

StockGuru: Smart Way to Predict Stock Price Using Machine Learning

Anjana Rajeev¹, Padmanayana², and Harshitha D³

^{1,3} Students, Department of Computer Science and Engineering, Srinivas Institute of Technology, Valachil, Karnataka, India

² Associate Professor, Department of Computer Science and Engineering, Srinivas Institute of Technology, Valachil, Karnataka, India

Correspondence should be addressed to Anjana Rajeev: anjanamalamal@gmail.com

Copyright © 2021 Made Anjana Rajeev et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT- Stock price prediction is a trending concepts in today's world. Proposed work use Twitter data to predict public mood and use the predicted mood to predict the stock market movements. The ceaseless use of social media in the contemporary era has reached unprecedented levels, which has led to the belief that the expressed public sentiment could be correlated with the behaviour of stock prices. Here we develop a system which collects past tweets, processes them further, and examines the electiveness of various machine learning techniques such as Naive Bayes classification and XgBoost algorithm, for providing a positive, negative or neutral sentiment on the tweet corpus. Subsequently, work employ an equivalent machine learning algorithms to analyse how tweets correlate with stock market price behaviour. Finally, examine our prediction's error by comparing our algorithm's outcome with next day's actual close price. Here proposed work takes data from Twitter and also to improve the accuracy proposed work also takes stock data from newspapers and yahoo finance also. The final results seem to be promising as we found correlation between sentiment of tweets and stock prices.

KEYWORDS- Sentimental Analysis, Machine Learning, Twitter API, Yahoo Finance

I. INTRODUCTION

Stock market prediction (SMP) is a method for predicting stock prices, the target of proposed work is a research of text extracted from social media, like Twitter, and analysed to work out the sentiment and emotions of the text, predict stock exchange movements [1]. Contrary to previous studies using daily data, proposed work specialize in high-frequency intraday data to accommodate investment decisions that are made continuously throughout the day. The proposed method is predicated on the utilization of XGBoost classifier. XGBoost provides a wrapper class to permit models to be treated like classifiers or repressors within the scikit-learn framework. The XGBoost model for classification is named XGBClassifier, will create and fit it to our training dataset. Models are fit using the scikit-learn API and therefore the model. Various data processing techniques are frequently involved to unravel this problem. But technique using machine

learning will give more accurate, precise and straightforward thanks to solve such issues associated with stock and market prices. "Stock Price Prediction Using Twitter Sentiment Analysis" a way for predicting stock prices is developed using tweets. The changes available prices of a corporation, the rises and falls, are correlated with the general public opinions being expressed in tweets that company. Positive tweets in social media on companies would definitely encourage people to buy the stocks of that company and as a result the stock price of that company would increase. A prediction model for locating and analysing correlation between contents of tweets and stock prices then making predictions for future prices are often developed by using machine. In the existing work, propose that a company's performance in terms of its stock work movement is foreseen by internal communication patterns. to urge early warning signals, they have a tendency to believe that its vital for patterns in company communication networks to be detected earlier for the assembly of great stock. There's no such application which may give the warning about the stock prediction. However only a few researchers have investigated the thought of making such application. The proposed work aims to gather Twitter data from search API and stock value from yahoo finance website. A machine learning algorithm is applied to the labelled dataset, and patterns are extracted, which successively perform sentiment analysis. Machine learning classification algorithm, namely naïve bayes, and therefore the XGBoost classifier are wont to classify the tweets. The classifier is fed with the dataset to gauge the accuracy of the stock price. The obtained value is compared with the important time-input to predict the longer-term stock movements [2].

II. RELATED WORK

The later work is designed to recognize the correlation between the stock market and the twitter data& use them to predict the future behaviour of various stock prices. Firstly, work consider the dataset of yahoo! Finance and also the collection of tweets from Twitter's search API and pre-processes it in order to become suitable for relabel analysis. Tokenization of the tweets is done where the tweets are divided by spaces and forming a list of individual tweets,

later each word in the tweet is used to train the classifier [3]. Next, they remove stop words from the list of words. Once the processing of data is completed, the sentiment analyser must be trained. Work turns the tweets as training data by converting them into positive and negative dataset for training, by using Twitter search API. Then uses external lexicon AFINN, which is a list of English words rated for valance with an integer between minus five and plus five. Next step is to construct and train a classifier. Naive Bayes Bernoulli and support vector machine are the two classifiers used for prediction. To calculate accuracy, it makes use of use k-fold cross validation method. In this method they train on all days up to a specific day and test for the next k days. K-fold cross validation splits training dataset into k smaller datasets. One of these subsets is left out and used as a test set to measure accuracy and the other k-1 subsets and uses to train the classifiers. They use support vector machine to achieve a prediction about future stock movements. Later the result will be compared with the upcoming stock exchanges. The achieved average accuracy is 87% concerning correct stock's movement prediction. Another research includes, investigating the impact of sentiments extracted from a large-scale collection of tweets from StockTwits for five Companies (AAPL, AMZN, GE, MSFT, and TGT). Work performs the sentiment analysis on collected StockTwits data using three machine learning methods (Naïve Bayes, SVM, and logistic regression) and five featurization techniques (bag of words, bigram, trigram, TF-IDF, and LSA). The combination of logistic regression and TF-IDF is able to achieve reasonably high accuracy level, between 75% and 85%, for all the five companies. The latter yields improved accuracy, with magnitude of the improvement higher for companies with more sentiments data. This provided a reasonable evidence that sentiments data has a positive impact on the accuracy of stock price change prediction. Another previous work, applies sentiment analysis and machine learning principles to find the correlation between "public sentiment" and "market sentiment". In this paper, they test a hypothesis based on the premise of behavioural economics that the emotions and moods of individuals affect their decision-making process, thus, leading to a direct correlation between "public sentiment and market sentiment". Recent studies have shown that social media sources can have an observable effect on investors opinions towards financial markets [4]. Lima et al. [5] improved the accuracy of predicting stock trends using SVM with an overall public sentiment attribute. In [6] the author observed the changes in the stock market related to the public sentiment. The author makes use of word2vec and N-gram textual representations, to analyse the tweets. In this paper they have applied sentiment analysis and machine learning algorithm to predict the stock market value. At the end of this paper, it shows a correlation between the stock market changes with in twitter tweets. In paper [7] the author analyses the predictability of stock market value using the data extracted from social media naïve bayes, K-nearest neighbours and the support vector machine are the three algorithms used for classification. They made use of evaluation metrics such as recall, precision and F-1 score to compare and test the performance of these algorithms. Later, K-Fold cross validation method is used to validate the results of machine learning models. Finally, they indicated that the sentiment and emotion obtained from tweets are strong predictions of stock market movements. John kordonis

research project had a gal to recognize the tweets related to stock market to predict the future stock prices. In this paper they develop a system which collects the tweets and perform further process and use machine learning algorithms such as naïve bayes Bernoulli's classification and support vector machine. Finally, the predictions error is examined by comparing the outcome of algorithm with the next days actual close price. Finally, the overall results found to be accurate with correlation between sentiment of tweets and the stock price [8].

III. SYSTEM IMPLEMENTATION

The system is implemented in following architecture, the user has to input a company name like Google, Tesla, HDFC etc. The proposed system will fetch the twitter tweets corresponding with names of the input company name, also the news headlines corresponding to the company also fetched using Twitter API and fin viz API respectively. The stock data like open close principle, etc. are being fetched from Yahoo finance API. The fetched twitter and news headlines are pre-processed in order to remove special character like #, \$," also emojis are being discarded.

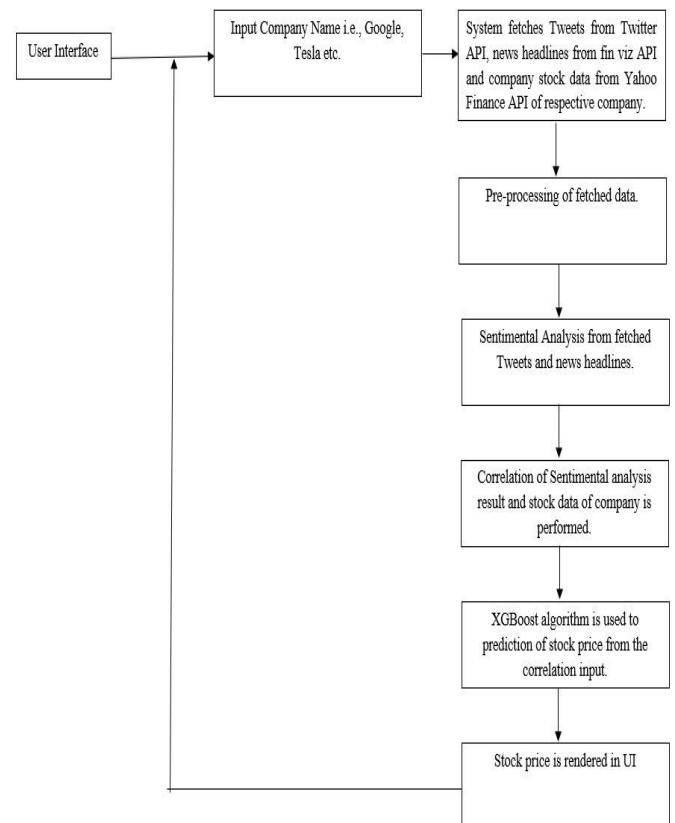


Fig. 1: Block Diagram of Proposed Methodology

The pre-processed data is then passed for sentimental analysis, the proposed system (figure 1) uses a Naïve Bayes model to perform sentimental analysis and the outcome is categorized into 3 class i.e., positive, negative and neutral of respective data fetched. This outcome is then merged with fetched stock data and passed to XGBoost model. This model predicts a stock price based the correlation data passed between the current market mood from sentimental analysis and current stock price of company. This price is rendered to UI.

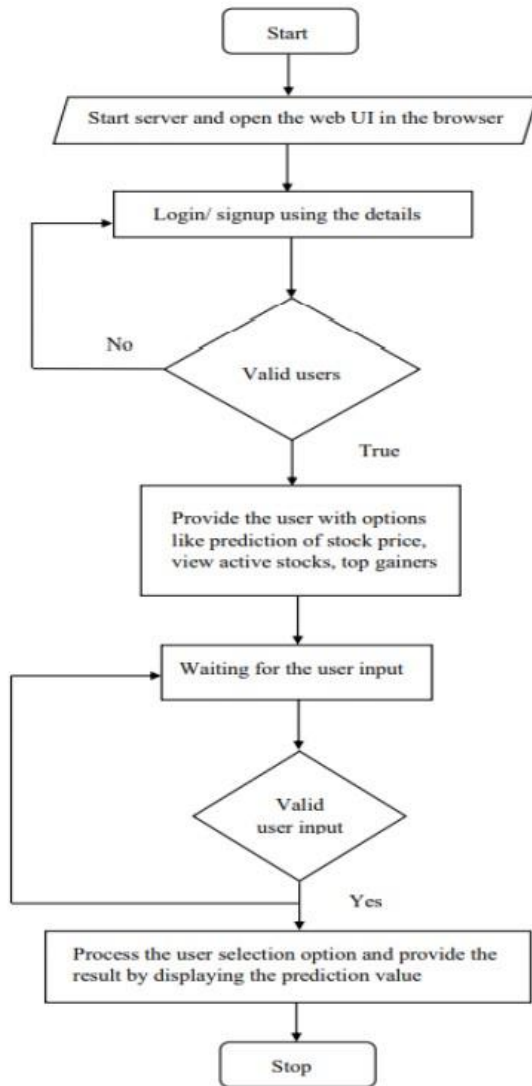


Fig. 2: Flowchart of User Interface of proposed system

The figure 2 describes the flowchart of user interface of proposed system.

Initially the user has to create an account in order to use the system. The existing user can login with their user credentials.

One's logged in user can perform various operation like get the stock price prediction, analyses over stock behaviour of particular at certain time interval, top gainers and losers of stock market etc.

IV. EXPERIMENTAL RESULT

The system is being designed and following outputs have been obtained as results. In figure 3 shows the user options to login or create an account in order to access the proposed system portal

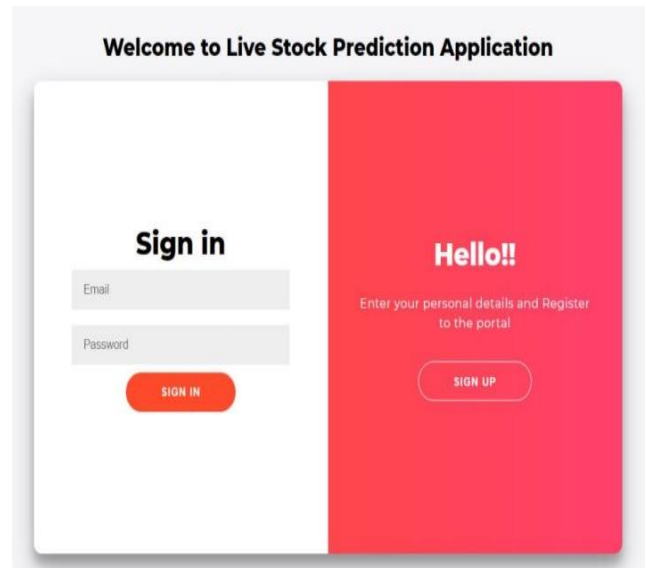


Fig. 3: Login/Signup Page

In figure 4 and 5 the homepage and description pages of user interface which contains details like project description, methodology etc.

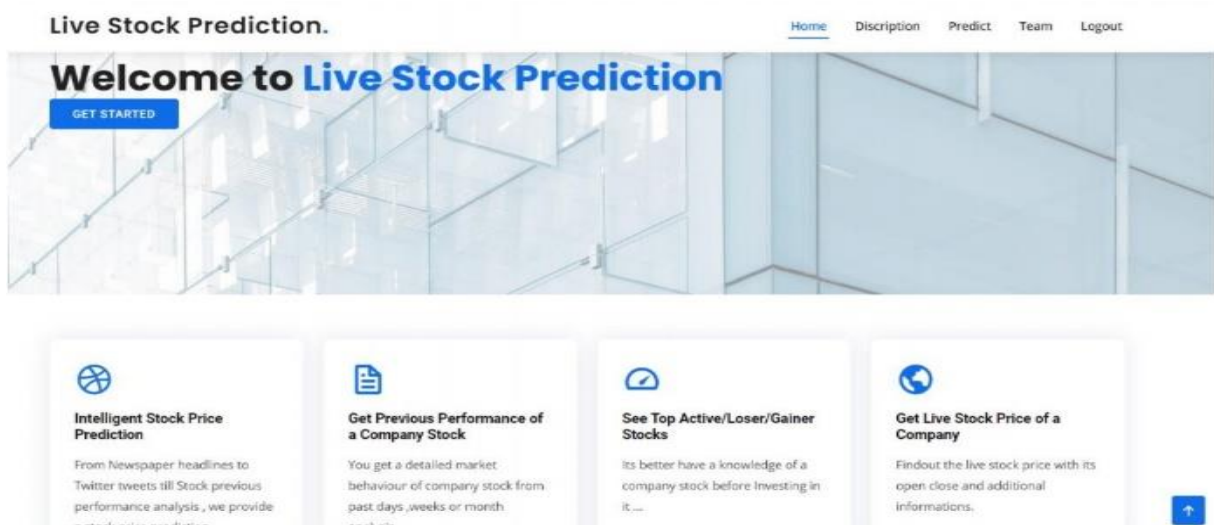


Fig. 4: Home Page

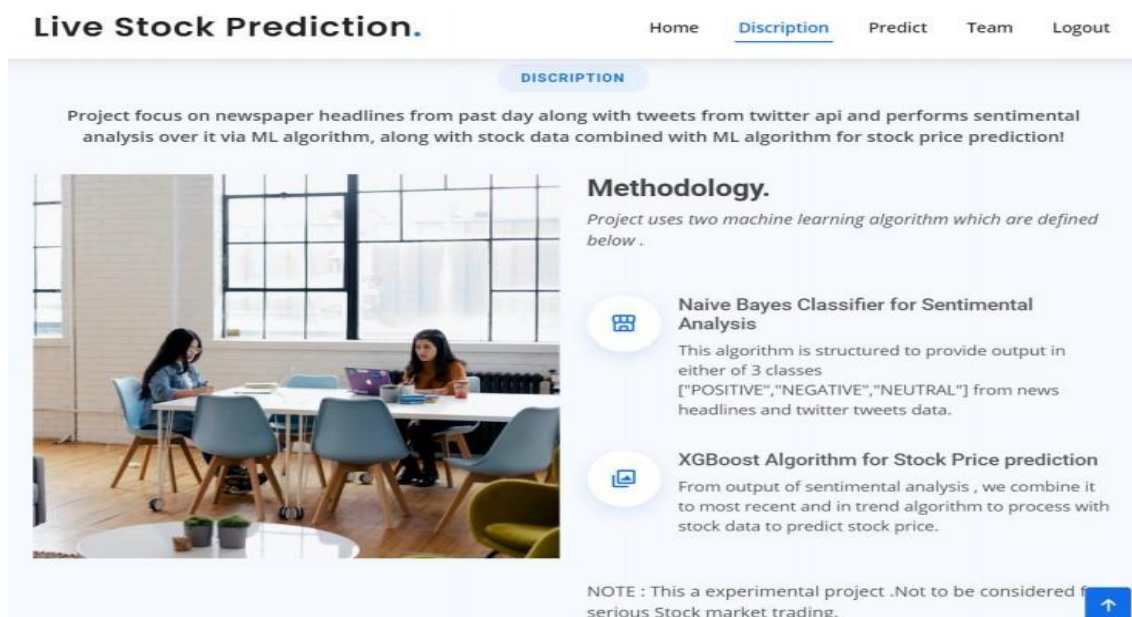


Fig. 5: Description Page

In figure 6 user option section where user can provide input to various feature like current stock price fetching, current and time interval-based analysis of company stock etc.

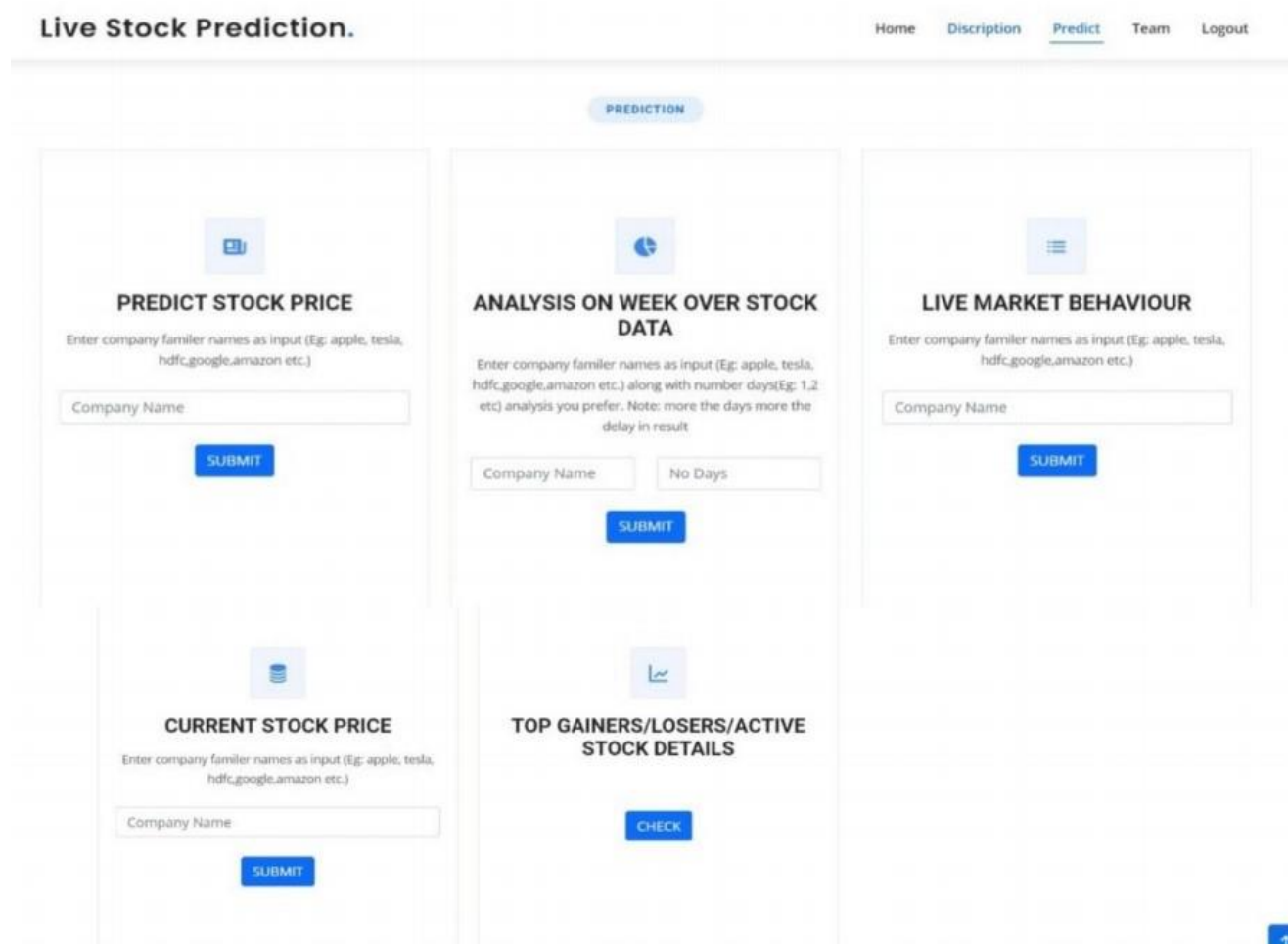


Fig. 6: User option page

In figure 7 shows the result of prediction done by proposed system for company stock price.

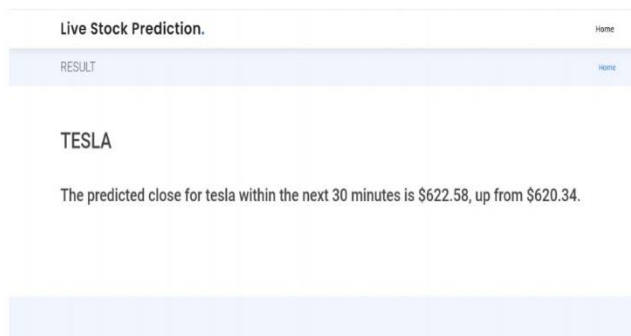


Fig. 7: Result Page

V. CONCLUSION

Stock market prediction using sentiment analysis makes use of prediction model to predict the stock price. In this project we make take into consider the live twitter data along with the stock data to predict the accuracy. Our results show that changes in the public sentiment can affect the stock market. That means that we can indeed predict the stock market with high chances. The accuracy obtained for predicting the stock price is 89%. Our dataset really extracts the real public sentiment from twitter API. We performed the sentiment analysis on collected StockTwits data using two machine learning algorithms i.e., Naïve Bayes and XGBoost classifier for prediction. The combination of these two is able to achieve reasonably high accuracy level, between 85% and 95%, for all the companies. The correlation analysis and stock price change prediction are done to measure the impact of sentiments on stock price. Comparison is done between using stock price and as only input and using a combination of stock price and sentiments. The latter yields improved accuracy, with the improvement higher for companies with more sentiments data. The future research in the field could investigate the importance of further developing the sentiment analysis to take more parameters in consideration as described in the conclusion. When analysing social media, new trends such as the use of emoticons, emojis and language slang must be taken into account in order to get satisfiable accuracy of the sentiment analysis. Furthermore, gathering of social media data during a longer period of time would be of interest. Extending the data mining to gather information from financial resources and newspapers could serve as an extension to traditional stock market prediction approaches on financial data only.

REFERENCES

- [1] R. Ahuja, H. Rastogi, A. Choudhuri and B. Garg, "Stock market forecast using sentiment analysis", 2nd International Conference on Computing for Sustainable Global Development, pp. 1008-1010, 2015.
- [2] S. Urolagin, "Text mining of tweet for sentiment classification and association with stock prices," Proceedings of 2017 International Conference on Computer and Applications, pp. 384-388, 2017.
- [3] T. O. Sprenger, A. Tumasjan, P. G. Sandner, and I. M. Welp. Tweets and trades: The information content of stock microblogs. European Financial Management, 20(5):926–957, 2014.
- [4] P. Wei and N. Wang, "Wikipedia and stock return: Wikipedia usage pattern helps to predict the individual stock movement," Proceedings of the 25th International Conference Companion on World Wide Web, pp. 591-594, 2016.
- [5] M. L. Lima, et. al., "Using sentiment analysis for stock exchange prediction", International Journal of Artificial Intelligence & Applications, vol. 7, pp. 59-67, 2016.
- [6] <https://arxiv.org/pdf/1610.09225>
- [7] <https://www.econstor.eu/bitstream/10419/215436/1/GLO-DP-0502.pdf>
- [8] https://www.researchgate.net/publication/311843931_Stock_Price_Forecasting_via_Sentiment_Analysis_on_Twitter

ABOUT THE AUTHORS



Anjana Rajeev, Under Graduate Final Year Student seeking her degree in Computer Science and Engineering at Srinivas Institute of Technology, Valachil, Karnataka, India



Padmanayana, Associate Professor in Computer Science & Engineering Department at Srinivas Institute of Technology, Valachil, Karnataka, India



Harshitha D, Under Graduate Final Year Student seeking her degree in Computer Science and Engineering at Srinivas Institute of Technology, Valachil, Karnataka, India