



INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Sentiment-Driven Stock Price Forecasting: Optimized Deep Learning For Investor-Informed Predictions

Dr. J. Rajendra Prasad¹, Yalagandula Baby Ramya², Sammeta Hemanth Venkata Sai Vaibhav³, Yarra

Lohith Naga Sai⁴

#1Head Of The Department in Department of IT, NRI INSTITUTE OF TECHNOLOGY,
AGIRIPALLI.

#2#3#4 B.Tech with Specialization of Information Technology in NRI INSTITUTE OF
TECHNOLOGY, AGIRIPALLI.

Abstract: The study introduces the MS-SSA-LSTM model, which amalgamates multi-source data, sentiment analysis, swarm intelligence algorithms, and deep learning methodologies to improve stock price forecasting. This model integrates sentiment analysis derived from East Money forum postings, establishing a distinctive sentiment lexicon and computing a sentiment index. This provides significant insights into the impact of market sentiment on stock prices. The Sparrow Search Algorithm (SSA) is employed to refine LSTM hyperparameters, enhancing prediction accuracy. •Experimental findings demonstrate the MS-SSA-LSTM model's exceptional performance. It is an essential

instrument for precise stock price forecasting. Designed for India's unpredictable financial market, the model specialises in short-term stock price forecasts, providing insights for agile decision-making by investors. A hybrid LSTM+GRU model was developed for stock sentiment classification. A comprehensive ensemble approach was implemented, utilising a Voting Classifier (AdaBoost + RandomForest) for sentiment analysis and a Voting Regressor (LinearRegression + RandomForestRegressor + KNeighborsRegressor) for stock price forecasting. These ensembles were smoothly merged with existing models (MLP, CNN, LSTM, MS-LSTM, MS-SSA-LSTM), thereby improving overall

predictive performance. A user-friendly Flask framework with SQLite support was designed to enhance user engagement and testing, simplifying the signup, signin, and model assessment processes.

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

1. INTRODUCTION

As India's stock market matures and Internet finance proliferates, many individuals recognise the significance of investment and opt to engage in the financial market. Nonetheless, the stock market is marked by extensive data and significant volatility. Numerous retail investors lack sufficient data-mining skills to generate profit. Consequently, precise stock price forecasting can mitigate investment risks and enhance returns for both investors and corporations.

Early scholars employed statistical techniques to develop a linear model that aligns with the stock price time series pattern. The conventional methodologies encompass ARMA, ARIMA, GARCH, among others. The ARMA model is utilised to do a time series stock analysis [1]. The ARIMA model, derived from the ARMA, forecasts stock price trend fluctuations [2]. The ARIMA model may incorporate wavelet analysis to enhance the fitting precision of the Shanghai Composite Index [3]. The GARCH model offers novel insights for forecasting stock time series within a specified time frame [4]. Simultaneously, certain scholars have integrated ARMA and

GARCH to develop a novel predictive model, offering theoretical support for the volumetric price analysis of multivariate stocks [5]. Typically, these traditional methods exclusively capture regular and structured data. Nevertheless, conventional forecasting techniques necessitate assumptions that are atypical in reality. Consequently, it is difficult to characterise nonlinear financial data by statistical methods.

Subsequently, numerous studies endeavour to predict stock prices employing machine learning methodologies, including Support Vector Machines (SVM) and Neural Networks. The fundamental concept of machine learning is to employ algorithms to analyse data, derive insights, and forecast outcomes for fresh data. Due of its distinct advantages in managing limited samples, high-dimensional data, and nonlinear scenarios, several researchers used SVM for stock forecasting. Hossain and Nasser [6] determined that the SVM technique outperforms statistical methods in stock prediction accuracy. Chai et al. [7] proposed a hybrid SVM model to predict fluctuations in the HS300 index, concluding that the least squares SVM integrated with the Genetic Algorithm (GA) yielded superior results. Nonetheless, the application of SVM to extensive training samples consumes significant memory and computational time, potentially constraining its capacity to predict substantial stock data. Subsequently, Artificial Neural Networks (ANN) and multi-layer ANN tackle financial time series challenges. The experimental data indicates that ANN offers rapid convergence and elevated accuracy [8], [9], [10]. Moghaddam and Esfandiyari

[11] assessed the impact of several feedforward artificial neural networks on forecasting market stock prices via experimental methods. Liu and Hou [12] enhanced the Back Propagation (BP) neural network with the application of Bayesian regularisation. Nonetheless, the conventional neural network approach presents several opportunities for enhancement. The capacity for generalisation is limited, resulting in rapid overfitting and entrapment in local optimisation. As several samples require training, superior models must be identified to address these issues.

This research proposes a novel model for predicting stock prices, termed MS-SSA-LSTM, which aligns the attributes of multi-source data with LSTM neural networks and employs 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. Investors and traders acquire data on specific equities, including past transaction records and commentary from stock market shareholders, and enter this information into the MS-SSA-LSTM model. The model autonomously generates a stock price trend chart and predicts the stock price for the following day.

2. LITERATURE SURVEY

2.1 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:

<https://www.sciencedirect.com/science/article/abs/pii/S0378437116002776>

ABSTRACT: We examined the existence and alterations in long memory characteristics in the returns and volatility dynamics of the S&P 500 and the London Stock Exchange utilising the ARMA model. Recently, multifractal analysis has emerged as a significant method for elucidating the complexity of financial markets, which linear approaches of efficient market theory struggle to characterise. The weak variant of the efficient market theory in financial markets suggests that price returns exhibit serial uncorrelation. Prices need to exhibit a random walk pattern. The random walk hypothesis is assessed in comparison to alternatives that incorporate either unifractality or multifractality. Numerous research indicate that the return volatility of equities demonstrates long-range dependency, hefty tails, and clustering phenomena. Due to their long-range dependence and large tails, it has been proposed that self-similar stochastic processes be utilised to represent these features in return volatility modelling. This study employs monthly and annual forecasting of Time Series Stock Returns for the S&P 500 and London Stock Exchange utilising the ARMA model. The statistical examination of the S&P 500 indicates that the ARMA model surpasses the London Stock Exchange and is proficient in

forecasting medium to long-term horizons utilising actual known values. The statistical examination of the London Stock Exchange indicates that the ARMA model for monthly stock returns surpasses the annual model. A comparison of the S&P 500 and the London Stock Exchange indicates that both markets exhibit efficiency and maintain financial stability amid economic expansions and contractions.

2.2 Gold price forecasting using ARIMA model:

https://www.researchgate.net/publication/322224066_Gold_Price_Forecasting_Using_ARIMA_Model

ABSTRACT: This paper provides an in-depth examination of the application of the ARIMA time series model to predict future gold prices in India, utilising historical data from November 2003 to January 2014, aimed at reducing the risk associated with gold purchases. Therefore, to provide guidance for investors regarding the optimal timing for purchasing or selling the yellow metal. This financial instrument has gained significant traction recently due to the constraints faced by the Indian economy, including a shifting political landscape, global indicators, and elevated inflation. Consequently, researchers, investors, and speculators are seeking alternative financial instruments to mitigate risk through portfolio diversification. Historically, gold in India was mostly acquired during marriage or ceremonial occasions; however, it has now garnered significance among investors, necessitating the development of appropriate methods for predicting gold prices.

2.3 Analysis and prediction of Shanghai composite index by ARIMA model based on wavelet analysis:

<https://www.semanticscholar.org/paper/Analysis-and-Prediction-of-Shanghai-Composite-Index-Hong-ya/4786aef4d7de870f79896d79db8f44e719950599>

ABSTRACT: Stock price prediction garners significant attention, despite the existence of numerous forecasting methods. However, they frequently encounter issues such as restricted predictive accuracy and a propensity to converge on local minima, among others. To enhance the precision of stock price predictions, a modified ARIMA model utilising wavelet analysis for stock price forecasting has been constructed. Subsequently, employ this technique to analyse the monthly average closing price of the Shanghai Composite Index. Additionally, the prediction results were compared with alternative methodologies. The findings demonstrate the efficacy of the suggested approach.

3. METHODOLOGY

i) Proposed Work:

The project presents the MS-SSA-LSTM model, an advanced approach for forecasting stock prices. This methodology mixes multi-source data, sentiment analysis, and swarm intelligence algorithms smoothly. [14, 15, 16, 30] By optimising LSTM hyperparameters using the Sparrow Search Algorithm, the system demonstrates remarkable accuracy in anticipating stock values. Experimental results confirm its

advantage over alternative models, emphasising its universal applicability and potential to improve prediction performance. This model is juxtaposed with MLP, CNN, LSTM, and MS-LSTM. A hybrid LSTM and GRU model was developed for stock sentiment classification. A comprehensive ensemble approach was implemented, utilising a Voting Classifier (AdaBoost + RandomForest) for sentiment analysis and a Voting Regressor (LinearRegression + RandomForestRegressor + KNeighborsRegressor) for stock price forecasting. These ensembles were smoothly incorporated with existing models (MLP, CNN, LSTM, MS-LSTM, MS-SSA-LSTM), thereby improving overall predictive performance. A user-friendly Flask framework with SQLite support was designed to enhance user engagement and testing, simplifying the signup, signin, and model assessment processes.

ii) System Architecture:

The research introduces the MS-SSA-LSTM model, a sophisticated method for predicting stock prices. This methodology mixes multi-source data, sentiment analysis, and swarm intelligence algorithms smoothly. [14, 15, 16, 30] By optimising LSTM hyperparameters with the Sparrow Search Algorithm, the system exhibits exceptional accuracy in predicting stock prices. Experimental results validate its superiority over competing models, highlighting its universal applicability and potential to enhance prediction performance. This model is compared with MLP, CNN, LSTM, and MS-LSTM. A hybrid model combining LSTM and GRU was created for stock

sentiment classification. An extensive ensemble methodology was employed, incorporating a Voting Classifier (AdaBoost + RandomForest) for sentiment analysis and a Voting Regressor (LinearRegression + RandomForestRegressor + KNeighborsRegressor) for stock price prediction. These ensembles were seamlessly integrated with existing models (MLP, CNN, LSTM, MS-LSTM, MS-SSA-LSTM), hence enhancing overall predictive efficacy. A user-centric Flask framework with SQLite integration was developed to improve user interaction and testing, streamlining the signup, signin, and model evaluation procedures.

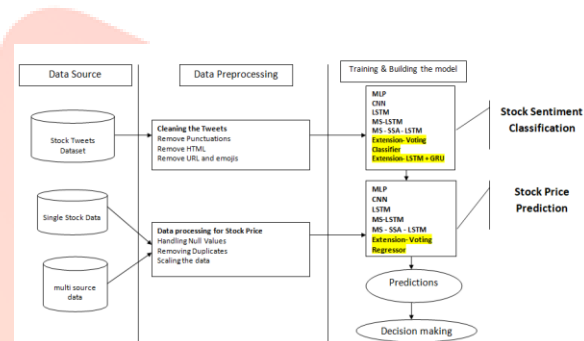


Fig 1 Proposed architecture

iii) Dataset collection:

STOCK TWEETS DATASET

The "Stock Tweets" collection comprises social media posts pertaining to equities and financial markets. We utilised it to comprehend individuals' emotions and responses to market news [1,4,7,8]. This facilitated the development of tools for stock trading and investments. We want to examine the influence of social media on stock prices and market movements to assist investors and traders. So, these are the top 5 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

Fig 2 Stock tweets dataset

ALL STOCK DATASET

The "All Stock Dataset" is an extensive compilation of financial data from multiple sources. It offers an abundance of information for comprehensive stock market analysis. We utilised this dataset to improve our stock price prediction model in our project. Our objective was to enhance the precision of stock price predictions by utilising many data sources, thereby benefiting 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 converting unrefined data into usable information for enterprises. Data scientists typically engage in data processing, encompassing the collection, organisation, cleansing, verification, analysis, and transformation of data into comprehensible formats such as graphs or papers. Data processing can be accomplished through three methods: manual, mechanical, and electronic. The objective is to enhance the value of information and streamline decision-making. This allows enterprises to enhance their operations and execute fast strategic decisions. Automated data processing

technologies, including computer software programming, are pivotal in this context. It may transform extensive datasets, especially big data, into significant insights for quality management and decision-making.

v) Feature selection:

Feature selection is the process of identifying the most consistent, non-redundant, and pertinent features for model development. Systematically minimising dataset dimensions is crucial while the magnitude and diversity of datasets persist in expanding. The primary objective of feature selection is to enhance the efficacy of a predictive model while minimising the computational expenses associated with modelling. Feature selection, a fundamental aspect of feature engineering, involves identifying the most significant features for input into machine learning algorithms. Feature selection strategies are utilised to diminish the amount of input variables by removing redundant or unnecessary features, hence refining the set to those most pertinent to the machine learning model. The primary advantages of conducting feature selection beforehand, as opposed to allowing the machine learning model to determine the significance of characteristics autonomously.

4. EXPERIMENTAL RESULTS

Precision: Precision assesses the proportion of accurately classified cases among those identified as positive. Consequently, the formula for calculating precision is expressed as:

Precision = True positives/ (True positives + False positives) = TP/(TP + FP)

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

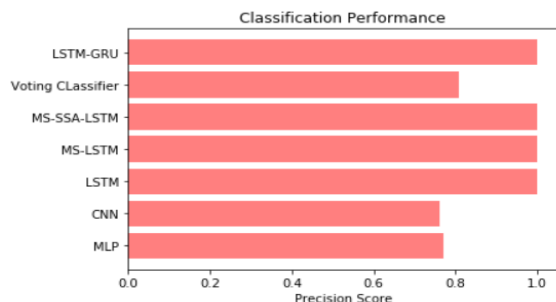


Fig 12 Precision comparison graph

Recall: Recall is a metric in machine learning that assesses a model's capacity to recognise all pertinent instances of a specific class. It is the ratio of accurately predicted positive observations to the total actual positives, offering insights into a model's efficacy in identifying occurrences of a specific class.

$$\text{Recall} = \frac{TP}{TP + FN}$$

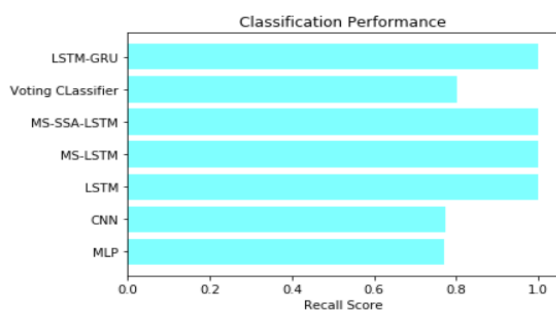


Fig 13 Recall comparison graph

Accuracy: Accuracy denotes the ratio of correct predictions in a classification task, assessing the overall correctness of a model's predictions.

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

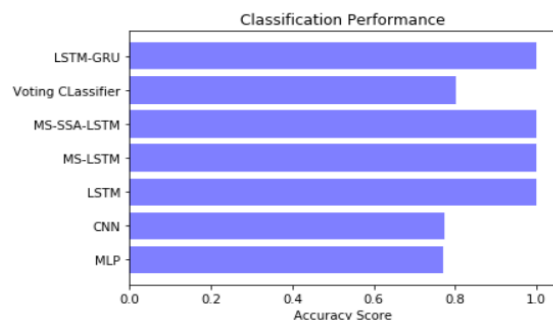
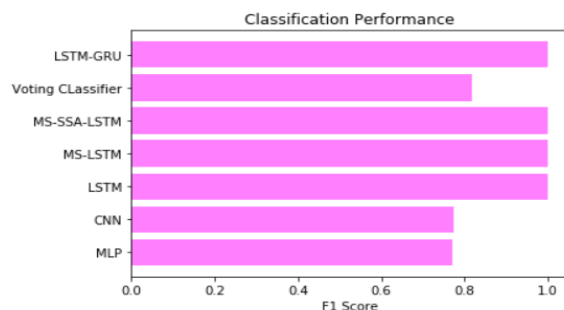


Fig 14 Accuracy graph

F1 Score: The F1 Score is the harmonic mean of accuracy and recall, providing a balanced metric that accounts for both false positives and false negatives, rendering it appropriate for imbalanced datasets.

$$\text{F1 Score} = 2 * \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} * 100$$



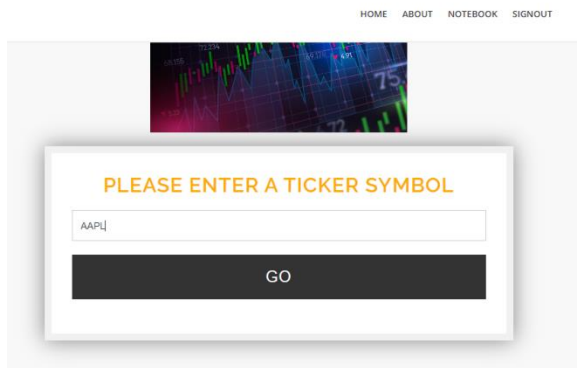


Fig 15 User input

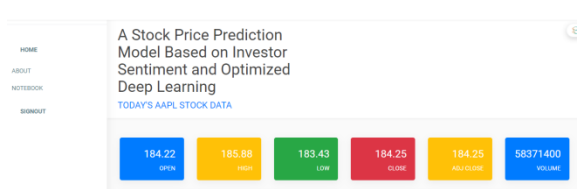


Fig 16 Result

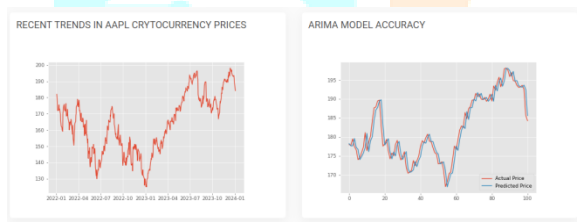


Fig 17 Graphs

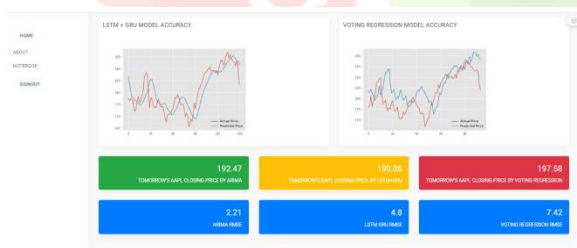


Fig 18 Graphs

5. CONCLUSION

The project sought to improve stock market forecasts, concentrating on the MS-SSA-LSTM model. The study examined multiple models, highlighting the importance of sentiment analysis and advanced algorithms for enhanced forecasting

[26]. The MS-SSA-LSTM model excelled in both stock price prediction and sentiment classification. Utilising varied data sources and sophisticated methodologies, it provided a holistic strategy for risk mitigation and enhanced returns. Current models (MLP, CNN, LSTM, MS-LSTM) exhibited proficiency, however the MS-SSA-LSTM model displayed exceptional performance, especially in short-term forecasts for India's volatile market. The introduction of ensemble models (Voting Classifier, LSTM+GRU, Voting Regressor) during the extension phase enhanced the predictive capabilities. LSTM and GRU excelled in sentiment categorisation, while the Voting Regressor demonstrated superior performance in stock price prediction, providing dependable options. The Flask addon enabled intuitive interactivity, including the entry of ticker symbols for precise predictions. LSTM and GRU models for sentiment analysis, along with a Voting Regressor for stock price predictions, were effectively deployed, improving accessibility for users and investors. Investors, traders, and enterprises are poised to gain from the project's sophisticated prediction models and intuitive interface. The MS-SSA-LSTM model and its extensions provide significant insights, mitigate investment risks, and improve decision-making in the volatile environment of the Chinese financial market.

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. Esfandiyari, "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. Weihang, 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.

Author Profile:

Dr. J. Rajendra Prasad working as a professor and HoD of IT Department, NRI Institute of Technology Agiripalli, Andhra Pradesh, India. He received Best teacher award in the year 2010 from JNTUK, Kakinada. He guided 2 PHD's in computer science and Engineering. He published more than 70 publications in international reputed journals. He wrote a text book: Mathematical foundation of computer science.

Email- rp.rajendra@rediffmail.com



I am Yalagandula Baby Ramya, currently pursuing a B.Tech in Information Technology at NRI Institute of Technology, Agiripalli. My areas of interest include Data Science, Deep Learning, Web Development, and Cloud Computing. I have earned certifications such as NPTEL's "Joy of Computing using Python", "Cloud Computing" and the Java Full Stack Developer certification by Wipro, where I gained expertise in front-end and back-end development. I have also completed internships in Data Science, DevOps, AI, Cloud Computing, and Data Visualization, which provided hands-on experience and practical skills in cutting-edge technologies, enabling me to contribute to innovative and impactful projects.



I am Sammeta Hemanth Venkata Sai Vaibhav, currently pursuing a B.Tech in Information Technology at NRI Institute of Technology. I have a strong interest in Full Stack Development, Cloud Computing, and Cybersecurity. I have completed certifications in Java Full Stack Development, Generative AI, Cyber Security, and Ethical Hacking, along with internships in Cloud Computing and Machine Learning. Additionally, I have gaining hands-on experience in Java, Python, AWS, and web technologies. I am passionate about building scalable applications and solving real-world problems through technology.



I am Yarra Lohith Naga Sai, currently pursuing a B.Tech in Information Technology at NRI Institute of Technology. I am deeply interested in Web Development, Servicenow Tool and cloud computing. I have completed certifications in Joy Of Computing using Python ,Data Structure and Algorithms using Java along with internships in Data science and DevOps. Additionally, I have learned coding experience in Java, python and I am knowledgeable about various AI tools and their applications across different domains. I am passionate about building scalable applications and solving real-world problems through technology.