

## Stock Market Prediction Analysis

Riya Sudam Bote<sup>1</sup>, Shikha Virender Chandel<sup>2</sup>, Mital Krushna Donadkar<sup>3</sup>, Prof. Narendra Gawai<sup>4</sup>

<sup>1,2,3</sup> Students, Dept of Computer Science and Technology, Usha Mittal Institute of Technology, Mumbai, India

<sup>4</sup> Professor, Dept of Computer Science and Technology, Usha Mittal Institute of Technology, Mumbai, India

\*\*\*

**Abstract** - Stock market has been playing a vital role in financial market. Even a small commodity has some or the other effects due to change in stock market. One needs investors for the growth of the company who are attracted by the stock price or market value of the company. So, the precise prediction of a stock price can help one to gain noteworthy benefits. Hence, this topic has been taken by researchers and many models and studies have been made through many years. In this paper, a new stock price prediction framework is proposed utilizing various models and making a hybrid model from the same; Long Short Term Memory (LSTM) model, Support Vector Regression (SVR) model, Linear Regression model along with Sentimental Analysis. From the simulation results, it can be noted that using this hybrid model with proper hyperparameter tuning, our proposed scheme can forecast future stock trends with high accuracy. The assessments are conducted by utilizing online datasets for stock markets having open, high, low, and closing prices.

**Key Words:** LSTM, SVR, Linear Regression, Sentimental Analysis, Stock Market Prediction.

### I. INTRODUCTION

Stock investments prove to be an alternate source of income in today's time. Some invests as a retirement source while some as a primary source. Even companies bond their workers by keeping their shares in their vicinity in the name of the company. Even though stocks are volatile in nature, visualizing share prices on different parameters can help the keen investors to carefully decide on which company they want to spend their earnings on. Data visualization is one of the prominent fields in today's technology. The stock market analysis and forecasting has been an important and emerging trend since developments in the field of machine learning. The financial institutions, brokerage firms, banking sectors and other sections use such analysis methods in order to gain knowledge about stock scores. The companies uses this to protect their companies from the risk of drop of their shares due to their investments. Even government bodies whether it is developed countries or developing countries use stock analysis for their growth as it effects the market price of other things also. The necessary factor involved in the monitoring has been providing security assurances to the volatile firms. The basic objective of stock market forecasting has been, in aiding to make the buy, sell or

hold decision. This application using the Streamlit, we can create dynamic plots of the financial data of a specific company by using the tabular data. On top of it, we can use a machine learning algorithm to predict the upcoming stock prices. This web application can be applied to any company (whose stock code is known) of one's choice.

### II. PROPOSED SYSTEM

In our proposed system the data will be taken from yfinance which will be online data. An ensembling model using the shown algorithms will be created i.e Linear Regression, SVR & LSTM. The algorithms are chosen as per how better they worked which is concluded from literature survey given forward. Stocks are considered to be combination of linearity and non-linearity so different algorithms are used to increase accuracy. Along with this model, sentiment analysis is also been done for testing polarity. As we can gain knowledge from positive and negative tweets which can help us to get expert advice besides it also works better for SVM.

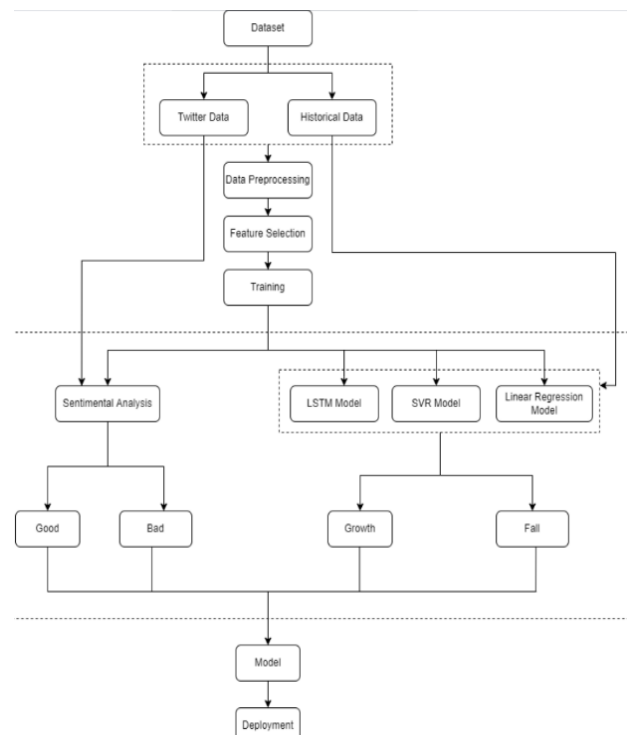


Fig.1 Proposed System

## II. RELATED WORK

Sonawane Nikita, Sonawane Abhijit, Sharma Aayush, Shinde Sagar, Prof.A.S Bodhe[1] introduced an LSTM (long short time memory) algorithm for training a stock prediction model. The paper gave an outstanding result in spite of the huge dataset. It recommended that stock advertisement classifying its classification depends on a particular company. The results obtained were better than other algorithms. The result can be considered uncommonly promising since it has illustrated able to predict well compared to other approaches.

Rahul Mangalampalli, Pawankumar Khetre , Vaibhav Malviya , Vivek Pandey[2] presented a different method containing Hybrid model of ARIMA and GRU. In this paper the author provided two approaches. The first approach was build on the news content, while the second approach was build on market data such as stock price at the time of news release, close price and change in indicator values were included in the classifier input. If we analyzed the market around the publication of the news it may lead to more precise labels.

Azadeh Nikfarjam, Ehsan Emadzadeh, Saravanan Muthaiyah[3]presented different methods to show the impact of financial news and tweets on prediction of stock market. They offered two approaches in which the first approach was based on news content, And the second approach was based on market data with parameters opening of stock price, closing of stock price and change indicator values were included in the classifier. The stock prices were recorded in a time interval around the broadcasting of the news.

Eunsuk Chonga, Chulwoo Hanb, and Frank C. Park[4] offered a structured analysis of the usage of deep learning networks for analyzing and predicting the stock market. The main characteristics such as extracting features from a large data set without depending on any external factors makes deep learning potentially attractive for prediction of stock market, they used intraday stock returns of high frequency as their input data, they also examined the effects of unsupervised feature extraction methods which were principal component analysis, autoencoder, and the restricted Boltzmann machine. Their study presented practical insights on how deep learning networks can be productively used for prediction and analysis of stock market.

Venkata Sasank Pagolu, Kamal Nayan Reddy Challa ,Ganapati Panda[5] presented a paper where they have employed two different textual representation that are word2vec and Ngram for evaluating the sentiments of public in tweets. They have used sentiment analysis and supervised machine learning algorithms to the tweets which were extracted from twitter and calculating the

connection between stock market, sentiments and movement of a particular company. As we all know positive tweets and news in social media of a respective company will inspire people to invest in that company. Atlast, the conclusion of the paper states that there is a strong relation between the rise and fall of the stock price and the public sentiment.

Ashwini Pathak, Sakshi Pathak[6] presented a paper which focused on applying various machine learning algorithms such as Support Vector Machine(SVM), KNN, Logistic Regression and Random Forest algorithm on the dataset. They calculated the performance metrics like recall rate, precision, accuracy and fscore. Their main purpose was to identify the best algorithm for stock market prediction.

G) Wu, M.C., Lin, S.Y., and Lin, C.H researched about stock market prediction using decision tree algorithm. The main approach was to modify the filter rule which was done combining three decision variables which were linked with fundamental analysis. The empirical test, which used the stocks of an electronic company of Taiwan, showed that the approach used outperformed the filter rule. In this work they only considered the past information; the future information was not taken into consideration at all. Their research had aimed to upgrade the filter rule by considering both future as well as past information.

## IV. IMPLEMENTATION

Irjet Template sample paragraph .Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

As it is said that stock market prediction is a complex problem since it has many factors. By using previous data and relating it to the current data with help of proper ML techniques one can easily predict the stock. For prediction of stock the techniques we used are SVR, LSTM and Linear Regression paired with sentimental analysis

The implementation of the project is divided into following steps:

### A. Data Pre-processing

The dataset which we used for training and testing the model was picked up from yfinance(Yahoo Finance). It contained around 9 lakh record of a particular company stock. The variable which we considered for pre-processing date, closing price and opening price.

```
Out[3]: {'zip': '94043',
'sector': 'Communication Services',
'fullTimeEmployees': 156500,
'longBusinessSummary': 'Alphabet Inc. provides online advertising services in the United States, Europe, the Middle East, Africa, the Asia-Pacific, Canada, and Latin America. The company offers performance and brand advertising services. It operates through Google Services, Google Cloud, and Other Bets segments. The Google Services segment provides products and services, such as ads, Android, Chrome, Google Maps, Google Play, Search, and YouTube; subscription-based products; and Fitbit wearable devices, Google Nest home products, Pixel phones, and other devices, as well as in-app purchases and digital content. The Google Cloud segment offers infrastructure, platform, and other services; Google Workspace that include cloud-based collaboration tools for enterprises, such as Gmail, Docs, Drive, Calendar and Meet; and other services for enterprise customers. The Other Bets segment sells health technology and internet services, as well as licensing and research and development services. The company was incorporated in 1998 and is headquartered in Mountain View, California.',
'city': 'Mountain View',
'phone': '650 253 0000',
'state': 'CA',
'country': 'United States',
'companyOfficers': [],
'website': 'https://www.abc.xyz',
'majority': 1,
```

Fig.2 Company Description

The data of yfinance which was used was accessible in csv format, which was then converted into a dataframe with the help of pandas library. The data was sorted and the null values were removed.

```
Out[31]: Open      0
High      0
Low       0
Close     0
Adj Close  0
Volume    0
dtype: int64
```

Fig.3 Data Preprocessing

Out[32]:	Open	High	Low	Close	Adj Close	Volume
count	2302.000000	2302.000000	2302.000000	2302.000000	2302.000000	2.302000e+03
mean	1092.339148	1102.388738	1081.780454	1092.414679	1092.414679	2.187185e+06
std	637.921285	644.124860	631.145349	637.715507	637.715507	1.392407e+06
min	350.350342	353.023010	348.108093	351.786774	351.786774	4.656000e+05
25%	590.047546	593.674973	585.560013	589.362000	589.362000	1.355050e+06
50%	956.234985	980.989990	947.385010	955.565002	955.565002	1.735050e+06
75%	1252.007507	1265.414978	1240.102509	1254.269958	1254.269958	2.517430e+06
max	3025.000000	3030.929932	2977.979980	2996.770020	2996.770020	2.310967e+07

Fig.4 Data Description

### B. Building and Training Model

For prediction of stock we selected characteristics such as RSI, Volume and the Closing price which will be used for selection of features.

The dataset has been divided into two modules training dataset and testing dataset. The model is trained with three different techniques that are Linear Regression, Simple Vector Regression and LSTM.

With the help of feature selection and combining the three algorithms the trend in the stock of a particular company will be predicted.

The accuracy we obtained by training the model are as following:

Linear Regression:

The accuracy of the model is 99.97912698861636

Fig.5 Accuracy of Linear Regression

SVR:

Train Accuracy= 53.165630299604295

Test Accuracy= 52.88135593220339

Fig.6 Accuracy of SVR model

LSTM:

Accuracy=	Date
2017-11-21	98.388214
2017-11-22	100.072749
2017-11-24	99.699376
2017-11-27	98.621119
2017-11-28	101.045454
...	
2019-12-20	99.890606
2019-12-23	99.422054
2019-12-24	99.901966
2019-12-26	98.072250
2019-12-27	100.399244
Length: 528, dtype: float64	

Fig.7 Accuracy of LSTM model

We trained our model further with sentiment analysis. The data from the twitter was obtained using snsrape module dating from 01-01-2019. Total 100 tweets were classified using sentiment analysis.

	tweet
0	'\$santosh41 &Tsla https://t.co/ye7ww9Yoyt'
1	'#Block (formerly #Square) and the #Bitcoin in...
2	'@darealbrianpark @SawyerMerritt @Tesla awesom...
3	'@Tesla may want to explore mining and product...
4	'\$HYMC Northern Nevada Lithium Project. Look h...

Fig.8 DataFrame of tweets

	tweet	probability	sentiment
0	'\$santosh41 & Tesla https://t.co/ye7ww9Yoyt'	0.965072	POSITIVE
1	'#Block (formerly #Square) and the #Bitcoin in...	0.850386	POSITIVE
2	'@darealbrianpark @SawyerMerritt @Tesla awesom...	0.978011	POSITIVE
3	'@Tesla may want to explore mining and product...	0.711631	NEGATIVE
4	'\$HYMC Northern Nevada Lithium Project. Look h...	0.956264	POSITIVE
...	...	...	...
95	'@NakedBakers @Tesla @elonmusk Nikola Tesla an...	0.962112	POSITIVE
96	'\$TSLA All that you need to know. Shorts haven...	0.998223	NEGATIVE
97	'A ciência	0.996266	POSITIVE
98	'@EverydayJohn_ @elonmusk @Tesla Elon be Spong...	0.914120	POSITIVE
99	'@MrJGBanks \$TSLA'	0.897744	NEGATIVE

100 rows x 3 columns

Fig.9 Probability of Sentiment Analysis

### C. Prediction

For prediction of the future stock price model is trained using hybrid algorithms. Predictions will be obtained by comparing previous stock prices with current stock prices of a particular company. The accuracy is obtained using accuracy\_score.

The prediction of stock trend i.e the stock value will be displayed in our model as follows:-

Prediction by LSTM:

```
# Predict the prices with the model
predicted_y = model.predict(X_test)
predicted_y = scaler.inverse_transform(predicted_y)
```

```
X_test.shape
```

```
(794, 36, 1)
```

```
predicted_y.shape
```

```
(794, 1)
```

```
predicted_y
```

Fig.10 Prediction by LSTM

Prediction by SVR:

```
svr_pred = model.predict(X_test).reshape(-1,1)
svr_pred = scaler.inverse_transform(svr_pred)
```

```
svr_pred.shape
```

```
(794, 1)
```

Fig.11 Prediction by SVR

### D. Visualization

The matplotlib library of python is used for visualization of stock graph.

The below graph (Fig.12) is obtained by comparing the predicted price with the actual price using Linear regression algorithm.

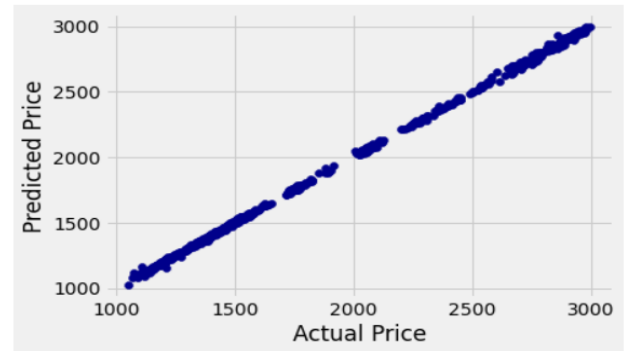


Fig.12 Graph Obtained using Linear Regression

The below graph (Fig.13) is obtained by comparing closing price with date using LSTM algorithm.

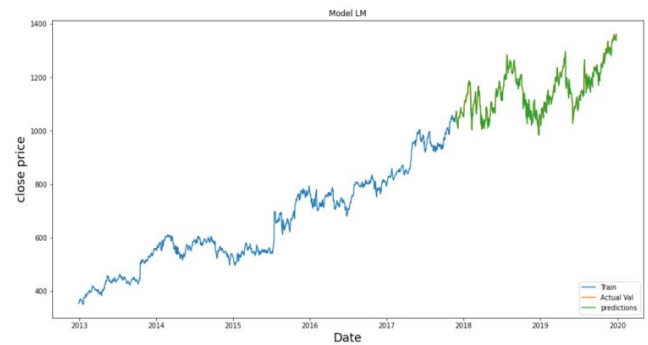


Fig.13 Graph Obtained using LSTM

The below graph (Fig.14) is obtained by predicted price with actual price using SVR algorithm.



Fig.14 Graph Obtained using SVR



The final output of our model is visualized as follows:

Prediction by LSTM in hybrid model:

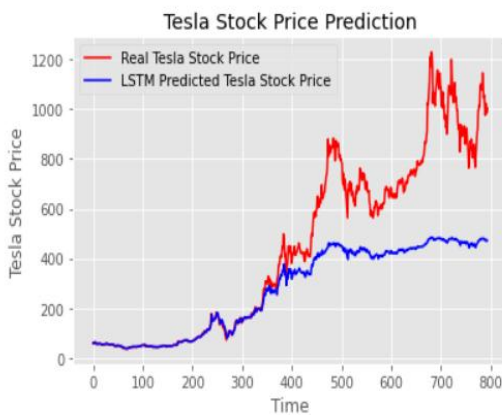


Fig.15 Prediction in hybrid model LSTM

Prediction by SVR in hybrid model:

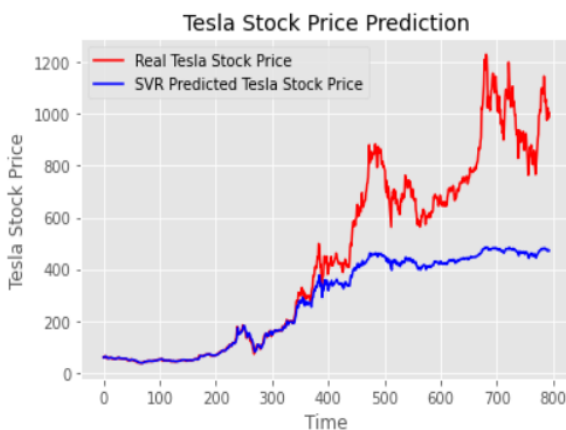


Fig.16 Prediction in hybrid model SVR

Prediction by hybrid model:-

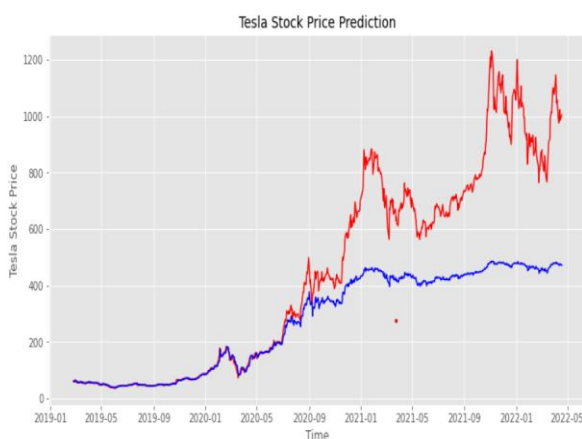


Fig.17 Prediction by hybrid model

## V. CONCLUSION

Through this research we can see that deep learning algorithms have great impact on the technologies for developing time series based prediction models. The accuracy level obtained by these algorithms are highest compared by any other different regression model for stock prediction. This model can help keen investors to obtain financial benefit while maintaining a viable environment in stock market. We faced difficulty in assembling sentiment analysis with our hybrid model and converting it into an API. In future we plan to overcome these difficulties by analyzing the data from different market of different categories by using and combining different algorithms.

## REFERENCES

- [1] Nikita, S., Abhijit, S. and Aayush, S., 2021. STOCK MARKET PREDICTION USING LSTM. [online] Irjet.net. Available at: <<https://www.irjet.net/archives/V6/i3/IRJET-V6I3219.pdf>> [Accessed 30 November 2021].
- [2] Mangalampalli, R., Khetre, P. and Malviya, V., 2021. STOCK PREDICTION USING HYBRID ARIMA AND GRU MODELS. [online] Ijert.org. Available at: <<https://www.ijert.org/research/stock-prediction-using-hybrid-arima-and-gru-models-IJERTV9IS050550.pdf>> [Accessed 30 November 2021].
- [3] Nikfarjam, A., Emadzadeh, E. and Muthaiyah, S., 2021. Text mining approaches for stock market prediction. [online] <https://www.academia.edu>. Available at: <[https://www.academia.edu/10400600/Text\\_mining\\_approaches\\_for\\_stock\\_market\\_prediction](https://www.academia.edu/10400600/Text_mining_approaches_for_stock_market_prediction)> [Accessed 30 November 2021].
- [4] Chonga, E., Hanb, C. and Park, F., 2021. Deep Learning Networks for Stock Market Analysis and Prediction: Methodology, Data Representations, and Case Studies. [online] durham-research-online. Available at: <<https://vdocument.in/durham-research-online-droduracukdroduracuk21533121533pdf-deep-learning.html>> [Accessed 30 November 2021].
- [5] arxiv.org. 2021. Sentiment Analysis of Twitter Data for predicting stock market moments. [online] Available at: <<https://arxiv.org/pdf/1610.09225.pdf>> [Accessed 30 November 2021].
- [6] www.ijert.org. 2021. Study of Machine learning Algorithms for Stock Market Prediction. [online] Available at: <<https://www.ijert.org/study-of-machine-learning-algorithms-for-stock-market-prediction>> [Accessed 30 November 2021].
- [7] Academia.edu. 2021. STOCK MARKET PREDICTION LITERATURE REVIEW AND ANALYSIS. [online] Available at: [Accessed 18 September 2021]