Abstract
Nowadays, data of social medial websites are getting more and more popular to be used as one of the most important data source for the data mining from which we can find the useful and interesting patterns. In this project, base on twitter data set that I collect using twitter API, I performed the sentimental mining and topic modeling. In the data collection phase, I used keywords such as the candidates’ name to filter the related data decreasing the noise to the most extend. To accomplish the sentimental mining, I chose Naïve Bayes algorithm and Support vector machine Model(SVM) two of the most commonly used algorithms that can be used as the classifier in the sentimental analysis. Then I trained these classifiers using a data set which was also from twitter and was related to 2016 presidential election from Kaggle and made the predication using twitter data set that I collected. Besides, Latent Dirichlet Allocation model was used to fulfill the topic modeling analysis finding the most frequent topics from the data of presidential election related tweets. At last, I evaluated the performance of each classification algorithm.

Problem Definition
Firstly, using sentimental analysis, based on the tweeter data, the popularities of each candidate will be come up with. In this phase, two calcific algorithms Naïve Bayes and Random Forest will be used as the classifier and performance of them will be evaluated at the end.
Secondly, I will compare the results of the analysis of tweeter data with the results of polls and contributions to see if they are comparable and if they roughly have the same trend.
Besides, I will use Latent Dirichlet allocation(LDA) model extracting frequent topics related to 2016 presidential election using tweeter data set. For better illustration of the results, I will visualize the results in World Cloud.

Design and Implementation
To find the best parameters for these classification algorithms, I used cross-validation method for the parameter tuning. As can be seen below, throw this way, the best kernel parameter for SVM and the best alpha parameter for Naïve Bayes were found.

Result
We can see that several interesting things from the result. Firstly, the overwhelming number of tweets are Negatives. Secondly, Donald Trump was more popular than the other candidates. Last, although Trump was more popular, Hilary was more possible win the election.