Lab 1: Sentiment Classification
with Naïve Bayesian Classifier (NBC) and SVM

Aim:

- Understand Naïve Bayes Classification and the SVM technique

- Become familiar with the data analysis process, including data cleaning

- Learn how to use Python for 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. Use Python to create your own Naïve Bayes and SVM classifiers, and train them to classify data into two sentiment classes: 'positive' and 'negative.'

2. Some data may contain HTML tags, hashtags, mentions, emojis, etc., so pre-processing is required.

3. Test your classifiers using the test dataset.

4. Apply your classifiers to evaluate a dataset of 5,000 Yelp reviews, which belongs to a different category and is not part of the training or test data.

Write a lab report of a few pages that includes: a) the title, b) names and team information, c) a systematic diagram of your sentiment analysis process, including training and testing, d) feature extraction, e) evaluation, f) results, and g) the code or a link to your code.