Assignment 2

Assignment No: 2

(Total points- *60*, Evaluation weightage/marks – )

Date of Assignment: 22/10/2022

Date of Submission: 11/11/2022

XIM University

School of Computer Science and Engineering



Name: Subhashis Suara & Siddharth Das

Roll Number: UCSE19012 & UCSE19010

Submitted to: Dr. Sourav Mandal, Assistant Professor

**Evaluation**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Questions | Q-1 | Q-2 | Q-3 | Q-4 | Q-5 | Q-6 | Q-7 | Q-8 | Q-9 | Q-10 |
| Points/ Marks | 50 | 10 |  |  |  |  |  |  |  |  |
| Points acquired |  |  |  |  |  |  |  |  |  |  |

|  |  |
| --- | --- |
| Total points and final marks: |  |
| Comments: |  |
| Faculty Signature: |  |

Student(s) Signature:

1. **SMS Spam Classification**

**Points- 50**

**Your Answer:**

**Report**

**SMS Spam Detector**

A web application that tells you if a given SMS is spam or not spam. It uses a ML model trained on the UCI SMS Spam Collection Dataset and is built using Flask. Project made by Subhashis Suara & Siddharth Das.

**About Dataset**

The SMS Spam Collection is a set of SMS tagged messages that have been collected for SMS Spam research. It contains one set of SMS messages in English of 5,574 messages, tagged according being ham (legitimate) or spam.

**How does it work?**

Initially we have dropped the unnecessary columns i.e. 'Unnamed: 2', 'Unnamed: 3', 'Unnamed: 4'. We have renamed columns i.e. v1 and v2 to label and message respectively. For classification later we replaced non spam to 0 and spam to 1. We first started to preprocess our data. All the special characters are removed from the string. We have also removed the stop words. We then used Porter Stemmer for stemming the words. After joining the stemmed words we have created a corpus of messages. The next step is to create Bag of Words model which we have done using CountVectorize function. We have then fitted the multinomial Naive Bayes to the training set and then saved the model in a file.

For deployment purpose we have used Flask. Using pickle library we load the Multinomial Naive Bayes model and CountVectorizer object. First the home page is displayed. The next step is to fill the sms(message) in the text field. After clicking the detect button , flask uses the Multinomial model to verify if the message is spam or not. The output is then displayed on the webpage.

**Code Repository:** [**https://github.com/subhashissuara/sms-spam-detector**](https://github.com/subhashissuara/sms-spam-detector)