

FINAL PROJECT

A. Sentimental Classification

1. Firstly, I downloaded an online dataset Tweets.csv (<https://www.kaggle.com/crowdflower/twitter-airline-sentiment>) and use this dataset as training group for our supervised classification model.
2. Then, I checked the F1 score, precision and recall for my rule-based classification model by using the ground truth in out of sample data. The result is not satisfied. F1 for rule-based classification is 0.5424036.
3. Then, I build three models: Naive Bayesion, Maximum Entropy, and Support Vector Machine. I train these models by using out of sample data and compare their F1 score for non-complaint class and AUC.
4. Finally, I find out that Maximum Entropy model is the best one.
5. I use Maximum Entropy model to predict my own data and come out with non-complaints. Then, I use rule-based data to make a deeper clean on my outcome.