Lab 3\_2 Marking Questionnaire

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Exercise | Criteria | Marks Available | Marks Given |
| 1 | Exercise 1 | Has one of the most\_frequent words or over\_threshold classifiers been   * instantiated * trained on the training data * tested on the test data * given a score in the output? | 2 |  |
| 2 | Has the code been updated in utils.py and imported into the notebook? | 1 |  |
| 3 | Ex 2.1 | Has a ConfusionMatrix been instantiated correctly to compare the predictions from a movie\_classifier and the gold\_standard labels? | 1 |  |
| 4 | Ex 2.2 | Correct code for precision | 1 |  |
| 5 | Correct code for recall | 1 |  |
| 6 | Correct code for F1 | 1 |  |
| 7 | Has a ConfusionMatrix been instantiated and precision, recall and F1 calculated (even if code wrong) for at least 2 classifiers? | 1 |  |
| 8 | Interpretation of results? | 2 |  |
| 9 | Ex 3.1 | Code to compute average accuracy across a number of runs: is the training and testing data resplit on each iteration? | 1 |  |
| 10 | Ex 3.2 | Construction of confusion matrix and recording of precision, recall and F1 | 1 |  |
| 11 | Ex 3.3 | Has accuracy been calculated for at least one classifier for a reasonable number of different sample\_sizes? | 1 |  |
| 12 | Have precision, recall, and F1 been calculated for at least one classifer for a reasonable number of different sample sizes? | 1 |  |
| 13 | Are results averaged over at least 3 runs? | 1 |  |
| 14 | Have 2 or more classifiers been considered? | 1 |  |
| 15 | Are results displayed in a table and/or graph? | 1 |  |
| 16 | Is the table/graph used a good way of visualising the results? | 1 |  |
| 17 | Interpretation of results? | 2 |  |
| TOTAL | | | 20 |  |