**Project 3: Classification Algorithms**

**CSE 601: Data Mining and Bioinformatics**

**Team members:**

Arvind Thirumurugan athirumu 50289656

Saketh Varma Pericherla sakethva 50288206

Vijay Jagannathan vijayjag 50290947

**Part 1: Implementing Classification Algorithms:**

**1.0 Results:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Dataset** | **Measures** | **Algorithms** | | | |
| **Accuracy** | **Precision** | **Recall** | **F1** |
| **1** | **K-nn** |  |  |  |  |
| **Naïve Bayes** |  |  |  |  |
| **Decision Tree** | 0.917418 | 0.884163 | 0.901467 | 0.889868 |
| **Random Forest** |  |  |  |  |
| **2** | **K-nn** |  |  |  |  |
| **Naïve Bayes** |  |  |  |  |
| **Decision Tree** | 0.599814 | 0.424779 | 0.467053 | 0.441297 |
| **Random Forest** |  |  |  |  |
| **3\_train** | **K-nn** |  |  |  |  |
| **Naïve Bayes** |  |  |  |  |
| **Decision Tree** | 0.8875 | 0.845714 | 0.787380 | 0.810158 |
| **Random Forest** | 1.0 | 0.7 | 0.7 | 0.7 |
| **3\_train** | **K-nn** |  |  |  |  |
| **Naïve Bayes** |  |  |  |  |
| **Decision Tree** | 1.0 | 1.0 | 1.0 | 1.0 |
| **Random Forest** | 1.0 | 1.0 | 1.0 | 1.0 |
| **4** | **K-nn** |  |  |  |  |
| **Naïve Bayes** |  |  |  |  |
| **Decision Tree** | 1.0 | 1.0 | 1.0 | 1.0 |
| **Random Forest** | 1.0 | 1.0 | 1.0 | 1.0 |

The above table shows the results obtained by our implementation of K-nearest neighbours, Decision Tree, Naïve Bayes and Random forest algorithms for the four datasets provided.

* Datasets 1 and 2 are trained using 10-fold cross validation while dataset 3 is trained with a separate training set and tested on a separate testing set provided. Dataset 4 is trained and tested on the same dataset due to the lack of sufficient data and a separate testing set.
* The various hyper-parameters chosen for these different algorithms to get these final results and further analysis on the results are discussed in the respective sections below.
  1. **K-Nearest Neighbors:**
  2. **Naïve Bayes:**
  3. **Decision Tree:**

**1.3.0 Algorithm Description:**

**1.3.1 Parameters:**

**1.3.2 Analysis:**

**1.3.3 Pros and Cons:**

* 1. **Random Forest:**

**1.4.0 Algorithm Description:**

**1.4.1 Parameters:**

**1.4.2 Analysis:**

**1.4.3 Pros and Cons:**