International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 11 Issue: 04 | Apr 2024 www.irjet.net p-ISSN: 2395-0072

Stock Market Prediction Using LSTM and Machine Learning Techniques

Dr. Chethan H K , Ms. Chaya Devi SN , Mr. Vinod Kumar C ,Mr. Vijay M

[1] *Professor, Dept. of Computer Science and Engineering, Maharaja Institute of Technology, Thandavapura*
[2,3,4, 5] *Students, Dept of Computer Science and Engineering, Maharaja Institute of Technology, Thandavapura*

Abstract - The financial market has seen significant involvement from the stock market. A minor commodity can experience varying effects as a result of stock market fluctuations. Investors drawn to the company by its market value or stock price for the company's expansion. Therefore, making an accurate stock price prediction might be quite beneficial. As a result, scholars have focused on this subject and have created numerous models and studies over the years. This research proposes a new framework for stock price prediction that combines many models (Long Short Term Memory (LSTM) model, Support Vector Regression (SVR) model, Linear Regression model, and Sentimental Analysis) into a hybrid model.

It is evident from the simulation results that our suggested strategy can predict future stock trends with this hybrid model when the hyper parameters are adjusted appropriately. Online datasets for stock markets with open, high, low, and closing prices are used to conduct the assessments.

Key Words: **Long Short Term Memory (LSTM), Support Vector Regression (SVR), Linear Regression, Sentimental Analysis.**

1. INTRODUCTION

These days, investing in stocks seems to be an additional source of revenue. While some invest as a primary source of income, others do it as a retirement plan. By retaining their shares locally under the company's name, they strengthen their bonds with their employees. Even though stocks are prone to volatility, discerning investors can make more informed decisions about which firm to invest their money in by visualizing share prices across a range of characteristics. With today's technology, one of the most prominent fields is data visualization. Since advances in machine learning, stock market analysis and forecasting have become a significant and developing trend. These analysis techniques are used by brokerage businesses, financial institutions, the banking industry, and other sectors in order to obtain familiarity with stock scores. Businesses employ this to shield themselves from the possibility that their investments will cause their share price to decline. Government agencies in both developed and emerging nations utilize stock analysis to boost their economy since it influences the pricing of other goods on the market. Giving the unstable enterprises security assurances has been a crucial component of the monitoring.

Assisting in the decision to purchase, sell, or hold has been the main goal of stock market forecasting. By utilizing the Streamline in this application we can, make dynamic graphs of a particular company's financial data. Furthermore, we can forecast future stock prices using a machine learning algorithm. Any corporation of one's choosing (with a recognized stock code) may use this online program.

1.1 Objective

Our stock market prediction web app aims to give consumers precise predictions and insights so they can make wise investing choices. Our platform uses real-time data analysis and sophisticated machine learning algorithms to forecast future stock movements with high accuracy. This allows users to take advantage of opportunities and reduce risk in the ever-changing financial markets. Furthermore, our intuitive interface and adaptable features guarantee smooth navigation and customized suggestions based on individual investing objectives and risk tolerances, enabling customers to successfully traverse the intricacies of the stock market.

2. PROPOSED SYSTEM

The data for our proposed method will come from yfinance, which is an internet source. Using the presented approaches, an assembling model (i.e., Linear, SVR, LSTM, and regression. The algorithms are selected based on how well they performed, as determined by the results of the literature review. Since stocks are thought to be a combination of linear and non-linearity, several techniques are applied to improve accuracy. Sentiment analysis has also been conducted in conjunction with this model to test polarity. In addition to being more beneficial for SVM, we can learn from both positive and negative tweets, which can assist us in obtaining professional guidance.

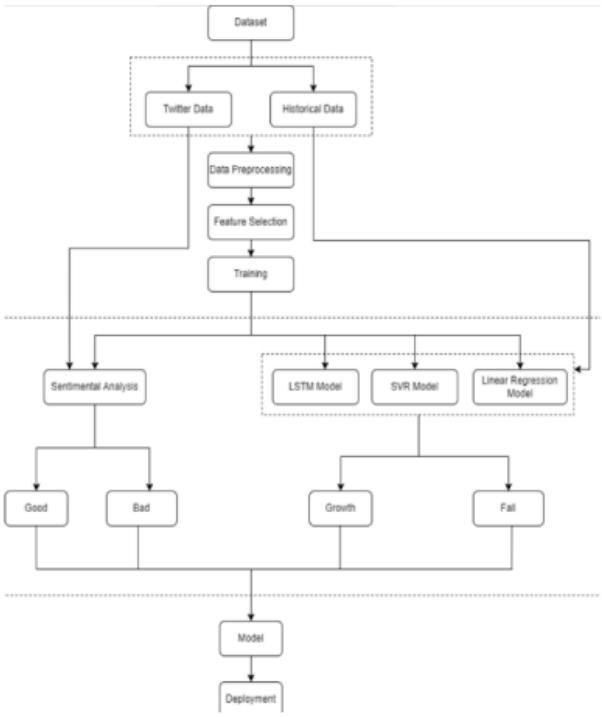

Fig 1: Proposed System

3. LITERATURE SURVEY

An LSTM (long short time memory) algorithm was introduced by Sonawane Nikita, Sonawane Abhijit, Sharma Aayush, Shinde Sagar, and Prof. A.S. Bodhe[1] in order to train a stock prediction model. Despite the large dataset, the paper produced an excellent result and suggested that stock advertisements be classified according to a specific company. The results obtained were superior to those of other algorithms, and the outcome can be deemed exceptionally promising as it has demonstrated the ability to predict well in comparison to other methods.

Several techniques were provided by Azadeh Nikfarjam, Ehsan Emadzadeh, and Saravanan Muthaiyah[2] to demonstrate the influence of financial news and tweets on stock prediction.

marketplace. They presented two methods: the first was based on news content, and the second was based on market data, with the classifier taking into account characteristics like stock price opening, closing, and change indicator values. The stock prices were noted during a window of time that included the newscast.

A research by Ashwini and Sakshi Pathak [3] concentrated on using different machine learning methods, including Support Vector Machine (SVM), KNN, and Logistic using the Random Forest technique and regression on the dataset.

Performance parameters such as recall rate, accuracy, precision, and fscore were computed. Finding the optimal algorithm for stock market prediction was their primary goal.

Using a hybrid model of ARIMA and GRU, Rahul Mangalampalli, Pawankumar Khetre, Vaibhav Malviya, and Vivek Pandey [4] introduced an alternative approach. The author offered two methods. While the second strategy was based 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, the first approach was based on the news content. We could get more accurate labels if we examined the market around the news's release.

4. IMPLEMENTATION

Implementation contains the following steps:

4.1 Data Pre-processing

We took the dataset from yfinance (Yahoo Finance) and utilized it for model testing and training. It had about 9 lakh records for that specific company. The variables that we took into account were the opening price, closing price, and pre-processing date.

Fig 2: Company Description

The yfinance data that was used was available in csv format, which the pandas software was used to transform into a data frame. After sorting the data, the null Values were eliminated.

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

Fig 3: Data Pre-processing

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	960.989990	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

4.2 Building and Training Model

We chose traits like RSI, volume, and closing price to be employed in the feature selection process in order to predict.

Two modules, a training dataset and a testing dataset, comprise the dataset. Three distinct methods are used to train the model: LSTM, Simple Vector Regression, linear regression. The trend in the stock of a specific firm will be forecasted with the aid of feature selection and the combination of the three algorithms.

The accuracy of the models are as follows:

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 using SVR

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 using LSTM

We utilized sentiment analysis to better train our model. The snscrape module was used to extract the data from Twitter, using a date of January 1, 2019. One hundred tweets in all were categorized.

	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:Data Frames of Tweets

	tweet	probability	sentiment
0	'$santosh41 &Tsla https://t.co/ye7ww9Yoyt'	0.965072	POSITIVE
1	'#Block (formerly #Square) and the #Bitcoin in...	0.850386	POSITIVE
2	'@darealbrianpark @SawyerMerritt @Tesla awesom...	0.978011	POSITIVE
3	'@Tesla may want to explore mining and product...	0.711631	NEGATIVE
4	'$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 × 3 columns

Fig 9: Probability of Sentimental Analysis

4.3 Prediction

Hybrid algorithms are used to train the model used to predict future stock prices. By contrasting current stock prices with past stock prices, predictions will be made. Accuracy score is used to determine accuracy.

In our model, the stock value, or the predicted stock trend, will be shown as follows:

Prediction using 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 using 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

4.4 Visualization

The stock graph is visualized using the Python matplotlib tool.The graph (Fig.12) below is produced by contrasting the connecting the expected price using the linear regression approach with the real price.

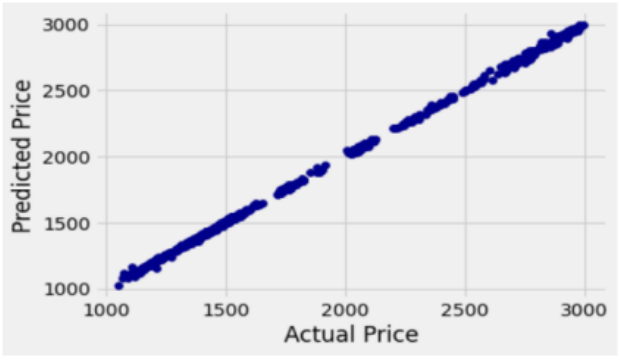

Fig 12:Graph Obtained using Linear Regression

The LSTM algorithm is used to compare the closing price with the date, resulting in the graph (Fig. 13) below.

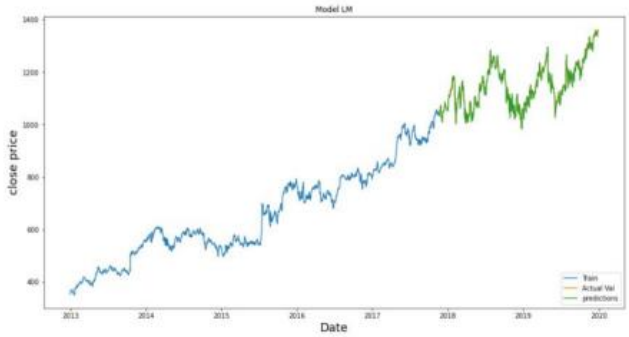

Fig 13:Graph Obtained using LSTM

The SVR method was used to generate the graph (Fig. 14) below, which compares the anticipated and real prices.

Fig 14: Graph Obtained using SVR

The following is a visualization of our model's final output: LSTM prediction in a hybrid model:

Fig 15: Prediction using hybrid model LSTM

SVR prediction in a hybrid model:

Fig 15: Prediction using hybrid model SVR

5. CONCLUSION

This study shows that deep learning algorithms have a significant influence on the technologies used to create prediction models based on time series data. When compared to other regression models, these methods yield the highest accuracy level for stock prediction. This methodology can assist astute investors in making money while preserving a healthy atmosphere in the stock market. Putting our hybrid model's sentiment analysis together and turning it into an API proved to be challenging. In the future, we intend to get over these challenges by employing and combining various algorithms to analyze data from various markets across various categories.

REFERENCES

[1] Nikita, S., Abhijit, S. and Aayush, S., 2021. STOCK MARKET PREDICTION USING LSTM. [online] Irjet.net. Available at: [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: [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: [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: [Accessed 30 November 2021].