

ENHANCED STOCK PRICE PREDICTION USING INVESTOR SENTIMENT AND MACHINE LEARNING

GUIDE : B. ROJA SRI M.Tech(CSE)
Assistant Professor, Department of CSE(AI&DS)
Eluru College of Engineering and Technology

TEAM LEAD : K. TEJASWI¹
TEAM MEMBERS: M. JITHIN VIJAY SAI²
L. AKANKSHA³
B. PAVAN KUMAR⁴

Eluru College of Engineering and Technology

ABSTRACT

In today's fast-evolving technological landscape, accurately predicting market trends plays a critical role in minimizing financial risks and maximizing potential returns. This work introduces a novel approach, the MS-SSA-LSTM model, designed to harness multi-source data in the prediction of stock prices. By incorporating sentiment analysis, swarm intelligence techniques, and deep learning, this method analyses data such as posts from the East Money forum to create a custom sentiment lexicon and calculate sentiment indices. These indices are then integrated with traditional market data, with the Sparrow Search Algorithm (SSA) optimizing the parameters of a Long Short-Term Memory (LSTM) network. The results show that the MS-SSA-LSTM model significantly improves forecasting accuracy, achieving an average R^2 increase of 10.74% compared to standard LSTM methods. Furthermore, the combination of sentiment indices and hyperparameter optimization enhances the model's performance, offering a robust solution for short-term stock price forecasting, especially in volatile markets like that of China.

INTRODUCTION

Stock price prediction has long been a critical area of research in financial markets, traditionally relying on historical price data, technical indicators, and fundamental analysis. However, with the rise of big data and advanced computing, machine learning (ML) has emerged as a powerful tool for forecasting stock prices with greater accuracy.

One promising approach to enhancing stock price prediction is incorporating investor sentiment, which reflects the collective emotions and opinions of market participants. Investor sentiment, derived from news articles, social media, financial reports, and other online sources, has been shown to influence market movements significantly. By integrating sentiment analysis with ML models, researchers and traders can capture market psychology, providing deeper insights into price trends and volatility.

This study explores how investor sentiment, when combined with advanced ML techniques, can improve stock price prediction accuracy. By leveraging natural language processing (NLP) for sentiment analysis and ML algorithms such as deep learning, ensemble models, and reinforcement learning, this approach aims to refine financial forecasting and

inform better investment.

Stock price prediction has been a focal point of financial research and investment strategies, aiming to anticipate market movements and optimize trading decisions. Traditional forecasting methods rely on historical price trends, technical indicators, and fundamental financial metrics. However, these approaches often fall short in capturing the complex, dynamic, and sometimes irrational nature of financial markets. With the rise of artificial intelligence (AI) and machine learning (ML), data-driven models have revolutionized predictive analytics, providing more sophisticated and adaptive techniques for stock price forecasting.

A critical factor that influences stock prices beyond historical data is investor sentiment—the collective mood and emotions of market participants. Market sentiment is often shaped by news reports, earnings announcements, social media discussions, expert analyses, and macroeconomic events. Psychological biases and investor emotions, such as fear and greed, frequently drive price fluctuations that are not always reflected in traditional quantitative models.

EXISTING SYSTEM

Yan et al. developed a high-precision prediction model using the LSTM deep neural network for short-term financial markets, showing superior prediction accuracy over BP neural networks and standard RNNs, successfully forecasting stock prices. Similarly, Nabipour et al. used ten technical indicators as inputs and found that LSTM outperformed other models like decision trees, random forests, Adaboost, XGBoost, ANN, and RNN in terms of model-fitting ability. Aksehir and Kilic proposed a CNN-based model to predict next-day trading behavior for Dow Jones Index equities, incorporating technical indicators, gold, and oil price data, with results 3-22% more accurate than other CNN-based models. However, the manual setting of hyperparameters in LSTM networks, which directly impacts model performance, is subjective and resource-intensive. To address this, scholars have turned to Swarm Intelligence (SI) algorithms to optimize these hyperparameters. SI can globally search for optimal solutions and reduce the need for extensive manual effort. Ji et al. demonstrated that an LSTM optimized using the Improved Particle Swarm Optimization (IPSO) model outperforms the standard LSTM, while Zeng et al. used an Adaptive Genetic Algorithm (AGA) to enhance prediction accuracy. The Sparrow Search Algorithm (SSA), inspired by the foraging and anti-predation behaviors of sparrows, offers robust optimization abilities in price prediction, surpassing traditional particle swarm and gray wolf algorithms. SSA's global search capabilities and adaptability to different problem sets make it effective for optimization. Moreover, real-time stock forum content provides valuable insights into investor sentiment, impacting investment decisions and stock price fluctuations. Analyzing this text data is essential, though sentiment classification methods such as machine learning and semantic analysis have their challenges. Machine learning offers high classification accuracy but requires manual training set classification, while semantic analysis is easier but struggles with applying standard dictionaries to the economic context, necessitating the development of specialized financial dictionaries.

DISADVANTAGES OF EXISTING SYSTEM

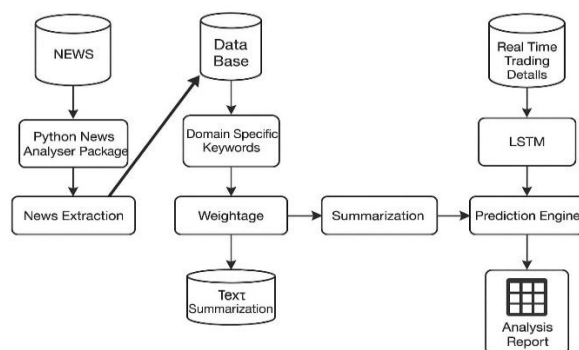
- Cannot Handle Complex Patterns – Linear and Logistic Regression assume simple relationships, making them ineffective for capturing complex air quality trends.
- Sensitive to Outliers and Imbalanced Data – These models can be easily affected by extreme values and struggle when one class is much larger than the other.
- Lower Accuracy – Due to their limitations, they often perform worse than advanced models like Decision Trees, which can handle more complex data.

The MS-SSA-LSTM model is introduced for predicting stock prices, integrating LSTM neural networks with the Sparrow Search Algorithm (SSA) to efficiently handle multi-source data. The model assists investors and traders by forecasting stock prices and generating stock price trend charts, enabling more informed investment decisions. It takes various inputs, including historical transaction data and comments from stock market shareholders, to predict the stock price for the following day. A key improvement in the model is the inclusion of sentiment indicators, which enhance prediction accuracy. However, using a general dictionary for sentiment analysis in the financial sector is not effective, highlighting the need for a specialized sentiment dictionary tailored specifically for individual stocks. Additionally, stock price series are complex, exhibiting characteristics such as nonlinearity, high noise, and strong time-variability, making the LSTM network particularly suitable for processing such time-series data. Hyperparameter tuning in LSTM networks significantly impacts prediction accuracy, but manually selecting the optimal hyperparameters is resource-intensive. To overcome this challenge, the Sparrow Search Algorithm, introduced in 2020, is used to optimize the LSTM model, improving its predictive capabilities while reducing the computational cost associated with manual hyperparameter selection.

ADVANTAGES PROPOSED SYSTEM

- More Accurate Predictions – Combines multi-source data and sentiment indicators for better stock price forecasting.
- Handles Complex Data – LSTM effectively processes nonlinear and time-variable stock price trends.
- Automated Optimization – The Sparrow Search Algorithm fine-tunes the model, reducing manual effort and improving performance.

SYSTEM ARCHITECTURE



The diagram represents a stock market prediction system that integrates news analysis with real-time trading data.

The process begins with a Python News Analyzer

Package, which extracts relevant information from news articles.

This extracted data is then stored in a database, where domain-specific keywords are identified and assigned weightage based on relevance. To streamline processing, text summarization condenses the extracted news. Simultaneously, real-time trading details are incorporated into the system. The summarized news and market data undergo further summarization before being processed by an LSTM (Long Short-Term Memory) neural network, which serves as the prediction engine. The final predictions are compiled into an analysis report, aiding investors and financial analysts in making informed decisions. By combining news sentiment analysis with real-time trading data, this model enhances stock market forecasting accuracy while reducing information overload.

Additionally, the integration of news sentiment analysis with real-time trading data ensures that the system remains responsive to market dynamics and investor sentiment, which are key drivers of stock prices. By leveraging advanced machine learning techniques and natural language processing (NLP), this model enhances stock market forecasting accuracy while reducing information overload. It eliminates the need for manual news analysis, making the prediction process more efficient and automated. Furthermore, the system can be expanded by incorporating alternative data sources such as social media sentiment, economic indicators, and macroeconomic factors, further improving its predictive capabilities. The adaptability of this framework allows it to be fine-tuned for different financial market.

Simultaneously, real-time trading details are fed into an LSTM model, which learns from past stock trends to improve prediction accuracy. The summarized news and LSTM outputs are combined in the prediction engine, which generates a final analysis report. This integration of news sentiment analysis and real-time trading data helps enhance stock market forecasting. By leveraging text summarization, keyword-based weighting, and LSTM's ability to detect patterns, the model provides a more efficient and accurate approach to stock market predictions.

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. 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: Sparrowsearch 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.tjyc.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.

Author's Pictures



Mrs.B.Roja Sri

M.Tech (CSE) working as Assistant Professor in department of CSE(Artificial Intelligence and Data Science), Eluru College of Engineering and Technology, Duggirala.

Email : bhagavataroajasri33@gmail.com



K. Tejaswi

B.Tech with specialization of CSE(Artificial Intelligence and Data Science), Eluru College of Engineering and Technology, Duggirala.

Email: tejaswikaranam27@gmail.com



B. Pavan Kumar

B.Tech with specialization of CSE(Artificial Intelligence and Data Science), Eluru College of Engineering and Technology, Duggirala.

Email: pawanbhimadolu38@gmail.com



M. Jithin Vijay Sai

B.Tech with specialization of CSE(Artificial Intelligence and Data Science), Eluru College of Engineering and Technology, Duggirala.

Email: jithinvijaysai@gmail.com



L. Akanksha

B.Tech with specialization of CSE(Artificial Intelligence and Data Science), Eluru College of Engineering and Technology, Duggirala.