IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Stock Market Prediction and Recommendation System

Aarti Puthran, Ashutosh Shukla, Devansh Shah , Khushboo Dattani

Teacher, Student, Student, Student
Information Technology
Shree L.R. Tiwari College of Engineering , Thane, India.

Abstract: Since its inception, the Stock market has become a sophisticated way of investing money and a reliable source of income. Performance of stock relies on multiple factors which makes it very unpredictable and volatile. Hence, Prediction can play an important and helpful role here. Predicting markets can offer great profit avenues, and much research is going on for it. For prediction, most researchers use either technical analysis or fundamental analysis. Technical analysis can be referred to as prediction on the basis of previous stock prices while fundamental analysis can be referred to as analysing the unstructured textual data, example: news, tweets, etc. Today the world is digitalized, the market information is accessible to common people as well, this takes it to text mining strategies that would extract significant information to analyze market behavior. In this project we will be using past prices of stock and news headlines together to predict the stock prices. The project will give recommendations on stock buy /sell on the basis of current news

Index Terms - Prediction, Analysis.

I. INTRODUCTION

Stock market acts as a fastest source of income for people. But due to the high risk factor, majorly the richer and upper middle class peoples were taking part in it. Stock Market is a place where people buy or sell shares of publicly listed companies. The first lockdown in April 2020, has affected lives of many people and also they lost their source of income. This situation forced the poor people to risk their money in the share market in order to make money for their livelihood. Nowadays, many people have been involved in Share Market and it has become an attractive topic to talk about. This project makes use of different algorithms and makes use of previous stock prices and news from google finance to predict whether the particular share should be bought or sold.

II. PROBLEM STATEMENT

The basic aim of the Stock market prediction and recommendation system is to design an effective and accurate system for predicting stock prices in future years. As a large amount of historical data is available, our system aims to exploit this data and process algorithms over the data to gain insights. Our system will analyze the data and aim to provide accurate predictions of the stock prices on the basis of previous year stock data and news headlines. It also aims to recommend that users should buy or sell the stock according to the predictions.

III. MOTIVATION

A large proportion of people invest in stocks, as it acts as a good source of income, but also comes with a risk of losing invested money, hence it creates a fear factor. To overcome this fear and to give people an insight into the future indexes of stock, so that they invest more confidently in the stock market and make better decisions. This will help people avoid making decisions based on their emotions and influence, and help them choose practically while investing in stocks. Basically, it will help people evolve financially. Hence, this is the main motivation to work on this project.

IV. IMPLEMENTATION

The system will be using the data analysis and news sentiment analysis together to give a better and accurate result. At first step, the data will be collected from yahoo finance for data analysis and it will be cleaned , similarly for the sentiment analysis data will be fetched from ticker finology site and it will be cleaned and bought in the format as required by the algorithm. We can then apply algorithms for data analysis and sentiment analysis, combining their results, we will be predicting stock price and recommending the user to buy or sell the stock.

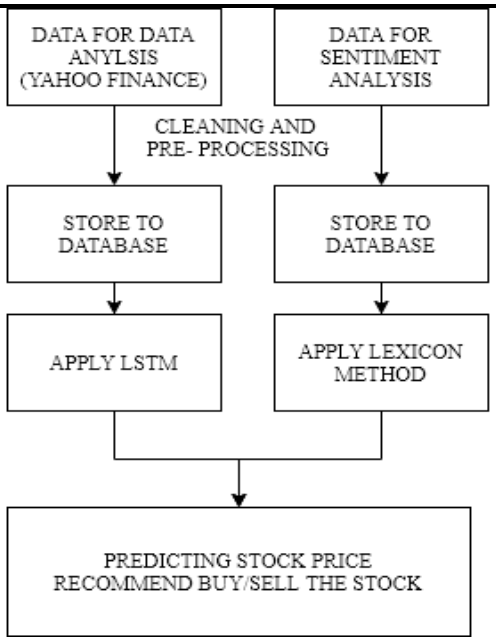


Figure 4.1 System Flowchart

A. Dataset Used

The dataset for price prediction and recommendation was collected from Yahoo Finance and Ticker Finology respectively. The dataset for price prediction contains everydays closing price of a share and dataset for recommendation contains headlines. The recommendation dataset was having irregularities and hence were removed in the data preprocessing step. Following is the snap of the dataset:

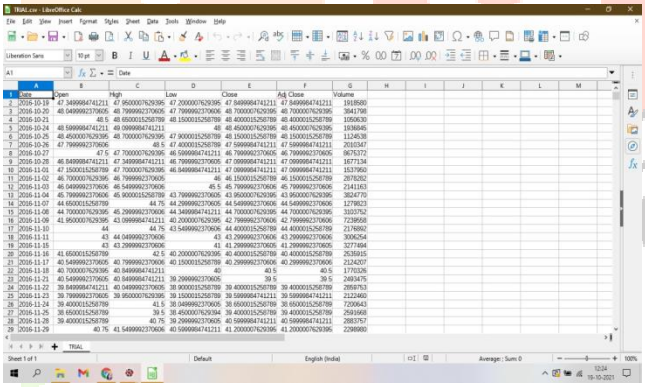
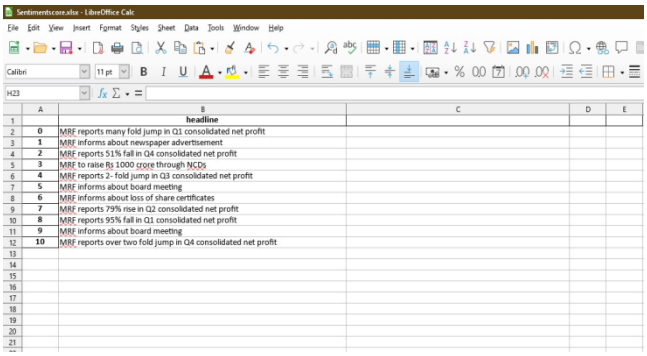


Figure 4.2 Data for Stock Price



	headline
0	MRF reports many fold jump in Q1 consolidated net profit
1	MRF informs about newspaper advertisement
2	MRF reports 51% fall in Q4 consolidated net profit
3	MRF to raise Rs 1000 crore through NCDs
4	MRF reports 2- fold jump in Q3 consolidated net profit
5	MRF informs about board meeting
6	MRF informs about loss of share certificates
7	MRF reports 79% rise in Q2 consolidated net profit
8	MRF reports 95% fall in Q1 consolidated net profit
9	MRF informs about board meeting
10	MRF reports over two fold jump in Q4 consolidated net profit

Figure 4.3 Data for Sentiment Analysis

B. Dataset cleaning and pre-processing

Since the algorithm requires proper sentences, we can't work with symbols popping up in the data. Hence, we need to clean the data before applying it on our model. So we convert it into expected format.

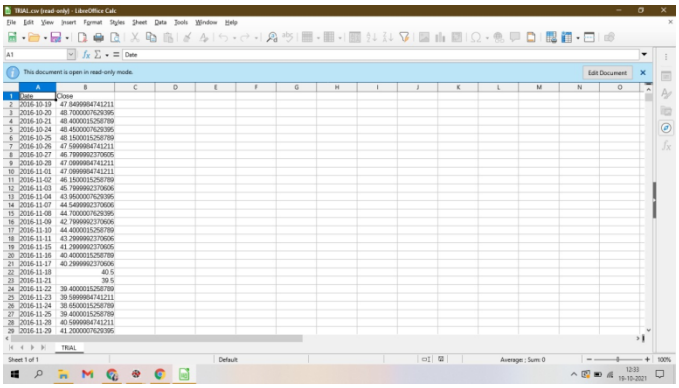


Figure 4.4 Clean Data for News Headlines

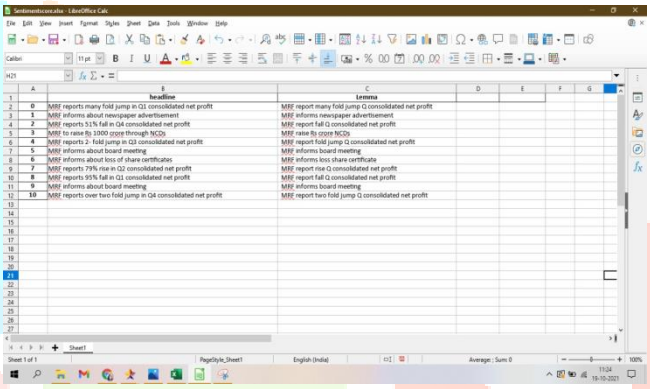


Figure 4.5 Clean Data for Stock Price 1

C. LSTM

In LSTM, the data will be segregated into two parts(65% training data and 35% testing data), training data will be used to train the model according to time step and then testing data will be used to test the accuracy. Long short-term memory (LSTM) units or blocks that are part of a recurrent neural network structure.[1] Recurrent neural networks use particular types of artificial memory processes that can help these artificial intelligence programs to more effectively emulate human thought.[1] The recurrent neural network uses LSTM blocks to provide context for the way the program takes inputs and generates outputs. The long short-term memory block is a complex unit with various components such as weighted inputs, activation functions, inputs from previous blocks and eventual outputs.[2] The unit is called a LSTM block because the program is using a structure founded on short-term memory processes to create longer-term memory.[3]

Deciding the information to be thrown away is the first step of our LSTM. A sigmoid layer called the "forget gate layer" makes this decision. It looks at ht−1 and xt, and outputs a number between 0 and 1 for each number in the cell state Ct−1. A 1 represents "completely keep this" while a 0 represents "completely get rid of this."
ft=σ(Wf.|ht-1,x|+bf)[4].

The next step is to make the decision about what new information is to be stored in the cell state. This consists of two parts:
1. A sigmoid layer called the "input gate layer" decides which values we'll update.
2. A tanh layer creates a vector of new candidate values, C~t, that could be added to the state. In the following step, we'll combine these two to create an update to the state.
it= σ(Wi.|ht-1,x|+bi)
Ct=tanh(Wc.|ht-1,x|+bc)[4]

It's now time to update the old cell state, Ct−1, into the new cell state Ct. The previous steps already decided what to do, we just need to actually do it. The old state is multiplied by ft, thus forgetting the things we decided to forget earlier. Then we add it∗C~t. This is the new candidate values, scaled by how much we decided to update each state value.

In the case of the language model, this is where we'd actually drop the information about the old subject's gender and add the new information, as we decided in the previous steps.

Ct = ft*Ct-1 + it*C't[4]

Old cell state Ct−1 is now updated into the new cell state Ct. The steps before this have already decided what to do, we just need to actually implement it.

We multiply the old state by ft, forgetting the things we decided to forget earlier. Then we add it∗C~t. This is the new candidate values, which is scaled by the extent to which we decide to update each state value.

Finally, we need to decide what we're going to output. This output will be based on our cell state, but will be a filtered version. First, we run a sigmoid layer which decides what parts of the cell state we're going to output. Then, we put the cell state through tanh (to push the values to be between −1 and 1) and multiply it by the output of the sigmoid gate, so that we only output the parts we decided to.

ot= σ(Wo.|ht-1,x|+bo)

ht = ot*tanh(Ct)

Data is divided into training and testing, we pass the training data into the LSTM model where it will learn the behaviour of data from input and output. After learning the behaviour testing data will be passed into the LSTM model and output for next 30 days will be given also the graph with respect to price will also be given to the user. [2]

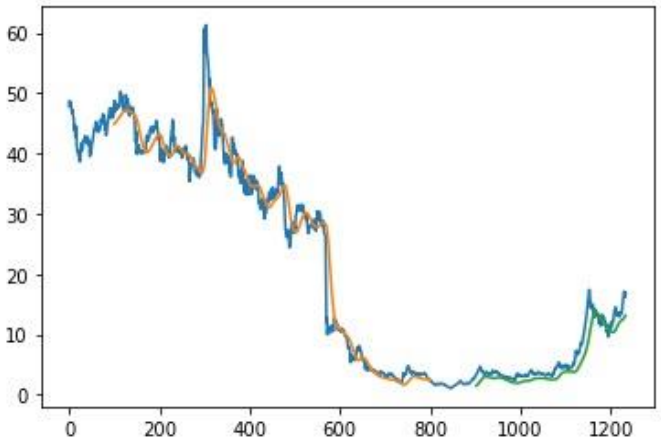
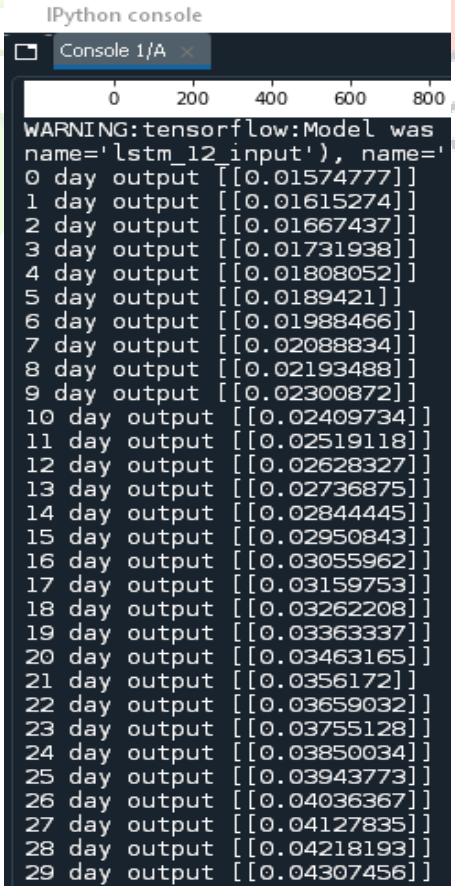


Figure 4.6 Graph of LSTM 1



4.7 Output for Stock Price

D. Sentiment Analysis

The Lexicon-based approach works on a predefined dictionary to score a sentences by taking the average of sentiment scores of all the words in the sentences.[5] The predefined sentiment lexicon should contain a word and corresponding sentiment score for it.[6] The negation form of vocabulary words should be added to the lexicon as individual entries, and they should be given higher precedence over the corresponding nonnegative terms. Simple rules can also be used to handle negation terms. There are several issues with this approach.[6] For instance, most of the time, in online reviews or any other online text source, the presence of more positive words does not ascertain the review to be positive or vice versa. In majority of the cases, it is very difficult to use the same lexicon for scoring documents of different domains.[7] To address this, a new set of words is added in sentiment lexicons based on the nature of the target domain. Research work is under progress to build domain-specific sentiment lexicons for specific target domains by bootstrapping from an initial smaller lexicon.[7] It involves calculating the sentiment from the semantic orientation of words or phrases that occur in a text. Dictionary of positive and negative words is required for this approach with a positive or negative sentiment value assigned to each of the words. Different approaches to creating dictionaries have been proposed, including manual and automatic approaches.[8] Generally speaking, in lexicon-based approaches a piece of text message is represented as a bag of words. From this dictionary each and every word of the sentences will be allotted a sentiment score within the message[6]. A combining function, such as sum or average, is applied in order to make the final prediction regarding the overall sentiment for the message.[8]

```
[5 rows x 6 columns]
                                                headline  ...  Analysis
0  MRF reports many fold jump in Q1 consolidated ...  ...  Positive
1          MRF informs about newspaper advertisement  ...   Neutral
2  MRF reports 51% fall in Q4 consolidated net pr...  ...   Neutral
3           MRF to raise Rs 1000 crore through NCDs  ...   Neutral
4  MRF reports 2- fold jump in Q3 consolidated ne...  ...   Neutral

[5 rows x 4 columns]
```

Figure 5.1 Output for Sentiment Analysis 1

5. Challenges

In our research we discovered that there are multiple factors that affect the future stock price apart from previous prices and the news headlines. The other factors that are responsible for change in price are inflation, growth in GDP, money supply, C/D ratio.etc. These factors affect the price of stock. Apart from these there are also factors which cannot be measured or tracked such as natural disasters, new government policies, political environment in a country, etc. All these factors have a potential to greatly impact the price and thus become a great challenge for our system.

6. Conclusion & Future work

A. Conclusion

In this work, we are predicting stock prices using old stock prices and financial news articles from the Ticker Finology website. The results suggest that there is a strong relationship between stock prices and financial news articles. And Finally, We will predict whether buying or selling stock will be more profitable to users or not.

B. Future work

- Train the model with more data so as to increase the accuracy.
- In this work, we are predicting stock prices using old stock prices and financial news articles from Yahoo Finance website.
- The results suggest that there is a strong relationship between stock prices and financial news articles.
- Lastly, We will predict whether buying or selling stock will be more profitable to users.

7. Acknowledgement

A few sublime human experiences defy expressions of any kind, and a feeling of true gratitude is one of them. We, therefore, find words quite inadequate to express our indebtedness to our Guide Prof. Aarti Puthran for her virtuous guidance, encouragement and help throughout this work. We express thanks to Prof. Roopali Loolage, HOD (IT), SLRTCE, Mira Road, for extending her support that she gave to truly help the progression of the project work. We express thanks to Prof. Deepali patil, Asst. Prof., SLRTCE, Mira Road for extending her support.

Our sincere thanks to Dr. Umesh S. Badede, Principal, SLRTCE, Mira Road for providing us the necessary administrative assistance in the completion of the work. We also wish to convey my appreciation to my friends who provided encouragement and timely support in the hour of need. Special thanks to our Parents whose love and affectionate blessings have been a constant source of inspiration in making this a reality.

We are extremely grateful to the celebrated authors whose precious works have been consulted and referred to in our project work. All the thanks are, however, only a fraction of what is due to the Almighty for granting us opportunity and the divine grace to successfully accomplish this assignment.

8.References

[1]Harsh Panday, V. Vijayarajan, Anand Mahendran, A. Krishnamoorthy, V.B. Surya Prasath, "Stock Prediction using Sentiment analysis and Long Short Term Memory", European Journal of Molecular & Clinical Medicine, 7-February 2020.

[2]Kunal Pahwa, Neha Agarwal, "Stock Market Analysis using Supervised Machine Learning", 9 International Conference on Machine Learning, 14th -16th Feb 2019.

[3]Sameer Yadav, "Stock Market Volatility – A Study Of Indian Stock Market", 4-April 2017.

[4]Sakarya University, "Stock Market Prediction using Training Algorithms of Neural Network", 2nd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE), 8-10 December 2016.

[5]Aditya Bhardwaj, Yogendra Narayan, Vanraj, Pawan, "Sentiment Analysis for Indian Stock Market Prediction Using Sensex and Nifty", 2015

[6]Ashish Pathak, Nisha P Shetty, "Indian Stock Market Prediction Using Machine Learning and Sentiment Analysis", 04 July, 2018.

[7]Adebayo Felix Adekoya, Benjamin Ashoobam Weyori, "Predicting Stock Market Price Movement Using Sentiment Analysis: Evidence From Ghana", May 2020.

[8]Saloni Mohan, Sahitya Mullapudi, Sudheer Sammeta, Parag Vijayvergia, David C. Anastasiu, "Stock Price Prediction Using News Sentiment Analysis", 2019.