Business, Management and Economics Engineering
ISSN: 2669-2481 / eISSN: 2669-249X
2022 Volume 20 Issue 2
https://doi.org/ 10.5281/zenodo.7195091

PREDICTION OF STOCK PRICES USING GENETIC ALGORITHM FEATURE SELECTION AND NEURAL NETWORK

Arun B Prasad
Associate Professor (Economics), Institute of Law Nirma University, Ahmedabad, India, Email:arunprasad16@gmail.com, Orcid ID:0000-0002-6108-9219

Manu Singh
Associate Professor, Department of Computer Science, ABES Engineering College Ghaziabad, India, Email:drmanuajaykumar@gmail.com

Manvi Chopra
Assistant Professor, Department of MBA, Dev Bhoomi Uttarakhand University, Dehradun, India, Email:chopramanvi4@gmail.com

Ram Bhawan Singh
Professor and Head,Department of Computer Science, Tula's Institute of Technology, Dehradun, India, India, Email:rambhawansingh@gmail.com

Abstract— A trustworthy data base for our stock price forecast may be found in the financial market's abundance of indicators that seek to explain changes in stock price. Due to their distinct industry sectors and geographical locations, individual stocks are cluttered with a variety of factors. As a result, it is vital to find a multi-element aggregate it really is ideal for a given inventory to evaluate its value. For function selection, the Long STM (LSTM) neural community inventory prediction version defined on these studies will appoint a green Genetic Algorithm (GA). First, we use the GA to rank the relevance of the aspects. The excellent aggregate of objects is then derived from this rating the usage of a trial-and-mistakes methodology. Finally, we integrate the excellent elements with the LSTM version to are expecting stocks. The CSI three hundred inventory dataset and in-intensity empirical investigations the usage of the China Construction Bank dataset display that the GA-LSTM version can forecast statistics better than the baseline methods.

Keywords- Genetic Algorithms (GA), Neural network, LSTM, financial market, Stock prediction, Neural Network.

I.INTRODUCTION

The number of publicly traded companies is growing quickly along with the social economy, making stocks one of the hottest issues in the finance industry. The path of diverse monetary behaviours is often motivated through the fashion of the inventory marketplace, so teachers are an increasing number of that specialize in a way to expect inventory costs. Because the inventory marketplace statistics consists of non-linear, excessive noise, complexity, and temporal features, teachers have studied the inventory prediction approach extensively. The

conventional approach for predicting inventory costs includes constructing a linear prediction version the use of beyond inventory statistics, with the ARIMA approach advocated as an alternative. Outliers withinside the examiner statistics even have a giant effect at the forecast, significantly decreasing the opportunity that the version might also additionally run into neighbourhood most appropriate difficulties. Nonetheless, in phrases of computational efficiency, this approach has a few advantages. Developed the SVM inventory prediction version, which significantly progressed the generalizability of the process. Fast operation rates, robust anti-interference capabilities, and high accuracy rates define ensemble learning. The experiment shown that the random forests stock prediction model outperforms SVM in terms of prediction accuracy after parameter optimization.

II LITERATURE SURVEY

The most important phase in any reasonable research is the literature review. We would want to review the prior research in the field in which we are working before we begin creating. Based on the idea of the study, we will predict or create the drawback and start working with the help of the prior research.A literature review or study forms an important a part of many projects.It shows that the researcher has read material on the topic of the project, which the researcher is comfortable with the term, concept and theories surrounding the chosen area of research and can make sure that the researcher is duplicating research that has before. This study identifies five major risk factors that influence stock and bond returns. There are three stock market factors: general market, firm size, and book-to-market equity. There are two aspects of the bond market. associated with default risks and maturity. Due to the stock market variables' common variance, which also exists within the bond market factors, stock returns and bond returns are related. excluding subpar corporations most important. The average returns on stocks and bonds appear to be clarified by the five criteria. The user can easily run to those locations because they are close by (e.g., 500 metres away).

Three measurement models (ARIMA, ARIMA-EGARCH, and ARIMA-EGARCH-M) are analysed in this study for hourly continuous power prices for all five Midwest Independent System Operator (MISO) centres, and their individual model execution is examined using example-based testing. The results of the ARIMA models show that autoregressive contingent heteroskedasticity is present. The EGARCH detail for the difference condition reveals the presence of a contrary influence impact in power costs for each centre, indicating the possibility of imbalanced time-shifting instability. As to estimates, no one model obviously rules the others as far as in-example anticipating execution upheld four figure assessment insights. In any case, the ARIMA-EGARCH-M model beats the contrary models (Michigan centre is that the exemption) as far as the out of test estimating execution.

Expectation of financial exchange patterns has been a space of extraordinary premium both to the individuals who wish to benefit by exchanging stocks the securities exchange and for specialists endeavouring to reveal the data stowed away in the securities exchange information. Uses of information digging procedures for securities exchange expectation, is a space of exploration which has been getting a great deal of consideration as of late. This work presents the plan and execution assessment of a half breed choice tree-harsh set-based framework for anticipating the following days‟ pattern in the Bombay Stock Exchange (BSESENSEX). Specialized markers are employed in the current investigation to identify the highlights from

the ordinary SENSEX data. The C4.5 choice tree is then utilised to select the critical highlights, and an unpleasant set-based architecture is used to start rules from the removed highlights. The execution of the mixed unpleasant set-based framework is contrasted with a fake neural organization-based pattern expectation framework, a trusting bayes-based pattern indicator, and the execution of the mixed unpleasant set-based framework. According to the results, the suggested framework performs better than neural organization-based pattern prediction frameworks as well as gullible bayes-based pattern forecasting frameworks.

III PROPOSED APPROACH

We encourage novel approaches, and this work suggests using GA to choose various components for the LSTM neural organisation stock forecast model. Through trial examination, this technique accomplished noteworthy outcomes in working on the precision of stock expectation. This depicts the approaches that are utilized in this examination. Portrays the GA-LSTM two-stage stock value forecast model. The best blend of highlight factors was controlled by the interaction of highlight choice and trial correlation. Sums up the discoveries and gives ideas to additional exploration

With the improvement of man-made reasoning innovation, profound learning has drawn in broad consideration because of its incredible exhibition in machine interpretation, voice feeling acknowledgment, picture acknowledgment and different perspectives. In contrast to the conventional measurable model, the deep neural network (DNN) may analyse a complicated nonlinear relationship by layering component representations, which is appropriate for the stock information investigation's multifarious impact and complex nonlinear problem. Tsantekidis and others built a inventory expectation version the usage of a convolutional neural network (CNN), then in comparison it to different traditional fashions to decide whether or not the convolution version will be used for marketplace forecasting. However, the convolutional neural company version isn't always the quality neural company version for inventory anticipation due to the country of marketplace information. Selvin et al. advanced 3 inventory forecast fashions using CNN, repeated neural company (RNN), and LSTM profound getting to know networks as their foundations. By predicting the inventory fees of recorded businesses, they investigated the presentation of the 3 fashions. Finally, it become concluded that LSTM neural company is normally suitable for comparing monetary change the usage of time collection because of its prolonged memorability.

System Diagram:

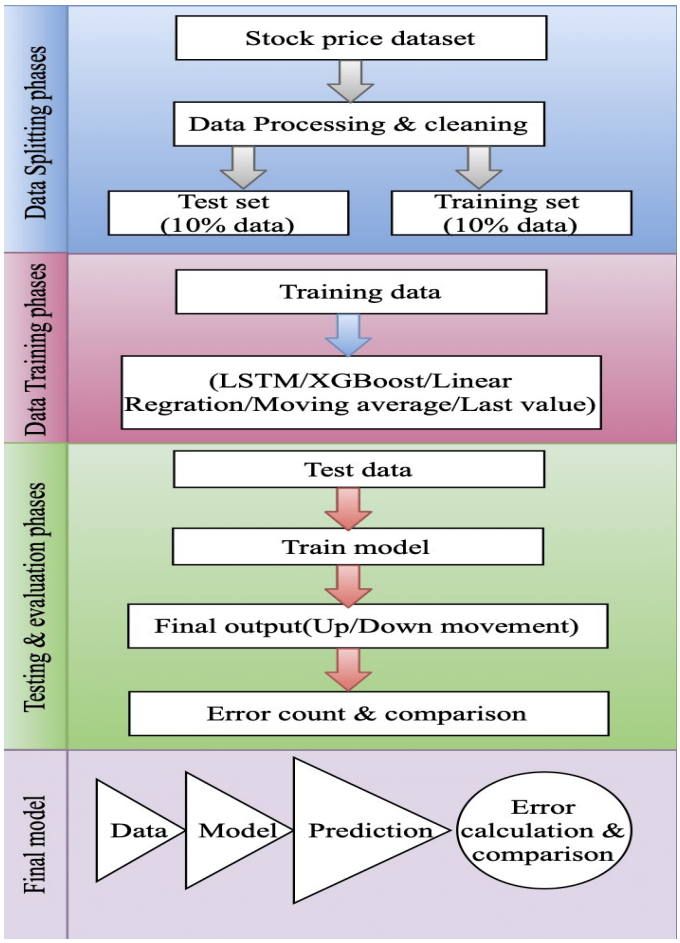
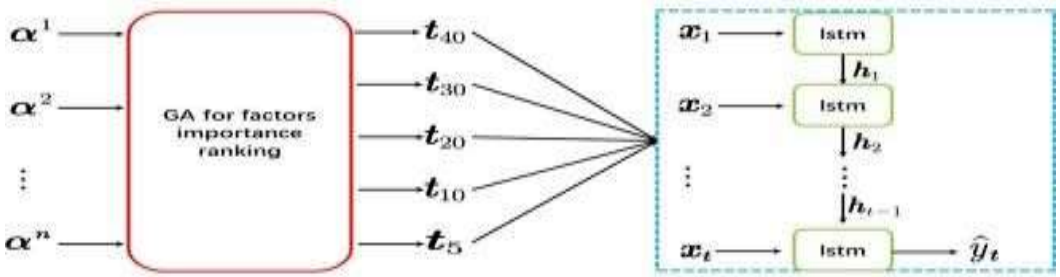


Fig 1. System Architecture

IV Proposed Algorithm:

A. Genetic Technique

It is an heuristic adaptive search strategy that is widely used to uncover approximations of optimal solutions to optimization problems with large search spaces. It may also be successfully used in the selection of optimization characteristics. Every person is a thing with chromosome-like traits, and GA encodes a potential resolution to a dispute into a personal. As a result of the algorithm gathering such a diverse group of individuals, the population is subjected to the GA optimization process.



IV EXPERIMENTAL RESULT

A. In this experiment, the LSTM model was used to normalize and fill in the mean of 2490 historical data points from the CSI 300 stock and the China Construction Bank.The MSE of the check set changed into accrued for the cause of version evaluation. With the last 20% of the data, the test is examined and trained. The unique component subset in addition to the pinnacle 30, 20, 10, and five component subsets withinside the component significance rating has been blanketed withinside the enter capabilities of the LSTM version for comparative testing. Four baseline fashions and our proposed version are as compared a good way to determine how properly our proposed approach works. Due to its long-time period memorability, the LSTM version stands proud from the relaxation and is specifically appropriate for time collection inventory marketplace forecasting. A traditional nonlinear inventory forecasting version with a excessive potential for generalization is the PCA-SVM version. After converting the parameters, the random forests inventory prediction version [8] can also additionally notably enhance forecast accuracy. A complex version for sequential prediction and function extraction called the dual-level attention-primarily based totally recurrent neural network (DARNN)Table three shows a contrast of numerous techniques' performance. In Table three, the quantity of things withinside the function set is denoted via way of means of the letter "k. "As proven in Table three, the recommended GA-LSTM version can outperform the baseline fashions on each dataset. When the pinnacle ten component subsets withinside the component significance rating is applied because the enter capabilities of the recommended GA-LSTM version at the China Construction Bank dataset, the prediction becoming diploma reaches its maximum level. The MSE is 0.0042 while it's far lowest. When the pinnacle 20 component subsets withinside the component significance rating is used as enter capabilities for the recommended GA-LSTM version at the CSI three hundred inventory dataset, the prediction becoming diploma reaches its maximum level. The MSE is 0.0039 while it's far lowest.as proven withinside the Figs. On the check set, our proposed GA-LSTM version plays admirably in 6 and 7.

B. **Comparison Graph:**

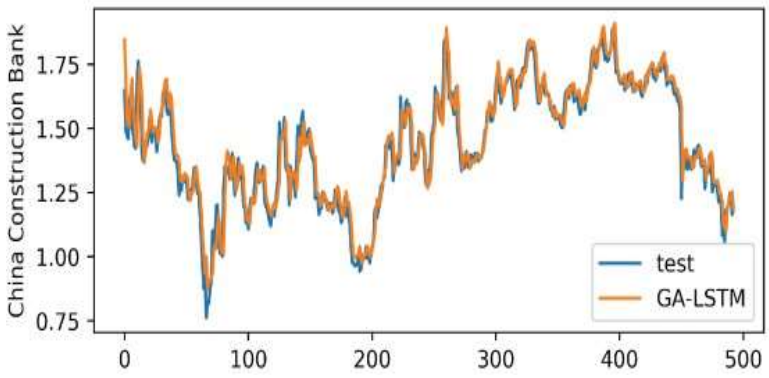


Fig 2. Graphs

C. **Comparison Table:**

model	China construction bank	CSI 300 stock
	MSE	MSE
PCA-SVM	0.0080	0,0096
Random forests	0.0078	0.0083
DA-RNN	0.0067	0.0070
LSTM	0.0072	0.0073
GA-LSTM(K=30)	0.0066	0.0054
GA-LSTM(K=20)	0.0053	0.0039
GA-LSTM(K=10)	0.0042	0.0043
GA-LSTM(K=5)	0.0047	0.0050

Table 1.comparative result

D. **Comparative discussion**

a. The customary stock forecast strategy is to assemble a straight expectation model dependent on the authentic stock information.

b. Suggested using the ARIMA technique to put together an autoregressive model to predict stock costs. Albeit this technique enjoys a few benefits in computational effectiveness, the suspicion of measurable dissemination and dependability of the exploration information restricts their capacity to demonstrate the nonlinear and non-fixed monetary time series, and the anomalies in the examination information likewise significantly affect the expectation incredibly lessen the chance of the model falling into nearby ideal issues.

Advantages and Disadvantages

1. LSTMs support a wide range of parameters, including learning rates and input and output biases.
2. The use of LSTMs reduces the complexity of updating each weight to O(1), which is comparable to BPTT.
3. GA searches a population of points rather than just one.
4. GA do not employ derivatives; they use reward (objective function) information.
5. multi-objective optimization is supported by GA.
6. Genetic algorithms employ probabilistic rather than deterministic transition principles.

Disadvantages

1. Using genetic algorithms still requires skill.
2. Genetic algorithms are less problem-specific, but it may be difficult to construct an objective function and select the right representation and operators.
3. It takes a long time to compute using the genetic algorithm.
4. LSTMs take longer to train

Application

1. Robot control.
2. Time series prediction.

3. Speech recognition.
4. Rhythm learning.
5. Music composition.
6. Grammar learning.
7. Handwriting recognition.
8. Human action recognition.

Conclusion

It introduces the multi-factor model to stock forecasting. There are numerous stock factors on the stock market that define changes in stock price. In this study, numerous typical stock factor selections are made. To choose the features that are most appropriate for the current scene, GA is suggested for feature selection because typicality does not imply that it can be applied to all circumstances. To predict stock values, the LSTM deep learning network model and the complex nonlinear relationship between factors and stocks are mined.

There are still several limitations, some of which are stated below, despite the stock price prediction model provided in this research having significant robustness and possibly improving prediction accuracy.For future research to include data from other stock markets, we first limit our experiment to Chinese stock data. Second, this study's design of model parameters frequently employs trial-and-error rather than a methodical approach to selecting the appropriate parameter size, such as the number of factors.The approach for development entails integrating with different device mastering technology to find the right parameters and enhance the model's interpretability. Additionally, while the GA (Genetic Algorithm) manipulate parameters, along with crossover rate, mutation rate, and quantity of element mixtures, are specified, a variety of perfect mixtures may be shaped to beautify the research's performance.

REFERENCES

[1] E. F. Fama and K. R. French, ``Common risk factors in the returns on stocks and bonds," J. Financial Econ., vol. 33, no. 1, pp. 3_56, Feb. 1993.

[2] P. Ding, Quantitative Investment: Strategy and Technology. Beijing, China:Electronics Industry Press, 2012, pp. 122_127. [3] N. Bowden and J. E. Payne, ``Short term forecasting of electricity prices for MISO hubs: Evidence from ARIMA-EGARCH models," Energy Econ., vol. 30, no. 6, pp. 3186_3197, Nov. 2008.

[4] B. B. Nair, V. P. Mohandas, and N. R. Sakthivel, ``A decision tree-rough set hybrid system for stock market trend prediction," Int. J. Comput. Appl., vol. 6, no. 9, pp. 1_6, Sep. 2010.

[5] H. Penman, ``Application study of BP neural network and logistic regression for stock investment," Expert Syst., vol. 5, no. 12, pp. 21_22, Dec. 1989.

[6] F. Li and C. Liu, ``Application study of BP neural network on stock market prediction," in Proc. 9th Int. Conf. Hybrid Intell. Syst., Beijing, China, 2009, pp. 2_5.

[7] L. J. Cao and F. H. Tay, ``Support vector machine with adaptive parameters in Financial time series forecasting," IEEE Trans. Neural Netw., vol. 14, no. 6, pp. 5_10 Nov. 2003.

[8] J. Deng and L. Li, ``Application of parameter optimization stochastic forest in stock prediction," Software, vol. 41, no. 1, pp. 178_182, Jan. 2020.

[9] M. R. Costa-jussà, A. Allauzen, L. Barrault, K. Cho, and H. Schwenk, ``Introduction to the special issue on deep learning approaches for machine translation," Comput. Speech Lang., vol. 46, pp. 367_373, Nov. 2017.

[10] H. M. Fayek, M. Lech, and L. Cavedon, ``Evaluating deep learning architectures for speech emotion recognition," Neural Netw., vol. 5, no. 3,

[11] J. Xing, K. Li, W. Hu, C. Yuan, and H. Ling, ``Diagnosing deep learning models for high accuracy age estimation from a single image," PatternRecognit., vol. 66, no. 1, pp. 106_116, Jun. 2017.

[12] D. Lv, D. Wang, M. Li, and Y. Xiang, ``DNN models based on dimensionality reduction for stock trading," Intell. Data Anal., vol. 24, no. 1, pp. 19_45, Feb. 2020.

[13] A. Tsantekidis, N. Passalis, A. Tefas, J. Kanniainen, M. Gabbouj, and A. Iosi_dis, ``Forecasting stock prices from the limit order book using convolutional neural networks," in Proc. IEEE 19th Conf. Bus.Informat. (CBI), Beijing, China, Jul. 2017, pp. 10_15.

[14] S. Selvin, R. Vinayakumar, E. A. Gopalakrishnan, V. K. Menon, andK. P. Soman, ``Stock price prediction using LSTM, RNN and CNN-sliding window model," in Proc. Int. Conf. Adv. Comput., Commun. Informat. (ICACCI), Beijing, China, Sep. 2017, pp. 234_239.

[15] U. F. Siddiqi, S. M. Sait, and O. Kaynak, ``Genetic algorithm for the mutualinformation-based feature selection in univariate time series data," IEEE Access, vol. 8, pp. 9597_9609, 2020.