Safe Trade – A Stock Recommender using Machine Learning Algorithms

Anitha Julian
Department of Computer Science and Engineering
Saveetha Engineering College
Chennai, India
cse.anithajulian@gmail.com

E GaneshReddy
Department of Computer Science and Engineering
Saveetha Engineering College
Chennai, India
edulaganeshreddy123@gmail.com

V R Dikshit Reddy
Department of Computer Science and Engineering
Saveetha Engineering College
Chennai, India
dikshitreddy17@gmail.com

D V V Somasekhar
Department of Computer Science and Engineering
Saveetha Engineering College
Chennai, India
somasekhardarisi7@gmail.com

2023 International Conference on Computer Communication and Informatics (ICCCI) | 979-8-3503-4821-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICCCI56745.2023.10128595

***Abstract*— The business and financial sector currently controls the majority of the global economy, and stock market trading is a key activity there. The project uses Google's Tensorflow's Keras API to create LSTM, training the algorithm on historical data to forecast stock prices. We focused on forecasting the trend in a stock's value over the subsequent 20 days after the prediction day. It is a method where one attempts to foresee the future worth of the company's present stocks to prevent loss or maybe make a profit. With the use of several quantities that will be discussed later in this report, this study will show how machine learning can be used to anticipate the same. To create somewhat accurate forecasts, we suggest a machine learning algorithm that will be taught using various datasets of certain firms that were available in the past. Shares market forecasting is a method for estimating the future value of a company's stock or another financial asset traded on an exchange. The only goal of stock market forecasting is a measurable substantial gain, along with, of course, avoiding large losses. Given that the efficient market hypothesis states that these forecasts cannot be produced with the data that is now available, some people may question the veracity of the findings of these predictions and conclude that they are fundamentally unexpected. However, a variety of instruments and technology make it possible to learn about future trends, which generates real income. The suggested approach detects frames with 85.5% accuracy.**

***Keywords* - LSTM, ANN, CNN, Stock Market**

I. INTRODUCTION

The business and finance sector is today the leader of the world's economy, and stock market trading [1] is a major practice in the finance sector. We propose a Machine Learning Algorithm which will be trained from different datasets of some companies available from the past to make near- effective predictions. Stock prices will be reliable indicators of true real enterprise value to the extent that this is true, to the extent that buyers and sellers are knowledgeable, and to the extent that when setting bid prices, they take into account all relevant factors, both current and future, relating not only to the specific enterprise and industry but also to general economic and political conditions. This procedure will increase the real value of better-run profitable companies. The allocation of new resources is significantly aided and guided by this in a private-ownership economy (or sector).

Practically all of the claims in the above paragraph must be qualified. Competition in the stock market isn't flawless. The highs and lows in this market are not smoothed out. Larger investors have greater influence.

II. LITERATURE SURVEY

Stock price prediction[5] is a popular yet challenging task and deep learning provides the means to conduct the mining for the different patterns that trigger its dynamic movement.

This project can predict the best stock to buy or sell.

It provides an efficient solution for easy investment in the stock and the most important thing: the audience will not miss any point of your presentation.

Machine learning is being used to aid in trading choices as financial organizations start to embrace artificial intelligence. Even though there is a plethora of stock data for machine learning models to train on, forecasting [4] the market is challenging for several reasons, including a high noise-to-signal ratio and the multiplicity of factors that influence stock prices. However, these models don't need to be highly accurate; even 60% accuracy might produce respectable

979-8-3503-4821-7/23/$31.00 ©2023 IEEE

profits. Using a long short-term memory neural network (LSTM)[2] for times series forecasting is one way to anticipate stock values.

A better variant of recurrent neural networks is called LSTMs (RNNs). RNNs are comparable to how people learn. Humans don't begin thinking from scratch every second when we think. Because we remember knowledge about previous words while reading sentences, for instance, we know that Bob is the one who plays basketball from the statement "Bob plays basketball." [5] Similar to this, RNNs are networked with loops in them so that the can utilize historical data before producing the intended result. RNNs can only link information that occurred recently; when the time gap widens, they are unable to do so. LSTM[11], a sort of RNN that remembers information over extended periods, comes into play in this situation, making them better suited for stock price prediction.

It has never been straightforward to invest in a portfolio of assets since complex models cannot accurately estimate future asset prices due to the irregularity of the financial market. The current prominent trend in scientific study is machine learning, which involves teaching computers to accomplish tasks that often require human intellect. To forecast [8] future stock market values, this article will use recurrent neural networks (RNN), particularly the Long-Short Term Memory model (LSTM). The major goal of this study is to determine how accurately a machine learning algorithm can forecast events and how many epochs can enhance our model.

III. Related Work

There are many different stock buyers and sellers on the stock market. The future scope of the market may be predicted using the stock market. It is important to construct a system that will operate with the highest degree of accuracy and that takes into account all significant variables that may have an impact on the outcome. There have previously been several stock market price prediction study studies. Over the business[7] and computer science areas, ranching is done sometimes Due to the numerous factors that might affect a share's profit or loss, the stock market performs well even when the economy is struggling. It is difficult to predict how a stock market will perform since many different elements are involved. Finding out how investors are feeling is the major goal. The research of national and worldwide events is generally laborious, and investors must be aware of the present price and obtain an extremely accurate prediction of the future price.

The greatest algorithm for forecasting future stock market performance is what we are trying to find. The institutions that support the stock market as well as investors will benefit greatly from the stock market's good forecast[9].

There are some mechanisms for stock price prediction that comes under technical analysis.

1. Statistical method

Before machine learning, statistical techniques were often employed. ARIMA, ESN, and Regression are the most often used approaches. Linearity and stationarity are the primary characteristics of the statistical technique. Linear discriminant analysis (LDA), regression algorithms, and quadratic discriminant analysis (QDA) are analyzed statistical techniques. It is done to analyze a frequently used method known as the ARIMA model. Auto-Regressive Moving Average is a method for using sinus time series as input variables (ARMA). Auto-Regressive-Essive Models (ARMA). In addition to being an extension of ARMA models, ARIMA may condense stationary and nonstationary series.

2. Pattern Recognition

This approach focuses on finding patterns. It thoroughly analyses the data and finds a pattern. In charts that display Open-High-Low-Close candlesticks, traders can locate buy and sell signals. Research is conducted on the trends in stock prices that may be used to forecast a stock's future. Analyzing patterns is done by looking at charts to provide forecasts for the stock market. To forecast future stock prices, chart patterns and market prices are compared.

3. Machine learning

Machine learning is used in many sectors. The stock market itself is one of the most well-liked methods. Algorithms for machine learning can be either supervised or unsupervised. In supervised learning, the object is used along with the learned machine learning data. Regression and classification are examples of supervised learning. It has a more tightly regulated environment. Unsupervised learning has a less regulated environment and unlabeled data. It analyses patterns and how they relate to clusters.

4. Sentiment analysis

Sentiment analysis is a method used to identify current trends. By examining social patterns like tweet activity, it tracks trends. Research is conducted on the use of segment signals from text to increase the effectiveness of models used to evaluate market trend.

Since many elements still need to be taken into consideration and the stock market prediction first doesn't seem statistical, it appears to be a hard problem. However, with the right machine learning techniques, it is possible to link old data to new data, teach the computer to learn from it and train it to make accurate assumptions. Although there are several models for machine learning as a whole, this paper concentrates on the two most crucial ones and uses them to make the prediction.

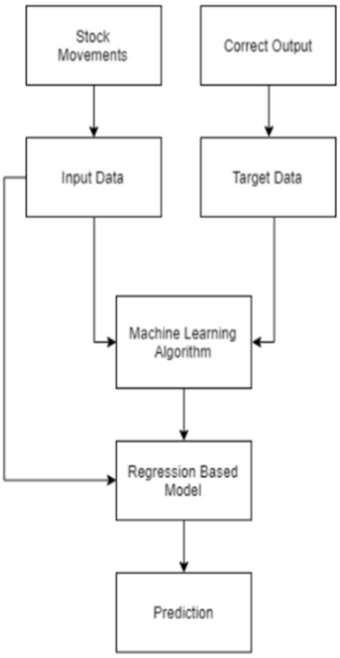

Fig. 1. Regression Model

IV. METHODOLOGY

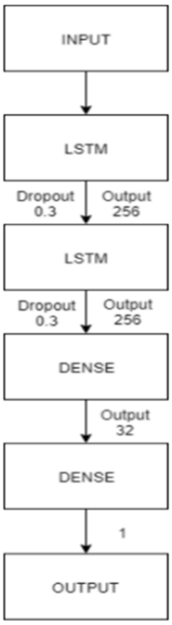

Fig.2. Long Short Term Memory(LSTM)

In the stock dataset prediction, two techniques LSTM and Regression have been applied. Positive outcomes have been achieved by both strategies due to an increase in forecast accuracy. The use of recently proposed machine learning approaches to the prediction of stochastic oscillations has shown encouraging results, marking its usage in successful exchange schemes. It has led to the conclusion that machine learning techniques can be used to anticipate the stock market more effectively and accurately.

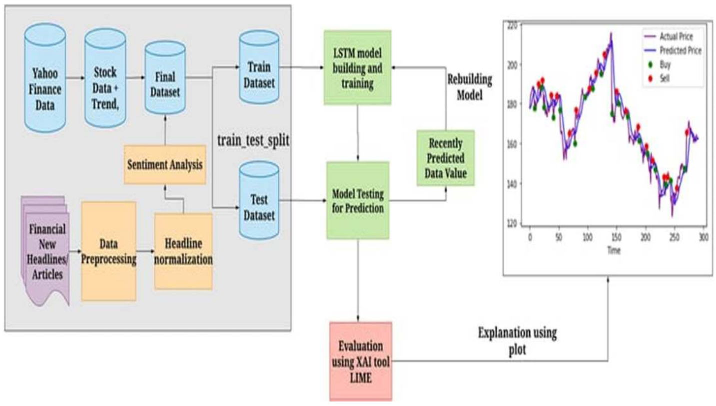

Fig. 3. System Architecture

5. Minmax Scaler

- For each feature X, we calculate the minimum value,(Xmin) and max value(Xmax).
- For each value in that feature X (Xi), calculate New $Xi = \frac{Xi - X_{\min}}{X\max - X_{min}}$
- Xmin=5
- Xmax=49

ORIGINAL	SCALED
5	0
10	0.113636
12	0.159091
14	0.204545
18	0.295455
23	0.409091
49	1

Table 1. Data for calculation

A. Existing System

Backpropagation Algorithm. [10] As a result, unnecessary data has been processed, resulting in time and memory waste. SVM and the Backpropagation Algorithm are less effective at predicting future stock prices since they analyze undesirable data. The present system's SVM and Backpropagation Algorithm are not very good at handling non-linear data. Therefore, LSTM (Long Short-Term Memory), which is more accurate for the next day than SVM and Backpropagation Algorithm, is used in our proposed future stock price prediction[7].

B. Proposed System

In the suggested approach, we aim to determine the closing value that will assist investors to buy or sell shares accurately the next day. Deep learning[8] uses an artificial neural network called Long Short-Term Table 2. Epochs For Google Dataset Using LSTM

Memory (LSTM). A memory cell in the advanced neural network LSTM saves a tiny quantity of data for later use. LSTM qualifies as a "general-purpose computer" thanks to its feedback linkages. A complete sequence of data, as opposed to just a single value, such as an image, may be processed using LSTM. The dropout mechanism used by the LSTM algorithm makes it significantly quicker than SVM and Backpropagation. Due to its ability to filter out unwanted data, the LSTM algorithm is more suited than the SVM and Backpropagation algorithms for forecasting future stock prices[3]. Due to the dropout process, less time and memory are used as compared to the exciting system. The LSTM algorithm handles non-linear data more effectively. We forecast the stock prices of 10 companies, tabulate them, and display them.

Deals with the classification of brain tumors accurately from the source, big data based on Hadoop framework with SVM classifier.

The proposed methodology consists of the following modules:

1. Elimination of Secondary attributes
2. Clustering and Classification

Table 3. Performance Measures

The performance of the algorithm has been evaluated using the metrics such as accuracy, MSE, and RSME

epochs	Accuracy	MSE	RMSE
10	92.5615	339.549	18.4269
20	94.2892	219.856	14.8276
30	94.6971	169.259	13.01
40	95.1746	141.106	11.8788
50	94.747	161.208	12.6968

	Date	Open	High	Low	Close	Volume
0	1/3/2012	325.25	332.83	324.97	663.59	7,380,500
1	1/4/2012	331.27	333.87	329.08	666.45	5,749,400
2	1/5/2012	329.83	330.75	326.89	657.21	6,590,300
3	1/6/2012	328.34	328.77	323.68	648.24	5,405,900
4	1/9/2012	322.04	322.29	309.46	620.76	11,688,800
...	...	...	...	...	...	...
1253	12/23/2016	790.90	792.74	787.28	789.91	623,400
1254	12/27/2016	790.68	797.86	787.66	791.55	789,100
1255	12/28/2016	793.70	794.23	783.20	785.05	1,153,800
1256	12/29/2016	783.33	785.93	778.92	782.79	744,300
1257	12/30/2016	782.75	782.78	770.41	771.82	1,770,000

1258 rows × 6 columns

Fig. 4.Google Stock Price Train Data

epochs	Accuracy
100	98.28213337528945
200	97.63336589796519
300	96.94409289369247
400	97.35454469043535

	Date	Open	High	Low	Close	Volume
0	1/3/2017	778.81	789.63	775.80	786.14	1,657,300
1	1/4/2017	788.36	791.34	783.16	786.90	1,073,000
2	1/5/2017	786.08	794.48	785.02	794.02	1,335,200
3	1/6/2017	795.26	807.90	792.20	806.15	1,640,200
4	1/9/2017	806.40	809.97	802.83	806.65	1,272,400
5	1/10/2017	807.86	809.13	803.51	804.79	1,176,800
6	1/11/2017	805.00	808.15	801.37	807.91	1,065,900
7	1/12/2017	807.14	807.39	799.17	806.36	1,353,100
8	1/13/2017	807.48	811.22	806.69	807.88	1,099,200
9	1/17/2017	807.08	807.14	800.37	804.61	1,362,100
10	1/18/2017	805.81	806.21	800.99	806.07	1,294,400
11	1/19/2017	805.12	809.48	801.80	802.17	919,300
12	1/20/2017	806.91	806.91	801.69	805.02	1,670,000
13	1/23/2017	807.25	820.87	803.74	819.31	1,963,600
14	1/24/2017	822.30	825.90	817.82	823.87	1,474,000
15	1/25/2017	829.62	835.77	825.06	835.67	1,494,500
16	1/26/2017	837.81	838.00	827.01	832.15	2,973,900
17	1/27/2017	834.71	841.95	820.44	823.31	2,965,800
18	1/30/2017	814.66	815.84	799.80	802.32	3,246,600
19	1/31/2017	796.86	801.25	790.52	796.79	2,160,600

Fig. 5. Google Stock Price Test Data

	Date	Open	High	Low	Close	Adj Close	Volume
0	2014-01-02	398.799988	399.359985	394.019989	397.970001	397.970001	2137800
1	2014-01-03	398.290009	402.709991	396.220001	396.440002	396.440002	2210200
2	2014-01-06	395.850006	397.000000	388.420013	393.630005	393.630005	3170600
3	2014-01-07	395.040009	398.470001	394.290009	398.029999	398.029999	1916000
4	2014-01-08	398.470001	403.000000	396.040009	401.920013	401.920013	2316500
...	...	...	...	...	...	...	...
1253	2018-12-24	1346.000000	1396.030029	1307.000000	1343.959961	1343.959961	7220000
1254	2018-12-26	1368.890015	1473.160034	1363.010010	1470.900024	1470.900024	10411800
1255	2018-12-27	1454.199951	1469.000000	1390.310059	1461.640015	1461.640015	9722000
1256	2018-12-28	1473.349976	1513.469971	1449.000000	1478.020020	1478.020020	8829000
1257	2018-12-31	1510.800049	1520.760010	1487.000000	1501.969971	1501.969971	6954500

1258 rows × 7 columns

Fig. 6. Amazon Train Data

	Date	Open	High	Low	Close	Adj Close	Volume
0	2019-01-02	1465.199951	1553.359985	1460.930054	1539.130005	1539.130005	7983100
1	2019-01-03	1520.010010	1538.000000	1497.109985	1500.280029	1500.280029	6975600
2	2019-01-04	1530.000000	1594.000000	1518.310059	1575.390015	1575.390015	9182600
3	2019-01-07	1602.310059	1634.560059	1589.189941	1629.510010	1629.510010	7993200
4	2019-01-08	1664.689941	1676.609985	1616.609985	1656.579956	1656.579956	8881400
5	2019-01-09	1652.979980	1667.800049	1641.400024	1659.420044	1659.420044	6348800
6	2019-01-10	1641.010010	1663.250000	1621.619995	1656.219971	1656.219971	6507700
7	2019-01-11	1640.550049	1660.290039	1636.219971	1640.560059	1640.560059	4686200
8	2019-01-14	1615.000000	1648.199951	1595.150024	1617.209961	1617.209961	6005900
9	2019-01-15	1632.000000	1675.160034	1626.010010	1674.560059	1674.560059	5998500
10	2019-01-16	1684.219971	1705.000000	1675.880005	1683.780029	1683.780029	6366900
11	2019-01-17	1680.000000	1700.170044	1677.500000	1693.219971	1693.219971	4208900
12	2019-01-18	1712.000000	1716.199951	1691.540039	1696.199951	1696.199951	6020500
13	2019-01-22	1681.000000	1681.869995	1610.199951	1632.170044	1632.170044	6416800
14	2019-01-23	1656.000000	1657.430054	1612.000000	1640.020020	1640.020020	5225200
15	2019-01-24	1641.069946	1657.260010	1631.780029	1654.930054	1654.930054	4089900
16	2019-01-25	1670.500000	1683.479980	1661.609985	1670.569946	1670.569946	4945900
17	2019-01-28	1643.589966	1645.000000	1614.089966	1637.890015	1637.890015	4837700
18	2019-01-29	1631.270020	1632.380005	1590.719971	1593.880005	1593.880005	4632800
19	2019-01-30	1623.000000	1676.949951	1619.680054	1670.430054	1670.430054	5783800
20	2019-01-31	1692.849976	1736.410034	1679.079956	1718.729980	1718.729980	10910300

Fig. 7.Amazon Test Data

Since the level of precision is compromised by omitting one dependence, the integration of the same is greater with other dependents. Since it is impossible to make an accurate prognosis for any fiscal day since the market is always shifting and turning the tables, the term "accuracy" is not used in the stock market. It becomes more possible and adaptable as component assets and dependencies are bigger, which makes it difficult to anticipate. The approximate value is taken into account while calculating the hit, profit, or gain rate for the same.

It is easier for users to see and understand the scenario and decide whether to invest and reap the benefits when several high-level machine learning algorithms are implemented and integrated into the project. The output is generated from the same and is made user-visible with the outputs in the form of a graph.

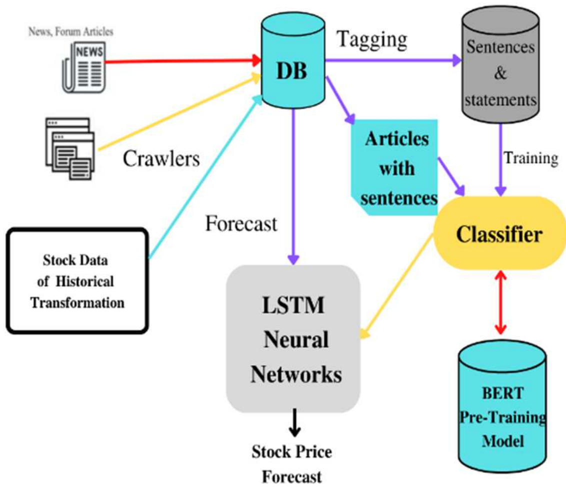

Fig. 8. Data Flow Diagram

V. MODULES IMPLEMENTATION

Fig. 9.Google Stock Prediction Graph

Stock Price	Time(in Days)
778	0
780	5
790	10
795	15
815	20

Table 4. Google Stock price data

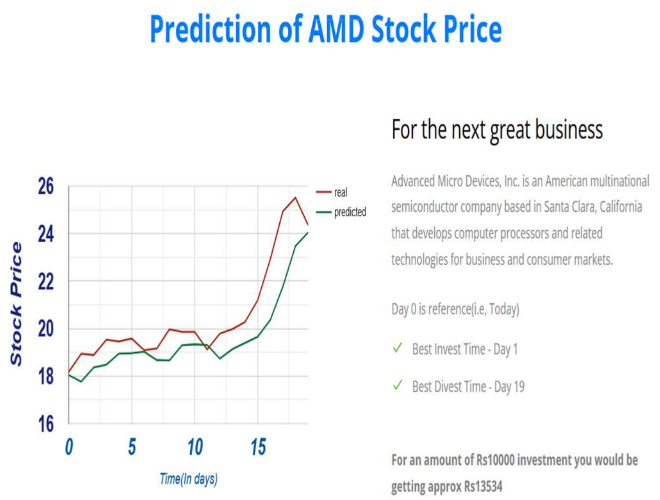

Fig. 10.Amazon Stock Prediction Graph

Stock Price	Time(in days)
18	0
19	5
20	10
21	15
22	15

Table 5. Amazon Stock data

These are graphs which are used to predict the accuracy of the google stock and amazon stock price data. And the accuracy of each algorithm will be compared and recommends the people to invest smartly in the stocks.

Numpy:

We can create and manage n-dimensional tensors with the aid of a fundamental Python 3 library for scientific computing. A matrix may be thought of as a 2-D tensor, a vector as a 1-D tensor, and so on.

Pandas:

It is used to manipulate and analyze data. It includes specific data structures and procedures for working with time series and mathematical tables. With the aid of Pandas, we can examine large data sets and draw conclusions based on statistical principles. Pandas can organize disorganized data sets, making them understandable and useful. In data science, relevant data is crucial.

Tensorflow:

A complete open-source machine learning platform is called TensorFlow. Researchers can advance the state-of-the-art in machine learning thanks to its extensive, adaptable ecosystem of tools, libraries, and community resources, while developers can simply create and deploy ML-powered apps.

To undertake machine learning and deep neural network research, the Google Brain team, a group of researchers and engineers inside Google's Machine Intelligence Research division, created TensorFlow. The system is broad enough to work in several different additional fields as well.

VI. CONCLUSION

To sum up, the stock is a system that follows the links in a chain, and its dependencies are also unexpected. It is described as a curve that is ever-changing and causes prices to fluctuate between low and high.

The accuracy is compromised by dependents since their level of integration with other dependencies is higher. Since it is impossible to make an accurate prognosis for any fiscal day since the market is always shifting and turning the tables, the term "accuracy" is not used in the stock market[6]. It becomes more possible and adaptable as component assets and dependencies are bigger, which makes it difficult to anticipate. The estimated values are considered while calculating the hit, profit, or gain rate for the same.

It is easier for users to see and understand the scenario and decide whether to invest and reap the benefits when several high-level machine learning algorithms are implemented and integrated into the project. The output is generated from the same and is made user-visible with the outputs in the form of a graph.

The suggested program does processing on the dataset's or.csv file's raw data set. Data is cleaned and purified before being processed further to produce useful results. Following the use of computational means, the result is shown on the screen as a graph.

Stocks are crucial to a company's success. They enable the corporation to quickly raise large sums of money, boost the company's reputation among the general public because anyone can now invest in it, and enable the initial investors to resell shares and profit from their initial investments. We offer a practical solution for simple stock market investing so that a layperson may also profit without a prior understanding of the technical aspects of the stock market. The website, SafeTrade, assists users in making wise investments by identifying and showing the number of different cases that fall under each class.

The best algorithm for predicting the market price of a stock based on various data points from historical data is Long Short-Term Memory (LSTM), which can handle a number of issues that earlier learning methods for recurrent neural networks could not. By comparing the accuracy of the various algorithms, we were able to determine which algorithm was most suitable. Nearly all sequence prediction issues can be solved most successfully with LSTMs. LSTMs are more efficient than RNNs and conventional feed-forward neural networks. They are able to recall certain patterns across extended periods of time, which accounts for this. The most effective algorithm for predicting the market price of a stock based on numerous data points from historical data is Long Short-Term, which we discovered by comparing the accuracy of the various algorithms.

REFERENCES

[1] Novel Algorithmic Trading Strategy Using Data-Driven Innovation_Stock Market Prediction Using Linear Regression and SVMPublisher: IEEE – 2020 Author: Bhawna Panwar, Gaurav.

[2] Deep Learning-Based Stock Price Prediction Using LSTM and Bi-Directional LSTM Model Publisher: IEEE – 2020 Author: Md. Arif Istiake Sunny, Mirza Mohd.

[3] Y. Guo, S. Han, C. Shen, Y. Li, X. Yin, and Y. Bai, ‘‘An adaptive SVR for high-frequency stock price forecasting,’’ *IEEE Access*, vol. 6, 11397 11404, 2018.

[4] Hadavandi, H. Shavandi, and A. Ghanbari, ‘‘Integration of genetic fuzzy systems and arti cial neural networks for stock price forecasting,’’ *Knowl-Based Syst.*, vol. 23, no. 8, pp. 800 808, 2010.

[5] M. Hagenau, M. Liebmann, and D. Neumann, ‘‘Automated news reading: Stock price prediction based on financial news using context-capturing features,’’ *Decis. Support Syst.*, vol. 55, no. 3, pp. 685 697, 2013.

[6] Hill, K. Cho, and A. Korhonen, ‘‘Learning zdistributed representations of sentences from unlabelled data,’’ in *Proc. NAACL-HLT*, 2016.

[7] Adrian Ţăran - Moroşan, “The relative strength index revisited”, African Journal of Business Management Vol.5(14), pp. 5855-5862, 18 July 2011.

[8] Yulius Hari and Lily Puspa Dewi, “Forecasting System Approach for Stock Trading with Relative Strength Index and Moving Average Indicator”, Journal of Telecommunication Electronic and Computer Engineering, Malaysia, Vol. 10 No. 2-3

[9] Pooja Talreja, “Forecasting Stock Market Trend Using Relative Strength Index”, IJLTEMAS, Volume III, Issue VI, June – 2014.

[10] Dr. Bhargavi. R, Dr. Srinivas Gumparthi, and Anith. R, “Relative Strength Index for Developing Effective Trading Strategies in Constructing Optimal Portfolio”, International Journal of Applied Engineering Research Volume 12, Number 19 (2017) pp. 8926-8936.

[11] Sepp Hochreiter and Jurgen Schmidhuber, “Long Short-Term Memory”, Neural Computation (1997), Volume – 9, Issue 2, pp. 1735-1780.