Stock Prediction using Sentiment Analysis of News Headlines and Artificial Neural Network
Team 02

Ishan Singhal
160010016

Avya Bansal
160010024

Vishal Jha
160010025

November 26, 2018

1 Introduction and Overview

Stock prices are volatile. They are hard to predict based only on some expertise through previous trends. Stock price fluctuation represents the current market trends and business growth among other factors that could be considered to sell or buy stocks. The daily news represents factual information about the companies and affect the stock market. Hence, we wish to create a model that will use news articles along with the historic trend of closing price to predict the change in stock prices of five companies namely **TCS, LT, Infosys, Yes Bank and Tata Steel**.

Related work

In the research paper[2], Naive Bayes classifier model is used to predict whether the stock price goes up/down based on sentiment analysis.

In the research paper[3], artificial neural networks for NASDAQ index prediction was developed and validated for two types of input dataset (four prior days and nine prior days). Model built for nine prior days gave better results.

2 Hypothesis

News articles are one of the most important factors which influence the stock market on a daily basis. We can use the sentiment of news articles to predict the stock price and hence test the "Efficient Market Hypothesis"that tells that the stock prices are very hard to predict only based on some expertise through previous trends and past stock prices.

3 Methods

The flowchart shown in fig1 gives an outline of the procedure followed.

3.1 Data Collection

Historical stock data has been collected from Yahoo Finance website. We performed web-scraping using **BeautifulSoup** to extract the news headlines from moneycontrol website. In fig2, we have plotted the closing prices of the 5 companies over a period of 13 years(2005-2018) to show that there is a lot of variation in the prices.

3.2 Data Processing

In the stock prices data, we have added an entry **label**, which is 0 or 1 depending on the change in stock price from previous day
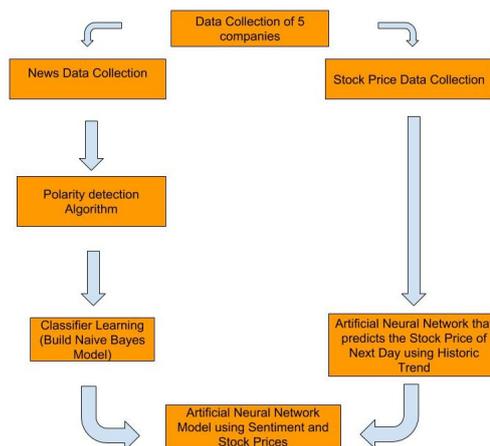

Figure 1: Steps for stock prediction based on sentiment analysis

i.e. 1 implies that the stock price has increased and 0 otherwise. We combined all headlines for same date into a single string. We then removed stopwords and performed stemming and lemmatization on these combined headlines using the NLTK package. For training the Naive Bayes Model, we also performed TF-IDF on our news headlines to give as input.
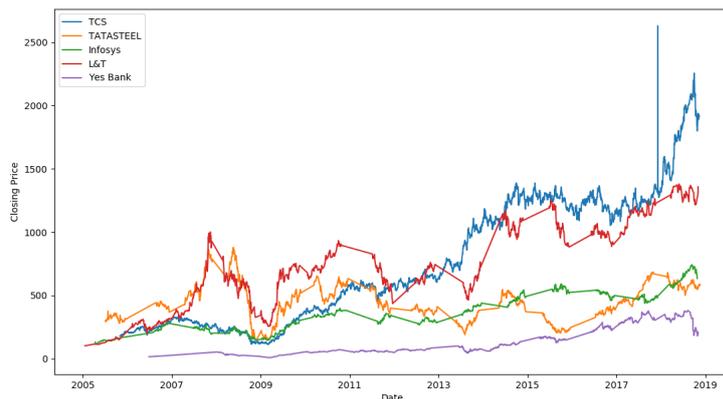

Figure 2: Variation in Closing prices

3.3 Sentiment Analysis:-

Now we tokenized the news headline into a vector of words. We required a list of financial words to evaluate the sentiment of our headlines. So we created a dictionary of 2432 positive and 7374 negative financial words using LoughranMcDonald dictionary[4]. We performed stemming and lemmatization on these words as well. Then we compared each word in the headline with the list of words to calculate the polarity score of each headline. **Polarity**

Score = count (positive matches) - count(negative matches)

If the score is greater than or equal to 0, we consider that the news has positive polarity else negative. Based on this we created a model that gives the sentiment value for a given headline. Many researches show that Naive Bayes classification algorithms perform well in text classification. Thus we classify the text and check accuracy of model. fig3 below shows the different accuracies we obtained for different companies. **Sentiment accuracy** represents the correlation between polarity score and label of each headline. **Bayes accuracy** represents the efficiency of our Bayes model.
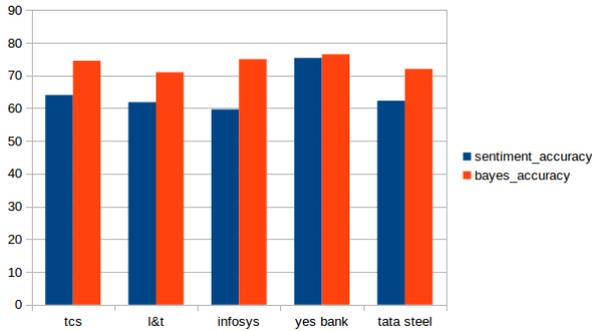

Figure 3: Comparison between accuracies for different companies
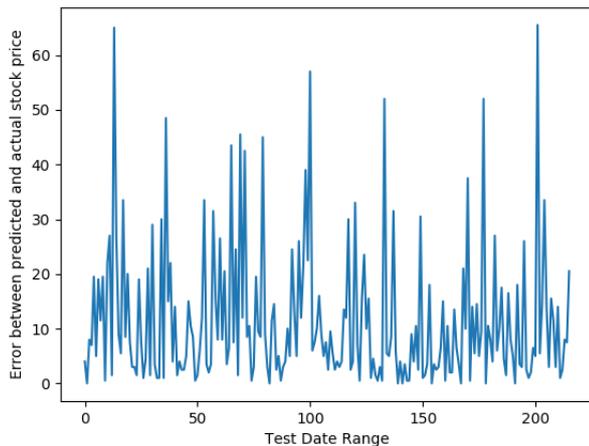
3.4 Neural Network

Figure 4: Without Sentiment

A neural network is created using only the historical trends for stock price prediction. It takes as input the stock value for prior 9 days and predicts the stock value of the next day. The number of neurons in input layer are 9 and the output layer consists of 1 neuron. Three hidden layers are created for the network with 100,250 and 400 neurons respectively.

Another neural network is created using historical trends as well as sentiment of news headlines. It takes as input the stock value and sentiment value for prior 9 days from the sentiment model previously created as well as the sentiment built on the next day's headline to predict the stock value of the next day. So, the number of neurons in input layer are 19 and the output layer consists of 1 neuron.

The result is verified using two error plots shown in fig4 and fig5. They represent the mean squared error of test dataset consisting of 300 days for the network without sentiment and with sentiment respectively for the company Yes Bank. The sum of errors has been found to decrease from 2730 to 2432 in fig5 as well as the

maximum error also decreased from 65 to 52. This shows that the prediction has improved by taking into account the sentiments of news headline.
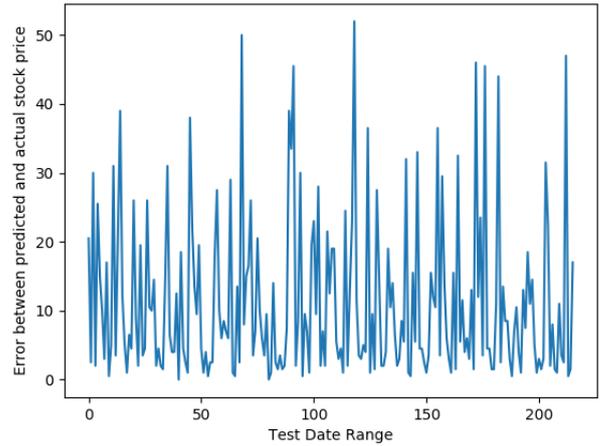

Figure 5: With Sentiment

4 Conclusion

As shown in fig3, the accuracy of sentiment prediction with stock price trend is as high as 70% for Yes Bank and of the model is 72%. We can conclude from this that there is definitely a correlation between news polarity and stock price trend. This was further tested in the neural network model for 5 different companies and the best results (obtained for Yes Bank) are shown in fig4 and fig5. From this, it is evident that adding sentiment to our neural network improves the accuracy which is in accordance with our hypothesis. Although these predictions still have large errors, the inclusion of sentiment values have reduced these errors.

5 Discussion and Future Directions

Possible sources of improvements

- Larger dataset (for training) of news headlines including tweets about various companies.

- We have considered a simplified model with stock prices of different companies independent of each other. This is not true in the real world. Thus we could build a model which considers these dependencies.

- Stock prices are governed by two factors - fundamental and technical. Here, we have considered only fundamental factors. We could improve our model by taking into account the technical factors as well.

References

[1] https://www.sciencedirect.com/science/article/pii/S207718861630024

[2] https://www.researchgate.net/publication/304995235_Stock_Trend_P

[3] https://www.sciencedirect.com/science/article/pii/S207718861630024

[4] http://www.ijritcc.org/download/browse/Volume_4_Issues/November_11-2016.pdf

[5] http://www3.nd.edu/ mcdonald/Word_Lists.html