



# Machine Learning Techniques on Mobile SMS Spam Detection

Megha Birthare<sup>1\*</sup>, Neelesh Jain<sup>2</sup>, Alpana Meena<sup>3</sup>

<sup>1</sup> SAM Global University, Bhopal India  
\*<sup>1</sup>meghabirthare@gmail.com

<sup>2</sup> SAM Global University, Bhopal India<sup>1</sup>  
sam.neelesh@gmail.com

<sup>3</sup> SAM Global University, Bhopal India  
alpana1102@gmail.com

**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.

## 1 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.

© The Author(s) 2025

S. Bhalerao et al. (eds.), *Proceedings of the International Conference on Recent Advancement and Modernization in Sustainable Intelligent Technologies & Applications (RAMSITA-2025)*, Advances in Intelligent Systems Research 192,

[https://doi.org/10.2991/978-94-6463-716-8\\_7](https://doi.org/10.2991/978-94-6463-716-8_7)

According to social websites experts, 40% of social website accounts are fraudulently utilized 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 categorize 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.

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 generalization, 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.

## 1.1 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 emphasizes 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 organizations 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 (Table 1).

### **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 utilized 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.

## 2 Literature Survey

Dubey, G., Navaney, P., and Rana, A. [1], By using supervised machine learning models, "SMS Spam Filtering Using Supervised Machine Learning Algorithms," which was presented at the 2018 8th International Conference on Cloud Computing, Data Science & Engineering (Confluence) in Noida, India, significantly aids in the detection of SMS spam. By applying these models, this study aims to improve spam filtering in SMS communications by more successfully differentiating spam from legitimate messages. Navaney, Dubey, and Rana describe their work as a valuable resource for individuals working in mobile communication security by providing in-depth insights into how supervised machine learning approaches might improve SMS spam filters through extensive experimentation. Their research supports a more safe and convenient mobile experience by improving the quality of SMS spam identification.

Gaikwad, S., and Ubale, G. [2], Term Frequency-Inverse Document Frequency (TFIDF) and a Voting Classifier are used in "SMS Spam Detection Using TFIDF and Voting Classifier," a presentation given at the 2022 International Mobile and Embedded Technology Conference (MECON) in Noida, India, to investigate SMS spam detection. This study demonstrates how well TFIDF performs feature extraction, distinguishes between spam and authentic communications, and applies a Voting Classifier to improve accuracy through ensemble learning. By combining various strategies, the study highlights how effective it is to combine several approaches for increased spam detection robustness and reliability.

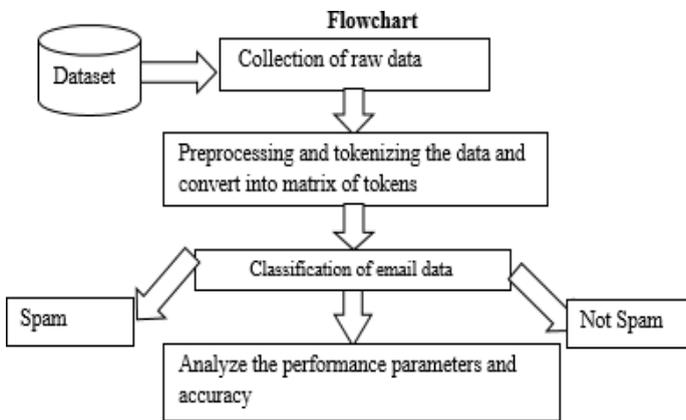
Alzahrani, S., Aljuhani, A., Subasi, A., and Aljedani, M. [3], Conducted a comparative analysis of decision tree algorithms for email spam filtering, which was presented at the 2018 1st International Conference on Computer Applications & Information Security (ICCAIS) in Riyadh, Saudi Arabia. The study compares the performance of various decision tree algorithms in filtering spam emails, evaluating the strengths and weaknesses of each approach. The research provides essential

performance benchmarks, offering a practical guide for selecting and improving spam filtering algorithms in email security systems.

### 3 Methodology Used

Fig. 1 shows the flowchart for identifying SMS spam-

- i. Data preprocessing
- ii. Exploratory data analysis
- iii. Feature extraction
- iv. Evaluation of model prediction



**Fig. 1.** Flowchart for Identifying SMS Spam

#### 3.1 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 data specified

#### 3.2 Data Processing

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

Among the advantages of data cleaning are:

- i. Enhanced ability to make decisions
- ii. Time conservation
- iii. Increased output
- iv. streamlined methods of doing business

v. Increased income

### 3.3 Operation for Preprocessing Your Dataset

**STEP 1:** Remove any unnecessary data from the data set (Table 2).

**Table 2.** After Preprocessing Data Table

	v1	v2
760	spam	Romantic Paris. 2 nights, 2 flights from â€79 .....
196	ham	Romantic Paris. 2 nights, 2 flights from â€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...

**STEP 2:** Upon renaming the data (shown in table 3) appropriately, we proceed with the task.

**Table 3.** After Labeling of target column

	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 Label Encoder 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 (Fig. 2).

	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...

**Fig 2.** Dataset Instance

**STEP 5:** We eliminate duplicate data from the provided dataset throughout this procedure(as shown in figure 3).

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

```
Out[36]: 403
```

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

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

```
Out[38]: 0
```

**Fig 3.** Python Steps for Step 5

### 3.4 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 analyses are the two categories into which exploratory data analysis is divided.

1. The proportion of spam and ham in our data set is displayed in the provided pie graphic (Fig. 4).
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

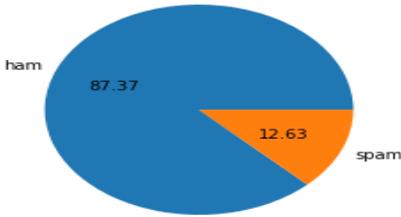


Fig 4: Pie plot of Ham and Spam Percentage

3. The provided information about Ham in dataset is shown in Fig. 5

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

Fig. 5 Ham Dataset

4. The provided information about Spam in dataset (Fig. 6)

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

[34]:

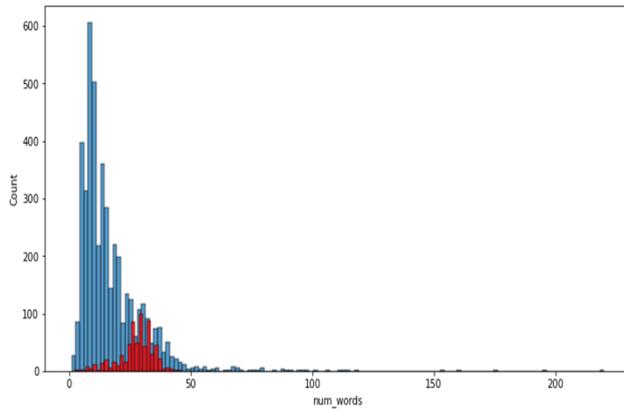
	num_char	num_words	num_sent
count	4516.000000	4516.000000	4516.000000
mean	70.459256	17.123782	1.820195
std	56.358207	13.493970	1.383657
min	2.000000	1.000000	1.000000
25%	34.000000	8.000000	1.000000
50%	52.000000	13.000000	1.000000
75%	90.000000	22.000000	2.000000
max	910.000000	220.000000	38.000000

```
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

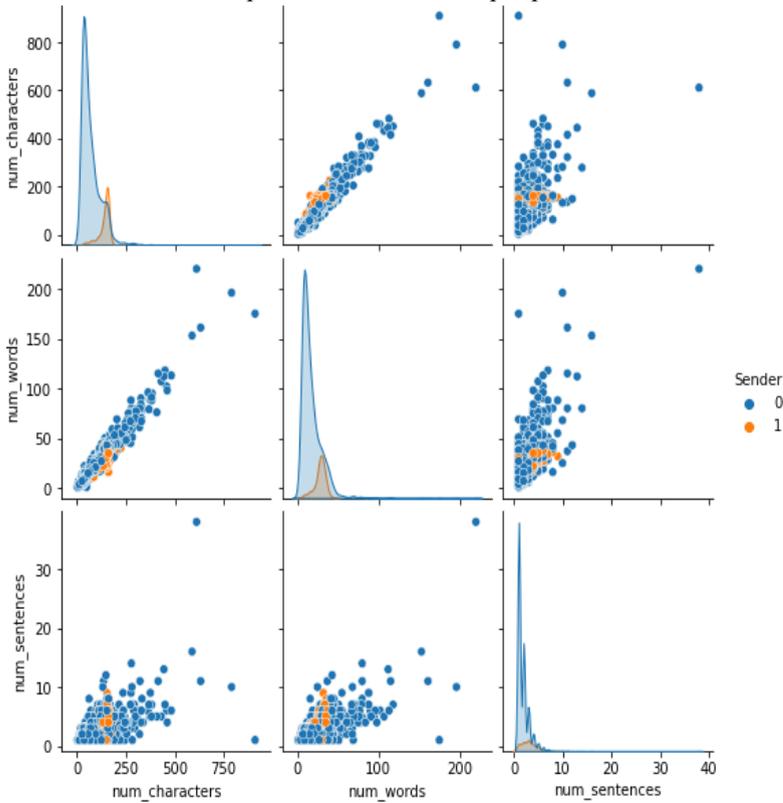
**Fig. 6** Ham Dataset

5. Histogram based on the word count (shown in Fig. 7).



**Fig. 7.** Based on the quantity of words used in spam and ham, a histoplot

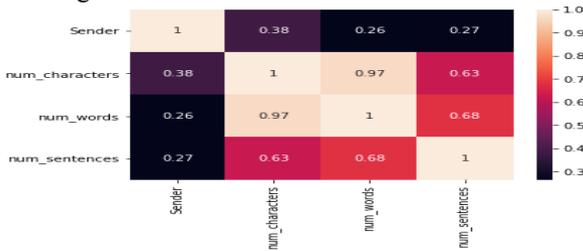
### 6. Use the Spam and Ham filter to pairplot



**Fig 8:** Pairplot using filter Spam or Ham

1. Using a heatmap (Fig. 8) 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 (Fig. 9). Following are correlations of 0.26 and 0.27. Num\_characters and num\_words show the strongest association.





1. converting textual input into arrays for modeling purposes (Fig. 12)

```
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
```

**Fig. 12.** Bernoulli Naive Bayes Accuracy, Precision Score with Confusion Matrix

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 (Fig.13 and Fig. 14) . This process helps determine the best model for spam classification.

#### 4.1 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
```

**Fig. 13.** Multinomial Naive Bayes Accuracy, Precision Score with Confusion Matrix

## 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

```

**Fig. 14** Bernoulli Naïve Bayes

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).

## 5 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 (Fig. 15), and it does not produce false positive values.

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

**Fig.15:** Algorithm Accuracy and Precision score on This Dataset

## 6 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 accounts 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.

## References

1. P. Elchouemi, W. C. Prasad, A. Alsadoon, and M. K. Chae, "Spam filtering and SMS categorization using gain and graph mining methods," *Proc. 7th IEEE Annu. Comput. Commun. Workshop Conf.*, 2017, pp. 101-106.
2. J. Brownlee, "Logistic regression for machine learning," *Machine Learning Mastery*, Apr. 1, 2016. [Online]. Available: <https://machinelearningmastery.com>
3. R. Roman, J. Lopez, I. Zhou, and W.-Y. Chin, "An effective multi-layered defense framework against spam," *Inf. Secur. Tech. Rep.*, vol. 12, no. 1, pp. 45-58, Jan. 2007.
4. K. R. Dhanaraj and V. Palaniswami, "SMS spam classification in a distributed context using Firefly and Bayes classifier," *Austr. J. Basic Appl. Sci.*, vol. 9, no. 10, pp. 25-31, 2015.
5. T. S. Guzella and W. M. Caminhas, "A review of machine learning techniques for spam filtering," *Appl. Expert Syst.*, vol. 35, no. 2, pp. 295-306, 2016.
6. G. Paliouras, K. Chandrinou, J. Androutsopoulos, and J. Koutsias, "Experimental comparison of Naïve Bayesian and keyword-based anti-spam filtering using personal SMS communications," *Comput. Intell.*, vol. 22, no. 3, pp. 487-500, 2017.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

