**19CSE453\_Natural Language processing:**

**Project Documentation:**

**About project:**

1. **Title of the Project:**

*Predict Stock Movement from Articles/News Headlines.*

1. **Description:**

**Objective:**

Main Objective of the project to Predict Stock Market Movement by taking reference of News Headlines, Articles.

To Implement our Project, we using NATURAL LANGUAGE PROCESSING Technology’s which a man with his ideas can interact with computer Intelligence called as Artificial Intelligence and this Artificial Intelligence is used to people personalized Prediction’s in the world.

Example like Optimizing products, Voice Sensor’s, logistics.

**Implementation:**

Through **Sentiment Analysis** We Implementing Our Prediction of Stock’s Project it’s a Sub-filled topic in Natural language Processing.

In order to proceed with this objective, we needed to understand what Sentimental Analysis is. Sentimental Analysis is an analytical method that the computer uses to understand a natural language and deduce if the message is positive, neutral or negative.

In our case, Sentimental analysis refers to the deduction of the news headlines if they increase the stock or reduce it. By doing so, we end up with the ‘emotional’ status of the data which is what sentimental analysis gives its user.

**General Prediction:**

Generally Stock Holders Can Predict using analysing the growth of a company and they can invest based on the future that how a company can be improving in business ideas, based on their new start-ups, based on sector like gold/petroleum etc.

**N Grams:**

N-grams of texts are extensively used in text mining and natural language processing tasks. They are basically a set of co-occurring words within a given window and when computing the n-grams you typically move one word forward

1. **Applications of the project:**

* Prediction of Stock Prices
* Helping In Analysing of Articles
* Analysis of Movement in Company Growth

**Rule Based/ Statistical Based:**

This is a Statistical based Approach Project, We are Making the cleaned text data to the numerical data so we can utilize in the various types ML Algorithm Implementations.

Along with this using some Datasets with the information daily growth, yearly grow, growth in their Stock-Holders, Growth In new Start-ups we can also predict Future.

**Sample Demo:**

1. **Data-Set:**

We using the dataset with the details of paper Publication dates and we can extend the data in dataset as taking 25 more popular headlines from different Publishers.

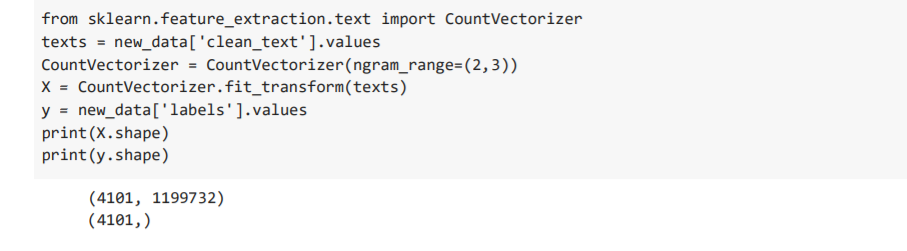
With the raw data, we cannot proceed much further until we manipulate the data to suit our analysis and convert the data into vectors that are much easier to work on. For this, we use Word2Vec.

This raw data is manipulated using python. We first split the data into lists of words but these lists are flooded with HTML tags and punctuations. We cleaned up the data and removed all HTML tags and punctuations.

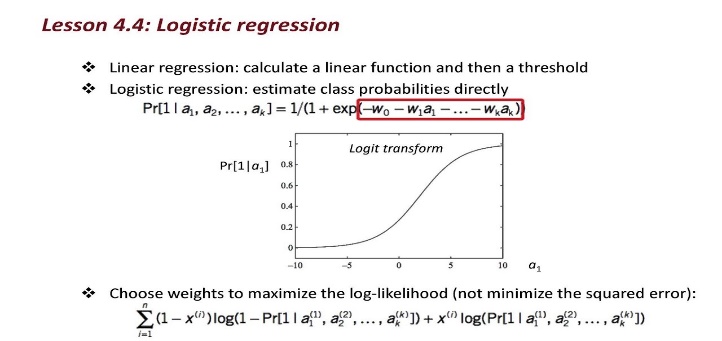
Then we moved forward with removing stop words. Stop words are words that do not contribute to the meaning or sentiment of the data such as ‘the’, ‘is’, ‘and’, etc. We have also converted all the letters to lowercase to make a more even data set to play with.

1. **Pre-processing And vectorization:**





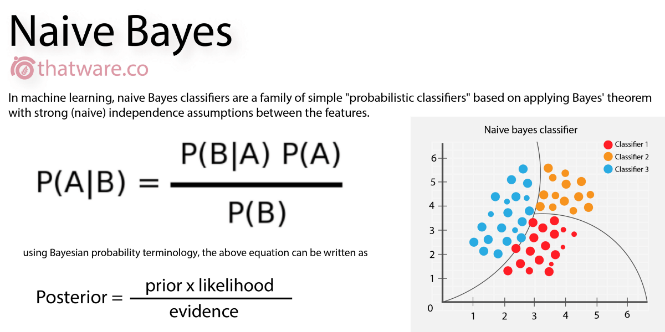
1. **Techniques used:**
2. *Logistic Regression:(Gopi Krishna-359)*



Logistic regression is a statistical analysis method used to predict a data value based on prior observations of a data set.

  It predicts a dependent data variable by Analysing the relationship between one or more existing independent variables.

1. *Naive Bayes: (Abhinay-311)*



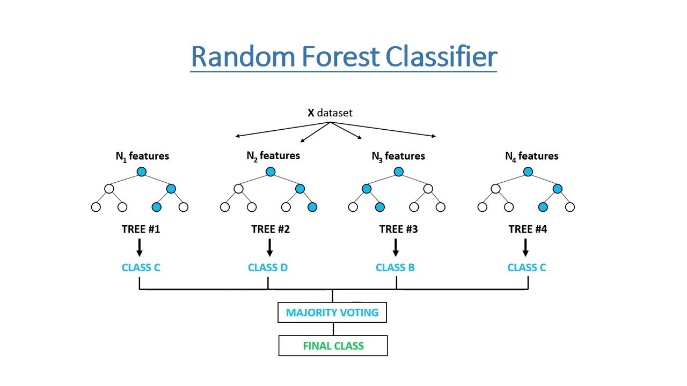
Naive Bayes is a classification algorithm that is suitable for binary and multiclass classification.

It is a supervised classification technique usedto classify future objects by assigning class labels to instances/records using conditional probability.

Naive Bayes is an eager learning classifier and it is sure fast. Thus, it could be used for making predictions in real time.

 It is the most popular choice for text classification problems.

1. *Random forest Classifier:(Teja-341)*



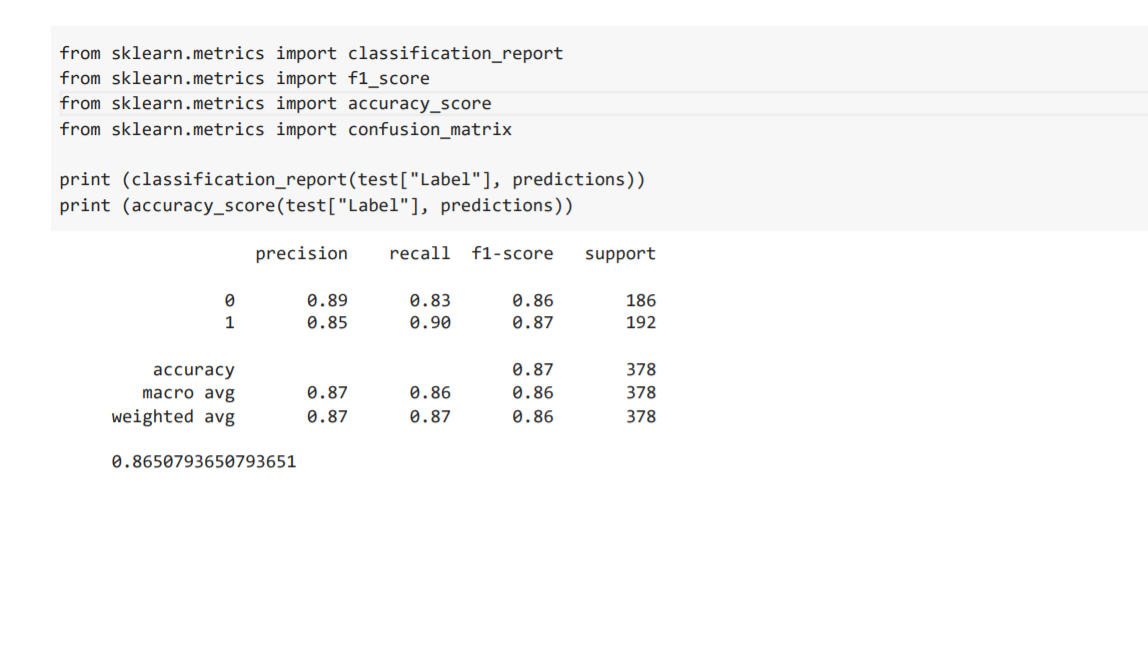
The random forest is a classification algorithm consisting of many decisions trees.

It Provides solutions to complex problems.

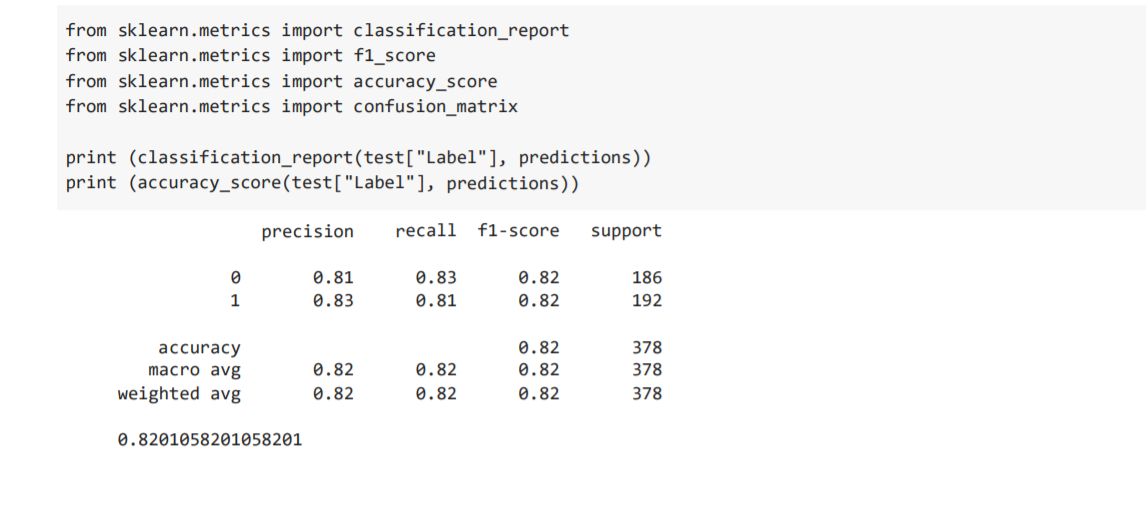
1. **Evaluation of the techniques:**

Accuracy We got is more than 70% So, as a fresher’s we got a >70% Accuracy that we feeling good and we try to improve it. Here we can Placing the screenshots of our accuracy according to techniques used individually.

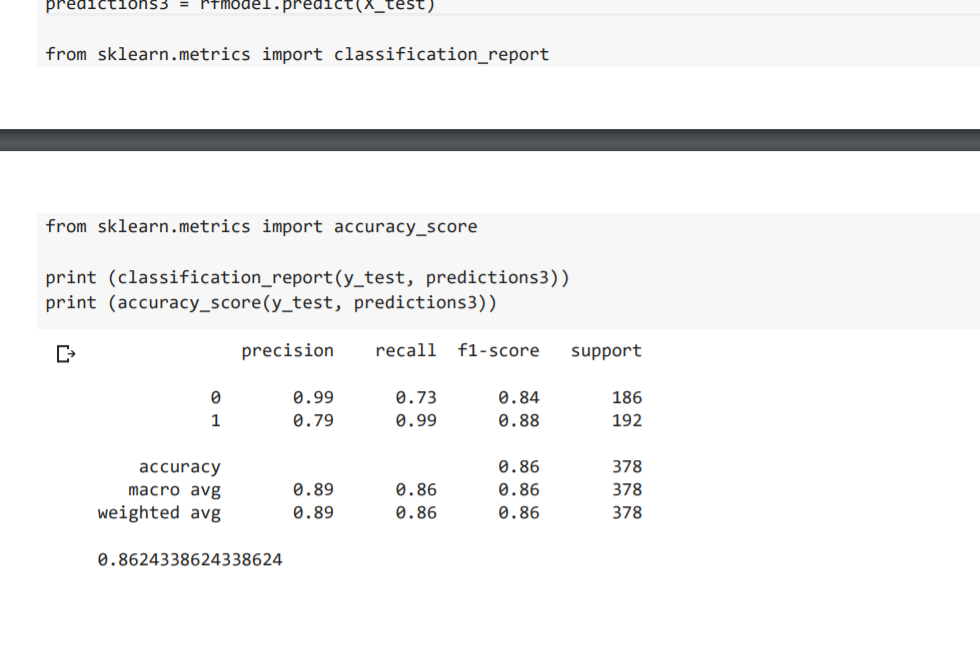
1. **Logistic Regression:**



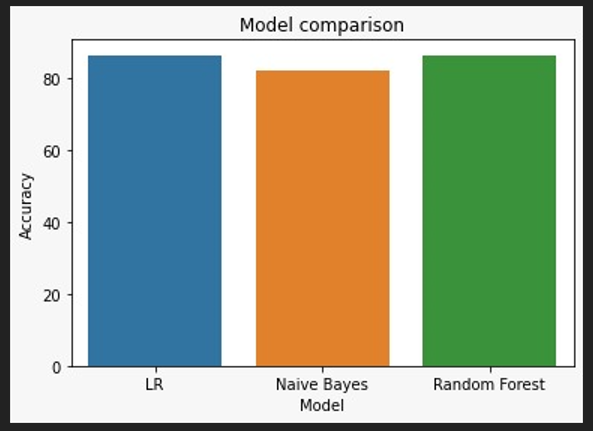
1. **Naïve bayes:**



1. **Random forest Classifier:**



**Model comparison with Graph:**



**Conclusion:**

It has led to the conclusion that it is possible to predict stock market with more accuracy and efficiency using machine learning techniques.

**Team Member’s:**

**B. Abhinay – AM.EN. U 4CSE19311**

(Documentation along Naïve bayes Approach Implementation)

**P. Teja Venkata Subbareddy – AM.EN. U 4CSE19341**

(Information and Analysis along Logistic Regression Implementation)

**Gopi Krishna – AM.EN. U 4CSE19359**

(Introduction to Project, Analysis along Implementation in Random-forest Classifier)

**\*\*THANK YOU\*\***