International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

A STOCK PRICE PREDICTION MODEL BASED ON INVESTOR SENTIMENT AND OPTIMIZED DEEP LEARNING

D V V DEEPTHI
Assistant Professor
Department of Computer Science and Engineering
TKRCET Autonomous ,
Hyderabad, India
dvvdeepthi@tkrcet.com

T.Sairam
Student,
Department of Computer Science and Engineering
TKRCET Autonomous,
Hyderabad, India
22k91a05p1@tkrcet.com

V. Kavya Sudha
Student,
Department of Computer Science and Engineering
TKRCET Autonomous ,
Hyderabad, India
22k91a05p9@tkrcet.com

S.Sathwika
Student,
Department of Computer Science and Engineering
TKRCET Autonomous ,
Hyderabad, India
shettysathwika6@gmail.com

U.Abhinaya
Student,
Department of Computer Science and Engineering
TKRCET Autonomous ,
Hyderabad, India
uppuabhinaya468@gmail.com

Abstract: The research presents a model, MS-SSA-LSTM that combines multi-source data, sentiment analysis, swarm intelligence algorithms, and deep learning models to enhance stock price predictions. This model combines sentiment analysis based on postings on the East Money forum, creating a unique sentiment lexicon and calculating a sentiment index. This gives major information about the effect of market sentiment on the price of stocks. Sparrow Search Algorithm (SSA) is used to optimize the hyperparameters of LSTM, and it improves the accuracy of prediction. The efficacy of the MS-SSA-LSTM model is outstanding as proven by experiments. It is a priceless tool in accurate stock price prediction. The model is tailored to the volatile financial market in China, and is focused on forecasting stock prices in the short term, which can be used to make fast decisions by investors. A stock sentiment classification hybrid LSTM and GRU model was created. An ensemble strategy was adopted, considering a Voting Classifier (AdaBoost + RandomForest) to perform sentiment analysis and Voting Regressor (LinearRegression + RandomForestRegressor + KNeighborsRegressor) to predict the price of stock. They were seamlessly integrated with the models already (MLP, CNN, LSTM, MS-LSTM, MS-SSA-LSTM) and enhanced the overall predictive capabilities. An easy to use Flask structure that supports SQLite was developed to increase user interaction and testing and make the signup, signin, and model evaluation process easier.

***"Index terms** - Deep learning, LSTM model, stock price prediction, sentiment analysis, sentiment dictionary, and sparrow search algorithm".*

1. INTRODUCTION

With the maturity of the stock market in China and the rapid growth of the Internet finance, most people find the importance of investment and they will participate in the financial market. However, the stock market is characterized by a lot of data and a lot of volatility. Many retail investors do not have adequate data-mining expertise to make money. Therefore, accurate prediction of stock prices can reduce the risks in investment and increase returns of investors and corporations.

The initial researchers used statistical methods to design a linear model, which conforms to the stock price time series pattern. The traditional methodologies include ARMA, ARIMA, GARCH, and others. The ARMA is designed to do a time series stock analysis [1]. The ARMA is used to obtain the ARIMA model, which predicts the fluctuations of stock prices. Wavelet analysis could be used in ARIMA model to improve the accuracy of fitting the Shanghai Composite Index [3]. GARCH model provides new ideas on how to predict a stock time series in a given time span [4]. At the same time, some researchers have combined ARMA and GARCH to create a new predictive model, which provided the theoretical framework of the volumetric price analysis of multivariate stocks [5]. These conventional methods normally only record common and structured data. However, the traditional forecasting methods require assumptions which are hardly found in the real world. As a result, it is tedious to describe nonlinear financial data using statistical tools.

Many scholars, therefore, attempt to forecast stock prices with the help of machine learning methods, such as Support Vector Machines (SVM) and Neural Networks. Machine learning is based on the principle that it uses algorithms to process data, extract insights, and predict the results of new data. The SVM shows clear strengths when working with limited samples, high dimensional data and non linear situations and many experts use SVM in predicting stocks. Hossain and Nasser [6] concluded that SVM method proves to be more effective than the statistical methods in predicting stocks. Chai et al. [7] have suggested a hybrid SVM model to forecast the variation of the HS300 index and concluded that the least squares SVM combined with the Genetic Algorithm (GA) was the most effective. Nevertheless, SVM consumes a lot of memory and computation time when trained on large-scale training samples and this may limit its ability to predict large volumes of stock data. Then, Artificial Neural Networks (ANN) and multi-layer ANN address issues of financial time series. Experimental evidence reveals that ANN has a high convergence rate and high accuracy [8], [9], [10]. Moghaddam and Esfandyari [11] evaluated the effects of some feedforward artificial neural networks to predict market stock prices through experimental means. Liu and Hou [12] improved the Back Propagation (BP) neural network by making use of Bayesian regularization. However, there are certain improvements that can be made with the traditional neural network method. It has a poor generalization ability and quickly overfits and gets stuck in local minima. Since multiple samples will need to be trained, it will be necessary to find better models to mitigate these problems.

The present study suggests a new stock price prediction model, called MS-SSA-LSTM, which combines the characteristics of multi-source data with LSTM neural networks and uses the Sparrow Search Algorithm. The MS-SSA-LSTM stock price prediction model can anticipate stock prices, aiding investors and traders in making better informed investment choices. It is the investors and traders who obtain data about particular equities such as past transaction data and comments of shareholders in the stock market and input this information into the MS-SSA-LSTM model. The model automatically creates a stock price trend chart and forecasts the stock price in the next day.

2. LITERATURE SURVEY

The presence and modifications of long memory properties in return and volatility dynamics of S&P 500 and London Stock Exchange, done through the ARMA model [1]. In recent years, multifractal analysis has been recognized as an important technique to explain the intricacy of financial markets, which is challenging to describe in the linear frameworks of efficient market theory. The weak form of the efficient market theory in financial markets is that the returns of prices are uncorrelated serially. Prices must display a random walk, in other words. The hypothesis of random walk is tested against the alternatives that include either the unifractality or the multifractality. Most research shows that stock returns volatility exhibits long-range dependency, heavy tails and clustering. Because of these long-range dependence properties and large tails, the use of self-similar stochastic processes has been suggested to capture these properties in self-similar models of return volatility. This paper uses monthly and annual prediction of Time Series Stock Returns of London Stock Exchange and S&P 500 using ARMA model. The information is up-to-date to October 2023. The statistical analysis of S and P 500 shows that ARMA model outperforms London Stock Exchange and is highly skilled in predicting medium to long term horizons using the real known values. The statistical analysis of the London Stock Exchange reveals that ARMA model of monthly stock returns is better than the annual model. An evaluation of the S&P 500 and the London Stock Exchange shows that the two markets are efficient and that they have financial stability during economic boom and recession periods.

The article gives an analysis on how the ARIMA time series model was utilized to forecast the future gold prices in India using the historical data collected between November 2003 and January 2014 to minimize buying risks involved with gold. Thus, in order to give investors information on the best time to buy or sell gold. 2This financial tool has been on the rise in the recent past because of the limitations experienced by the Indian economy such as changing political environment, its global signals and high inflation rates. Researchers, investors and speculators are therefore looking to find alternative financial instruments to fall back on in order to diversify their portfolios to reduce risk. In the past, gold could only be obtained during marriages or rites in India but currently, it has acquired importance among investors thus the need to predict the gold prices using proper methodologies.

GARCH model and its many variations have been widely applied in the practice and literature of finance. Quasi maximum likelihood estimation assumes that the innovations in GARCH processes are identically and independently distributed, have a mean of zero, and a variance of one (strong GARCH) [4]. With weaker conditions (the absence of unconditional correlation, weak GARCH), higher order dependency structures can be used to predict the GARCH innovations (and stock returns) ex-ante, and thus, ex-ante. This study uses rolling windows of stock returns to test the autonomy of consecutive GARCH

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

innovations. The results of rolling values of independence testing show the changes of the serial dependency over time and provide good information on the prediction of the one-step-ahead changes of the stock prices. The benefits of ex ante forecasting are documented of nonparametric innovation forecasts, especially when the sign of the innovation predictors is combined with independence tests (-values) and/or the sign of linear predictions of returns.

Computational intelligence methods (Support Vector Machine (SVM) and Relevance Vector Machine (RVM)) have been successfully used in financial forecasting using GARCH-type models, especially ARMA-GARCH, in the last few years. [2, 6]This study analyzes the application of ARMA-GARCH, recurrent SVM (RSVM) and recurrent RVM (RRVM) to volatility prediction. The two GARCH procedures (based on RSVM and RRVM) are used and compared to the parametric GARCH models (Pure and ARMA-GARCH) regarding their effectiveness in multi-period forecasting. There are four performance measures that are used in evaluating the models and they include Mean Squared Error (MSE), Mean Absolute Error (MAE), Deterministic Score (DS), and the coefficient of determination (R squared) of linear regression. The paper employs two composite indices of Asian stock markets; BSE SENSEX and NIKKEI225. This study also examines how the presence of outliers affects the modeling and prediction of volatility. As our experiment shows, the RSVM and RRVM have similar performance and better than GARCH-type models in predicting accuracy. The ARMA-GARCH model is superior to the pure GARCH and the RRVM with RSVM is the only one that has the strength of forecasting.

This paper aims at providing an EMD-LSSVM (empirical mode decomposition least squares support vector machine) model in analyzing the CSI 300 index. A WD-LSSVM (wavelet denoising least squares support machine) is provided to provide a reference point to evaluate the performance of EMD-LSSVM [7]. Since choosing the parameters is one of the most important tasks in models, various optimization methods are utilized, such as simplex, grid search (GS), particle swarm optimization (PSO), and genetic algorithm (GA). According to experimental results, the EMD-LSSVM model based on the use of GS algorithm is superior to other approaches in predicting the direction of stock market movements.

3. METHODOLOGY

i) Proposed Work:

The project introduces the MS-SSA-LSTM model that is the sophisticated method of stock price prediction. This model combines various sources of data, sentiment analysis, and swarm algorithms (intelligence) in a perfect blend. 14, 15, 16, 30 The system shows impressive accuracy in the stock price prediction after optimization of LSTM hyperparameters with the help of the Sparrow Search Algorithm. Its superiority over other models is proven by experimental results, which reveal the universal applicability and the possibility of enhancing predictive performance. This model is compared with MLP, CNN, LSTM, and MS-LSTM. The hybrid LSTM and GRU stock sentiment classification model was created. The overall ensemble methodology was used, and the sentiment analysis was performed with a Voting Classifier (AdaBoost + RandomForest), whereas the stock price prediction was done with a Voting Regressor (Linear Regression + Random Forest Regressor + K-Neighbors Regressor). These models were seamlessly combined with existing models (MLP, CNN, LSTM, MS-LSTM, MS-SSA-LSTM), thus improving overall predictive performance. An easy to use Flask server was developed to facilitate user interaction and testing making the sign-up, sign-in and model evaluation easier.

ii) System Architecture:

The initial step is importing datasets, which include Stock Tweets Dataset, Single Stock Data and Multi-Source Data. Sentiment analysis and stock price forecasting are based on these datasets. The Stock Tweets Dataset is undergoing a cleaning process which involves the removal of punctuation marks, HTML tags, URLs and emoticons. This step will prepare the text to be analyzed in terms of sentiments. Single Stock Data and Multi-source Data are handled to overcome the null values, remove duplicates and normalize the data. This standardizes the financial data to make predictions about stock prices. Sentiment classification is performed using multiple models, including MLP, CNN, LSTM, MS-LSTM, MS-SSA-LSTM, and extensions to them, like Voting Classifier and LSTM + GRU. They evaluate purified twitter data to obtain market sentiment. Another set of models is used to forecast the price of stocks, consisting of MLP, CNN, LSTM, MS-LSTM, MS-SSA-LSTM, and extended Voting Regression. They use processed financial data to make predictions about stock prices. After the training process is completed, there is generation of predictions using the models. Forecasts provide an insight into market sentiment in sentiment research. Models anticipate future stock prices for stock price forecasting. Sentiment analysis and prediction of stock prices via stock price models are necessary to help investors and traders make informed decisions. The combined results help consumers navigate complex landscape of the stock market, reduce the risks and maximize earnings on the investments.

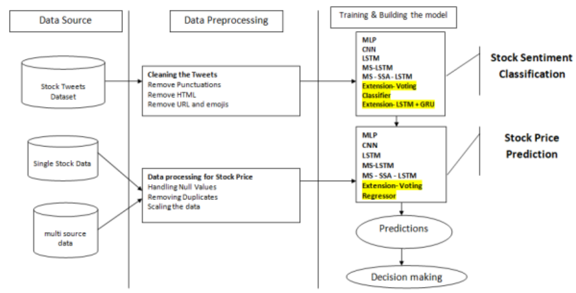


Fig 1 Proposed architecture

iii) Dataset collection:

STOCK TWEETS DATASET

The Stock Tweets dataset consists of social media updates related to stocks and financial markets. We used it to understand the feelings and reactions of people to the news in the market [1, 4, 7, and 8]. This helped in coming up with stock trading and investment tools. We desire to investigate how social media affects stock price and market movements to benefit investors and traders.

These are the top five rows of the dataset.

	Text	Sentiment
0	Kickers on my watchlist XIDE TIT SOQ PNK CPW B...	1
1	user: AAP MOVIE. 55% return for the FEA/GEED i...	1
2	user I'd be afraid to short AMZN - they are lo...	1
3	MNTA Over 12.00	1
4	OI Over 21.37	1

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Fig 2 Stock tweets dataset

ALL STOCK DATASET

The All Stock Dataset is a large compilation of financial information which includes several sources. It provides much information on an in-depth analysis of stock markets. In our project, we used this data to enhance our model of stock price prediction. Our aim was to improve the accuracy of stock price forecasting using different data sources, which would be of benefit to investors and enterprises.

THIS IS THE SAMPLE DATASET

Date	Open	High	Low	Close	Volume
2012-01-03	325.25	332.83	324.97	663.59	7,380,500
2012-01-04	331.27	333.87	329.08	666.45	5,749,400
2012-01-05	329.83	330.75	326.89	657.21	6,590,300
2012-01-06	328.34	328.77	323.68	648.24	5,405,900
2012-01-09	322.04	322.29	309.46	620.76	11,688,800

Fig 3 All stock dataset

iv) Data Processing:

Data processing is transforming raw data into insights that are useful to businesses. Data scientists are usually involved in data processing, which includes data collection, data organization, data cleansing, data validation, data analysis, and data transformation to easily understandable formats, e.g., graphs or papers. There are three ways to process data, namely, manual, mechanical, and electronic. The aim is to add value to information and optimize decision making. This enables businesses to optimize their operations and make quick strategic choices. This is where automated technologies of data processing such as computer software programming come in. It has the potential to convert large amounts of information and big data in particular to meaningful information that can be utilized in quality management and decision-making.

v) Feature selection:

The process of selecting the most consistent, non-redundant and relevant features to develop a model is known as feature selection. It is important to gradually reduce the size of datasets as the size and variety of datasets continue to grow. The main goal of feature selection is to increase the effectiveness of a predictive model and reduce the cost of computation of a model.

One of the core features of feature engineering is feature selection where one or two of the most important features are chosen to feed into machine learning algorithms. The feature selection methods are applied in order to reduce the size of the input variables by eliminating the redundant or irrelevant features, thus narrowing the list to the most relevant features to the machine learning model. The main benefits of performing feature selection prior to the machine learning model as opposed to letting the machine learning model decide on the importance of features on its own.

vi) Algorithms:

The Multilayer Perceptron (MLP) operates by processing the data on several levels. It begins with an input layer which receives data and then there are hidden layers where each neuron is the weighted sum of its inputs, the result of which is subjected to a non-linear activation function and passed on to the next layer. The weights among neurons are modified during training to enhance the network's capacity to discern intricate patterns in data. Terminal output layer generates forecasts or classifications. MLPs find application in a variety of tasks such as image recognition and financial prediction because they can model complex relationships in data.

A Convolutional Neural Network (CNN) is a deep learning algorithm that can be applied to various types of data other than image data. It processes data through layers which contain convolutions and pooling operations, thus enabling the network to automatically identify relevant patterns or features in the data. This makes CNNs beneficial to tasks that require sequential data or grids, like time series analysis or structured data processing. They are outstanding in drawing intricate relationships and hierarchies, making them better suited in different fields, such as natural language processing and financial forecasting.

Long Short-Term Memory (LSTM) is a type of recurrent neural networks (RNNs) that is designed to process sequential data. Unlike traditional RNNs, LSTMs are efficient at identifying and preserving relationships over long sequences, making them well-suited to scenarios that require complex, distant relationships between data points. LSTMs make use of particular memory cells and gates which enable them to store, update or delete information, therefore they can correctly model sequential patterns. This has found application in various fields, such as in natural language processing, speech recognition and financial time series analysis, where understanding the past and predicting the future trends are critical.

Multi-Source Long Short-Term Memory (MS-LSTM) is an improved variant of the traditional LSTM neural network, designed to simultaneously process an enormous number of sources. It is good at handling large amounts of information by consolidating data on multiple sources and, therefore, it is highly beneficial in complex tasks such as price prediction of stocks. MS-LSTM enriches the capacity of the model to identify and analyze intricate relationships and patterns using a vast quantity of data, therefore, improving the predictive power of the system in the situations where diverse sources of data are crucial.

An example of a sophisticated model of stock price prediction is the MS-SSA-LSTM model, or Multi-Source Sparrow Search Algorithm Long Short-Term Memory. It combines multi-source information of numerous sources, sentiment analysis, and optimizes the Long Short-Term Memory (LSTM) network with the Sparrow Search Algorithm (SSA). This advanced model is a skillful approach to addressing the complexities of financial forecasting because it offers a more accurate and robust way of forecasting the value of stocks. It outsmarts conventional models and has a wide-ranging universal applicability making it a priceless tool to investors and businesses dealing with volatile financial markets.

The Voting Regressor is an ensemble machine learning approach, which combines the predictions of different regression algorithms to improve the accuracy of forecasting. There are three different regressors in this example: Linear Regression, Random Forest Regressor, and k-Neighbors Regressor. It aims to create a more accurate and robust model of regression problems by combining their individual predictions. The methodology takes advantage of the strengths of each underlying regressor, such as the linearity of Linear Regression, the flexibility of Random Forest, and the proximity-

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

based learning of k-Neighbors Regression, to enhance the overall predictive performance.

LSTM+GRU is an advanced type of recurrent neural network (RNN) architecture that combines the capabilities of the Long Short-Term Memory (LSTM) and the Gated Recurrent Unit (GRU) cells. It enhances the model to detect sequential patterns on the data with the combination of memory storage properties of LSTM and the computational efficiency of GRU. This amalgamation is especially efficacious for tasks related to time series data, natural language processing, and sequential pattern recognition, as it mitigates the constraints of each cell type independently, yielding enhanced performance and training efficiency.

In this study, the Voting Classifier is a vital component of the sentiment classification that combines the strengths of AdaBoost and Random Forest (RF) [18, 39]. It makes use of the boosting mechanisms of AdaBoost, whereby a large number of weak learners are pooled to form a strong classifier, as well as the ensemble learning approach of RF, which pools predictions of multiple decision trees. By combining both of these algorithms, the Voting Classifier enhances the accuracy and stability of sentiment classification and makes it a powerful tool to evaluate market sentiment in our study.

4. EXPERIMENTAL RESULTS

Accuracy: The test accuracy refers to the ability of the test to correctly differentiate between the patient and the healthy cases. The ratio of the quantity of true positives and the quantity of true negatives with all the cases that have been assessed should be computed in order to determine the accuracy of a test. Mathematically this is:

$$Accuracy = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} (1)$$

Precision: Precision is a measurement of the percentage of correctly identified cases out of those that were defined as positive. As a result, the precision formula can be defined as:

$$Precision = \frac{\text{True Positive}}{\text{True Positive } + \text{False Positive}} (2)$$

Recall: Recall is a machine learning metric which evaluates the ability of a model to detect every relevant instance of a given class. It refers to the proportion of the correctly identified positive observations to the actual positives, provides information on the effectiveness of a model in detecting the instances of a particular classification.

$$Recall = \frac{\text{TP}}{\text{TP } + \text{ FN}} (3)$$

F1-Score: F1 score is a performance measure of a machine learning model. It combines precision and recall rates of a model. The accuracy measure goes to determine how often a model makes accurate predictions on the entire dataset.

$$F1\ Score = 2 * \frac{Recall \text{ X } Precision}{Recall + Precision} * 100 (1)$$

Table (1) evaluates the performance measures of the methods Accuracy, Precision, Recall, and F1 Score. LSTM-GRU consistently outperform all other algorithms in all criteria. The tables offer a

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

comparative study of the measures of the alternative methods.

Table (2) evaluates the performance metrics R2 Score, MSE, RMSE and MAE of each algorithm. The MS-LSTM uniformly outperforms all other algorithms by all measures. The tables give a comparative analysis of the measures of the alternative methods.

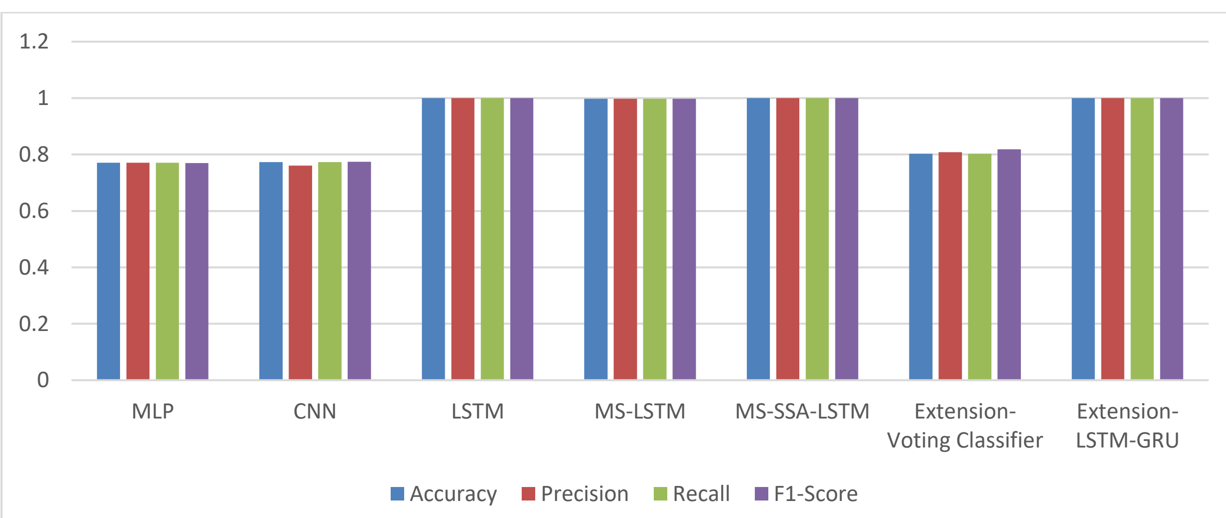
Table.1 Performance Evaluation Table - Stock sentiment classification

ML Model	***Accuracy***	***Precision***	***Recall***	***F1-Score***
MLP	0.771	0.771	0.771	0.770
CNN	0.773	0.761	0.773	0.774
LSTM	1.000	1.000	1.000	1.000
MS-LSTM	0.998	0.998	0.998	0.998
MS-SSA-LSTM	1.000	1.000	1.000	1.000
Extension- Voting Classifier	0.803	0.808	0.803	0.819
Extension- LSTM-GRU	**1.000**	**1.000**	**1.000**	**1.000**

Table.2 Performance Evaluation Table - Stock price prediction

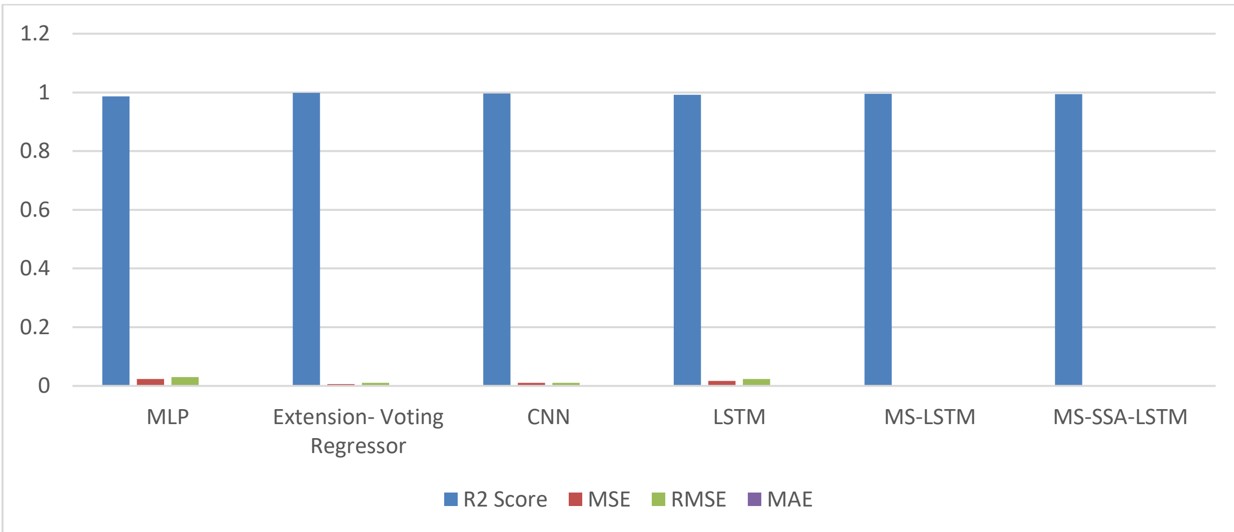
ML Model	**R2 Score**	**MSE**	**RMSE**	**MAE**
MLP	0.987	0.023	0.030	0.001
Extension- Voting Regressor	0.999	0.006	0.010	0.000
CNN	0.996	0.011	0.010	0.000
LSTM	0.992	0.017	0.024	0.001
MS-LSTM	**0.995**	**0.001**	**0.001**	**0.000**
MS-SSA-LSTM	0.994	0.001	0.001	0.000

Graph.1 Comparison Graph -Stock sentiment classification



International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Graph.2 Comparison Graph - Stock price prediction



The accuracy, precision, recall, and F1-Score are represented in blue, red, green, and purple in Graph (1). The LSTM-GRU has better performance with all criteria, and the highest values are achieved in comparison with the other models. These findings are graphically illustrated in the graphs above.

Graph (2) shows the R2 score in blue, MSE in red, RMSE in green and MAE in purple. The MS-LSTM shows a better performance in all measures, reaching the highest values, compared to the other models. These findings are represented graphically in the graphs above.

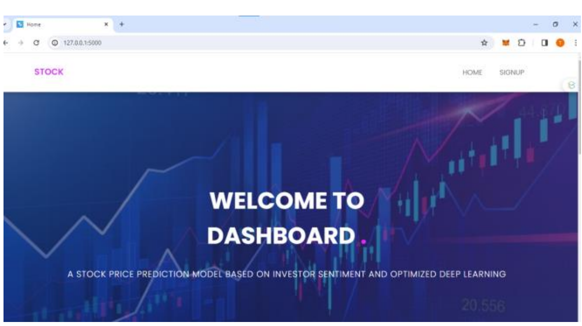


Fig 4 Home page

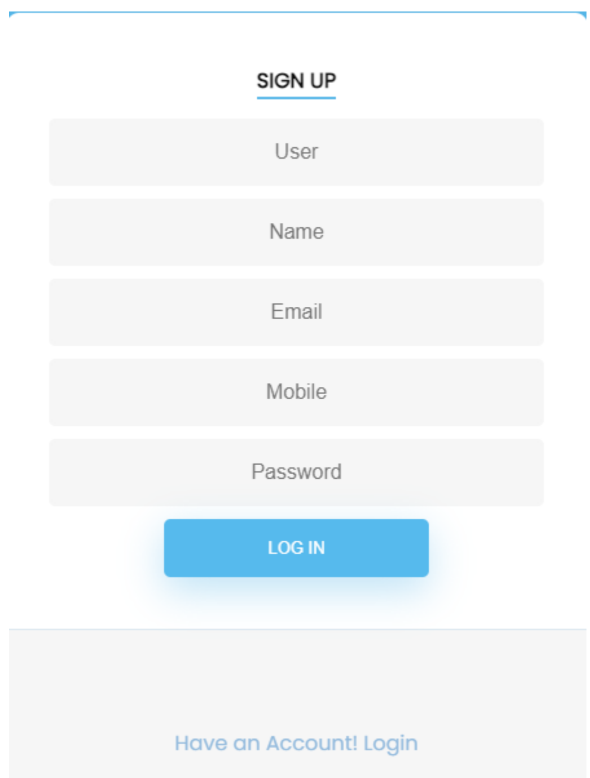


Fig 5 Signin page

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

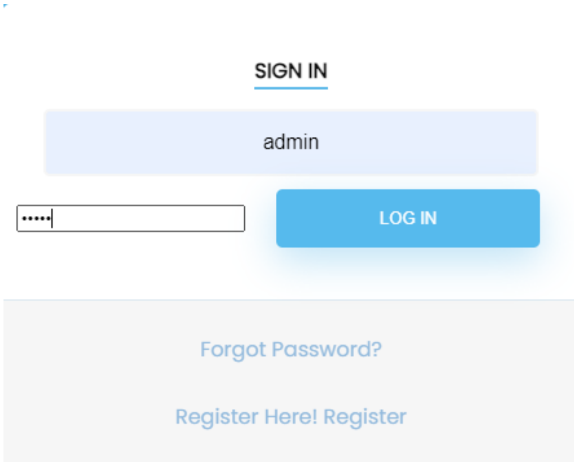


Fig 6 Login page

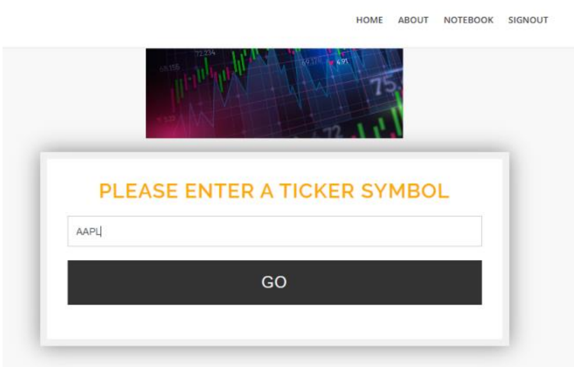


Fig 7 User input

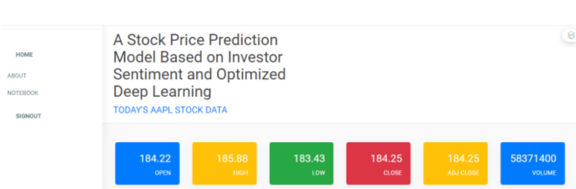


Fig 8 Result

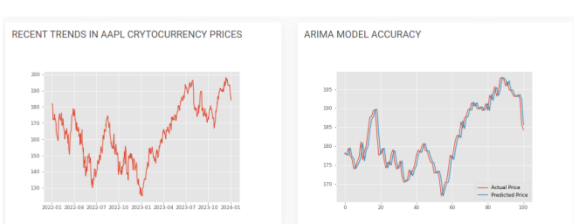


Fig 9 Graphs

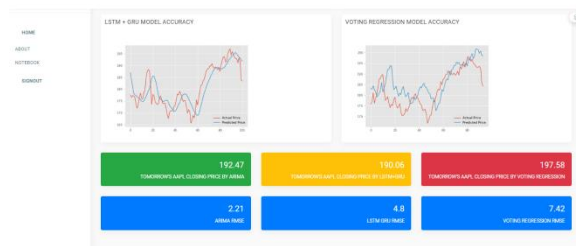


Fig 10 Graphs

5. CONCLUSION

The task aimed to enhance stock market predictions, and the focus was on the MS-SSA-LSTM model. The paper has reviewed several models, which emphasize the significance of sentiment analysis and sophisticated algorithms to make more accurate predictions [26]. MS-SSA-LSTM model was found to be very good in stock price prediction and sentiment classification. It made use of diverse sources of data and advanced methodologies to offer an integrated approach to risk reduction and improved returns. Existing models (MLP, CNN, LSTM, and MS-LSTM) were proficient, but the MS-SSA-LSTM model demonstrated superiority, particularly in short-term predictions of the volatile China market. The predictive toolset was improved with the introduction of ensemble models (Voting Classifier, LSTM+GRU, Voting Regressor) during the extension phase. Sentiment categorization was best performed by LSTM and GRU, whereas Voting Regressor was the most reliable in predicting a stock price. The Flask addon allowed the use of interactivity in an intuitive manner such as typing in ticker symbols to make specific predictions. Sentiment analysis LSTM and GRU models were efficiently deployed, as well as a

Voting Regressor predicting the stock price, making it easier to access by users and investors. The sophisticated prediction models and user-friendly interface of the project will benefit investors, traders, and enterprises. MS-SSA-LSTM model and its developments offer tremendous information, reduce risk of investment and enhance decision making in the unstable Chinese financial market.

6. FUTURE SCOPE

Increasing the usefulness of the model to handle real-time data flows can enable investors to make more timely decisions. This may be improved by adding sources of data which provide real-time information. [34] The better understanding of the market sentiment can be achieved by improving the sentiment analysis part by using natural language processing (NLP) and sentiment-specific machine learning models. Exploring and combining the information of multiple sources, such as social media, news feeds, and macroeconomic aspects, may give a comprehensive view of the market and potentially improve the accuracy of predictions. This could make the model more transparent and easier to use by creating tools or features that will explain the predictions of the model. Investors may gain advantages by comprehending the rationale underlying particular forecasts. It can be beneficial to improve the model capabilities with the risk evaluation and portfolio optimization to provide investors with an overall approach to the management of their investments. This can involve looking at the asset diversification and risk-adjusted returns.

REFERENCES

[1] M. M. Rounaghi and F. N. Zadeh, ‘‘Investigation of market efficiency and financial stability between S&P 500 and London stock exchange: Monthly and yearly forecasting of time series stock returns using ARMA model,’’ Phys. A, Stat. Mech. Appl., vol. 456, pp. 10–21, Aug. 2016, doi: 10.1016/j.physa.2016.03.006.

[2] G. Bandyopadhyay, ‘‘Gold price forecasting using ARIMA model,’’ J. Adv. Manage. Sci., vol. 4, no. 2, pp. 117–121, 2016, doi: 10.12720/joams.4.2.117-121.

[3] H. Shi, Z. You, and Z. Chen, ‘‘Analysis and prediction of Shanghai composite index by ARIMA model based on wavelet analysis,’’ J. Math. Pract. Theory, vol. 44, no. 23, pp. 66–72, 2014.

[4] H. Herwartz, ‘‘Stock return prediction under GARCH—An empirical assessment,’’ Int. J. Forecasting, vol. 33, no. 3, pp. 569–580, Jul. 2017, doi: 10.1016/j.ijforecast.2017.01.002.

[5] H. Mohammadi and L. Su, ‘‘International evidence on crude oil price dynamics: Applications of ARIMA-GARCH models,’’ Energy Econ., vol. 32, no. 5, pp. 1001–1008, Sep. 2010, doi: 10.1016/j.eneco.2010.04.009.

[6] A. Hossain and M. Nasser, ‘‘Recurrent support and relevance vector machines based model with application to forecasting volatility of financial returns,’’ J. Intell. Learn. Syst. Appl., vol. 3, no. 4, pp. 230–241, 2011, doi: 10.4236/jilsa.2011.34026.

[7] J. Chai, J. Du, K. K. Lai, and Y. P. Lee, ‘‘A hybrid least square support vector machine model with parameters optimization for stock forecasting,’’ Math.

Problems Eng., vol. 2015, pp. 1–7, Jan. 2015, doi: 10.1155/2015/231394.

[8] A. Murkute and T. Sarode, ''Forecasting market price of stock using artificial neural network,'' Int. J. Comput. Appl., vol. 124, no. 12, pp. 11–15, Aug. 2015, doi: 10.5120/ijca2015905681.

[9] D. Banjade, ''Forecasting Bitcoin price using artificial neural network,'' Jan. 2020, doi: 10.2139/ssrn.3515702.

[10] J. Zahedi and M. M. Rounaghi, ''Application of artificial neural network models and principal component analysis method in predicting stock prices on Tehran stock exchange,'' Phys. A, Stat. Mech. Appl., vol. 438, pp. 178–187, Nov. 2015, doi: 10.1016/j.physa.2015.06.033.

[11] A. H. Moghaddam, M. H. Moghaddam, and M. Esfandyari, ''Stock market index prediction using artificial neural network,'' J. Econ., Finance Administ. Sci., vol. 21, no. 41, pp. 89–93, Dec. 2016, doi: 10.1016/j.jefas.2016.07.002.

[12] H. Liu and Y. Hou, ''Application of Bayesian neural network in prediction of stock time series,'' Comput. Eng. Appl., vol. 55, no. 12, pp. 225–229, 2019.

[13] A. M. Rather, A. Agarwal, and V. N. Sastry, ''Recurrent neural network and a hybrid model for prediction of stock returns,'' Expert Syst. Appl., vol. 42, no. 6, pp. 3234–3241, Apr. 2015, doi: 10.1016/j.eswa.2014.12.003.

[14] A. Sherstinsky, ''Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network,'' Phys. D, Nonlinear Phenomena, vol. 404, Mar. 2020, Art. no. 132306, doi: 10.1016/j.physd.2019.132306.

[15] G. Ding and L. Qin, ''Study on the prediction of stock price based on the associated network model of LSTM,'' Int. J. Mach. Learn. Cybern., vol. 11, no. 6, pp. 1307–1317, Nov. 2019, doi: 10.1007/s13042-019-01041-1.

[16] X. Yan, W. Weihan, and M. Chang, ''Research on financial assets transaction prediction model based on LSTM neural network,'' Neural Comput. Appl., vol. 33, no. 1, pp. 257–270, May 2020, doi: 10.1007/s00521-020- 04992-7.

[17] M. Nabipour, P. Nayyeri, H. Jabani, A. Mosavi, E. Salwana, and S. Shahab, ''Deep learning for stock market prediction,'' Entropy, vol. 22, no. 8, p. 840, Jul. 2020, doi: 10.3390/e22080840.

[18] Z. D. Aksehir and E. Kiliç, ''How to handle data imbalance and feature selection problems in CNN-based stock price forecasting,'' IEEE Access, vol. 10, pp. 31297–31305, 2022, doi: 10.1109/ACCESS.2022. 3160797.

[19] Y. Ji, A. W. Liew, and L. Yang, ''A novel improved particle swarm optimization with long-short term memory hybrid model for stock indices forecast,'' IEEE Access, vol. 9, pp. 23660–23671, 2021, doi: 10.1109/ACCESS.2021.3056713.

[20] X. Zeng, J. Cai, C. Liang, and C. Yuan, ''A hybrid model integrating long short-term memory with adaptive genetic algorithm based on individual ranking for stock index prediction,'' PLoS ONE, vol.

17, no. 8, Aug. 2022, Art. no. e0272637, doi: 10.1371/journal.pone.0272637.

[21] J. Xue and B. Shen, ‘‘A novel swarm intelligence optimization approach: Sparrow search algorithm,’’ Syst. Sci. Control Eng., vol. 8, no. 1, pp. 22–34, Jan. 2020, doi: 10.1080/21642583.2019.1708830.

[22] J. Borade, ‘‘Stock prediction and simulation of trade using support vector regression,’’ Int. J. Res. Eng. Technol., vol. 7, no. 4, pp. 52–57, Apr. 2018, doi: 10.15623/ijret.2018.0704009.

[23] X. Li and P. Tang, ‘‘Stock price prediction based on technical analysis, fundamental analysis and deep learning,’’ Stat. Decis., vol. 38, no. 2, pp. 146–150, 2022, doi: 10.13546/j.cnki.tjyjc.2022.02.029.

[24] J. Heo and J. Y. Yang, ‘‘Stock price prediction based on financial statements using SVM,’’ Int. J. Hybrid Inf. Technol., vol. 9, no. 2, pp. 57–66, Feb. 2016, doi: 10.14257/ijhit.2016.9.2.05.

[25] J. B. De Long, A. Shleifer, L. H. Summers, and R. J. Waldmann, ‘‘Noise trader risk in financial markets,’’ J. Political Economy, vol. 98, no. 4, pp. 703–738, Aug. 1990, doi: 10.1086/261703.

[26] H. Cui and Y. Zhang, ‘‘Does investor sentiment affect stock price crash risk?’’ Appl. Econ. Lett., vol. 27, no. 7, pp. 564–568, Jul. 2019, doi: 10.1080/13504851.2019.1643448.

[27] R. P. Schumaker, Y. Zhang, C.-N. Huang, and H. Chen, ‘‘Evaluating sentiment in financial news articles,’’ Decis. Support Syst., vol. 53, no. 3, pp. 458–464, Jun. 2012, doi: 10.1016/j.dss.2012.03.001.

[28] M. Nofer and O. Hinz, ‘‘Using Twitter to predict the stock market,’’ Bus. Inf. Syst. Eng., vol. 57, no. 4, pp. 229–242, Jun. 2015, doi: 10.1007/s12599- 015-0390-4.

[29] P. Fan, Y. Yang, Z. Zhang, and M. Chen, ‘‘The relationship between individual stock investor sentiment and stock yield-based on the perspective of stock evaluation information,’’ Math. Pract. Theory, vol. 51, no. 16, pp. 305–320, 2021.

[30] Z. Jin, Y. Yang, and Y. Liu, ‘‘Stock closing price prediction based on sentiment analysis and LSTM,’’ Neural Comput. Appl., vol. 32, no. 13, pp. 9713–9729, Sep. 2019, doi: 10.1007/s00521-019-04504-2.

[31] X. Xu and K. Tian, ‘‘A novel financial text sentiment analysis-based approach for stock index prediction,’’ J. Quantum Technol. Econ., vol. 38, no. 12, pp. 124–145, 2021, doi: 10.13653/j.cnki.jqte.2021.12.009.

[32] C.-R. Ko and H.-T. Chang, ‘‘LSTM-based sentiment analysis for stock price forecast,’’ PeerJ Comput. Sci., vol. 7, p. e408, Mar. 2021, doi: 10.7717/peerj-cs.408.

[33] Y. Li and Y. Pan, ‘‘A novel ensemble deep learning model for stock prediction based on stock prices and news,’’ Int. J. Data Sci. Anal., vol. 13, no. 2, pp. 139–149, Sep. 2021, doi: 10.1007/s41060-021-00279-9.

[34] C. Kearney and S. Liu, ‘‘Textual sentiment in finance: A survey of methods and models,’’ Int. Rev. Financial Anal., vol. 33, pp. 171–185, May 2014, doi: 10.1016/j.irfa.2014.02.006.

[35] T. Wang and Z. Zhang, ‘‘Research on the construction method of emotional lexicon for movie review,’’ Comput. Digit. Eng., vol. 50, no. 4, pp. 843–848, 2022, doi: 10.3969/j.issn.1672-9722.2022.04.031.

[36] Y. Rao, J. Lei, L. Wenyin, Q. Li, and M. Chen, ‘‘Building emotional dictionary for sentiment analysis of online news,’’ World Wide Web, vol. 17, no. 4, pp. 723–742, Jun. 2013, doi: 10.1007/s11280-013-0221-9.

[37] L. Yang and T. Zhai, ‘‘Research on sentiment tendency analysis of video reviews based on sentiment dictionary,’’ Netw. Secur. Technol. Appl., vol. 255, no. 3, pp. 53–56, 2022.

[38] A. Fathy, T. M. Alanazi, H. Rezk, and D. Yousri, ‘‘Optimal energy management of micro-grid using sparrow search algorithm,’’ Energy Rep., vol. 8, pp. 758–773, Nov. 2022, doi: 10.1016/j.egyr.2021.12.022.

[39] Y. Chen, Z. Liu, C. Xu, X. Zhao, L. Pang, K. Li, and Y. Shi, ‘‘Heavy metal content prediction based on random forest and sparrow search algorithm,’’ J. Chemometrics, vol. 36, no. 10, Sep. 2022, Art. no. e3445, doi: 10.1002/cem.3445.

[40] S. Hochreiter and J. Schmidhuber, ‘‘Long short-term memory,’’ Neural Comput., vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.