Project Report on

Spam Text detection

at

U. V. Patel College of Engineering

Prepared by:

Madhavi Patel (18012011072)

Guide:

Prof. Ketan Sarvakar

B.Tech Semester VII
(Computer Engineering)
December 2021

Submitted to,
Department of Computer Engineering
U.V. Patel College of Engineering
Ganpat University, Kherva - 384 012

Abstract:

The number of people using mobile devices increasing day by day. SMS (short message service) is a text message service available in smartphones as well as basic phones. So, the traffic of SMS increased drastically. The spam messages also increased. The spammers try to send spam messages for their financial or business benefits like market growth, lottery ticket information, credit card information, etc. So, spam classification has special attention. In this paper, we applied various machine learning and deep learning techniques for SMS spam detection. we used a dataset from UCI and build a spam detection model. Our experimental results have shown that our LSTM model outperforms previous models in spam detection with an accuracy of 98.5%. We used python for all implementations.

Problem Description:

In three seconds, more than 90% of people read an SMS, the chances of readability are extremely high. SMS may help increase customer engagement, promote products and services, or provide viewers with urgent updates. SMS messaging is also designed to enable companies to send messages to any contact number worldwide from any website or service using an API. Thanks to its omnipresent nature, SMS spamming has become a major nuisance for mobile subscribers. It includes substantial costs in terms of lost productivity, use of network bandwidth, administration, and personal privacy attack.
So, we are going to create an SMS spam detection model which will help you to find whether an SMS is spam or not using LSTM.

Notebook used:

Colab - Colab allows anybody to write and execute arbitrary python code through the browser, and is especially well suited to machine learning, data analysis and education. Colab notebooks allow you to combine executable code and rich text in a single document, along with images, HTML, Latex and more.

Tool, Technology and Library requirements:

- **Pandas**

  pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series.

- **NumPy**

  NumPy is a library for the Python programming language, adding support for large, multi- dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays.

- **Matplotlib.pyplot**

Matplotlib.pylot is a state-based interface to matplotlib. It provides an implicit, MATLAB-like, way of plotting. It also opens figures on your screen, and acts as the figure GUI manager.

- **Keras**

Keras is an open-source software library that provides a Python interface for artificial neural networks. Keras acts as an interface for the TensorFlow library.

- **LSTM model**

LSTM networks were designed specifically to overcome the long-term dependency problem faced by recurrent neural networks RNNs.

- **Sequencial model**

Sequential is the easiest way to build a model in Keras. It allows you to build a model layer by layer. Each layer has weights that correspond to the layer the follows it.

Work Flow:

1. First we import all the required libraries such as import pandas as pd , numpy as np and matplotlib.pyplot as plt
2. data reading.
3. data preparation
4. applying RNN
5. applying LSTM

Dataset Name:

SPAM text message 20170820 - Data.csv

Dataset location:

https://www.kaggle.com/kentata/rnn-for-spam-detection/data

Github link:

https://github.com/madhavi12345678/spam_detection.git