International Journal of Analysis of Basic and Applied Science **ISSN: 2457-0451**

Vol. No.6, Issue II, Apr-Jun, 2022 http://bharatpublication.com/current-issue.php?jID=30/IJABAS

Developing a Smart, Integrated Stock Price Prediction Model for Enhancing the Efficacy of Forecasting Stock Price

Drishti Arora

Goel, Gupta, Maheshwari Associates, Gurugram, Haryana, India

ABSTRACT

Determining the future value of a company's stock is the primary objective of stock price prediction, which is influenced by various factors such as industry trends and market conditions. However, the complexity of stock data poses a challenge for machine learning models due to the high dimensionality and correlated attributes, which can impact the accuracy of predictions. Principal Component Analysis (PCA) is employed to address this challenge by reducing dimensionality, thereby improving the suitability of linear regression algorithms for predicting future stock prices. This study investigates the impact of PCA on Tesla stock price data both before and after applying linear regression algorithms. The resultsdemonstrate that PCA enhances the performance of machine learning models by reducing correlation and selecting principal components that capture essential information while minimizing data redundancy. Evaluation metrics such as root mean square error and R-square value are utilized for assessment.

INTRODUCTION

Stock price prediction is a significant topic in the financial domain, as it directly influences economic and social behaviours. Traditional methods like the autoregressive integrated moving average (ARIMA) model are commonly used to construct forecast models. However,nonlinear forecasting models, such as decision trees based on rough sets, suffer from overfitting due to the large amount of noise in the datasets.

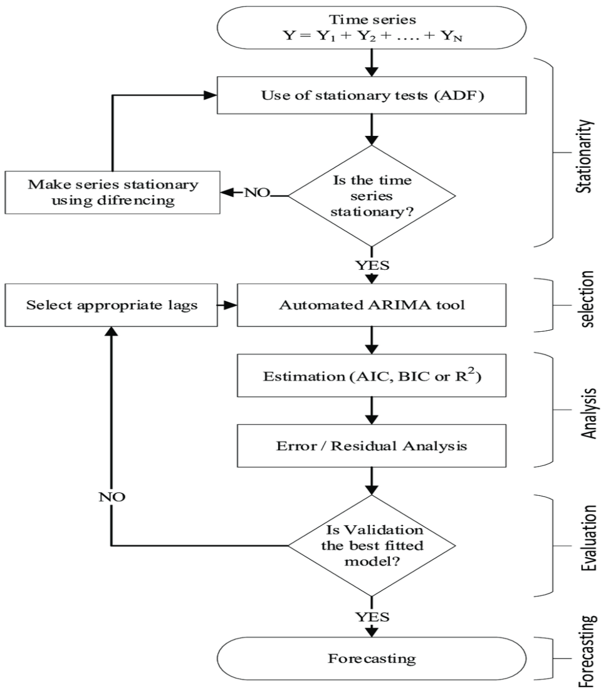

Fig 1: Stock price prediction

To address this issue, artificial neural networks (ANNs) are preferred for their ability to handle nonlinear data models, particularly in time series prediction. ANNs are employed to mitigate local optimal problems, while support vector machines (SVMs) improve model generalizability. Further optimization of parameters results in better performance from random forests compared to SVMs.

Deep neural networks (DNNs) analyse layered feature representations to capture complex, deep nonlinear relationships in the data. The objective is to predict stock price trends using linear regression as a classification model, augmented by PCA to optimize results by reducing dimensionality and eliminating redundancy.

TECHNIQUES

1. Support Vector Machine (SVM):

SVM is a versatile algorithm that addresses the issue of linear indivisibility through kernel functions. Its primary objective is to find a hyperplane that separates data points effectively. SVM is particularly suited for small datasets, nonlinear patterns, and pattern recognitiontasks. It often outperforms neural networks in prediction accuracy.

2. Random Forest Algorithm (RF):

RF is built on the foundation of decision tree algorithms, where records are split into smaller features. It is commonly utilized for both classification and regression analyses. By utilizing random subsets of data from the stocks, RF trains models to achieve higher accuracy in prediction.

3. Artificial Neural Networks (ANN):

ANNs are essential for modelling nonlinear relationships. Inspired by the human brain, this algorithm comprises input, output, and hidden layers. Input features feed into the model, and outputs are computed through calculations in the hidden layers. ANNs employ nodes and utilize the error backpropagation algorithm, which involves steps like parameter initialization, forward propagation, error calculation, and error backpropagation.

4. Empirical Mode Decomposition (EMD):

EMD enhances accuracy while reducing complexity. This method relies on Fourier transformation to decompose signals based on time scales and identify regular patterns intime series data. ARIMA (Auto Regressive Integrated Moving Average) is commonly used inconjunction with EMD to improve prediction accuracy.

5. Long Short-Term Memory (LSTM):

LSTM stands out as one of the most effective algorithms used in numerous stock market projects. It involves various operations followed by model training. SVM is often employed alongside LSTM to address small design, nonlinearity, and pattern recognition issues, ultimately leading to superior prediction performance compared to neural networks.

METHODOLOGY

Predicting Tesla's closing price is the primary focus of this system. Employing various Machine Learning Algorithms, we train the machine using historical data points to forecast future stock prices.

1. Data Description:

The dataset spans four years from June 29, 2010, to February 3, 2020, for Tesla. Attributes include High, Low, Open, Close, Adjacent Close, and Volume, with daily closing prices extracted.

2. Data Pre-processing:

Data pre-processing is crucial to extract important details from the dataset. Since data maynot always be clean and structured, we perform pre-processing to remove outliers, noise, and missing values. Normalization is applied to ensure numerical values are transformed without altering their ranges.

3. Principal Component Analysis (PCA):

PCA is employed to reduce the dimensions of various variables in the dataset. It aims to derive uncorrelated variables from correlated ones through mathematical procedures. By reducing the data's dimensionality, PCA helps focus on the most influential components whileensuring independence among variables, which is essential for linear models.

4. Linear Regression:

Linear regression is a supervised technique used to fit a best-fit line based on the relationship between independent variable X and dependent variable Y. The equation for linear regressionis represented as:

i. y^ = a + bx (I)

ii. y = b0 + b1x1 + b2x12 + b2x13 +.......................bn x1n (II)

CONCLUSION

1. Stock market investment is a complex task due to various factors. A reliable prediction system assists investors in making accurate and profitable investment decisions by providing insights into future stock price directions.
2. The utilization of high-dimensional data enhances the performance of Machine Learning models, especially in the classification of PCA. This paper explores the impact of features on PCA and investigates the selection of optimum principal components.
3. This study contributes to understanding the effects of implementing PCA on highly correlated datasets and aids in selecting optimal principal components for different techniques, such as k for k-NN, kernel parameters in SVM, and the number of PCs.
4. Sentiment analysis significantly influences future stock prices, indicating the potential for highly efficient predictions with a combined approach. However, existing propositions may struggle to predict dynamic and rapidly changing patterns in stock price movements.
5. Many studies evaluate their machine learning models using various techniques, contributing to ongoing advancements in stock price prediction.

REFERENCES

1. Shah D, Isah H, Zulkernine F, “Predicting the effects of news sentiments on the stock market”, In2018 IEEE International Conference on Big Data (Big Data) 2018 Dec 10 (pp.4705-4708) IEEE.
2. Mehtab S, Sen J, Dutta A, “Stock price prediction using machine learning and LSTM- based deep learning models. symposium on Machine Learning and MetaheuristicsAlgorithms, and Applications” , Springer, Singapore.2020 Oct 14 (pp. 88-106).
3. NekoeiQachkanloo H, Ghojogh B, Pasand AS, Crowley M, “Artificial counselor system for stock investment”, InProceedings of the AAAI Conference on Artificial Intelligence 2019Jul 17 (Vol. 33, No. 01, pp. 9558- 9564).
4. Mehtab S, Sen J. A robust ,“Predictive model for stock price prediction using deep learning and natural language processing.” Available at SSRN 3502624. 2019 Dec 12.

5. Chen S, Zhou C, “Stock prediction based on genetic algorithm feature selection and long short-term memory neural network”, IEEE Access. 2020 Dec 24;9:9066-72.
6. R. Sathya, Prateek Kulkarni, Momin Nawaf Khalil, Shishir Chandra Nigam, “Stock PricePrediction using Reinforcement Learning and Feature Extraction”, IJRTE Volume-8 Issue-6, March 2020
7. “Pattern Recognition in Stock Market”, D. Sorna Shanthi*1; T. Aarthi*2; A.K. Bhuvanesh*3; RA. Chooriya Prabha*4.2020.
8. Chacón HD, Kesici E, Najafirad P, “Improving financial time series prediction accuracy using ensemble empirical mode decomposition and recurrent neural networks” ,IEEE Access.2020 May 25;8:117133-45
9. Yuan X, Yuan J, Jiang T, Ain QU, “ Integrated long-term stock selection models based onfeature selection and machine learning algorithms for China stock market”, IEEE Access. 2020 Jan 24;8:22672-85.
10. Ketter W, Collins J, Gini M, Gupta A, Schrater P, “A Computational Approach to Predicting Economic Regimes in Automated Exchanges”, 2005 Jan 1 (pp. 147-152).
11. Li X, Li Y, Yang H, Yang L, Liu XY, “DP-LSTM: Differential privacy-inspired LSTM for stock prediction using financial news.”, arXiv preprint arXiv:1912.10806. 2019 Dec 20.
12. Cho P, Park JH, Song JW., “Equity Research Report-Driven Investment Strategy in Korea Using Binary Classification on Stock Price Direction.” ,IEEE Access. 2021 Mar 22;9:46364-73.
13. Blohm I, Riedl C, Füller J, Köroglu O, Leimeister JM, Krcmar H, “The effects of prediction market design and price elasticity on the trading performance of users: an experimental analysis.” , arXiv preprint arXiv:1204.3457. 2012 Apr 16.
14. Chou JS, Truong DN, Le TL, “Interval forecasting of financial time series by accelerated particle swarm optimized multi-output machine learning system.” ,IEEE Access. 2020 Jan 10; 8:14798-808.
15. Shen J, Shafiq MO, “Short-term stock market price trend prediction using a comprehensive deep learning system.”, Journal of Big Data. 2020 Dec;7(1):1-33