*Government arts and science collage thiruvennai nallur,villupuram.*

**OPTIMIZING SPAM FILTERING WITH MACHINE LEARNING(Project)**

**Abstract:** Machine learning is a field of inquiry devoted to understanding and building methods that "learn" – that is, methods that leverage data to improve performance on some set of tasks. It is seen as a part of artificial intelligence. Now we are going to train the system to filter the spam SMS using Machine Learning methods and techniques.

In the modern world where digitization is everywhere, SMS has become one of the most vital forms of communications, unlike other chatting-based messaging systems like Facebook, Whatsapp etc.

SMS does not require active internet connection at all. As we all know that hackers/Spammers tries to intrude in mobile Computing device. And SMS support for mobile devices had become vulnerable, as attacker tries to intrude to the system by sending unwanted link, with which on clicking those link the attacker can again remote access over the mobile computing device. So, to identify those messages Authors have developed a system which will identify such malicious messages and will identify whether or not the message is SPAM or HAM (malicious or not malicious).

Filtering SMS Spam and Implementation in Android. We implement these techniques in an Android application. In this project, we attempt to find a technique to filter spam SMSs by studying spam features and the way in which spammers send spam messages. So we filter spam messages based on the length of sender number and content of the SMS. We also check if a message contains spam features such as spam keywords, some special characters or URLs and using machine learning algorithms. To check if an SMS contains spam keywords, we first create a list of common spam keywords and assign weight to each of them according to their frequency and the android application will optimize the spam SMS and helps us to view the relevant SMS.

Keywords: SMS, SPAM and HAM, Machine Learning, Text Classification, Data Set.

**MILESTONE 1: Define Problem / Problem Understanding.**

I ) Specify the business problem:

Most of the SMS spam includes commercial advertisements and promotions of service plans. Such SMS distract user, waste their time in reading and deleting them and flood the user inbox. Some spam messages are sent with the intent of hacking or malware attack which imposes higher risk to smartphone users as smartphone are connected to internet and Bluetooth devices that contains authentic information. So, there is risk of financial fraud and malware downloads. Over the last year, on average, there have been: 376,032,773 spam texts per day. Our Business problem aims at filtering the Spam SMS by using the Machine Learning Technique. Spam Filtering process is analyzed in two Folds A) Service Providers B) User with machine learning algorithm we indent to develop a Android application that filters the Spam SMS by effective predictive model.

II) Business requirements:

The basic business requirement for optimizing Spam Filtering can filter a spam messages with highest efficiency using android application This model includes various factor such aslength of mobile numbers, content of the SMS, spam keywords and URL.

III) Literature survey:

[1]In "Sneha Dalvi University of Georgia , Georgia, USA" paper titled "Filter SMS Spam and Implementation in Android" In this research project, They Studied and worked different spam features such as spam keywords, special characters ,URLs and used them to determine whether an SMS is spam or not. It was found that the accuracy of the system was 55.96% and false positive rate was 1.05%.

[2]In “Suparna DasGupta , Soumyabrata Saha and Suman Kumar Das”

That paper titled” Spam SMS Detection Using Machine learning. From the above discussion and experiment authors have a concluded that machine learning algorithms can play a vital role in identifying SPAM SMS. The accurancy obtained in this work is more than 70%.

[3] Gmez Hidalgo.J.M., Cajigas Bringas.G., Puertas Sanz.E., Carrero Garca,F presented detailed note on Content Based SMS Spam Filtering. Proceedings of the 2006 ACM Symposium on Document Engineering (ACM DOCENG’06), Amsterdam, The Netherlands 10-13, 2006.

IV) Social or Business Impact:

Optimizing Spam SMS filter Using Machine Learning models aims at filtering the Spam SMS and presenting only the required SMS for the users. This model enable the people to avoid viewing the unwanted SMS messages and all the filtered messages stands important.

Business Impact:

By filtering the Spam SMS, Essential SMS and Detail alone are viewed which we promote the service provider business aspect using android application.