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?**

Demo Website: http://sms-spam-detector.centralindia.cloudapp.azure.com:5000

From the dataset, we dropped unnecessary columns i.e., 'Unnamed: 2', 'Unnamed: 3', 'Unnamed: 4'. Then we renamed the columns i.e., v1 and v2 to label and message respectively. For classification purpose we replaced not spam label with 0 and spam with 1. We preprocessed our data by removing all special characters and stopwords. We then used Porter Stemmer for stemming the words. After joining the stemmed words, we have created a corpus of messages. The next step was to create a Bag of Words model which we accomplished using CountVectorize function from sklearn library. We then fit the Multinomial Naive Bayes to the training set and then saved the model using the pickle module.

For deployment purpose we have used Flask. Using the pickle module, we loaded the Multinomial Naive Bayes model and the CountVectorizer object. The Flask app has 2 routes. First route is the home page route, and second route is the detect route. The home page consists of a text field to input the SMS message and a button to detect the SMS type. Upon clicking the detect button, the Flask app uses the Multinomial model to classify if the message is spam or not spam. The detect page then displays the appropriate output on the webpage.

**How to run the project?**

Ensure that you have Python installed and enter the following commands in your terminal:

git clone https://github.com/subhashissuara/sms-spam-detector

pip install -r requirements.txt

python app.py

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