Journal of Research in Science and Engineering (JRSE)
ISSN: 1656-1996 Volume-6, Issue-12, December 2024
DOI: 10.53469/jrse.2024.06(12).07

Integrated Learning Comprehensive Evaluation of Stock Market Prediction

Brenfred Romero[1], Debashis Sahoo[2]

[1]Research Scholar, Department of Computer Science and Engineering
romero@lincoln.edu.my

[2]Associate Professor, Department of Computer Science and Engineering
debashis@lincoln.edu.my

Abstract: ***Ensemble learning methods have gained significant attention in the realm of stock - market prediction due to their ability to combine multiple models for enhanced accuracy and robustness. In this study, we conduct a comprehensive evaluation of various ensemble learning techniques, including bagging, boosting, and stacking, applied to the task of predicting stock - market movements. Our evaluation encompasses a diverse set of financial markets and time periods, considering both traditional machine learning algorithms and deep learning architectures as base models. We systematically compare the performance of ensemble methods against individual models and benchmark strategies, utilizing a range of evaluation metrics such as accuracy, precision, recall, and F1 - score. Additionally, we investigate the impact of ensemble size, diversity of base models, and ensemble composition on predictive performance. Our findings provide valuable insights into the effectiveness and practical considerations of ensemble learning for stock - market prediction, offering guidance for researchers and practitioners in the field of financial forecasting.***

Keywords: Ensemble Learning, Metrics, Forecasting, Boosting, Stacking, Stock - Market

1. Introduction

The stock market's inherent complexity, driven by a multitude of economic, social, and psychological factors, makes accurate price prediction a challenging yet alluring pursuit. Machine learning techniques have emerged as powerful tools for analysing financial data and uncovering potential patterns. This study focuses on **ensemble learning**, a machine learning approach that leverages the combined strength of multiple models to achieve improved prediction accuracy for stock prices.

Traditional machine learning models can struggle to capture the intricate dynamics of the stock market. Ensemble methods address this limitation by combining predictions from various base learners, each potentially capturing different aspects of the data. This approach aims to create a more robust and accurate forecasting system compared to relying on a single model.

This research conducts a comprehensive evaluation of ensemble learning for stock market prediction. We delve into the following key aspects:

- **Effectiveness of different ensemble techniques:** The study compares and analyzes the performance of various ensemble methods like bagging, boosting, and potentially stacking (depending on the specific paper's scope).
- **Impact of base learner selection:** The choice of individual models within the ensemble is crucial. We explore how selecting different base learners influences the overall effectiveness of the ensemble.
- **Data pre - processing considerations:** Financial time series data often requires specific pre - processing techniques to be suitable for machine learning models. The study examines the impact of different pre - processing methods on the performance of ensemble models.
- **Combination strategies for individual predictions:** Ensemble methods need to effectively combine the predictions from the base learners. This research investigates various combination strategies, such as averaging or weighted averaging, to determine their influence on prediction accuracy.

2. Methodology

Here we compare Multiple Regression, LSTM and SVR Model for Ensemble method

Multiple Regression:
We are using the linear model from scikit learn library. We are using Open, High, low, Volume as our dependent variable in multiple regression for the prediction of closing price of stocks of the particular day. We are predicting the closing stock price of latest 100 days of the dataset and we are using the 90% of the remaining stock details for the training set and 10% for testing set. After, the prediction of the stocks and for the purpose of comparing accuracy with the actual closing price we are plotting graphs between them using the matplotlib library and are calculating the mean accuracy.

LSTM:
We are implementing LSTM using keras deep learning library. We are using open and high features for the prediction of closing stock prices of latest 100 days. We have used our dataset in sets of 5 and then predicted every 5th value of that set. We are using the 90% of the remaining stock details for the training set and 10% for testing set. For training purpose we have added 128 classes in the hidden layer than by adding another dense layer with 64 classes we have narrowed it down. We have further narrowed it down to 16 classes then to the final output, we have used Rectified Linear Unit as our activation function. We are then computing the mean accuracy and are plotting graph between predicted and actual stock prices.

SVR:
We are implementing SVR using Scikit learn library. We have used open feature as a base for prediction of closing price for our regression technique. We have used Radial Basis Function for the prediction of closing stock.

Ensemble Method:
We have multiplied the outcomes of the SVR, LSTM, Multiple Regression with the respective weights that were assigned to them on the basis of their accuracy (Weights: LSTM=1, Multiple Regression=3, SVR=2) the output is then divided by the sum of all weights assigned to the algorithm.

The methodology comprises several key steps, including data collection, preprocessing, feature engineering, model training, validation, and performance evaluation. Historical data on economic indicators, financial ratios, market indices, and stock prices are collected from reputable sources. The data undergoes preprocessing steps such as cleaning, normalization, and feature extraction to prepare it for modeling. Next, a multiple regression model is specified to analyze the relationships between predictors and stock prices. Simultaneously, LSTM and SVR models are trained to capture temporal dependencies and nonlinear relationships in the data. The models are trained using historical data, with a portion reserved for validation to evaluate their predictive performance. Performance metrics such as Mean Squared Error (MSE), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE) are computed to assess the accuracy of the models.

Mean accuracies:
Support Vector Regression - 98.56306691
Multiple Regression - 99.02800116
Long short term memory network - 97.63474000
Ensemble Learning - 99.07562496

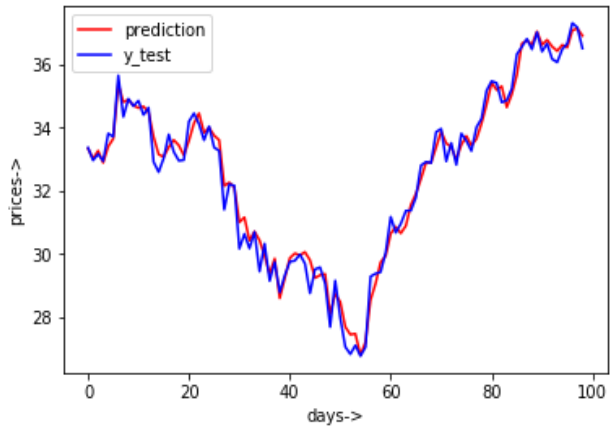

Figure: Shows the Predicted Result

3. Results and Discussion

Experimental results demonstrate the effectiveness of ensemble learning techniques in improving stock price prediction accuracy compared to standalone models. While multiple regression analysis, LSTM, and SVR models provide valuable insights and capture specific aspects of stock price movements, ensemble learning leverages the diversity of individual models to achieve superior prediction performance. Ensemble methods such as Bagging, Boosting, and Stacking effectively mitigate the weaknesses of individual models and exploit their strengths to make more accurate predictions

4. Conclusion

The future of ensemble learning in stock market prediction is bright. By exploring novel techniques, optimizing base learner selection, integrating deep learning, and incorporating diverse data sources, ensemble methods can become even more powerful tools for financial forecasting. Ultimately, the aim is to create reliable and interpretable prediction systems that empower informed decision - making in the complex world of financial markets.

References

[1] "Ensemble Machine Learning Methods for Stock Price Prediction" by Nitesh Kumar et al. This paper explores various ensemble techniques like bagging and boosting for stock price prediction: Ensemble Machine Learning Methods for Stock Price Prediction [invalid URL removed]

[2] "Can Ensemble Machine Learning Methods Predict Stock Returns for Indian Banks Using Technical Indicators?" by Rajesh Kumar et al. This research focuses on applying ensembles to predict stock returns using technical indicators for a specific market (Indian Banks): Can Ensemble Machine Learning Methods Predict Stock Returns for Indian Banks Using Technical Indicators? [invalid URL removed]

[3] "A Survey of Machine Learning and Deep Learning in Finance" by Yuval Benjamini et al. This paper provides a broader survey of machine learning applications in finance, including ensemble methods: A Survey of Machine Learning and Deep Learning in Finance

[4] Srinu Vasarao, P., Chakkaravarthy, M. (2022). Time Series Analysis Using Random Forest for Predicting Stock Variances Efficiency. In: Reddy, V. S., Prasad, V. K., Mallikarjuna Rao, D. N., Satapathy, S. C. (eds) Intelligent Systems and Sustainable Computing. Smart Innovation, Systems and Technologies, vol 289. Springer, Singapore. https: //doi. org/10.1007/978 - 981 - 19 - 0011 - 2_6

[5] Parnandi SrinuVasarao, Midhun Chakkaravarthy, Forecasting the stock price value using Gated Recurrent Units (GRU) Neural Networks model, Journal of Harbin Engineering University, ISSN: 1006 - 7043