15	1

3. The provided information about Ham in dataset

```
ds[ds['target'] == 0][['num_char','num_words','num_sent']].describe()
```

[34]:

	num_char	num_words	num_sent
count	4516.000000	4516.000000	4516.000000
mean	70.459256	17.123782	1.820195
std	56.358207	13.493970	1.383657
min	2.000000	1.000000	1.000000
25%	34.000000	8.000000	1.000000
50%	52.000000	13.000000	1.000000
75%	90.000000	22.000000	2.000000
max	910.000000	220.000000	38.000000

4. The provided information about Spam in dataset

```
ds[ds['target'] == 1][['num_char','num_words','num_sent']].describe()
```

	num_char	num_words	num_sent
count	653.000000	653.000000	653.000000
mean	137.891271	27.667688	2.970904
std	30.137753	7.008418	1.488425
min	13.000000	2.000000	1.000000
25%	132.000000	25.000000	2.000000
50%	149.000000	29.000000	3.000000
75%	157.000000	32.000000	4.000000
max	224.000000	46.000000	9.000000

5. Histogram based on the word count.

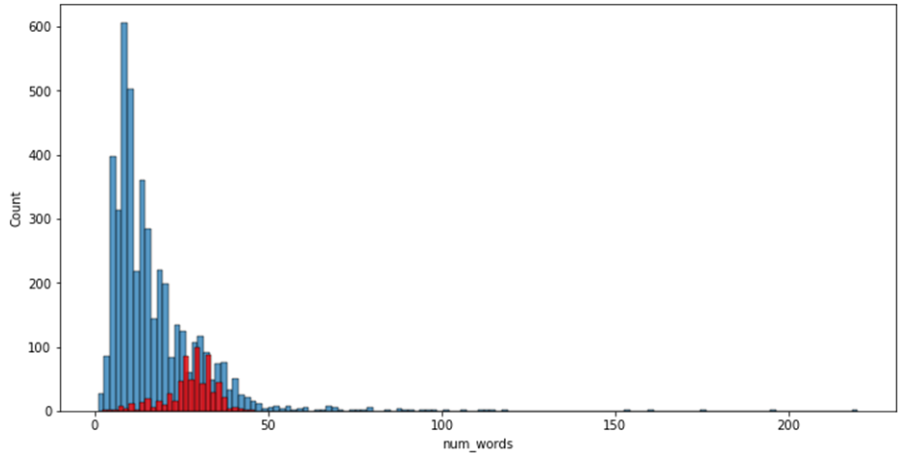


Figure 3: Based on the quantity of words used in spam and ham, a histoplot

6. Use the Spam and Ham filter to pairplot

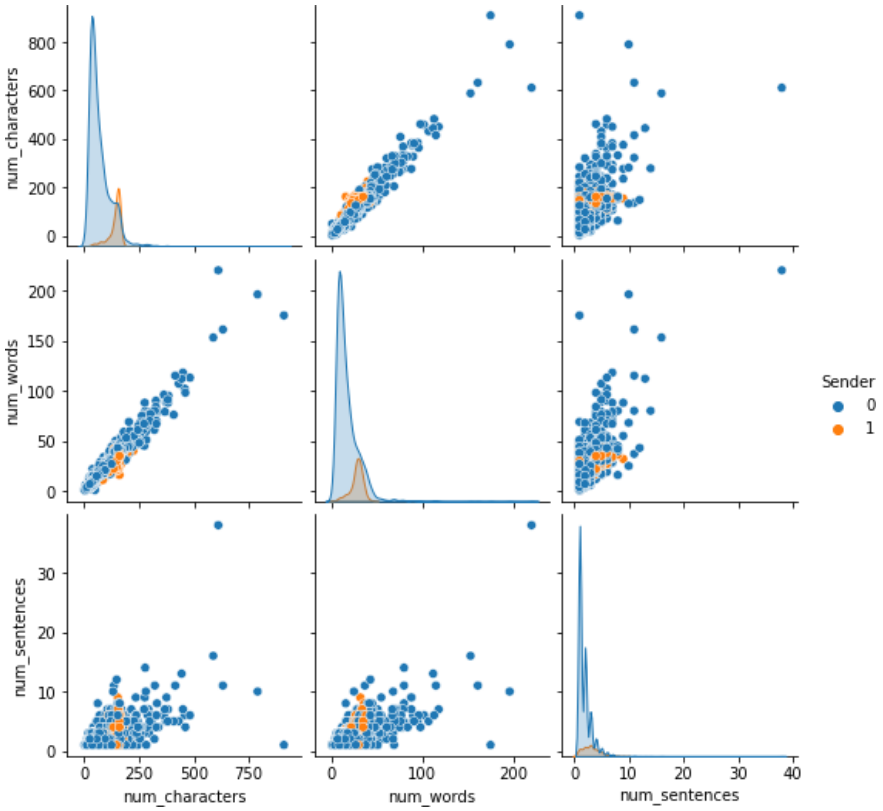


Figure 4: Pairplot using filter Spam or Ham

1. Using a heatmap to examine the relationship

According to the given heatmap, the probability of the message being spam increases with the quantity of words, with a correlation of 0.38 regarding the sender. Following are correlations of 0.26 and 0.27.

Num_characters and num_words show the strongest association.

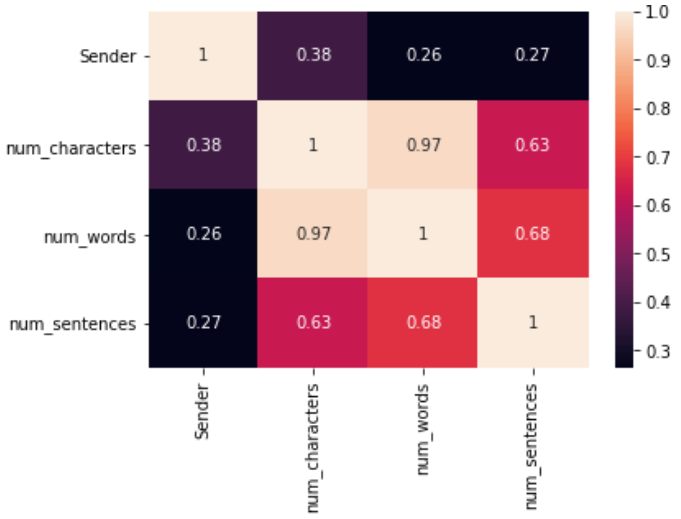
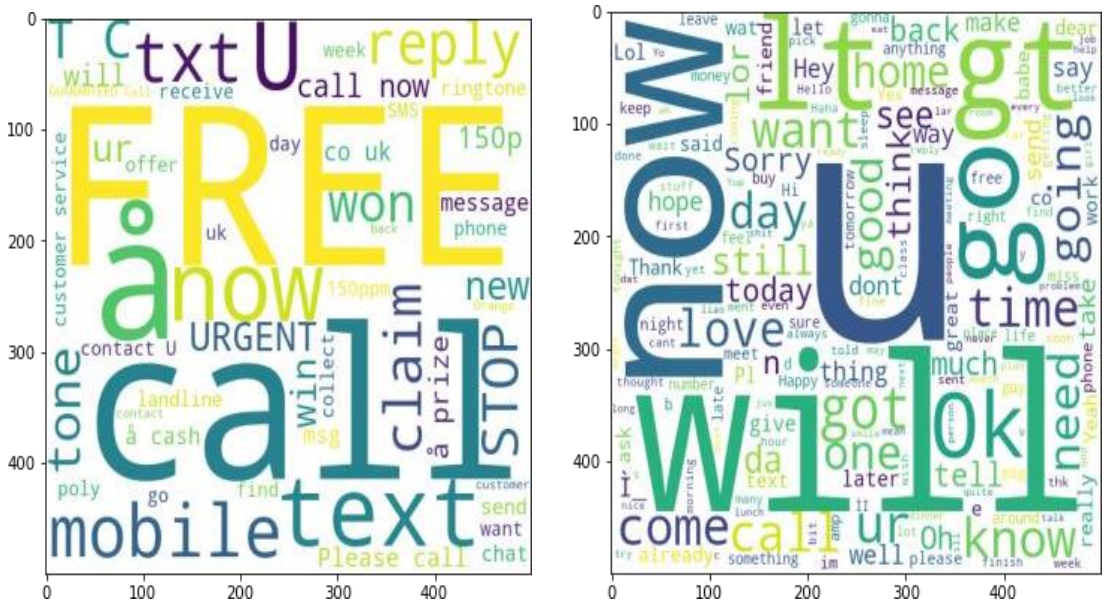


Figure 5: Correlation Analysis Heatmap

2. To determine which terms are most often used in spam and ham, respectively, we employ wordclouds.



SPAM **HAM**

Figure 6: Search Wordclouds for the Most Common Ham and Spam Words.

It turns out that the most commonly used terms in spam are free, call, now, and mobile text, while the most frequently used terms in ham are will, u, and go.

3. This highlights the quantity of commonly used terms in spam texts.

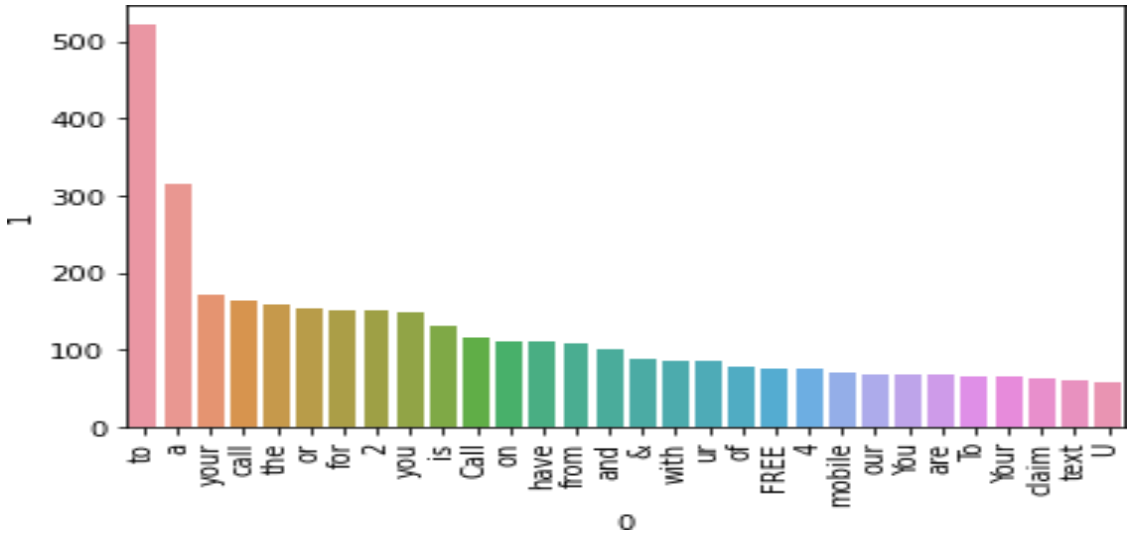


Figure 7: quantity of frequently used terms in spam texts.

4. This highlights the quantity of commonly used terms in Ham SMS.

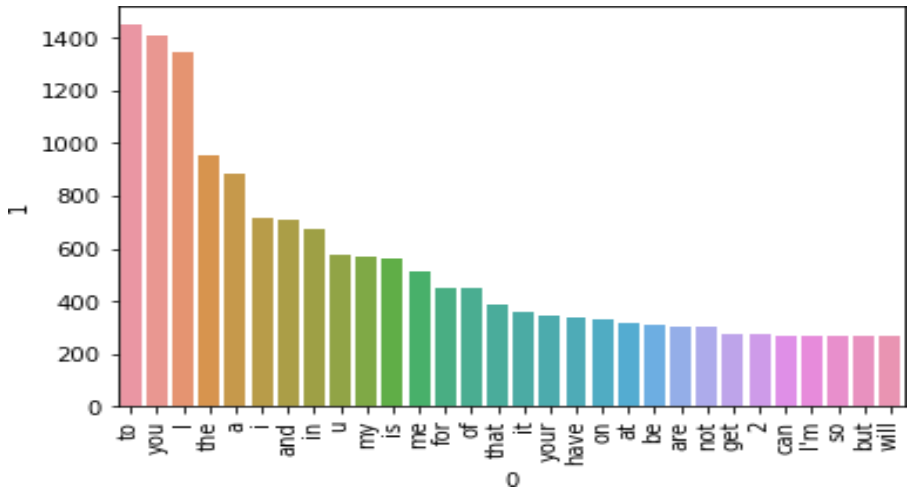


Figure 8: quantity of frequently used terms in spam texts.

MODEL BUILDING

1. converting textual input into arrays for modeling purposes.

```
In [100]: from sklearn.feature_extraction.text import CountVectorizer,TfidfVectorizer
          cv = CountVectorizer()
          tfidf = TfidfVectorizer(max_features=3000)

In [102]: X = tfidf.fit_transform(df['Text']).toarray()

In [103]: X.shape
Out[103]: (5169, 3000)

In [105]: y = df['Sender'].values
```

2. We employ train-test split alongside various algorithms, such as MultinomialNB (mnb) and BernoulliNB (bnb), to calculate the model's accuracy, confusion matrix, and precision. This process helps determine the best model for spam classification.

Bernoulli Naive Bayes

```
bnb.fit(x_train,y_train)
y_pred3 = bnb.predict(x_test)
print(accuracy_score(y_test,y_pred3))
print(confusion_matrix(y_test,y_pred3))
print(precision_score(y_test,y_pred3))
```

```
0.9819471308833011
[[1358    2]
 [  26  165]]
0.9880239520958084
```

Figure 10: Bernoulli Naive Bayes Accuracy, Precision Score with Confusion Matrix

Multinomial Naive Bayes

```
In [112]: mnb.fit(X_train,y_train)
          y_pred2 = mnb.predict(X_test)
          print(accuracy_score(y_test,y_pred2))
          print(confusion_matrix(y_test,y_pred2))
          print(precision_score(y_test,y_pred2))

          0.9738878143133463
          [[896   0]
           [ 27 111]]
          1.0
```

Figure 10: Multinomial Naive Bayes Accuracy, Precision Score with Confusion Matrix

We observe that the Multinomial model does not produce any false positives, and its accuracy is moderate. Therefore, this model is ideal for spam detection as it never misclassifies ham as spam. However, we also evaluate other algorithms to find a potentially better model for our problem. These algorithms include KNeighborsClassifier (KN), MultinomialNB (NB), DecisionTreeClassifier (DT), RandomForestClassifier (RF), AdaboostClassifier (ADABOOST), ExtraTreesClassifier (ETC), and GradientboostingClassifier (GBDT).

IV. RESULT ANALYSIS

The NB Algorithm is the most effective algorithm for identifying spam since it proposed the greatest accuracy and precision combined. As a result, a high number of spam SMS messages are recognised, the algorithm's accuracy rises, and it does not produce false positive values.

	Algorithm	Accuracy	Precision
1	KN	0.909091	1.000000
2	NB	0.971631	1.000000
3	NB2	0.981947	0.988024
0	SVC	0.974855	0.975000
5	RF	0.973565	0.974684
7	ETC	0.977434	0.969880
8	GBDT	0.950999	0.932331
6	adaBoost	0.958736	0.915033
4	DT	0.934881	0.821429

Figure 11: Algorithm Accuracy and Precision score on This Dataset

V. CONCLUSION

In today's connected world, SMS has become an essential means of communication due to its global message-sending capability. Every day, some 270 billion SMS texts are sent and received, of which 57% are spam. Spam texts, often referred to as "non-self," are uninvited, malicious, and can compromise personal information such as bank account or financial details. They can also cause harm to people, groups, or communities. Ads or links to websites that house malware or engage in phishing schemes to get user information could be included in them. Spam is a serious problem because it puts users' security at risk, annoys them, and costs money.

The spam detection feature of this project can recognise texts that include particular data. Scam text messages can be recognised by their reliable and validated domain names. For the purpose of categorizing texts and determining whether or not they are spam, the classification of spam texts is crucial. Naive Bayes has low false positive spam detection rates, which are often acceptable to consumers, making it a baseline technique for regulating spam to the unique SMS requirements of individual users. The accuracy of the entire classification process is increased by further optimizing the parameters of the Naive Bayes technique. The Naive Bayes Classifier can increase spam detection's precision.

VI. REFERENCES

1. Elchouemi, P. W. C. Prasad, A. Alsadoon, and M. K. Chae. Gain and the graph mining method are used in spam filtering and sms categorization (sfecm). The 7th IEEE Annual Computing and Communication Workshop and Conference will be held in 2017.
2. "Logistic Regression for Machine Learning," by Jason Brownlee April 1, 2016, The Machine Learning Mastery. Logistic regression using machine learning is available at https://machinelearningmastery.com.
3. ianying Zhou, Wee-Yung Chin, Rodrigo Roman, and Javier Lopez, (2007) "An Effective MultiLayered Defense Framework against Spam", Information Security Technical Report 01/2007.
4. For sms spam classification in a distributed context, K.R. Dhanaraj, V. Palaniswami, Firefly, and Bayes classifier, Aust.J. BasicAppl
5. A review of machine learning techniques to spam filtering by Guzella, T. S., and Caminhas, W. M. Appl. Expert Syst.
6. An experimental comparison of naïve Bayesian versus keyword-based anti-spam filtering using personal sms communications by Androutsopoulos, J. Koutsias, K. Chandrinos, and C.