

MES COLLEGE OF ENGINEERING, KUTTIPPURAM
DEPARTMENT OF COMPUTER APPLICATIONS
20MCA246 – MAIN PROJECT

PRO FORMAT FOR THE APPROVAL OF THE FOURTH SEMESTER MAIN PROJECT

(Note: All entries of the pro forma for approval should be filled up with appropriate and complete information. Incomplete Pro forma of approval in any respect will be rejected.)

Main Project Proposal No : _____
(Filled by the Department)

Academic Year : 2020-2022


Year of Admission : 2020

1. Title of the Project : SMS SPAM DETECTION USING MACHINE LEARNING

2. Name of the Guide : Mr. MOHAMMAD JABIR C

3. Number of the Student: MES20MCA-2015

4. Student Details

Name (in BLOCK LETTERS)	Roll Number	Signature
1. <u>DEEPIKA BALAKRISHNAN C</u>	<u>15</u>	

Date: 16/04/2022

Approval Status : Approved / Not Approved ____

Signature of
Committee Members }

Comments of The Mini Project Guide

Dated Signature

Initial Submission :

First Review :

Second Review :

Comments of The Project Coordinator

Dated Signature

Initial Submission:

First Review

Second Review

Final Comments:

Dated

Signature of HOD

SMS SPAM DETECTION USING MACHINE LEARNING

DEEPIKA BALAKRISHNAN C

INTRODUCTION

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 / Spammer 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 gain 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). Authors have created a dictionary using the TF-IDF Vectorizer algorithm, which will include all the features of words a SPAM SMS possess, based on content of message and referring to this dictionary the system will be classifying the SMS as spam or ham.

OBJECTIVES

The daily traffic of Short Message Service (SMS) keeps increasing. As a result, it leads to dramatic increase in mobile attacks such as spammers who plague the service with spam messages sent to the groups of recipients. Mobile spams are a growing problem as the number of spams keep increasing day by day even with the filtering systems. Spams are defined as unsolicited bulk messages in various forms such as unwanted advertisements, credit opportunities or fake lottery winner notifications. Spam classification has become more challenging due to complexities of the messages imposed by spammers. Hence, various methods have been developed in order to filter spams.

TOOLS / PLATFORM, HARDWARE AND SOFTWARE REQUIREMENT

Hardware Requirements

- Input Device : Mouse, Keyboard
- Output Device : Monitor
- Memory : 4 Gb Ram (Minimum)
- Processor : Intel core i3 or above

Software Requirements

- Operating System : Windows 8 /10for Better Performance
- Front End : Python (Flask)
- Back End : MySQL
- Software Used : PyCharm

PROBLEM DEFINITION

Using the previous experience of the works done, we changed the course of our research. It was decided to take the direction of automating the process of finding texts containing a humiliating slope. Further, after that, we began to study whether there are similar works on the territory of Kazakhstan or at least on the territory of the former USSR countries. As it turned out, there are several similar works in concept, but even they did not meet our requirements. Since we decided to create an autonomous information system that could automatically detect and take actions to prevent them

BASIC FUNCTIONALITIES

- Exploring the media space where textual information is most used
- Use the methods of data collection in these portals and make an information system that makes it possible to analyze text information in real time.
- After the data has been collected, do manual marking of the data
- Experiment with conventional machine learning algorithms (Linear Regression, SVM, KNN...)
- Create a collection of words from the data set to create the core of the language
- Collect a large amount of training data
- Make a training of deep learning algorithms for natural language processing