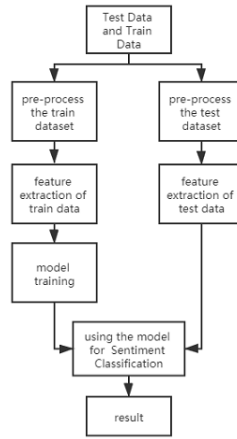# Sentiment Classification with Naive Bayesian Classifier

GROUP 6
Members: Xin Sun, Yuan Wo, Yuxin Meng

1 Systematic Diagram

This system diagram shows the emotional analysis of the data.

2 Feature Extraction

In order to extract features, we choose to use the text feature extraction function CountVectorizer(). The function takes into account the frequency of each word, and then forms a feature matrix, with each line representing word frequency statistics for a training text. By converting text words into vectors, we can get the frequency of the key words about emotion, which can be used to train the model.

3 Evaluation

To evaluate the performance of our model, we used the same methods to pre-process the evaluation data. We got the 'eval_features' and 'eval_label' for our model. and then tested the model on the dataset and calculated the accuracy to see the performance of the model.

4 Result

To achieve a accuracy above 70 % , we tried different combination of parameters, finally, we got an accuracy of 76.76 % on test dataset and an accuracy of 75.38 % on evaluation dataset. The result proves that we can use the Bayesian Classifier to achieve sentiment classification tasks, and the generalization ability of our model is good. The training speed of the model is fast, but the accuracy of the model is not very high.

Here attach the link of our lab.
https://github.com/eve76/big-data-lab