1. **Introduction:**

Millions of people around the world use their mobile devices to send and receive text messages on a daily basis. Due to a lack of adequate message filtering mechanisms, this form of communication is insecure. Spam is a major factor in the insecurity of mobile message communication. It is widely accepted that spam is a major issue in email and other message services. Spam is a type of unwanted email. Spam e-mails and messages, which are sent to users without their permission, are unwanted for the recipients. Different types of content, such as pornographic material or advertisements for products or services, can be found here. Spam has increased as a result of the widespread use of mobile devices for sending and receiving email and text messages. About 85 percent of the emails and texts that mobile users receive are spam. Senders pay very little for mail and messages, but recipients pay a lot. Spam's cost to service providers and end users can be calculated by looking at how much time and important mail is lost as a result of spam. Value-able e-mails are negatively impacted by spam mails and messages because users have limited Internet services, short periods of time and memory. Standard communication protocols enable the transmission of short text messages between fixed line and mobile phone devices through the Short Message Service (SMS), a text-based component of phone, web and mobile communication systems. Text message spam is becoming more and more affordable for spammers because of the growing popularity of text messaging in China and other countries, where charges have dropped to as low as US$ 0.001 or even free of charge for messages. Mobile phone spam varies greatly from region to region. In North America, for example, spam accounted for less than 1% of all SMS messages in 2010, whereas in Asia, spam accounted for up to 30% of all SMS messages. Due to the fact that some people pay for text messages, there is an inherent problem with SMS spam. Furthermore, spam-filtering software for mobile phones is hard to come by. Another issue to be concerned about is the possibility of messages of an emergency nature being blocked because they are important and legitimate. Despite this, many service providers provide ways for their customers to reduce the number of spam SMS messages they receive.

Academic researchers in this field are having difficulties in dealing with SMS spam in the same way that carriers are. For example, the evaluation of various approaches can be hampered by the absence of real and public databases. Even though significant effort has been made to generate public benchmark datasets for anti-spam filtering, mobile spam filtering still has very few corpora, usually of small size, in contrast to email spam, which has available a large variety of datasets. Since the standard SMS message is limited to 140 bytes, or 160 characters, established email spam filters may suffer a significant decrease in performance when they are directly employed in dealing with mobile spam. They also use numerous idioms and abbreviations in their writing, making it difficult to understand. If you're looking to fill an important void, this report provides the largest publicly accessible collection of real, non-encrypted SMS spam ever assembled, as far as we're aware! To provide a good baseline for future comparison, we also compare the performance of two well-established machine learning methods.

1. **Description Of Selected Dataset:**

Any scientific study would be incomplete without accurate data. The evaluation and comparison of methods can be hindered if there is a lack of representative data. The lack of publicly available data has a greater impact on recent studies because of this. There are few databases of legitimate SMS messages on the Internet, so it is difficult to find real samples of mobile phone spam for spam filtering studies. Several sources of data were employed in the process of compiling the data sets for this project.

The Grumbletext website was manually mined for its cache of 425 SMS spam messages. Most people who post complaints about spam SMS messages on this UK forum don't report the exact message they received. Many web pages had to be carefully scanned to identify spam messages in the claims, which was a time-consuming and difficult process. The NUS SMS Corpus (NSC) is a dataset of approximately 10,000 legitimate messages collected for research at the National University of Singapore's Department of Computer Science. Students at Nanyang Technological University (NTU) are the majority of those who have posted messages. Volunteers who were made aware that their contributions would be made public submitted these messages. Finally, the SMS Spam Corpus v.0.1 Big has been incorporated. Publicly accessible, it contains 1,002 ham and 322 spam SMS messages.

The dataset “SMS spam collection dataset” contains 5572 instances and two attributes v1 and v2. The v2 is the input messages which are either spam or nonspam. The predicted label v1 has two classes: 0 = nonspam and 1 spam. In the data, 4900 are nonspam samples and 672 are spam samples.

|  |  |  |
| --- | --- | --- |
| Message | Amount | Percentage |
| Spam | 4900 | 86.60 |
| Not A Spam | 672 | 13.40 |
| Total | 5572 | 100 % |

Table :1 Description Of Message Collection dataset

1. **Objectives Of Project:**

SMS spam messages have increased dramatically as the number of mobile phone users has grown. Because of the low SMS rate and the limited supply of spam-filtering software for mobile phones, combating mobile phone spam is difficult to execute. Spam detection is a major problem in mobile message communication, making it unsafe. By spam messages our cellphones can be hacked, and we can also loss our credit card credentials as hackers nowadays are attempting to hack our devices by spam messages. To combat this issue, a precise and accurate method for spam detection in mobile message communication is required.

1. **Evaluation Methodology:**

We have worked on the publicly available dataset which include two classes of spam and not a spam. Firstly we have evaluated the total number of messages and labels we have for the dataset. After this we have preprocessed the data by removing the stop words and punctuations from the messages. After the complete preprocessing we have extracted the features from the dataset. After feature extraction we have employed the machine learning algorithms to evaluate our model in terms of accuracy, specificity and sensitivity. We have employed two algorithms. The first one is Naïve bayes algorithm and second one is support vector machine. Among two classifiers the naïve bayes outperforms.

1. **Implementation:**

We have employed three basic steps. First one is preprocessing in which we have preprocessed the data for better results. Secondly we employed the feature extraction and lastly we have employed the classification between two classes, which message is a spam message or not a spam message.

1. **Preprocessing**

In preprocessing we have converted the upper case letters to lower case letters, secondly we employed the removal of stop words. In order to visualize the most frequently used words in Spam and Ham messages, we will create a WordCloud for Spam and Ham messages.Using NLTK lemmatization, stemming, or part-of-speech distinction, we can now normalize the text (POC). However, due to the widespread use of abbreviations and shorthand in text messages, these techniques aren't always effective. Use "IDK" to stand for "I don't know" or "wut" to stand in for "what". As a result, these methods will not be used to process the text.

1. **Feature Extraction**

After the preprocessing we have extracted the features from the preprocessed message or SMS data. Since SciKit Learn's algorithm models need a vector to work with, the messages will remain in their current state of being lists of tokens.

The bag-of-words (BOW) model will help us accomplish this in three easy steps:

* In each message, count how many times a word appears (Known as term frequency - TF)
* To make frequent tokens less weighted, count the number of times they have been used (inverse document frequency - IDF)
* To remove any traces of the original text, convert the vectors to unit length (L2 norm)

1. **Classification:**

Once the messages represented as vectors, we can finally train our spam/ham classifier. After the successful feature extraction we have then passed these features to our classifiers by splitting the data into 70% for training and 30 % for testing. We have employed the state of art machine learning classifiers which are random forest, SVM and naïve bayes. Among all the classifiers the naïve bayes outperforms.

1. **Evaluation Methodology:**

We will evaluate our proposed model for spam message classification using different performance metrics like classification report, accuracy, F1 score, recall and support .

1. **Classification Approach:**

The classification of the messages in terms of spam and not a spam is a very important factor in this era of technology because now a days hackers are attempting to hack the phones and other gadgets by spamming techniques. We have employed the state of art machine learning classification algorithm which are naïve bayes, random forest and support vector machine. The features extracted by TF-IDF were passed to the classifiers and performance was evaluated.

1. **Evaluation:**

We have evaluated our model for message/SMS classification in terms of “spam” and not spam by accuracy, specificity and sensitivity.

Classification Report Of SVM in term of accuracy, specificity, sensitivity and recall.

Chart, treemap chart

Description automatically generated

We have evaluated our model for message/SMS classification in terms of “spam” and not spam by accuracy, specificity and sensitivity.

* **Confusion Matrix Of Naïve Bayes Algorithm:**

A picture containing text

Description automatically generated

* **Accuracy Of Model by Naïve Bayes:**

A picture containing logo

Description automatically generated

1. **Summary & Conclusion:**

To protect email and message communications, spam detection is critical. Detecting spam is a major problem, and researchers have come up with a variety of methods to do so. These methods, on the other hand, lack the ability to identify spam accurately and efficiently. We've developed a spam detection method based on machine learning predictive models to address this problem. Experiments have shown that the proposed technique is highly effective at finding spam. When compared to other methods, the proposed one had a 96 percent accuracy rate. To sum it all up: The results show that the proposed method is more reliable for spam identification, both in terms of its accuracy and timeliness.