

Discovering Market Movements for Making Informed Stock Decisions

¹Ananya Rajesh, ²Bryan Philip, ³Maria Johnson, ⁴Anand Pardeshi

^{1,2,3}UG Student, ⁴Assistant Professor, Department of Information Technology, Fr. Conceicao Rodrigues Institute of Technology, Vashi, India

¹rajeshananya9@gmail.com, ²bryanphil31@gmail.com, ³yvaine.mjm@gmail.com, ⁴anand.pardeshi@fcrit.ac.in

Abstract - Over the years stock markets have been a puzzle that small individual investors have failed to crack. A company's stock price is proportional to its growth and way of carrying over the years, stock markets have been seen as a gamble by many. This paper aims to provide a solution to this issue. It introduces machine learning algorithms such as LSTM, SVR and performs Sentiment Analysis using NLP. This paper proposes to revolutionize the investment process for individual/ retail investors in the stock market. It presents a combination of intricately crafted formulae and machine-learning models to predict the longevity of stock in a competitive market. Another important aspect of an investment is tracking the user's portfolio and rebalancing it from time to time to optimize returns from the investments. This paper further focuses on providing a visual and easy to understand representation of the user's portfolio and provides a dashboard that gives graphical insights to users to help them efficiently analyze their holdings across markets.

Keywords – LSTM, Machine Learning, NLP, Portfolio, Sentiment Analysis, SVR

I. INTRODUCTION

The decision of whether to buy or sell shares in the stock market has become an area of interest for researchers. Around the world, there are discussions being made on such important decisions on a daily basis. Economists and investors are now trying their level best to work on the idea regarding the prediction of stock markets.[2]

There are three approaches related to the information required to make a prediction. The first approach, technical analysis, is based on the premise that the future behavior of a financial time series is conditioned to its own past. The second approach, fundamental analysis, is based on external information as political and economic factors. This information is taken from unstructured data like news articles, financial reports or even published microblogs by analysts. In some cases, investors tend to buy after positive news resulting in the stress of buying and higher stock prices; and after negative news, they sell, resulting in a decrease of prices. Finally, the third approach considers as relevant all information coming from both, financial time series and textual data.[3]

II. BACKGROUND THEORY

A. LSTM

Long-Short-Term Memory Recurrent Neural Network belongs to the family of deep learning algorithms. It is a recurrent network because of the feedback connections in its

architecture. It has an advantage over traditional neural networks due to its

capability to process the entire sequence of data. Its architecture comprises the *cell*, *input gate*, *output gate* and *forget gate*.

$$i_t = \sigma(w_i[h_{t-1}, x_t] + b_i)$$

$$f_t = \sigma(w_f[h_{t-1}, x_t] + b_f)$$

$$o_t = \sigma(w_o[h_{t-1}, x_t] + b_o)$$

i_t → represents the input gate.

f_t → represents the forget gate.

o_t → represents output gate.

σ → represents sigmoid function.

w_x → weight for the respective gate(x) neurons.

i_t → represents the sigmoid function.

h_{t-1} → output of the previous LSTM block(at timestamp $t - 1$).

x_t → input at current gate.

b_x → biases for the respective gates(x).

Long term stock prices can be predicted using LSTM.

B. SVR

Support Vector Regression (SVR) is the combination of Support Vector Machines and Regression. Minimizes the margin error. It can be applied to predict values within a nonlinear threshold. It works well when needed to predict stock prices from 1-22 days ahead and are preferable for intraday traders and swing traders.

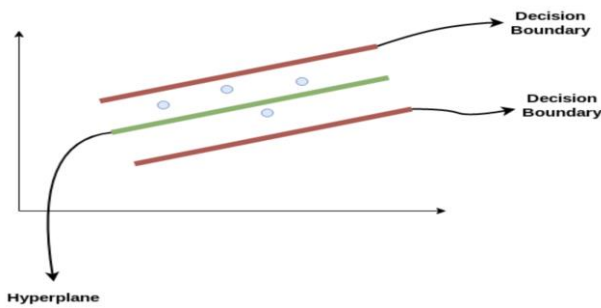


Fig. 1. SVR

As shown in Fig.1 The objective, when moving on with SVR, is to basically consider the points that are within the decision boundary line. The best fit line is the hyperplane that has a maximum number of points.

Assuming that the equation of the hyperplane is as follows:

$$Y = wx + b \text{ (equation of hyperplane)}$$

Then the equations of decision boundary become:

$$wx + b = +a$$

$$wx + b = -a$$

Thus, any hyperplane that satisfies our SVR should satisfy:

$$-a < Y - wx + b < +a$$

C. SENTIMENT ANALYSIS

Instead of the need to go through each headline for every stock a user is interested in, Python can be used to parse this website data and perform sentiment analysis (i.e. assign a sentiment score) for each headline before averaging it over a period of time.

The idea is that the averaged value may give valuable information for the overall sentiment of a stock for a given day (or week if you decide to average over a week's news).

To calculate the document sentiment score, each positive word counts as + 1 and each negative word as - 1.

III. RELATED WORKS

This paper experiments on the performance of DNN with a dataset that is combined with finance tweets sentiment, stock history price and stock price technical indicator. We find the finance tweets that are posted from market closure till market open the next day has more predictive power on next day stock movement. The outcomes of attention-based LSTM models have improved over conventional LSTM on an aggregate dataset. In the individual stock dataset, the results are very interesting. The best result we get is near 65%, which is a decent result. However, the distribution of accuracy is very similar to Gaussian Distribution and it raises a lot more interesting questions to be answered.[1]

In this paper, the LSTM was adopted to predict the “buy” and “sell” recommendations for the stock trading in order to support investors in minimizing their risks when they decide

to buy or sell the stocks in the stock market. This research found that LSTM can achieve the highest accuracy of all stocks. The performance of the proposed model is evaluated and compared to SVM, Logistic Regression, Random Forest, Decision tree, KNN and MLP. The opening, closing, highest and lowest prices are used as the features of all inputs.[8]

Feed Forward Neural Network is the simplest form of deep neural network, where multiple neurons are chained together in the form of multiple hidden layers, each that processes the input and outputs to the next layer as its input. The size of the original dataset dictates the accuracy of the output. The bigger datasets have higher accuracy. The input is sometimes condensed to reduce the number of dimensions using techniques like linear regression, max pooling and 2D-PCA (Principal component analysis). Another way to achieve this is to use a framework called a Recurrent Neural Network (RNN). This is in the form of data from the present as well as the past. To implement such a mechanism, a feedback loop is created such that the output at any instant also acts as the input to the adjoining moment. Long Short-Term Memory is an alteration to the concept of Recurrent Neural Network that carries special gates. These gates consist of a Sigmoid (σ) function that helps it select how much of the input should be considered for the yield, then passed through a tanh function to set the values between (-1, 1).[6] This helps the framework to correct previous mistakes by overwriting them with newer information deemed more accurate.[3]

IV. PROPOSED SYSTEM

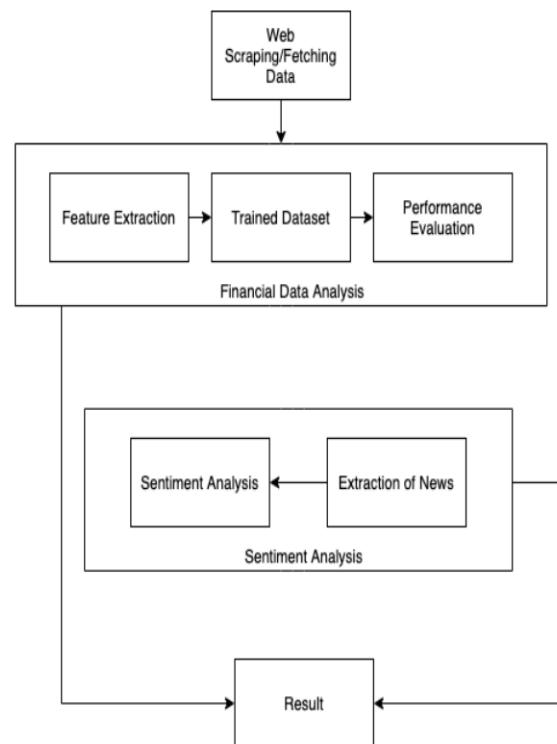


Fig. 2. Schematic Diagram

The following are its components:

A. Financial Data Analysis:

Indian stocks will be used for the implementation of the project. Stock data of the past three years will be used with the help of APIs. Training and testing of the model will be done using the indicators from the API. The system will be evaluated using various evaluation techniques.

B. Sentiment Analysis:

1. Tokenizing the news
2. Removing stopwords from the tokens
3. Do POS (Part of Speech) tagging of the tokens and select only significant features/tokens
4. Pass the tokens to a sentiment classifier which classifies the tweet sentiment as positive, negative or neutral by assigning it a polarity between -1.0 to 1.0

V. RESULT

The financial data and news sentiment analysis is combined to provide recommendations to users regarding the buying, selling or holding of a stock.

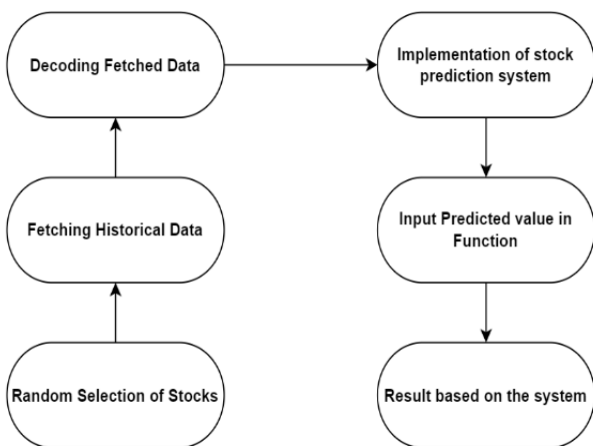


Fig. 3. Flow Chart

The figure depicts the workflow of our project. It will first randomly select stocks and then will fetch its previous data. It will then decode the fetched data and implement the stock prediction system. It will further input the predicted value into the function and will finally conclude with the predicted value based on the function value.

VI. IMPLEMENTATION

This paper is divided into the following modules:

A. Authentication of User Login/Registration

The Sign Up page would facilitate gathering user details and generating User Login credentials for future use. The Home page would be designed to provide the user with an overview of services they would be provided with.

B. Stock Recommendation using SVR and LSTM

Efficient suggestions of stocks to provide optimal returns to investors without having to do tedious research and selection process for value investing. The models covered requirements for short term(SVR) as well as long term(LSTM) transactions.



Fig. 4. Stock prediction for Reliance

C. Stock Recommendation using NLP News

Many studies done by economists have proved that there is a correlation between the general mood of the public and various fluctuating price movements in the stock markets.

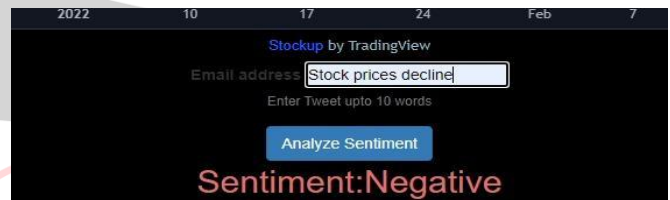


Fig. 5. Twitter Sentiment Analysis

D. Portfolio Tracker/ Balancer

To appropriately diversify and track portfolios for each user.



VII. CONCLUSION

Stock markets are highly volatile. It was observed that during the coronavirus pandemic induced lockdowns, there was a major spike in the number of working class individuals and students entering the markets. But this influx of retail investors demanded a system to help simplify/optimize the investing and stock picking process for those without any prior knowledge of the markets. This paper suggests stocks based on the analysis of the above mentioned online sentiments along with historical data related to a particular company like its net profit/loss, price action, stock MA (Moving Average), growth percentages, traded volumes etc. to forecast its performance in the near future. Along with it a portfolio tracker as mentioned before, can help users track the returns from their investments.

REFERENCES

- [1] P. Vats and K. Samdani, "Study on Machine Learning Techniques In Financial Markets," *2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN)*, 2019, pp. 1-5, doi: 10.1109/ICSCAN.2019.8878741.
- [2] D. Shah, H. Isah and F. Zulkernine, "Predicting the Effects of News Sentiments on the Stock Market," *2018 IEEE International Conference on Big Data (Big Data)*, 2018, pp. 4705-4708, doi: 10.1109/BigData.2018.8621884.
- [3] M. R. Vargas, B. S. L. P. de Lima and A. G. Evsukoff, "Deep learning for stock market prediction from financial news articles," *2017 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)*, 2017, pp. 60-65, doi: 10.1109/CIVEMSA.2017.7995302.
- [4] Shah, Parita & Swaminarayan, Priya. (2021). Sentiment Analysis—An Evaluation of the Sentiment of the People: A Survey. 10.1007/978-981-15-4474-3_6.
- [5] Shah, Bhumi & Shah, Margil. (2021). A Survey on Machine Learning and Deep Learning Based Approaches for Sarcasm Identification in Social Media. 10.1007/978-981-15-4474-3_29.
- [6] A. Menon, S. Singh and H. Parekh, "A Review of Stock Market Prediction Using Neural Networks," *2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN)*, 2019, pp. 1-6, doi: 10.1109/ICSCAN.2019.8878682.
- [7] Y. Xu and V. Keselj, "Stock Prediction using Deep Learning and Sentiment Analysis," *2019 IEEE International Conference on Big Data (Big Data)*, 2019, pp. 5573-5580, doi: 10.1109/BigData47090.2019.9006342.
- [8] T. Sanboon, K. Keatruangkamala and S. Jaiyen, "A Deep Learning Model for Predicting Buy and Sell Recommendations in Stock Exchange of Thailand using Long Short-Term Memory," *2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS)*, 2019, pp. 757-760, doi: 10.1109/CCOMS.2019.8821776.
- [9] A. Sharma, D. Bhuriya and U. Singh, "Survey of stock market prediction using machine learning approach," *2017 International conference of Electronics, Communication and Aerospace Technology (ICECA)*, 2017
- [10] Obthong, Mehtabhorn & Tantisantiwong, Nongnuch & Jeamwaththanachai, Watthanasak & Wills, Gary. (2020). A Survey on Machine Learning for Stock Price Prediction: Algorithms and Techniques. 10.5220/0009340700630071.