Lab 1: Sentiment Classification with Naïve Bayesian Classifier

Aim:

- Understanding Naïve Bayesian Classification technique

- Get familiar with data analysis process including data cleaning.

- Learning to employ Python to do data analysis

Materials:

We provide you with three datasets we got from a social media site:

1) training dataset, train.csv

2) test dataset, test.csv

3) evaluation dataset, evaluation.csv

Requirements:

1. Using Python to create your own Naïve Bayesian classifier and train the network to classify data into two sentiment classes: "positive" and "negative".

2. Some of data include HTML tags, hashtags, mentions, emojis etc. so you need to do pre-processing.

3. Test your classifier with the test dataset

4. Applying your classifier to evaluation the dataset of 5000 yelp reviews which is a different category and the data there do not appear in train/test data.

Write one-page lab report containing a) title b) names and team 3) systematic diagram of your sentiment analysis process, including training and testing; c) feature extraction; d) evaluation; e) results.