

Course Project

**Student Name:** Piao Chuxin, Ye Xiaoxing

**Student ID:** 115010058, 115010270

---

## Contents

<b>1</b>	<b>Abstract</b>	<b>2</b>
<b>2</b>	<b>Introduction</b>	<b>2</b>
<b>3</b>	<b>Model</b>	<b>2</b>
3.1	Sentiment Analysis . . . . .	2
3.1.1	Naive Bayes . . . . .	3
3.1.2	Linear Model . . . . .	3
3.1.3	Support Vector Machine . . . . .	3
<b>4</b>	<b>Conclusion and Findings</b>	<b>3</b>

# 1 Abstract

Twitter is a great firehouse of real-time information. Using the data from twitter, people can gain information such as political leaning and other tendencies. Through the classification process of the data before the election day, the more competitive candidate with a higher approval rating can be obtained.

## 2 Introduction

American President Election is one of the catchiest affairs all over the world. To predict the candidate with higher approval rate beforehand, social networks can be effective in obtaining the political leaning of citizens in the United States.

Twitter is one of the influential social networks that can be used in predicting the political leaning of people who use twitter. Having downloaded the data of one day from the twitter website, the emotion scores have been computed, represented by a number from -1 to 1, in which 1 represents the positive emotion while -1 represents the negative emotions. After obtaining the score of emotion, the keywords which could represent different candidate have been searched. For example, Trumpor Donald could represent Donald Trump, and Hillary or Clintoncould represent Hillary Clinton. Compared the twitter score which contains different key words to the average emotional score of all the twitter score, the rough political leaning of different candidates can be obtained.

To tell the political leaning of people, we use sentiment analysis. Sentiment Analysis, sometimes known as opinion mining, is a field of study combining the use of natural language processing, text analysis, and linguistics, to identify and study affective and subjective information. Typically, it analyses peoples opinions towards entities, such as products, companies, and parties. Thanks to the rapid-developing social media, this area has now been a hot topic with a constant source of textual data.

## 3 Model

### 3.1 Sentiment Analysis

To evaluate the tweets' sentiment, we first trained a series of models. They are...

1. Naive Bayes
  - (a) Simple Naive Bayes
  - (b) Multinomial Naive Bayes
  - (c) Bernoulli Naive Bayes
2. Linear Model
  - (a) Logistic Regression
  - (b) Stochastic Gradient Descent
3. Support Vector Machine
  - (a) Support Vector Classification (deprecated)
  - (b) Linear SVC

In this part, we would like to introduce the models one by one, and then our implementation.

### 3.1.1 Naive Bayes

Naive Bayes classifier is a simple probabilistic classifier based on Bayes' Theorem, while assuming the independence between features. This technique has been studied for over 60 years, and it is now hot since it is simple but effective. It is still a popular and the baseline method for text classification. With proper pre-processing, it is even competitive with advanced methods like SVM.

Naive Bayes is a simple technique for constructing classifiers. Class labels are assigned to instances, where the labels is a finite set. It is a family of algorithms based the principle that all assume the independence of feature, given the class variable. Despite the oversimplified assumptions, the Naive Bayes classifiers works well in many real-world cases. And, one of the advantage is that, only a small training set is required, to guess the essential parameters. [1]

### 3.1.2 Linear Model

### 3.1.3 Support Vector Machine

The second bullet point is easily explained by the fact that, under the hood, scikit-learn relies on different C libraries. In particular SVC() is implemented using libSVM, while LinearSVC() is implemented using liblinear, which is explicitly designed for this kind of application.

## 4 Conclusion and Findings

## References

- [1] Wikipedia contributors. Naive bayes classifier — Wikipedia, the free encyclopedia, 2018. [Online; accessed 29-April-2018].