National Journal of Commerce and Management (NJCM)
Peer-Reviewed/Refereed Journal | ISSN 2394-6342 (O)
Volume 11, Issue 02, July December 2024
http://njcm.pratibha-spandan.org

FORECASTING DAILY STOCK MARKET TRENDS: A RANDOM FOREST APPROACH FOR PREDICTING NEXT-DAY RETURN DIRECTION

SHANTNU SOOD

Research Scholar**,** Himachal Pradesh University Business School, Summer Hill, Shimla

DR. PUNEET BHUSHAN

Assistant Professor, Himachal Pradesh University Business School, Summer *Hill, Shimla*

ABSTRACT

Predicting stock prices has long posed a significant challenge due to the inherent complexities and stochastic nature of financial time series data. Traditional time series models such as Autoregressive Integrated Moving Average (ARIMA), Generalized Autoregressive Conditional Heteroskedasticity (GARCH), and Vector Autoregressive (VAR) have been extensively employed to model and predict stock market behaviour, particularly in capturing linear relationships and stationary data. However, with the advent of big data and enhanced computational power, machine learning techniques have emerged as powerful tools for financial forecasting, capable of handling large datasets and capturing complex, non-linear relationships in market data. Therefore, this paper investigates the effectiveness of one such modern machine learning technique called Random Forest Classifier, specifically focusing on its ability to predict the sign of the next day's return in financial markets. The findings contribute to the ongoing discourse on the potential of machine learning to outperform traditional methods in stock market prediction, offering new insights into an enduring challenge.

Keywords*: Stock Market Trends, Random Forest Approach, Next-Day Return.*

INTRODUCTION

Predicting stock prices has long been a challenging endeavour due to the inherent complexities and stochastic nature of financial time series data. The fluctuating behaviour of markets, influenced by a myriad of unpredictable factors such as geopolitical events, macroeconomic indicators, and investor sentiment, makes accurate forecasting of stock prices particularly difficult. However, while predicting the exact future price of a stock is notoriously challenging, research has shown that it is relatively more feasible to forecast the direction or sign of the next day's return in financial markets.

Historically, traditional time series models, such as Autoregressive Integrated Moving Average (ARIMA), Generalized Autoregressive Conditional Heteroskedasticity (GARCH), and Vector Autoregressive (VAR) models, have been employed to model and predict stock market behaviour. These methods have proven effective in certain contexts, particularly when dealing with linear relationships and stationary data. ARIMA models, for example, are well-suited for capturing linear trends and seasonality, while GARCH models are adept at modelling volatility clustering, a common feature in financial markets.

However, the advent of big data and the exponential increase in computational power have opened new avenues for stock market prediction. Machine learning techniques, which excel at handling large datasets and capturing complex, non-linear relationships, have become increasingly popular in financial forecasting. Methods such as Support Vector Machines (SVMs), Decision Trees, and Random Forests have gained prominence due to their flexibility and robustness in analysing financial data. Unlike traditional models that rely heavily on statistical assumptions, machine learning algorithms can learn patterns directly from the data, making them particularly effective in capturing the intricate and often non-linear dynamics of financial markets.

The purpose of this research paper is to examine the effectiveness of one such modern machine learning called Random Forest Classifier in predicting the sign of the next day's return in financial markets.

REVIEW OF LITERATURE

Numerous studies have explored stock trading investments using various models. For instance, **Abe and Nakayama (2018)** utilized deep learning techniques to forecast returns one month ahead in the Japanese stock market, contributing to the body of work on deep learning in finance. Researchers have also compared the performance of various methodologies such as Neural Networks, Support Vector Machines, and Random Forests in predicting stock returns. Other notable studies in this domain include those by **Ma et al. (2021)** and **Kyriakou et al. (2019),** which employed different machine learning strategies for stock market predictions.

Several works have specifically examined and compared the effectiveness of different deep learning models, such as the research by **Kohli et al. (2018).** In addition, **Nelson et al. (2017)** assessed the accuracy of various deep learning approaches in financial forecasting. **Leung et al. (2021)** investigated the performance of the Gradient Boosting Machine against traditional regression models, concluding that the machine learning approach delivered superior results. Similarly, **Fieberg et al. (2022)** found that while complex machine learning models notably outperformed simple linear regression models, especially in bear markets, they did not result in additional losses for investors compared to traditional methods.

Wong et al. (2020) developed a time-varying neural network that incorporates factors like size, momentum, and industry metrics to predict stock returns. **Wang et al. (2018)** introduced an adaptive model using deep reinforcement learning techniques to ensure consistent returns even in volatile market conditions. Furthermore, studies like **Nabipour et al. (2020)** have examined the comparative performance of deep learning and machine learning algorithms in forecasting stock market trends. Beyond stock return predictions, machine learning models have also been applied to predict market volatility **(Filipović & Khalilzadeh, 2021)** and cryptocurrency returns **(Akyildirim et al., 2020).**

Routledge (2019) illustrated the potential of machine learning techniques for stock selection. **Mehtab and Sen (2019)** proposed an LSTM-based model that integrates public sentiment analysis with long-term historical price data to predict future stock prices. **Nevasalmi (2020)** introduced a multinomial classification approach focusing on predicting significant stock returns while disregarding minor fluctuations near zero return.

RESEARCH METHODOLOGY

OBJECTIVES OF THE STUDY:

- To predict the sign of next day's return for different companies that were part of the NSE 500 index for the given period.
- To measure the accuracy of Random Forest Classifier in predicting the sign of next day's return.

DATA AND PROCESSING:

- The study has been conducted on the constituents of NSE 500 index for the period of 1/04/2013 to 31/03/2023. The data has been taken from CMIE Prowess IQ database. There was a total of 956 companies during the entire period. The target variable for the study is the sign of the next day's return.

FEATURES

This study uses 13 features that can be categorized into technical, fundamental and momentum features.

Table 1: Features

Technical	Fundamental	Momentum
Relative Strength Index, Moving Average Convergence Divergence, Average True Range	EV PBITDA, Turnover, PE Ratio, PB Ratio, Market Capitalisation	Return 5 day, return 10 days, return 21 days, return 42 days, Return 63 day

MACHINE LEARNING MODEL USED

RANDOM FOREST CLASSIFIER:

The Random Forest algorithm is a machine learning technique that employs an ensemble approach for solving classification and regression problems. It combines the strengths of Decision Trees and Bagging (Bootstrap Aggregation) to address the issue of overfitting commonly seen in decision trees. Bagging is particularly effective in reducing the variance of algorithms with high variance, like CART-based decision trees. Random Forest consists of a collection of unpruned classification and regression trees, which are generated from subsets of the training dataset. The algorithm first creates subsets from the training dataset, and then a CART model is applied to each subset, resulting in predicted outputs, after that the bagging process is used to form an ensemble of the predicted outputs from each model. Through this procedure, the ensemble's predicted output demonstrates a lower prediction error compared to individual models.

FINDINGS AND DISCUSSIONS

This section presents the experimental results and the evaluation metrics applied within the proposed prediction framework. The experiments were conducted using Python 3 in conjunction with Scikit-Learn. To evaluate the performance of the machine learning models and their predictions, classification metrics such as F1-score, accuracy, precision, and recall were employed. The F1-score and accuracy were determined by analyzing precision and recall, which are based on true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). A summary of these measurements is provided in Table 2.

Table 2: Metrics For Evaluation of The Model

Accuracy	(TN+TP)/(TN+TP+FN+FP)
F1-Score	(2 * Precision * Recall)/ (Precision + Recall)
Precision	TP/(TP+FP)
Recall	TP/(TP+FN)

In Table 4 the out-of-sample performance of each machine learning model that was employed in this study has been mentioned. The models have generated the results based on 13 features in total out of which 3 were technical and 5 were fundamental in nature and 5 were momentum based. Because of these features, the below-mentioned hyperparameters provide the best results for the random forest model.

Table 3: Best Hyperparameters Selection Using GridSearchCV for Random Forest

Number of estimators	200
Max depth	20
Min samples leaf	2
Min samples split	10
Random state	42

Table 4: Model Performance

Metrics	Random Forest
Accuracy	0.675589
Precision	0.65237
Recall	0.659817
F1 Score	0.656073

The out-of-sample accuracy of 67.56% indicates that the model performs better than random guessing, which would yield an accuracy of around 50% for a binary classification problem. However, the moderate precision and recall values suggest that while the model is reasonably effective at identifying the direction of the next day's return, there is still a significant margin for improvement.

The F1 Score of 0.6561 further confirms that the model has a balanced trade-off between precision and recall, making it a reliable tool for predicting the sign of the next day's return within the context of the dataset used. However, the model's performance could potentially be enhanced by further tuning, incorporating additional features, or exploring alternative modelling approaches.

These metrics provide valuable insights into the model's predictive capabilities and highlight areas where additional refinement may be necessary to improve forecasting accuracy in financial markets.

CONCLUSION

The objective of this investigation was to evaluate the predictive efficacy of the Random Forest Classifier in predicting sign of next day's return. Notably, this research uniquely incorporated fundamental, technical and momentum indicators as concurrent predictors for classifying future stock returns.

Moving forward, there are opportunities for expansion in the study. Firstly, the inclusion of additional Machine Learning models could enrich the analysis. Furthermore, exploring alternative input data, such as macroeconomic variables, could potentially enhance the performance assessment of these models.

REFERENCE

Abe, M., & Nakayama, H. (2018). Deep learning for forecasting stock returns in the Cross-Section. In *Lecture Notes in Computer Science* (pp. 273–284). https://doi.org/10.1007/978-3-319-93034-3_22

Akyildirim, E., Goncu, A., & Sensoy, A. (2020). Prediction of cryptocurrency returns using machine learning. *Annals of Operations Research*, *297*(1–2), 3–36. https://doi.org/10.1007/s10479-020-03575-y

Can neural networks predict stock Market? - Analysis, ratings and forecasts. (n.d.). AC Investment Research. https://www.ademcetinkaya.com/p/how-does-rating-model-work.html

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., ... & Zhou, T. (2015). Xgboost: extreme gradient boosting. *R package version 0.4-2*, *1*(4), 1-4.

Fieberg, C., Metko, D., Poddig, T., & Loy, T. (2022). Machine learning techniques for cross-sectional equity returns' prediction. *OR Spectrum*, *45*(1), 289–323. https://doi.org/10.1007/s00291-022-00693-w

Filipović, D., & Khalilzadeh, A. (2021). Machine Learning for Predicting Stock Return Volatility. *Social Science Research Network*. https://doi.org/10.2139/ssrn.3995529

GeeksforGeeks. (2023). Boosting in Machine Learning Boosting and AdaBoost.

GeeksforGeeks.https://www.geeksforgeeks.org/boosting-in-machine-learning-boosting-and-adaboost.

GeeksforGeeks. (2021). How to Visualize a Neural Network in Python using

GeeksforGeeks. (2023a). XGBoost. *GeeksforGeeks*. https://www.geeksforgeeks.org/xgboost/

Kohli, P. P. S., Zargar, S., Arora, S., & Gupta, P. (2018). Stock prediction using machine learning algorithms. In *Advances in intelligent systems and computing* (pp. 405–414). https://doi.org/10.1007/978-981-13-1819-1_38

Kyriakou, I., Mousavi, P., Nielsen, J. P., & Scholz, M. (2019). Forecasting benchmarks of long-term stock returns via machine learning. *Annals of Operations Research*, *297*(1–2), 221–240. https://doi.org/10.1007/s10479-019-03338-4

Leung, E., Lohre, H., Mischlich, D., Shea, Y., & Stroh, M. (2021). The promises and pitfalls of machine learning for predicting stock returns. *The Journal of Financial Data Science*, *3*(2), 21–50. https://doi.org/10.3905/jfds.2021.1.062

Ma, Y., Han, R., & Wang, W. (2021). Portfolio optimization with return prediction using deep learning and machine learning. *Expert Systems with Applications*, *165*, 113973. https://doi.org/10.1016/j.eswa.2020.113973

Machine Learning Random Forest Algorithm - JavatPoint. (n.d.). www.javatpoint.com.

Mehtab, S., & Sen, J. (2019). A robust predictive model for stock price prediction using deep learning and natural language processing. *Social Science Research Network*. https://doi.org/10.2139/ssrn.3502624

Nabipour, M., Nayyeri, P., Jabani, H., Shahab, S., & Mosavi, A. (2020). Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms Via Continuous and Binary Data; a Comparative Analysis. *IEEE Access*, *8*, 150199–150212. https://doi.org/10.1109/access.2020.3015966

Nelson, D., Pereira, A., & Oliveira, R. (2017, May 1). *Stock market's price movement prediction with LSTM neural networks*. IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/7966019

Nevasalmi, L. (2020). Forecasting multinomial stock returns using machine learning methods. *The Journal of Finance and Data Science*, *6*, 86–106. https://doi.org/10.1016/j.jfds.2020.09.001

R, M. C., Marimuthu, B., & Ghosh, B. (2023). Analysis of Nifty 50 index stock market trends using hybrid machine learning model in quantum finance. *International Journal of Power Electronics and Drive Systems*, *13*(3), 3549. https://doi.org/10.11591/ijece.v13i3.pp3549-3560

Routledge. (2019, May 13). *Machine Learning for Stock Selection*. Taylor & Francis. https://www.tandfonline.com/doi/abs/10.1080/0015198X.2019.1596678

Support Vector Machine (SVM) Algorithm - JavatPoint. (n.d.). www.javatpoint.com.
https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm

Wang, J., Sun, T., Yuan, B., Cao, Y., & Wang, D. (2018, December 1). *Financial Markets Prediction with Deep Learning*. IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/8614047

Wong, S. Y. K., Chan, J., Azizi, L., & Da Xu, R. Y. (2020). Time-varying neural network for stock return prediction. *arXiv (Cornell University)*. https://arxiv.org/pdf/2003.02515.pdf