KT project2 Peer-Reviewing:

Summary:

The report chose to use Naïve Bayes classifier with the help of Weka for classifying a large amount of tweets into different sentiment classes. Author gave some logical reasons and added 5 new features, which is *Exclamatory, Reaction, Positive Words, Negative Words, http(s),* respectively. Then, combined these new features with the original 46 features and use some metric-----accuracy, precision, recall and F1-sorce, in terms of multiple tables, to compare this new system with the original one. Also, gave abundant discussions, evaluations and assumptions with the statistical results, which are reasonable and convincing.

Advantages:

There are many things the author did pretty well as below:

1. The 5 new features are novel and suitable for sentiment analysis tweets. It considers people’s behavior on tweets and concludes some specific words to represent positive and negative sentiment.
2. Author assumes some reasons for the misclassifications, shows some examples for confirming those assumptions discuss some of the misclassification situations and gives a way for how to improve it.
3. Use many metrics for evaluation and compare those metrics with statistical result. For example, explain why accuracy increases and why precision and recall decreases.

Weak points

1. Just use Naïve Bayes classifier, which lacks comparison with other classifiers.
2. Author mentioned “single tokens being more effective than groups of tokens”. It is better to add those 5 new features one by one separately into original features than straightly combine all new features with original features. Because it is hard to evaluate each new feature’s impact on the accuracy.
3. For future improvement, the number of words in features “Positive Words” and ”Negative Words” is not enough (one of the reason for accuracy increases slightly), and it’s better to give more sentiment words for these two features. One suggestion is that find some third library online about sentiment words, import them into you program for analysis.