OVERVIEW :

In this project we are analyzing the sentiment of the people. For analyzing the sentiment we are using social media data sets (twitter). We collected the live tweets (datasets) from twitter using the twitter API.

This type of sentiment analysis problem is come under the classification problems. The classification is done based on positive and negative tweets in the twitter. (It shows how many percentage of people think about the product positive, negative and neutral)

The project is classified into two phases.

In phase 1, the implementation is done based on Naïve Bayesian classifier and in the phase 2 we are using logistic regression.

Programming language used for phase 1 is Python and for phase 2, Python with Spark is used.

The basic concept of the project is that, we need to train the Machine Learning model with the training dataset. Then by passing the testing datasets into the trained model, we get the predicted result of the testing datasets.

By comparing the predicted result of the testing dataset with the predefined label of the testing dataset, the accuracy can be calculated.

It is found that Naïve Bayesian Classifier provides 89% accuracy for sentiment analysis. By Using Logistic Regression the efficiency can be increased to 97%.

INPUT AND OUTPUT EXPLANATION:

The system asks for the keyword and the number of tweets to be generated. The keyword may be a product or the service or a brand.

We need to provide the keyword and the number of tweets.

After processing all the inputs system provides the output that, how many percentage of people think positive about the keyword and negative about the keyword.

Contact : Rajik Ahamed M (mmsrajikahamed@gmail.com)