

Stock Prices and Risk Prediction with Sentiment Analysis Using Machine Learning Technology

Prof. Siddhesh Khanvilkar, Mr. Aditya Jadhav, Mr. Aniket Khaladkar, Mr. Sahil Lamture

Department of Information Technology,
Pillai HOC College of Engineering and Technology,
Panvel, Maharashtra, India

skhanvilkar@mes.ac.in, adityajadhav1234@gmail.com, aniketvsk5@gmail.com, sahillamture20@gmail.com

Abstract- “Stock Prices Prediction” is a key topic in the present situation. Therefore, the curiosity about this topic is increasing among the researchers as well as retail investors. After multiple research one thing has everyone noticed that past data of stock market, Search Engine optimization Queries and the data which is generated from sources like Twitter and other financial platforms, has a close relationship with the future price of any stock. Previously a decade ago there was problem of lack of related information in the field of stock market. Therefore, in order to fill that information bridge between market and investors and predict an future, we discuss an simple method to analyse different information source. To do so, Long Short Term (LSTM), Linear Regression and Autoregressive Integrated Moving Average (ARIMA) etc. algorithms were utilized to examine separate sources, which helps users or investors precise results of stock prices of next 7 days or so. Through this method, new investors or retail investors can survive in the stock market.

Keywords- Sentiment Analysis (Text Blob), LSTM, Linear Regression, ARIMA, etc.

I. INTRODUCTION

Global economy has strong impact on stock market which is considered as highly dynamic market. There are many factors behind it which affect the market directly or indirectly [6]. Economic factors like Interest rate and inflation, market psychology, government policies, natural disasters, and other direct factors conduct a strong effect on stock markets.

In recent years, to gain high profit from stock market researchers and retail investors have keen interest in predicting the stock prices [11]. Due which rise of new techniques for data pre-processing and data modelling and with the help of data mining, big data, ML, AI investors have likely to use different webapp resources to predict stock market fluctuations such as company’s fundamental and technical performance [5][6].

Because of globalization and the financial market volatility [1], it is very difficult to predict stock price actions only using various theories and information present in the market. For effective result, above mentioned things should be considered for prediction [6].

II. EXSITING SYSTEM

1. Fundamentals of Analysis of Stock:

Fundamental analysis approach focuses on a company’s previous performance and results. Performance such as ROI (Return on Investment), Dividend, Revenue, earning, Compounded Annual Growth Rate (CAGR), P/B ratio,

etc., [2]. are used to analyze stocks which may or may not results a positive price action [4][8]. This approach totally based on fundamentals of particular company or stock.

2. Technical Analysis of Stock:

Technical approach represents that, the prediction of future prices done by applying time series analysis on past data or historical data [5]. As well Statistical techniques such as Bollinger band, RSI, MACD types of indicators are applied to predict the recent trend in the company’s stock [9]. Whether it is upward or downward depending upon market trend.

III. PROPOSED METHOD

1. By Qualitative Analysis of stock:

Company related news in the stock market highly affect the market price action trend. Presently, twitter has become the most effective and fastest way of used media [4]. Every one shares their positive or negative sentiments on this platform [7] [9] [11]. Therefore, with combined resources of news feed as well as twitter data analysis of such high scale data can be easily done with the help of Sentiment analysis.

2. By Quantitative Analysis of stock:

Historical data or dataset of company is now easily available over internet for stock markets [3]. So, by applying multiple machines learning models on this dataset gives user an accurate result for future investments [4]. These ML models can be trained on basis of particular stocks as well on sectorial basis.

IV. METHODOLOGY

1. Data Collection:

Historical data is cumulated from yahoo finance and another related format [7]. Along with this we are using tweet API to collect and analyse tweets related for sentiment analysis.

2. Data Pre-processing:

Here different techniques were utilized, which pre-process the tweet data collected from tweet API [6]. Filtering, Data cleaning, tokenization, Data transformation, Data reduction.

3. Sentiment Analysis:

Based on the sentiment analysis, tweets are break down as positive, negative and neutral. These tweets are studied by users and denoted as 0 for Neutral, 1 for Positive and 2 for Negative [10].

4. Display Output:

The produced output is displayed on webpage after performing sentiment analysis as well as timeseries prediction [1]. Along with this, it will display future predicted prices of stock.

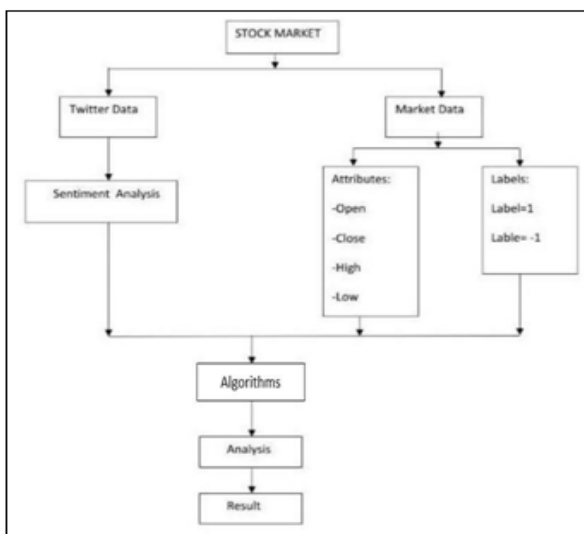


Fig 1. Flow Diagram.

5. Data:

To predict the stock prices, there is a need of predicting variable and these variables are nothing but closing price of stock, return values of different stock market [5][9]. The data used to predict the stock price values. Live stock prices and historical market data were piled up using Yahoo finance API [8].

For example, in order to download around 0.8 million tweets for any company Python's web crawling app is written. Collected Tweets were English, and company's ticker symbol or tag was the keyword used to search

different Tweets [11]. e.g., AMZN for Amazon.com Inc [8]. Twitter sentiments are created with each tweet and the result of sentiment analysis is calculated based on positive or negative sentiments. Cash tag helps to filter news, means it will eliminate the irrelevant company names e.g., we can ignore apple fruit or any other contrast text from the received search results, etc [2].

6. Procedure:

Now, data collected from previous part needs to be pre-processed [1]. As we know, stock market is closed on Saturday and Sunday. So, for weekends empty or null spaces were come with the corresponding weekday closed stock values. Particularly, on Saturday and Sunday. For every fraction of seconds, there are multiple tweets generated for a company [9]. The sentiment scores were calculated after pre-processing all that tweets [6].

As well as sentiment value mean score value and the Twitter volume of each day of stock is calculated based on available data. Positive, Negative or neutral is defined using relative kind of polarity presented during time interval [1]. Sentiment analysis is a very important part in this analysis. Tweeter data is depending upon the accuracy of the polarity score and to strengthen the prediction. A widely used tweeter text processing library of python called Text Blob is used to perform Sentiment analysis [2][3][4].

The different kind of algorithms like ARIMA, LSTM (Long short-term memory) etc. were used, since predicting stock prices is our goal [5]. Among these algorithms, LSTM gives comparatively best results.

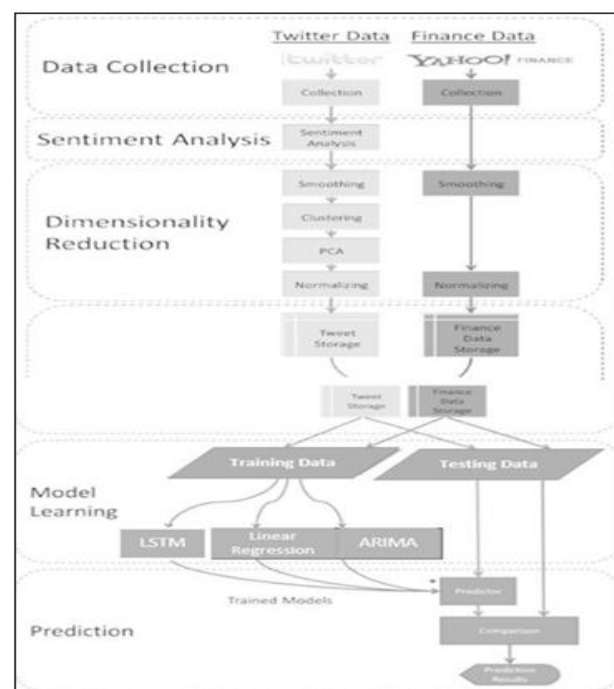


Fig 2. Collection and Processing.

7. Algorithms:

7.1 Long Short-Term Memory (LSTM):

In the below (fig 3), actual price is represented by red or orange line and the predicted price is represented by blue line. The LSTM model efficiency is revealed by fig. 3 graph [2]. For the predicting the real trend approximately, an amount of time should pass the trend. The model result in a Trained Score of 0.050 RMSE value and a Test Score of 0.10 RMSE [10][11]. The LSTM based Models has greater accuracy than the Regression based Models. Again, it varies from stock to stock [5].

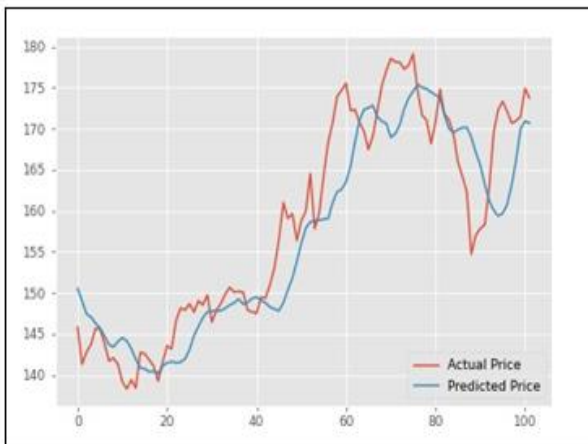


Fig 3. Actual vs. Predicted Trend of LSTM.

7.2 Linear Regression:

The figure shown below (fig 4), is output of linear regression algorithm which predict varying prices with respect to the time on the given data [2]. The Root mean-square test give result in a confidence score of 0.87825 [11].

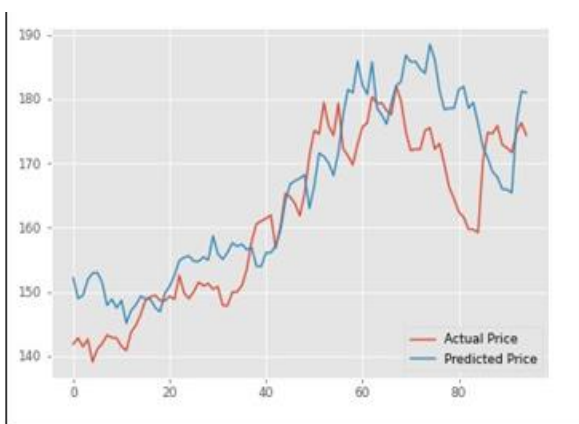


Fig 4. Actual vs. Predicted Linear Regression.

7.3 Autoregressive Integrated Moving Average (ARIMA):

The results shown in the graph (fig 5) is from the ARIMA model, which is widely used in predicting stock market prices and has a tremendous capability for prediction specially in case of derivative market [9].

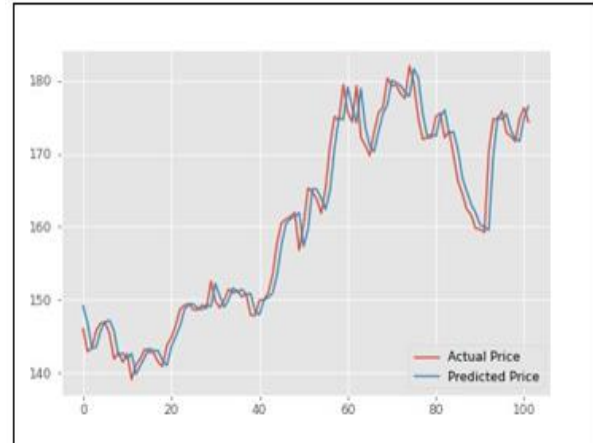


Fig 5. Actual & Predicted Trend of ARIMA.

V. FUTURE SCOPE

Prediction accuracy percentage is high. It also helpful to investors to invest for long term/Swing trade. We predict and compare stock prices and risk using different algorithms.

The model will help a lot of new investors in deciding when to buy or sell a particular stock. In future we can include derivative market stock based on future demand by the investors or retailers.

VI. CONCLUSION

The main agenda of this study is to estimate or predict stock prices of any company from national as well as international market (NSE, Dow Jones 30, etc) [8]. by using algorithms, which were introduced in this paper [10]. As mentioned, both the historical data as well as user-generated data that is tweets were gone through different stages of pre-processing to fit different algorithmic models [4]. And the model weightage was different with the prediction available time range.

As it will be inversely proportional with each other. With the increase in time data range, the weightage for the twitter sentiment prediction decreases and with the decrease in prediction of time data range of twitter model shows increasing of weightage [11].

From this analysis we can conclude that, the twitter is beneficial for swing trading than for long-term predictions [9].

VII. ACKNOWLEDGEMENT

Authors thank Prof. Siddhesh Khanvilkar, Associate professor at University of Mumbai and Head of Department, Dr. Divya Chirayil at Pillai HOC College of Engineering and Technology, Rasayani, Panvel.

REFERENCES

- [1] Sulaiman Olaniyi, A., Adewole, K. and Jimoh, R. (2011). Stock Trend Prediction Using Regression Analysis – A Data Mining Approach. [online] <https://www.researchgate.net>. Available at: https://www.researchgate.net/publication/277409163_Stock_Trend_Prediction_Using_Regression_Analysis_A_Data_Mining_Approach [Accessed 11 Jan. 2019].
- [2] S. Liu, G. Liao and Y. Ding, "Stock transaction prediction modelling and analysis based on LSTM," 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA), Wuhan, 2018, pp. 2787-2790.
- [3] Trans. Syst., D. D. Wu, L. Zheng, and D. L. Olson, "A decision support approach for online stock forum sentiment analysis," *IEEE Trans. Syst., Man, Cybernet., Syst.*, vol. 44, no. 08, pp. 1077-1087, Jan 2014. Stock Twits. <https://stocktwits.com/>
- [4] N. Oliveira, P. Cortez, and N. Areal, "Proceedings of the ACM International Database Engineering & Applications Symposium," pp. 115-123, 2014. Automatic creation of stock market lexicons for sentiment analysis using Stock Twits data.
- [5] "The Behavior of Stock-Market Prices," *The Journal of Business*, E. F. Fama, vol. 38, no. 1, pp. 34-105, 1965. Available: https://www.researchgate.net/publication/313473231_Stock_Price_Change_Prediction_Using_News_Text_Mining. [Accessed: 03-Mar-2019] M. Beckmann, "STOCK PRICE CHANGE PREDICTION USING NEWS TEXT MINING", 2017. [Online].
- [6] Dassanayake, W. and Jayawardena, C. (2017). Determinants of stock market index movements: Evidence from the New Zealand stock market. [online] <https://www.researchgate.net>. Available at:
- [7] https://www.researchgate.net/publication/314668127_Determinants_of_stock_market_index_movements_Evidence_from_New_Zealand_stock_market [Accessed 2 Jan. 2019].
- [8] Mao, H., Counts, S. and Bollen, J. (2011). Predicting Financial Markets: Comparing Survey, News, Twitter and Search Engine Data. [online] [Arxiv.org](https://arxiv.org). Available at: <https://arxiv.org/pdf/1112.1051.pdf> [Accessed 11 Jan. 2019].
- [9] B. Qian and K. Rasheed, "Stock market prediction with multiple classifiers," *Applied Intelligence*, vol. 26, pp. 25-33, 2007.
- [10] R. F. Stambaugh, J. Yu, and Y. Yuan, "The short of it: investor sentiment and anomalies," *Journal of Financial Economics*, Vol. 104, pp 288-302, 2012.