Japneet Singh 2019MT10696 Semester 1 2022-23

**Report: COL774 Assignment 2 (Naïve Bayes & SVM)**

**Part 1. Text Classification**

(1a) Simple Naïve Bayes

(1b) Random and Fixed Guessing

(1c) Confusion Matrices

(1d) Naïve Bayes with **Stop Word Removal** and **Stemming**

(1e) Feature Engineering

1. Unigrams + Bigrams
2. Additional Features
   1. Unigrams + Bigrams + Trigrams
   2. Unigrams + Trigrams
3. Comparison of various feature sets

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model Description | Train Set Accuracy | Test Set Accuracy | Precision | Recall | F1 Score |
| Unigrams only |  |  |  |  |  |
| Random Prediction |  |  |  |  |  |
| Always Predicting Positive |  |  |  |  |  |
| Unigrams only (with stop word removal and stemming) |  |  |  |  |  |
| Unigrams + Bigrams |  |  |  |  |  |
| Unigrams + Bigrams +Trigrams |  |  |  |  |  |
| Unigrams + Trigrams |  |  |  |  |  |

**Part 2. Binary Image Classification**

(2a) CVXOPT with Linear Kernel

(2b) CVXOPT with Gaussian Kernel

(2c) scikit-learn SVM – Linear and Gaussian Kernels

**Part 3. Multi Class Image Classification**

(3a) CVXOPT Gaussian Kernel One-vs-One Classification

(3b) scikit-learn SVM Gaussian Kernel One-vs-One Classification

(3c) 5-fold cross validation