**Scoring Rubric – Programming Assessment 6 – Naïve Bayes**

**Criterion Name:** *Part 1, Task 1.1: 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* | 2 or more reviews and 2 or more stop words have been printed out from the dataset. |

**Criterion Name:** *Part 1, Task 1.2: 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* | 2 or more reviews have been printed out from the dataset and they do not contain stop words, punctuation marks, numbers, unwanted symbols, hyperlinks, and usernames. The reviews should also all be in lowercase. |

**Criterion Name:** *Part 1, Task 1.3: 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 40000 reviews and test data contains 10000 reviews. |

**Criterion Name:** *Part 1, Task 1.4: Create Naïve Bayes Classifier*

|  |  |  |
| --- | --- | --- |
| **Point Value** | **Option Name** | **Description Explanation** |
| ***3*** | *Weak* | Task is incomplete/code has not been written/cell has not been run/code contains errors. |
| ***15*** | *Satisfactory* | Task has been attempted, the code has been written and the cell has been run; however, the implementation is not fully correct. |
| ***30*** | *Strong* | The vocabulary has been made and the classifier is running perfectly by returning the argmax of the likelihood. |

**Criterion Name:** *Part 1, Task 1.5: Implement Evaluation Functions*

|  |  |  |
| --- | --- | --- |
| **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 3 requirements have been coded and the results are correctly displayed with an accuracy of greater than 70%. |

**Criterion Name:** *Part 2: Naïve Bayes Classifier Using Scikit-learn*

|  |  |  |
| --- | --- | --- |
| **Point Value** | **Option Name** | **Description Explanation** |
| ***5*** | *Weak* | Task is incomplete/code has not been written/cell has not been run/code contains errors. |
| ***20*** | *Satisfactory* | Task has been attempted, the code has been written and the cell has been run; however, the implementation is not fully correct. |
| ***40*** | *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 5. |