**Scoring Rubric – Programming Assessment 2 – KNN**

**Criterion Name:** *Part 1, Task 1: Dataset*

|  |  |  |
| --- | --- | --- |
| **Point Value** | **Option Name** | **Description Explanation** |
| ***1*** | *Weak* | Task is incomplete/code has not been written/cell has not been run/code contains errors. |
| ***3*** | *Satisfactory* | Task has been attempted, the code has been written and the cell has been run; however, the implementation is not fully correct. |
| ***5*** | *Strong* | Task has been completed i.e., 2 or more tweets and 2 or more stop words have been printed out from the dataset. |

**Criterion Name:** *Part 1, Task 2: Data Preprocessing*

|  |  |  |
| --- | --- | --- |
| **Point Value** | **Option Name** | **Description Explanation** |
| ***2*** | *Weak* | Task is incomplete/code has not been written/cell has not been run/code contains errors. |
| ***5*** | *Satisfactory* | Task has been attempted, the code has been written and the cell has been run; however, the implementation is not fully correct. |
| ***10*** | *Strong* | Task has been completed i.e., 2 or more tweets have been printed out from the dataset and they do not contain stop words, punctuation marks, numbers, unwanted symbols, hyperlinks, and usernames. The tweets should also all be in lowercase. |

**Criterion Name:** *Part 1, Task 3: Splitting the Dataset*

|  |  |  |
| --- | --- | --- |
| **Point Value** | **Option Name** | **Description Explanation** |
| ***1*** | *Weak* | Task is incomplete/code has not been written/cell has not been run/code contains errors. |
| ***3*** | *Satisfactory* | Task has been attempted, the code has been written and the cell has been run; however, the implementation is not fully correct. |
| ***5*** | *Strong* | The size of training and test data is correct i.e. training data contains 1304 tweets and test data contains 326 tweets. |

**Criterion Name:** *Part 1, Task 4: Feature Extraction*

|  |  |  |
| --- | --- | --- |
| **Point Value** | **Option Name** | **Description Explanation** |
| ***2*** | *Weak* | Task is incomplete/code has not been written/cell has not been run/code contains errors. |
| ***5*** | *Satisfactory* | Task has been attempted, the code has been written and the cell has been run; however, the implementation is not fully correct. |
| ***10*** | *Strong* | The vocabulary has been printed out and the bow representation for a random tweet follows the format given in question. |

**Criterion Name:** *Part 1, Task 5, 6, and 7: Create, Implement and Cross Validation*

|  |  |  |
| --- | --- | --- |
| **Point Value** | **Option Name** | **Description Explanation** |
| ***5*** | *Weak* | Task is incomplete/code has not been written/cell has not been run/code contains errors. |
| ***17*** | *Satisfactory* | Task has been attempted, the code has been written and the cell has been run; however, the implementation is not fully correct. |
| ***35*** | *Strong* | The results of cross validation are graphs for each version of the classifier contain k values on the x-axis and classification accuracy on the y-axis and a second set contains values on the x-axis and classification f1 score on the y-axis. (4 plots total, each one having values for accuracy f1 plotted for values of k from 1 to 10). |

**Criterion Name:** *Part 1, Task 8: Classification*

|  |  |  |
| --- | --- | --- |
| **Point Value** | **Option Name** | **Description Explanation** |
| ***2*** | *Weak* | Task is incomplete/code has not been written/cell has not been run/code contains errors. |
| ***5*** | *Satisfactory* | Task has been attempted, the code has been written and the cell has been run; however, the implementation is not fully correct. |
| ***10*** | *Strong* | The final accuracy after classifying the test set is above 75 and f1 score is above 60. |

**Criterion Name:** *Part 2: kNN classifier using scikit-learn*

|  |  |  |
| --- | --- | --- |
| **Point Value** | **Option Name** | **Description Explanation** |
| ***3*** | *Weak* | Task is incomplete/code has not been written/cell has not been run/code contains errors. |
| ***12*** | *Satisfactory* | Task has been attempted, the code has been written and the cell has been run; however, the implementation is not fully correct. |
| ***25*** | *Strong* | The accuracy score, classification reports and confusion matrix are shown for both distance metrics and values for accuracy and F1 are similar (+-5%) to those obtained in Part 1 Task 8. |