|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Sr.No | Test Case Description | Expected Results | Actual Results | Pass/ Fail | Comments |
| 1 | Scrape websites using BeautifulSoup and Selenium and get data from them in form of paragraphs | Scraped data stored in a csv file | Scraped data stored in a csv file | Pass | BeautifulSoup did not work for dynamic websites so Selenium was used in place of it.  Authorisation issues which prevented scraping. |
| 2 | Scrape comments from Quora and Reddit | Scraped comments stored in a csv file | Comments in a csv file | Fail | Authorisation issues. We had to result to “copy and paste” |
| 3 | Scrape comments from LinkedIn | Scraped comments stored in a csv file | Scraped comments stored in a csv file | Pass | Not all comments were scaped |
| 4 | Scrape PdFs | Stored data in a csv file | Stored data in a csv file | Pass | Extra data cleaning as the headers and footers of the page were also scraped |
| 5 | Data Cleaning  -changing text to lower case  -Removing punctuations/numbers  - removing stop words  -Tokenization  -Lemmatization | Cleaned dataset | Cleaned dataset | Pass | Everything went well |
| 6 | Data Labelling  -Comparing two pretrained sentiment analyzer and picking the best(TextBlob and VADER) | Best sentiment analyzer | VADER – best sentiment analyzer | Pass | Everything went well.  Both were not 100% accurate, however, VADER was better. |
| 7 | Exploratory Data Analysis  -getting description about the dataset  -removing duplicates  -removing NaN Values  -word cloud for visualization  -top 10 positive/negative words. | Meet all requirements | All requirements was satisfied | Pass | The word cloud showed some words that did not show any sentiment |
| 8 | Machine Learning stages  -Feature Extraction (BOW, TF-IDF)  - Changing parameters for potential increase in accuracy | Extracting features with two methods  -Improved accuracy | Features were extracted with the two methods  Improved accuracy | Pass | The “max-features” parameter was increased from 100 to 1000 and there was an increase in accuracy of the models |
| 9 | Balance Dataset | Improved Accuracy | Accuracy reduced compared to the unbalanced dataset | Fail | Reduction in dataset could have led to lesser accuracy. |
| 10 | Changing the ration of training to testing dataset  --Train/Test split (90:10, 80:20, 70:30, 60:40) | Picking the best ration | The best ratio 90:10 | Pass | The more the training dataset, the better the accuracy |
| 11 | Try different models to compare accuracy  -Model Building using Naïve Bayes Classifier, Support Vector Classifier, Logistic regression, Random Forest | Bar plot showing comparison | Bar plot showing comparison | Pass | Everything went well |
| 12 | Run codes on a different IDE (PyCharm) | Codes running perfectly in Pycharm | Codes running perfectly in pycharm | Pass | Everything went well |