



ISSN: 2454-9940



**INTERNATIONAL JOURNAL OF APPLIED
SCIENCE ENGINEERING AND MANAGEMENT**

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

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

Komireddy Shreyas¹, Lingampally Ganesh²,

Mangali Vamshi³, Dr. A. Sudhakar⁴

^{1,2,3} UG Student, Dept. of ECE, CMR Institute of Technology, Hyderabad

⁴ Associate Professor, Dept. of ECE

CMR Institute of Technology, Hyderabad

ABSTRACT

Accurate prediction of stock prices can reduce investment risks and increase returns. This paper combines the multi-source data affecting stock prices and applies sentiment analysis, swarm intelligence algorithm, and deep learning to build the MS-SSA-LSTM model. Firstly, we crawl the East Money forum posts information to establish the unique sentiment dictionary and calculate the sentiment index. Then, the Sparrow Search Algorithm (SSA) optimizes the Long and Short-Term Memory network (LSTM) hyperparameters. Finally, the sentiment index and fundamental trading data are integrated, and LSTM is used to forecast stock prices in the future. Experiments demonstrate that the MS-SSA-LSTM model outperforms the others and has high universal applicability. Compared with standard LSTM, the R^2 of MS-SSA-LSTM is improved by 10.74% on average. We found that: 1) Adding the sentiment index can enhance the model's predictive

performance. 2) The LSTM's hyperparameters are optimized using SSA, which objectively explains the model parameter settings and improves the prediction effect. 3) The high volatility of China's financial market is more suitable for short-term prediction.

INTRODUCTION

With the maturity of China's stock market and the rapid growth of Internet finance, many people realize the importance of investment and choose to enter the financial market. However, the stock market is characterized by massive data and enormous volatility. Many retail investors need more data-mining skills to make money. Therefore, accurate stock price prediction can reduce investment risks and improve investment returns for investors and enterprises. Early scholars used statistical methods to construct a linear model to fit the stock price time series trend. The traditional methods contain ARMA, ARIMA, GARCH, etc. The ARMA is established to conduct a time series stock analysis [1]. The ARIMA model is developed based on the ARMA and predicts the trend of stock price changes [2].

The ARIMA The associate editor coordinating However, when the SVM applies to large-scale the review of this manuscript and approving it for training samples, much memory and computing publication was Kuo-Ching Ying . model can also time will be consumed, which may limit its introduce wavelet analysis to improve the fitting development space in predicting a large amount of accuracy of the Shanghai Composite Index [3]. stock data. Then, Artificial Neural Networks The GARCH model provides innovative ideas for (ANN) and multi-layer ANN address financial stock time series prediction through a time time series issues. According to the experimental window [4]. At the same time, some scholars have data, ANN has the benefits of quick convergence combined ARMA and GARCH to build a new and high accuracy [8], [9], [10]. Moghaddam and prediction model, which provided theoretical Esfandyari [11] evaluated the effect of several support for the volumetric price analysis of feedforward artificial neural networks on the multivariate stocks [5]. Generally speaking, these market stock price forecast through experiments. classical methods only capture regular and Liu and Hou [12] improved the BP (Back structured data. However, traditional forecasting Propagation) neural network using the Bayesian methods require assumptions that are uncommon regularization method. Nevertheless, the in real life. Therefore, It is challenging to describe traditional neural network method has the nonlinear financial data using statistical methods. following areas for improvement. Generalization Subsequently, many researchers attempt to ability is not strong, quickly leads to overfitting, anticipate stock prices using machine learning and falls into local optimization. Since many approaches such as Support Vector Machines samples need to be trained, better models must be (SVM) and Neural Networks. Machine learning's found to solve these problems. A new model for core idea is to use algorithms to parse data, learn predicting stock prices is proposed in this paper from it, and make predictions about new data. (MS-SSA-LSTM), which matches the Because the SVM shows unique benefits in characteristics of multi-source data with LSTM dealing with limited samples,high-dimensional neural networks and uses the Sparrow Search data, and nonlinear situations, many scholars use Algorithm. The MS-SSA-LSTM stock price it in stock forecasting. Hossain and Nasser [6] forecast model can forecast the stock price in found that the SVM method is superior to the advance and help investors and traders make more statistical ones in stock prediction accuracy. informed investment decisions. Investors and Chai et al. [7] suggested a hybrid SVM model to traders obtain the data of individual stocks they anticipate the HS300 index's ups and downs and want to invest in, including historical transaction found that the least squares SVM combined with data and comment information of stock market the Genetic Algorithm (GA) performed better. shareholders, and input them into the MS-SSA-

LSTM model. The model automatically outputs a stock price trend chart and forecasts the stock price for the next day. Here, we can get the motivation for this paper. (1) Adding sentiment indicators to the model features will improve prediction accuracy.

However, applying a general dictionary in the financial field cannot achieve good results. Therefore, We need to construct a sentiment dictionary specific to individual stocks. (2) Stock price series has complicated characteristics such as nonlinearity, high noise, and strong time-variability. LSTM is effective at handling comprehensive time series data. So LSTM network, a deep learning method, is adopted. (3) In the LSTM network, hyperparameter variation directly affects the model's prediction accuracy. Artificial selection of appropriate hyperparameters for the network model will cost many resources. Therefore, LSTM can be optimized using the Sparrow Search Algorithm proposed in 2020. Here are the main contributions of this paper. (1) The SSA-LSTM model incorporates multi-source data that affects stock prices, such as historical trade data and stock forum comments. The MS-SSA-LSTM model outperforms the single data source model regarding prediction accuracy. (2) Analyze the stock forum comments text information. Then, we construct a sentiment dictionary based on authoritative dictionaries in the financial investment field suitable for individual stock sentiment analysis and sentiment indicator

calculation. (3) A Sparrow Search Algorithm-optimized LSTM model can acquire better hyperparameter values and forecast stock prices than a single LSTM network.

LITERATURE REVIEW

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

We investigated the presence and changes in, long memory features in the returns and volatility dynamics of S&P 500 and London Stock Exchange using ARMA model. Recently, multifractal analysis has been evolved as an important way to explain the complexity of financial markets which can hardly be described by linear methods of efficient market theory. In financial markets, the weak form of the efficient market hypothesis implies that price returns are serially uncorrelated sequences. In other words, prices should follow a random walk behavior. The random walk hypothesis is evaluated against alternatives accommodating either unifractality or multifractality. Several studies find that the return volatility of stocks tends to exhibit long-range dependence, heavy tails, and clustering. Because stochastic processes with self-similarity possess long-range dependence and heavy tails, it has been suggested that self-similar processes be employed to capture these

characteristics in return volatility modeling. The of news stories. This system is trained on present study applies monthly and yearly historical data to provide investors with forecasting of Time Series Stock Returns in S&P one- and five-days-ahead gold price 500 and London Stock Exchange predictions while achieving a highly using ARMA model. The statistical analysis of interpretable trading strategy in terms of S&P 500 shows that the ARMA model for S&P rule complexity. We demonstrate that the 500 outperforms the London stock exchange and proposed system is effective in terms of it is capable for predicting medium or long both prediction accuracy and horizons using real known values. The statistical interpretability compared with state-of-the-analysis in London Stock Exchange shows that art models, such as extreme learning the ARMA model for monthly stock returns machines and neural networks with deep outperforms the yearly. A comparison between learning. Our findings suggest that the S&P 500 and London Stock Exchange shows that component of news affect is particularly both markets are efficient and have Financial important for one-day-ahead predictions. Stability during periods of boom and bust.

Fuzzy Rule-Based Prediction of Gold Prices using News Affect

Because of gold's value, systems for predicting its price have attracted extensive interest in the scientific and industrial communities. Diverse artificial intelligence methods outperform traditional statistical methods in predicting short- and long-term gold price. However, previous research has neglected the transparency of these systems, nor have these systems incorporated the potentially important effect of media sentiment on investment decisions. Therefore, we here propose a fuzzy rule-based prediction system with a component that processes various aspects

We also show that the proposed system performs well in terms of average annual return while providing an interpretable set of linguistic trading rules. This has important implications for investors.

Gold has had gaining share of investment portfolios of both retail and institutional investors over the last decade ([Smales, 2014](#)). In the global economy, gold has been widely utilized as a 'store of value' and 'safe haven', as well as a derivative instrument and risk-diversification security, amongst other uses ([Ntim, English, Nwachukwu, & Wang, 2015](#)).

Therefore, factors affecting gold prices have been widely investigated, with a focus on financial and macroeconomic variables

(Qian, Ralescu, & Zhang, 2019). Linkages among gold and equity markets have also been assessed to demonstrate the gold's capacity to hedge equity losses (El Hedi Aroui, Lahiani, & Nguyen, 2015), its characteristic as a 'safe haven' (Creti, Joëts, & Mignon, 2013). Causalities have also been found between gold and other precious metal markets (Bhatia, Das, Tiwari, Shahbaz, & Hasim, 2018), and significant interconnections have been found between the gold market and various macroeconomic variables, such as inflation (Gangopadhyay, Jangir, & Sensarma, 2016).

A Comparative Analysis of the ARIMA and LSTM Predictive Models and Their Effectiveness for Predicting Wind Speed

Forecasting wind speed has become one of the most attractive topics to researchers in the field of renewable energy due to its use in generating clean energy, and the capacity for integrating it into the electric grid. There are several methods and models for time series forecasting at the present time. Advancements in deep learning methods characterize the possibility of establishing a more developed multistep prediction model than shallow neural networks (SNNs). However, the accuracy and adequacy of long-term wind speed prediction is not yet well

resolved. This study aims to find the most effective predictive model for time series, with less errors and higher accuracy in the predictions, using artificial neural networks (ANNs), recurrent neural networks (RNNs), and long short-term memory (LSTM), which is a special type of RNN model, compared to the common autoregressive integrated moving average (ARIMA). The results are measured by the root mean square error (RMSE) method. The comparison result shows that the LSTM method is more accurate than ARIMA.

Forecasting S&P-100 stock index volatility: The role of volatility asymmetry and distributional assumption in GARCH models

This study investigates the daily volatility forecasting for the Standard & Poor's 100 stock index series from 1997 to 2003 and identifies the essential source of performance improvements between distributional assumption and volatility specification using distribution-type (GARCH- N , GARCH- t , GARCH- HT and GARCH- SGT) and asymmetry-type (GJR-GARCH and EGARCH) volatility models through the superior predictive ability (SPA) test. Empirical results indicate that the GJR-GARCH model achieves the most accurate volatility forecasts, closely followed by the EGARCH model. Such evidence strongly demonstrates that modeling asymmetric components is more important than specifying error distribution for improving volatility

forecasts of financial returns in the presence of fat-tails, leptokurtosis, skewness and leverage effects. Furthermore, if asymmetries are neglected, the GARCH model with normal distribution is preferable to those models with more sophisticated error distributions.

Since the 1987 stock market crash, modeling and forecasting financial market volatility has received a great deal of attention from academics, practitioners and regulators due to its central role in several financial applications, including option pricing, asset allocation and hedging. In addition, the financial world has witnessed the bankruptcy or near bankruptcy of various institutions that incurred huge losses due to their exposure to unexpected market moves for more than a decade. These financial disasters have further highlighted the significance of volatility forecasting in risk management (calculating Value-at-Risk). Given these facts, the quest for accurate forecasts appears to still be ongoing in recent years.

International evidence on crude oil price dynamics: Applications of ARIMA-GARCH models

We examine the usefulness of several ARIMA-GARCH models for modeling and forecasting the conditional mean and volatility of weekly crude oil spot prices in eleven international markets

over the 1/2/1997–10/3/2009 period. In particular, we investigate the out-of-sample forecasting performance of four volatility models — GARCH, EGARCH and APARCH and FIGARCH over January 2009 to October 2009. Forecasting results are somewhat mixed, but in most cases, the APARCH model outperforms the others. Also, conditional standard deviation captures the volatility in oil returns better than the traditional conditional variance. Finally, shocks to conditional volatility dissipate at an exponential rate, which is consistent with the covariance-stationary GARCH models than the slow hyperbolic rate implied by the FIGARCH alternative.

The past few years have witnessed a renewed interest in modeling and forecasting crude oil prices and their volatility. A number of factors may have contributed to that interest. First, such information is useful for modeling oil price derivatives, forecasting future spot prices, calculating measures of risk as well as hedging (Hansen and Lunde, 2005). Second, oil price movements appear to have undesired consequences for the economies of both oil-importing and oil-exporting countries. A large and growing body of literature has studied the effect of oil shocks on inflation (Rothemberg and Woodford, 1996), output (Hamilton and Herrera, 2004, Mork, 1989, Ferderer, 1996, Yang et al.,

2002); investment (Bernanke, 1983, Elder and Serletis, 2009); monetary policy (Bernanke et al., 1997 and commodity prices (Chaudhuri, 2001).

EXISTING SYSTEM

In addition, Yan et al. [16] developed a high-precision prediction model based on the LSTM deep neural network in a short-term financial market. LSTM neural network has a higher prediction accuracy than BP neural network and standard RNN and can successfully forecast stock prices. Nabipour et al. [17] selected ten technical indicators as the prediction model's input. The experiment revealed that LSTM outperformed other algorithms in terms of model-fitting abilities, including decision tree, random forest, Adaboost, XGBoost, ANN, and RNN. Aksehir and Kilic [18] suggested a CNN-based model predicting the nextday trading behavior of Dow Jones 30 Index equities. Technical indicators, gold, and oil price data are fed into the model. The accuracy of the results is 3-22% higher than other models based on CNN.

Disadvantages

- The complexity of data: Most of the existing machine learning models must be able to accurately interpret large and complex datasets to detect Stock Price Prediction.

- Data availability: Most machine learning models require large amounts of data to create accurate predictions. If data is unavailable in sufficient quantities, then model accuracy may suffer.

- Incorrect labeling: The existing machine learning models are only as accurate as the data trained using the input dataset. If the data has been incorrectly labeled, the model cannot make accurate predictions.

Proposed System

A new model for predicting stock prices is proposed in this paper (MS-SSA-LSTM), which matches the characteristics of multi-source data with LSTM neural networks and uses the Sparrow Search Algorithm. The MS-SSA-LSTM stock price forecast model can forecast the stock price in advance and help investors and traders make more informed investment decisions. Investors and traders obtain the data of individual stocks they want to invest in, including historical transaction data and comment information of stock market shareholders, and input them into the MS-SSA-LSTM model. The model automatically outputs a stock price trend chart and forecasts the stock price for the next day. Here, we can get the motivation for this paper.

(1) Adding sentiment indicators to the model features will improve prediction accuracy. However, applying a general dictionary in the financial field cannot achieve good results. Therefore, We need to construct a sentiment dictionary specific to individual stocks.

(2) Stock price series has complicated characteristics such as nonlinearity, high noise, and strong time-variability. LSTM is effective at handling comprehensive time series data.

So LSTM network, a deep learning method, is adopted. (3) In the LSTM network, hyperparameter variation directly affects the model's prediction accuracy. Artificial selection of appropriate hyperparameters for the network model will cost many resources. Therefore, LSTM can be optimized using the Sparrow Search Algorithm proposed in 2020.

Advantages

(1) The SSA-LSTM model incorporates multi-source data that affects stock prices, such as historical trade data and stock forum comments. The MS-SSA-LSTM model outperforms the single data source model regarding prediction accuracy.

(2) Analyze the stock forum comments text information. Then, we construct a sentiment dictionary based on authoritative

dictionaries in the financial investment field suitable for individual stock sentiment analysis and sentiment indicator calculation.

(3) A Sparrow Search Algorithm-optimized LSTM model can acquire better hyperparameter values and forecast stock prices than a single LSTM network.

CONCLUSION

The stock price is affected by shareholder emotion, and the hyperparameters in the LSTM network are frequently chosen based on subjective experience. Therefore, this paper proposes the MS-SSA-LSTM model of stock price prediction. Six representative data sets of individual stocks in the Chinese financial market are selected to train and test the model. Moreover, four assessment indicators are employed to check the model's prediction performance. Through comparative analysis, we can draw the following results. First of all, the MS-SSA-LSTM model considers multiple data sources. On the one side, the data source is the characteristics of the historical transaction data, including the previous closing price, opening price, closing price, etc. On the other side, the data source is the sentiment of shareholders in the market. Adding a sentiment indicator enhances the model's predictive performance compared

with only using fundamental stock trading indicators as input. Thus, the stock bar can be used as a guiding platform for investor sentiment.

The platform managers can conduct public opinion management and early warning, reasonably guide shareholders to invest rationally and take countermeasures in advance for possible panic or riot. Ultimately, we should establish a good and orderly stock market environment. Secondly, the parameters of LSTM need to be adjusted artificially, which challenges obtaining the best prediction results. Therefore, the SSA is chosen to optimize the LSTM model's hyperparameters. This approach not only objectively explains the network structure and parameter setting of the model but also improves the model's adaptability and forecasting capabilities. In a challenging financial market, the LSTM model optimized by SSA can quickly and precisely comprehend data properties, provide high-precision price prediction, and decrease investors' risk. Thirdly, the comparative experiments of six individual stocks in different models verify that the MS-SSA-LSTM model has high accuracy, reliability, and adaptability to the stock market.

In the experiment, 5-10 time steps can make the prediction effect of the model

reach optimal, indicating that the enormous volatility of China's financial market is more suitable for short-term prediction. Meanwhile, this model can be applied to other time series problems. Finally, the MS-SSA-LSTM model is universal. Although we tested the model only on China's financial market, it is also suitable for foreign stock markets. The input values to the model are multi-source data, one is the technical stock data, and the other is the sentiment index of shareholders. These multi-source data exist in both Chinese and foreign financial markets. Social media platforms for discussing stock prices are diversified and open. As long as we can dig out the comments on the stock, we can calculate the sentiment index and use it as one of the model's input features.

In addition, establishing a comment platform for the stock market helps us collect comments easily and quickly.

In the MS-SSA-LSTM model, our multi-source data still have limitations. Regarding sentimental analysis, this paper only divides emotions into positive and negative. In addition, other variables, such as macroeconomic conditions and policy shifts, will also impact the stock price forecast. In the future, sentimental analysis needs to be further refined. We should extract different emotional indicators into our research, such as sadness, fear, anger,

and disgust. Meanwhile, more data sources, such as Weibo and WeChat official accounts, may be introduced to estimate market sentiments. In addition, we should use mining techniques to find other factors that predict stock prices.

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] Reddy, Kallem Niranjana, and Pappu Venkata Yasoda Jayasree. "Low Power Strain and Dimension Aware SRAM Cell Design Using a New Tunnel FET and Domino Independent Logic." *International Journal of Intelligent Engineering & Systems* 11, no. 4 (2018).

[12] Reddy, K. Niranjana, and P. V. Y. Jayasree. "Design of a Dual Doping Less Double Gate Tfet and Its Material Optimization Analysis on a 6t Sram Cells."

[13] Reddy, K. Niranjana, and P. V. Y. Jayasree. "Low power process, voltage, and temperature (PVT) variations aware improved tunnel FET on 6T SRAM cells." *Sustainable Computing: Informatics and Systems* 21 (2019): 143-153.

[14] Reddy, K. Niranjana, and P. V. Y. Jayasree. "Survey on improvement of PVT aware variations in tunnel FET on SRAM cells." In *2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC)*, pp. 703-705. IEEE, 2017

[15] Karne, R. K. ., & Sreeja, T. K. . (2023). PMLC- Predictions of Mobility and Transmission in a Lane-Based Cluster VANET Validated on Machine Learning. *International Journal on Recent and Innovation Trends in Computing and Communication*, 11(5s), 477–483.

<https://doi.org/10.17762/ijritcc.v11i5s.7109>

[16] Radha Krishna Karne and Dr. T. K. Sreeja (2022), A Novel Approach for Dynamic Stable Clustering in VANET Using Deep Learning (LSTM) Model. *IJEER* 10(4), 1092-1098. DOI: 10.37391/IJEER.100454.

[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