Stock Prediction using Sentiment and Technical Analysis

[1]Smit Kesarkar, [2]Kaustubh Khadilkar, [3]Khushi Singh, [4]Mitali Kasurde, [5]Pooja Mundhe

[1]Student, [2]Student, [3]Student, [4]Student, [5]Professor
[1] CSE Department,
MIT-WPU, Pune, India

Abstract: Nowadays, with the rapid increase in prices, having more than one source of income has become extremely important. One not only needs to save money but also needs to tackle inflation and make sure the money stored is not losing its value while kept in lockers and banks. Money which is not wisely invested will lose its purchasing power in the long run against inflation. Mutual funds ,public provident funds ,gold ,silver ,saving schemes ,real estate ,cryptocurrency and stock market are few places where one can invest money. It has been observed people try to shy away from stock market simply because of lack of their knowledge in this respective field which has also caused various misconceptions. For a layman all the information will be overwhelming if he doesn't possess the required tools and knowledge to analyze the data. This has been the case when it comes to stock market investing since a long time. But with the recent development in deep learning and natural language processing, the gap can finally be filled. These are fast processing models which provide technical as well as sentiment analysis on the stock data using LSTM and LDA which subsequently aids the user in making decisions in stock markets.

Index Terms - Time Series Forecasting, Stock Prediction, LSTM Networks, Sentiment Analysis, LDA.

I. INTRODUCTION

Of India's 1.36 billion people, only about 3.7 per cent invest in equities. A recent study said that Indians tend to save 27% of their income. This saved money tends to lose its purchasing power due to inflation. Everyone wants to earn the most returns that they can get for their money. More and more people are wanting to save and invest their money to have passive sources of income. But people try to stay away from the market due to their fear and ignorance, their lack of knowledge leads them to believe stock market is no less than gambling. With the myriad of data available on the stocks one is overwhelmed .The people who intend to invest in stocks, have limited knowledge on how to analyse the data. Many times people get in touch with financial consulting firms. These firms always charge people heavily share in their profits. This can be avoided if people could make sense of the data available and make informed decisions. If the people had the necessary tools and skillset to analyse the data , the middle men could be completely cut off from the equation. People could become retail investors and have more control and voice over their savings and investments.

When it comes to stock market analysis there is ramification into two different branches which are sentiment analysis and technical analysis. Stock traders using technical analysis work on one most important assumption which is 'History repeats itself'. Technical analysis is the methodology of predicting trend of stock prices based on the past market prices. Stock prices and the overall market is said to be influenced by the sentiments of buyers and sellers. Here is where sentiment analysis comes in, it refers to the overall consensus of stock prices or market as a whole.

The main purpose of both of these branches of analysis is to condense the vast amount of data into a format which is easy to interpret for investors. As a result, getting these analysis methods together will give the potential investor a clearer idea and he will be able to make an informed decision.

I. LITERATURE SURVEY

In the paper by Manaswi and Navin Kumar (2018) RNN and LSTM are used for time series forecasting. Traditionally RNNs were used for sequence and time series prediction. But this posed the challenge of exploding /vanishing gradient. Hochreiter and Schmidhuber came up with a solution to this problem which is LSTM in their paper in 1997. Exploding / vanishing gradient which was faced in RNNs is now solved with the help of LSTMs , because of the gates present in the LSTM cells. Siddharth Banyal, Pushkar Goel and Deepank Grover in their paper "Indian Stock-Market Prediction using Stacked LSTM AND Multi-Layered Perceptron" (2020) have used Multi Layer Perceptron and LSTM for time series prediction , and the results show that LSTMs perform better compared to MLP. Moghar A. & Hamiche M. in their paper "Moghar, A., & Hamiche, M. (2020). Stock Market Prediction Using LSTM Recurrent Neural Network" have also used LSTMs network on the stocks of Nike (NKE) and Google (GOOGL) and the model is shown to have promising results. Qiu, J., Wang B. & Zhou C. in their paper "Forecasting stock prices with long-short term memory neural network based on attention mechanism" (2020) have used LSTM and Gated Recurrent Unit(GRU)model for time series forecasting, where their model is predicting the opening price of the stock and have shown to have excellent results .Their paper also shows that between LSTM and GRU , LSTMs have better fitting degree and improved accuracy for the predictions.

A paper by Siami-Namini, S., Tavakoli, N., & Siami Namin, A on ARIMA vs LSTM'S (2018 ICMLA) did research on the two approaches of time series forecasting and concluded with the results that LSTM's was far more superior to ARIMA model. Sentiment analysis has Natural Language Processing at its backbone. In natural language processing, the Latent Dirichlet Allocation (**LDA**) is a generative statistical model that allows sets of observations to be explained by unobserved groups that explain why some parts of the data are similar. In Zhou Tong's paper "A TEXT MINING RESEARCH BASED ON LDA TOPIC MODELLING" (2016), an introduction of text mining and topic model LDA is represented. It proposed two experiments, which built up topic models on Wikipedia articles and Twitter users' tweets. A brief introduction of each experiment including overview, pre-processing and model training is given and analysed in this paper.

II. SYSTEM OVERVEIW

The basic purpose for developing this system is the prediction of trend of stocks using technical and sentiment analysis data. This paper describes an approach where both the sentiment as well as technical analysis have been incorporated together. This will be accomplished by building a website where the users can login or create an account on the website, after logging in user will be able to select the companies which he would like to predict and then be able to see prediction of trend using sentiment analysis and the general public opinion using sentiment analysis. This paper will be immensely helpful for people wanting to begin investing in stock markets also for the people already investing in stock markets will have analysis done for them to invest wisely
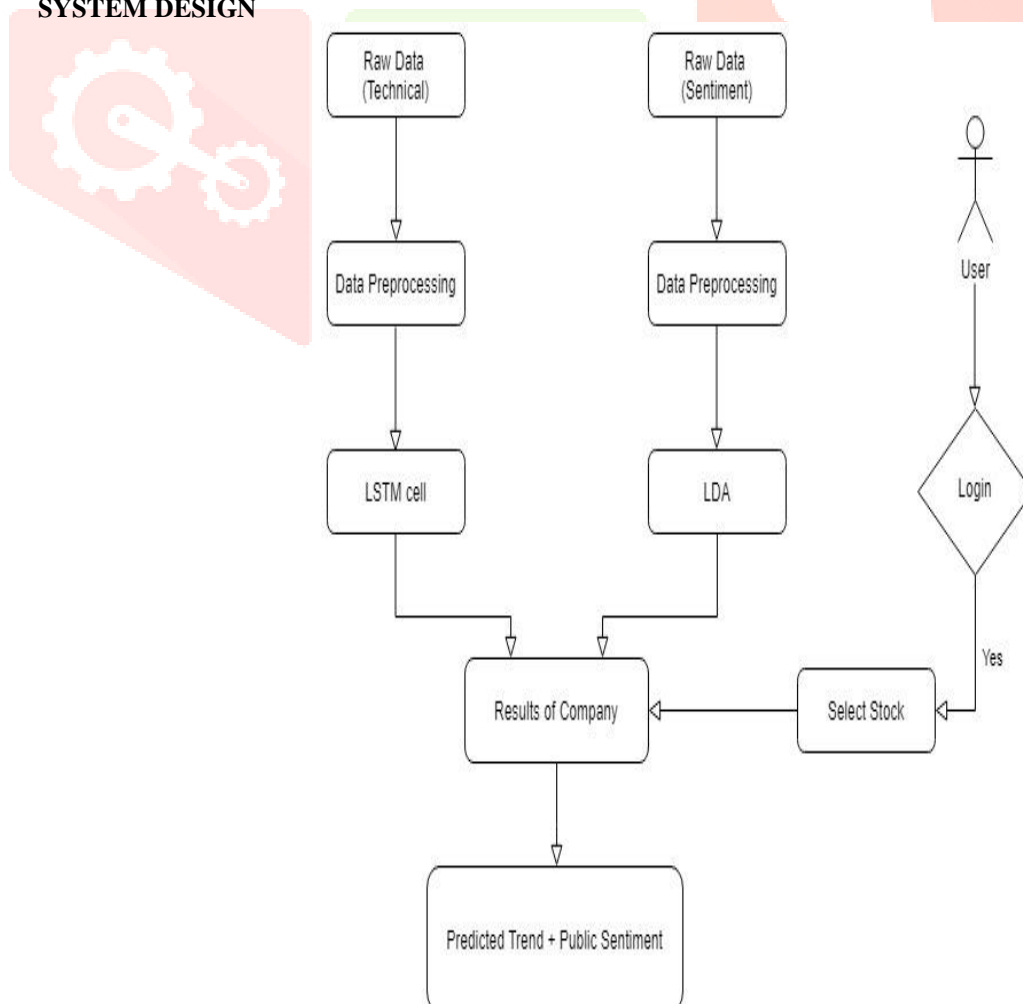
III. SYSTEM DESIGN

Fig-1 : System Flow Diagram

This system provides an interactive website where users can register on the website. After successful login he will be given access to the list of companies listed in the SENSEX. Users can select the company stock which they want to predict. Users will get results of both technical and sentimental analysis on the selected company stock price. Users can add stocks into his portfolio as a watch-listing feature.

IV. METHODOLOGY

4.1 Technical Analysis

The main objective of this analysis is to predict the trend of the stock price. This is done by using LSTM Neural network. As seen in the literature survey LSTM's have proven very efficient in handling time series data, which is exactly what stock prices are.
The dataset used is the price history and trading data of the company which is one of the 30 in the index SENSEX 30 from NSE (National Stock Exchange) India. Dataset is at a day-level with pricing and trading values split across .csv files for each stock along with a metadata file with some macro-information about the stocks itself. This dataset is gathered from the government NSE website [11]. It is necessary to load the data set once it has been identified. Here, Jupyter Notebook which allows developing a web application in an open source that uses python language, just by specifying the path where the data set is located the data is loaded from that specified path into the application by the python libraries.

	Date	Symbol	Series	Prev Close	Open	High	Low	Last	Close	VWAP	Volume	Turnover	Trades	Deliverable Volume	%Deliverble
0	2000-01-03	ASIANPAINT	EQ	361.20	370.0	390.00	370.00	385.0	381.65	380.54	3318	1.262617e+11	NaN	NaN	NaN
1	2000-01-04	ASIANPAINT	EQ	381.65	380.0	392.00	375.00	390.0	385.55	383.50	4818	1.847699e+11	NaN	NaN	NaN
2	2000-01-05	ASIANPAINT	EQ	385.55	371.5	390.00	371.50	383.0	383.00	379.81	2628	9.981384e+10	NaN	NaN	NaN
3	2000-01-06	ASIANPAINT	EQ	383.00	384.9	384.90	374.50	375.1	377.50	379.88	3354	1.274114e+11	NaN	NaN	NaN
4	2000-01-07	ASIANPAINT	EQ	377.50	376.0	390.00	370.00	389.0	385.70	383.38	9589	3.676275e+11	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
5301	2021-04-26	ASIANPAINT	EQ	2517.95	2530.0	2575.00	2530.00	2558.0	2557.90	2557.47	1103980	2.823400e+14	58557.0	335232.0	0.3037
5302	2021-04-27	ASIANPAINT	EQ	2557.90	2545.0	2579.90	2534.00	2571.0	2574.35	2560.55	866331	2.218288e+14	66988.0	281178.0	0.3246
5303	2021-04-28	ASIANPAINT	EQ	2574.35	2588.0	2620.25	2575.00	2612.0	2614.55	2605.91	1065561	2.776752e+14	52622.0	449241.0	0.4216
5304	2021-04-29	ASIANPAINT	EQ	2614.55	2630.0	2642.00	2570.00	2613.0	2613.45	2607.33	1295346	3.377394e+14	64039.0	447979.0	0.3458
5305	2021-04-30	ASIANPAINT	EQ	2613.45	2595.0	2605.80	2524.05	2529.0	2536.40	2558.66	1384907	3.543512e+14	91934.0	750496.0	0.5419

5306 rows × 15 columns

Fig-2 : Dataset of Asian Paints company

As shown in the fig.2 dataset consists of the Open , High ,Low, Close ,Date ,Volume ,Symbol ,Turnover ,Trade , etc. Generally in prediction of stock prices the closing price of the stock is considered. First step of preprocessing of the data would be to get rid of all the information apart from the closing price. By doing this we will be left with a dataset consisting the date and the closing price of the stock on that date. This data will be then fed to the LSTM neural network. Then LSTM will run this time series data and give us the prediction of the future closing price of the stock of the company.

4.2 Sentiment Analysis

LDA is a completely unsupervised algorithm that models each document as a mixture of topics. Then the model generates automatic summaries of topics in terms of a discrete probability distribution over words for each topic, and further infers per-document discrete distributions over topics. LDA makes the explicit assumption that each word is generated from one underlying topic [8].
Data scraping of news websites is done to get news about the stock company we would like to predict the stock of.
Preprocessing of the data is required for sentiment analysis. Preprocessing of the data includes like tokenization, removing stop-start words, stemming words. All these steps help reduce data redundancy which later helps in computation.
After this the data will be fed to the LDA model , which will later give predict whether the company stock has a overall positive or negative sentiment in public.

V. CONCLUSION

The methodology for the purpose of technical and sentimental analysis on the stock prices has been outlined in this paper. The methodology has been studied to predict the trend of the stock prices and general public sentiment of that stock. This prediction of the trend process is studied through the use of a machine learning approach, it can provide a highly accurate trend of the stock as well as the general public sentiment of the stock.. The Studied technique utilizes LSTM neural network for technical and LDA for sentiment analysis. We believe that if this approach is studied and refined more deeply it can be incorporated as a tool which will provide great assistance to people wanting to invest in the stock market.

REFERENCES

[1] **[1]** Qiu, J., Wang, B., & Zhou, C. (2020). Forecasting stock prices with long-short term memory neural network based on attention mechanism. PLOS ONE, 15(1), e0227222.

[2] Moghar, A., & Hamiche, M. (2020). Stock Market Prediction Using LSTM Recurrent Neural Network. Procedia Computer Science, 170, 1168–1173. doi:10.1016/j.procs.2020.03.049 .

[3] Manaswi, Navin Kumar (2018). Deep Learning with Applications Using Python || RNN and LSTM. , 10.1007/978-1-4842-3516-4(Chapter 9), 115–126.

[4] Deep Learning with Long Short-Term Memory for Time Series Prediction Yuxiu Hua, Zhifeng Zhao, Rongpeng Li, Xianfu Chen, Zhiming Liu, and Honggang Zhang.

[5] ISSN: 2278-3075, Volume-9 Issue-3, January 2020 1051 Published By: Blue Eyes Intelligence Engineering & Sciences Publication Retrieval Number: C8026019320/2020©BEIESP DOI: 10.35940/ijitee.C8026.019320 Indian Stock-Market Prediction using Stacked LSTM AND Multi-Layered Perceptron Siddharth Banyal, Pushkar Goel, Deepank Grover

[6] Siami-Namini, S., Tavakoli, N., & Siami Namin, A. (2018). A Comparison of ARIMA and LSTM in Forecasting Time Series. 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)

[7] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long Short-Term Memory. Neural Comput. 9, 8 (November 15, 1997), 1735–1780.

[8] Putri, I., & Kusumaningrum, R. (2017). Latent Dirichlet Allocation (LDA) for Sentiment Analysis Toward Tourism Review in Indonesia. Journal of Physics: Conference Series, 801, 012073.

[9] Shams, M., Shakery, A., & Faili, H. (2012). A non-parametric LDA-based induction method for sentiment analysis. The 16th CSI International Symposium on Artificial Intelligence and Signal Processing (AISP 2012).

[10] Learning, D. and Memory, E. (2019). Essentials of Deep Learning : Introduction to Long Short Term Memory. [online] AnalyticsVidhya
https://www.analyticsvidhya.com/blog/2017/12/fundamentals-of-deep-learning-introduction-to-lstm/#inbox/_blank

[11] https://www1.nseindia.com/products/content/equities/indices/historical_index_data.htm