# Report

|  |  |
| --- | --- |
| Figure 1 Topic Frequency Chart | Figure 2 Sentiment Frequency Chart |

1. (1 mark) Give simple descriptive statistics showing the frequency distributions for the sentiment and topic classes across the full dataset. What do you notice about the distribution?

Topic statistics

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Variable | N | N\* | Mean | SE Mean | StDev | Minimum | Q1 | Median | Q3 | Maximum |
| Frequency | 20 | 0 | 100.0 | 20.8 | 93.0 | 7.0 | 26.0 | 57.5 | 157.3 | 358.0 |

Sentiment Statistics

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Variable | N | N\* | Mean | SE Mean | StDev | Minimum | Q1 | Median | Q3 | Maximum |
| Count | 3 | 0 | 667 | 334 | 579 | 153 | 153 | 553 | 1294 | 1294 |

1. (2 marks) Vary the number of words from the vocabulary used as training features for the standard methods (e.g. the top *N* words for *N* = 100, 200, etc.). Show metrics calculated on both the training set and the test set. Explain any difference in performance of the models between training and test set, and comment on metrics and runtimes in relation to the number of features.
2. (2 marks) Evaluate the standard models with respect to baseline predictors (**VADER** for sentiment analysis, majority class for both classifiers). Comment on the performance of the baselines and of the methods relative to the baselines.
3. (2 marks) Evaluate the effect that preprocessing the input features, in particular stop word removal plus Porter stemming as implemented in **NLTK**, has on classifier performance, for the three standard methods for both sentiment and topic classification. Compare results with and without preprocessing on training and test sets and comment on any similarities and differences.
4. (2 marks) Sentiment classification of neutral tweets is notoriously difficult. Repeat the experiments of items 2 (with N = 200), 3 and 4 for sentiment analysis with the standard models using only the positive and negative tweets (i.e. removing neutral tweets from both training and test sets). Compare these results to the previous results. Is there any difference in the metrics for either of the classes (i.e. consider positive and negative classes individually)?
5. (6 marks) Describe your best method for sentiment analysis and your best method for topic classification. Give some experimental results showing how you arrived at your methods. Now provide a brief comparison of your methods in relation to the standard methods and the baselines.